Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 20–16

CONVEX TRANSFORM ORDER OF BETA DISTRIBUTIONS WITH SOME CONSEQUENCES

IDIR ARAB, PAULO EDUARDO OLIVEIRA AND TILO WIKLUND

ABSTRACT: The convex transform order is one way to make precise comparison between the skewness of probability distributions on the real line. We establish a simple and complete characterisation of when one Beta distribution is smaller than another according to the convex transform order. As an application, we derive monotonicity properties for the probability of Beta distributed variables exceeding the mean or mode of their distribution. Moreover, as a byproduct, we obtain a simple alternative proof of the mode-median-mean inequality for unimodal distributions that are skewed in a sense made precise by the convex transform order. This new proof also gives an analogous inequality for the anti-mode of distributions that have a unique anti-mode. Such inequalities for Beta distributions follow as special cases. Finally, some consequences for the values of distribution functions of Binomial distributions near to their means are mentioned.

KEYWORDS: convex transform order, Beta distribution, sign pattern, exceedance probability, skewness.
MATH. SUBJECT CLASSIFICATION (2010): 60E15, 33B20.

1. Introduction

How to order probability distributions according to criteria that have interpretable probabilistic consequences is a common question in probability theory. Naturally there will exist many different order relations, each one highlighting a particular aspect of the distributions. Classical examples are given by orderings that capture size and dispersion. In reliability theory some ordering criteria are of interest when dealing with ageing problems. These help decide, for example, which lifetime distributions exhibit faster ageing. An account of different orderings, their properties, and basic relationships may be found in the monographs of Marshal and Olkin [13] or Shaked and Shanthikumar [19].

In this paper we shall be interested primarily in two such orderings, known in the literature as the convex transform and the star-shape transform orders.

Received May 14, 2020.
This work was partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

These orders are defined by the convexity or star-shapedness of a certain mapping that transforms one distribution into another. The convex transform order was introduced by van Zwet [22] with the aim of comparing skewness properties of distributions. Oja [16] suggests that any measure of skewness should be compatible with the convex transform order, and that many such measures indeed are. Hence, this ordering gives a convenient formalisation of what it means to compare distributions according to skewness.

With respect to the ageing interpretation, the convex transform order may be seen as identifying ageing rates even in the case of lifetimes that did not start simultaneously, while the star-shape order requires the same starting point for the distributions under comparison[15], as described in Nanda et al. [15].

Establishing that one distribution is smaller than another is often difficult and tends to rely on being able to control the number of crossing points between various affine transforms of distributions functions. Based on such techniques explicit characterisations of the ordering relationships within the Gamma and the Weibull families are given in Arab and Oliveira [2, 3] and Arab et al. [4].

This family of Beta distributions is a two-parameter family of distributions supported on the unit interval. It appears in various context, for example in the study of order statistics and in Bayesian statistics as a conjugate prior for a variety of distributions arising from Bernoulli trials.

The main contribution of this paper is to characterise when one Beta distribution is smaller than another according to the convex- and star-shaped transform orders. This characterisation implies various monotonicity properties for the probabilities of Beta distributed random variables exceeding the mean or mode of their distribution. Using this allows one to derive, in some cases, simple bounds for such probabilities. These differ from concentration inequalities such as Markov's inequality or Hoeffding's inequality in that they study the probability of exceeding, without necessarily significantly *deviating* from, the mean or the mode.

A well known connection between Beta and the Binomial distributions allows us to translate these results into similar monotonicity properties for the family of Binomial distributions. Such probabilities of exceeding means have received attention in the context of studying properties of randomised algorithms, see for example Karppa et al. [12], Becchetti et al. [5], or Mitzenmacher and Morgan [14]. They have also found applications in dealing with specific aspects in machine learning problems, such as in Doerr [8], Greenberg and Mohri [9], with sequels in Pelekis [18] and Pelekis and Ramon [17], or Cortes et al. [7] for more general questions. Such an inequality for the Binomial random variables was also used by Wiklund [20] when studying the amount of information lost when resampling.

These properties also allow one to compare the relative location of the mode, median, and mean of certain distributions that are skewed in a sense made precise by the convex transform order. Such mode-median-mean inequalities are a classical subject in probability theory. While our condition for these inequalities to hold has previously been suggested by van Zwet [23], our proof appears novel. The proof also allows us to establish a similar inequality for absolutely continuous distributions with unique anti-modes, meaning distributions having densities with a unique minimum. For an account of the field we refer the interested reader to van Zwet [23] or, for more recent references to Abadir [1] or Zheng et al. [21].

This paper is organised as follows. In Section 2 we define the relevant concepts and definitions. The main results, characterising the order relationships within the Beta family is presented in Section 3. Consequences are discussed in Section 4, while proofs of the main results are presented in Section 5. Some auxiliary results concerning the main tools of analysis are given in the Appendix A.

2. Preliminaries

In this section we present the basic notions necessary for understanding the main contributions of the paper.

Let us first recall the classical notion of convexity on the real numbers.

Definition 1 (Convexity)**.** A real valued function $f\colon I \mapsto \mathbb{R}$ on an interval I is said to be *convex* if for every $x, y \in I$ and $\alpha \in [0,1]$ we have $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$.

We will also need the somewhat less well known notion of star-shapedness of a function on the real numbers.

Definition 2 (Star-shapedness)**.** A function $f\colon [0,a] \to \mathbb{R}$, for some $a > 0$, is said to be *star-shaped* if for every $0 \leq \alpha \leq 1$, we have $f(\alpha x) \leq \alpha f(x)$.

Star-shapedness could generally be defined on general intervals with respect to an arbitrary reference point. For our purposes it suffices to consider functions on the non-negative half-line, star-shaped relative to the origin.

It is immediate that a convex $f\colon I \to \mathbb{R}$ on an initial segment of the non-negative half-line that satisfies $f(0) \leq 0$ is star-shaped. Moreover, f is star-shaped if and only if $\frac{f(x)}{x}$ is increasing in $x \in I$. We refer the reader to Barlow et al. [6] for some more general properties and relations between these types of functions.

Our main concern in this paper is to establish certain orderings of the family of Beta distributions that are defined in $[0, 1]$.

Definition 3 (Beta distribution)**.** The Beta distribution $\mathrm{Beta}(a, b)$ with parameters $a, b > 0$ is a distribution supported on the unit interval and defined by the density given for $x \in [0, 1]$ by

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)}, \qquad \text{where } \mathrm{B}(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx. \tag{1}$$

We shall be interested in ordering with respect to orderings determined by the convexity or star-shapedness of a certain mapping. Primarily the following order due to van Zwet [22]. In order to avoid working with generalised inverses, we restrict ourselves to distributions supported on intervals.

Definition 4 (Convex Transform Order $\leq_c$)**.** Let P and Q be two probability distributions on the real line supported by the intervals I and J, with strictly increasing distribution functions $F\colon I \to [0, 1]$ and $G\colon J \to [0, 1]$, respectively. We say that $P \leq_c Q$ or, equivalently, $F \leq_c G$, if the mapping $x \mapsto G^{-1}(F(x))$ is convex. Moreover, if $X \sim P$ and $Y \sim Q$, we will also write $X \leq_c Y$ when $P \leq_c Q$.

It is immediate from the definition that if $X \sim P$ and $Y \sim Q$ then both $X \leq_c Y$ and $Y \leq_c X$ if and only if there exist some $a > 0$ and $b \in \mathbb{R}$ such that X has the same distribution as $aY + b$. In other words, the convex transform order is invariant under orientation preserving affine transforms.

Although this order relation is popular in reliability theory, the convex transform order was first introduced by van Zwet [22] to compare the shape of distributions with respect to skewness properties. The idea is roughly as follows. Let X and Y be random variables having, say, absolutely continuous distributions given by distribution functions F and G, respectively. Then $G^{-1}(F(X))$ has the same law as Y. Convexity of $x \mapsto G^{-1}(F(x))$ implies that the transformed distribution tends to be spread out on the right tail while being compressing on the left tail. In other words, Y will have a distribution more skewed to the right. Indeed, if ψ is an increasing function then $X \leq_c \psi(X)$ if and only if ψ is convex.

In the reliability literature the convex transform ordering is known as the *increasing failure rate* (IFR) order. Indeed, assuming that F and G are absolutely continuous distribution functions with derivatives f and g and failure rates $r_F = f/(1-F)$ and $r_G = g/(1-G)$ then $F \leq_c G$ is equivalent to

$$\frac{f(F^{-1}(u))}{g(G^{-1}(u))} = \frac{r_F(F^{-1}(u))}{r_G(G^{-1}(u))}$$

being increasing in $u \in [0,1]$.

The second order of interest is defined analogously to the convex transform order, but now with respect to star-shapedness.

Definition 5 (Star-shaped order $\leq_*$)**.** Let P and Q be two probability distributions on the real line supported by the intervals $I = [0,a]$ and $J = [0,b]$, for some $a, b > 0$, with strictly increasing distributions functions $F\colon I \to [0,1]$ and $G\colon J \to [0,1]$, respectively. We say that $P \leq_* Q$ or, equivalently, $F \leq_* G$, if the mapping $x \mapsto G^{-1}(F(x))$ is star-shaped. Moreover, if $X \sim P$ and $Y \sim Q$, we will also write $X \leq_* Y$ when $P \leq_* Q$.

If $X \sim P$ and $Y \sim Q$ for appropriate P and Q then $X \leq_* Y$ and $Y \leq_* X$ if and only if there exists an $a > 0$ such that X has the same distribution as aY.

The star transform order can be interpreted in terms of the average failure rate, which is why it is sometimes known as *increasing failure rate in average* (IFRA) order. In fact, $F \leq_* G$ is equivalent to $G^{-1}(u)/F^{-1}(u)$ being increasing in $u \in [0,1]$. Moreover

$$\frac{G^{-1}(x)}{F^{-1}(x)} = \frac{\overline{r}_F(F^{-1}(u))}{\overline{r}_G(G^{-1}(u))},$$

where $\overline{r}_F(x)$ and $\overline{r}_G(x)$ are known as the failure rates in average of F and G, respectively, and are defined by $\overline{r}_F(x) = -\ln(1-F(x))/x$ and $\overline{r}_G(x) = -\ln(1-G(x))/x$.

While the star-shaped order is strictly weaker than the convex transform order for distributions having support with a lower end-point at 0, such as the Beta distributions, it is of some independent interest and a useful intermediate step in proving ordering according to the convex transform order.

The stochastic dominance order is also known also as first stochastic dominance (FSD) in reliability theory, and captures the notion of one distribution attaining larger values than the other. It is generally easier to verify than the convex transform order or star-shaped order and will serve here primarily to establish necessity of the sufficient conditions for convex transform ordering between two Beta distributions.

Definition 6 (Stochastic dominance $\leq_{\mathrm{st}}$)**.** Let P and Q be two probability distributions on the real line with distributions functions $F\colon \mathbb{R} \to [0,1]$ and $G\colon \mathbb{R} \to [0,1]$, respectively. We say that $P \leq_{\mathrm{st}} Q$ or, equivalently, $F \leq_{\mathrm{st}} G$, if $F(x) \geq G(x)$, for all $x \in \mathbb{R}$. Moreover, if $X \sim P$ and $Y \sim Q$, we will also write $X \leq_{\mathrm{st}} Y$ when $P \leq_{\mathrm{st}} Q$.

3. Main results

The main results of this paper describe the stochastic dominance, star-shape transform order, and convex transform order relationships within the family of Beta distributions. The proofs are postponed until Section 5.

The stochastic dominance relationships within the family of Beta distributions are known and fairly straightforward to establish. In this paper they will serve to establish necessity of the sufficient conditions for being ordered according to the convex- or star-shaped transform order.

Theorem 7. *Let* $X \sim \mathrm{Beta}(a,b)$ *and* $Y \sim \mathrm{Beta}(a',b')$*, then* $Y \leq_{\mathrm{st}} X$ *if and only if* $a \geq a'$ *and* $b \leq b'$*.*

The star-shape ordering relationships within the family of Beta distributions has been addressed previously by Jeon et al. [11, Example 4], but only for the case of integer valued parameters that satisfy certain conditions. Here we extends this to a complete classification.

Theorem 8. *Let* $X \sim \mathrm{Beta}(a,b)$ *and* $Y \sim \mathrm{Beta}(a',b')$*, then* $X \leq_{*} Y$ *if and only if* $a \geq a'$ *and* $b \leq b'$*.*

As mentioned above, the star-shape transform order is of some independent interest and it serves as an intermediate step when establishing the ordering according to the convex transform order. It turns out to be the case that two Beta distributions are ordered according to the convex transform order if and only if they are ordered according to the star-shaped order.

Theorem 9. *Let* $X \sim \mathrm{Beta}(a,b)$ *and* $Y \sim \mathrm{Beta}(a',b')$*, then* $X \leq_{c} Y$ *if and only if* $a \geq a'$ *and* $b \leq b'$*.*

4. Some consequences of the main results

A first simple result follows from the invariance of the convex ordering under affine transformations. Recall that the family of Gamma distribution with parameters $\alpha, \theta > 0$, denoted $\mathrm{Gamma}(\alpha, \theta)$, is defined by the density functions given for $x \geq 0$ by

$$f(x) = \frac{x^{\alpha-1} e^{-x/\theta}}{\theta^{\alpha}\, \Gamma(\alpha)}, \qquad \text{where } \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx.$$

It is easily seen that if $X_b \sim \mathrm{Beta}(a, b)$ for $a, b > 0$, then, leaving a fixed, the distributions of bX_b converge weakly to $\mathrm{Gamma}(a, 1)$ as b tends to $+\infty$. The following proposition is an immediate consequence of the transitivity of the transform orders and Theorems 8 and 9.

Proposition 10. *Let $X \sim \mathrm{Beta}(a, b)$ and $Y \sim \mathrm{Gamma}(a, \theta)$ for $a, b, \theta > 0$, then $X \leq_* Y$ and $X \leq_c Y$.*

4.1. Probabilities of exceedance. It was noted already by van Zwet [22] that the probabilities of random variables being greater than (or smaller than) their expected values is monotone with respect to convex transform ordering of their distributions. This is essentially a immediate consequence of Jensen's inequality. The idea generalises directly to any functional that satisfies a Jensen-type inequality.

Theorem 11. *For any interval I, measurable function $h \colon I \to \mathbb{R}$ and $X \sim P$ with P supported in I denote the distribution of $h(X)$ by P_h.*

Let $\mathcal{F}$ be a set of continuous probability distributions on intervals in $\mathbb{R}$ and $T \colon \mathcal{F} \to \mathbb{R}$ a functional satisfying for all $P \in \mathcal{F}$ and h convex and increasing with $P_h \in \mathcal{F}$ that $h(T(P)) \leq T(P_h)$.

Then if $X \sim P$ and $Y \sim Q$ with distributions $P, Q \in \mathcal{F}$ such that $X \leq_c Y$ it holds that $\mathbb{P}(X \geq T(P)) \geq \mathbb{P}(Y \geq T(Q))$.

If T satisfies instead $h(T(P)) \geq T(P_h)$ then, under the same assumptions on X and Y, the conclusion becomes $\mathbb{P}(X \geq T(P)) \leq \mathbb{P}(Y \geq T(Q))$.

Proof: Assume T satisfies the first inequality, $h(T(P)) \leq T(P_h)$. Let F and G be the distribution functions of X and Y, respectively, and $h(x) = G^{-1}(F(x))$. Clearly h is increasing. The assumption $X \leq_c Y$ implies h is also convex so that $G^{-1}(F(T(P))) = h(T(P)) \leq T(P_h) = T(Q)$. Since G is increasing it follows that $F(T(P)) \leq G(T(Q))$.

The second statement, for T satisfying $h(T(P)) \leq T(P_h)$, follows by reproducing the same argument with the inequality reversed. ■

Note now that the standard Jensen inequality implies that we may take as T in Theorem 11 the expectation operator $T(P) = \mathbb{E}(X)$ for $X \sim P$. Hence the following is immediate.

Corollary 12. *Let X and Y be two random variables such that $X \leq_c Y$. Then $\mathbb{P}(X \geq \mathbb{E}(X)) \geq \mathbb{P}(Y \geq \mathbb{E}(Y))$.*

As a direct consequence of Theorem 9 and Corollary 12, we have the following monotonicity properties of Beta distributed random variables exceeding their expectation.

Corollary 13. *For each $a, b > 0$ let $X_{a,b} \sim \mathrm{Beta}(a, b)$. Then $(a, b) \mapsto \mathbb{P}(X_{a,b} \geq \mathbb{E}(X_{a,b}))$ is increasing in a and decreasing in b.*

This provides immediate bounds for the probabilities of Beta distributed random variables exceeding their expectation.

Corollary 14. *Let $X_{a,b} \sim \mathrm{Beta}(a, b)$, where $a, b \geq 1$. Then*

$$e^{-1} < \left(\frac{b}{1+b}\right)^b \leq \mathbb{P}(X_{a,b} \geq \mathbb{E}(X_{a,b})) \leq 1 - \left(\frac{a}{1+a}\right)^a < 1 - e^{-1}.$$

Proof: Compute $\mathbb{P}(X_{a,b} \geq \mathbb{E}(X_{a,b}))$ for $a = 1$ or $b = 1$, use the monotonicity given in Corollary 13, and, finally allow $a, b \to +\infty$ to find both numerical bounds. ■

Using Theorem 11 we may prove similar monotonicity properties for the probabilities of exceeding modes or anti-modes. Recall that an absolutely continuous distribution is unimodal if it has a continuous density with a unique maximum and uniantimodal if it has a continuous density with a unique minimum.

Corollary 15. *Let $X \sim P$ and $Y \sim Q$ be two real valued random variables with absolutely continuous distributions P and Q supported on some intervals I and J and such that $X \leq_c Y$.*

If P and Q are unimodal with modes $\mathrm{mode}(X)$ and $\mathrm{mode}(Y)$, respectively, then $\mathbb{P}(X \geq \mathrm{mode}(X)) \leq \mathbb{P}(Y \geq \mathrm{mode}(Y))$.

If P and Q are uniantimodal with anti-modes $\text{anti-mode}(X)$ and $\text{anti-mode}(Y)$, respectively, then

$$\mathbb{P}(X \geq \text{anti-mode}(X)) \geq \mathbb{P}(Y \geq \text{anti-mode}(Y)).$$

Proof: We prove only the result for modes, the statement about anti-modes being proved analogously. To prove the statement we shall rely on Theorem 11.

Define $\mathcal{F}$ as the set of absolutely continuous unimodal distributions supported in some interval in $\mathbb{R}$ and $T\colon \mathcal{F} \to \mathbb{R}$ the functional defined by $T(P)$ being equal to the unique mode of P, for every $P \in \mathcal{F}$. By Theorem 11 it suffices the prove that T satisfies $h(T(P)) \leq T(P_h)$, for every $P \in \mathcal{F}$, and h convex and increasing such that $P_h \in \mathcal{F}$. For this purpose, choose f a continuous and unimodal version of the density of P, and denote, for notational simplicity, the unique mode by m. Simple computation yields that $g(x) = \frac{f(h^{-1}(x))}{h'(h^{-1}(x))}$ is a density for P_h. Since P_h has some continuous density with a unique mode and h is increasing and convex, g must be a such a density. Denote the mode $T(P_h)$ by m_h.

Since m is a mode of P it follows that $f(m) \geq f(h^{-1}(m'))$ and, by the unimodality of P_h, it follows that

$$\frac{f(h^{-1}(m'))}{h'(h^{-1}(m'))} = g(m') \geq g(h(m)) = \frac{f(h^{-1}(h(m)))}{h'(h^{-1}(h(m)))} = \frac{f(m)}{h'(m)}.$$

Consequently $h'(h^{-1}(m')) \leq h'(m)$, which in turn implies that $m' \leq h(m)$, since h' and h are both increasing. The conclusion now follows immediately from Theorem 11. ■

Similarly to Corollary 13, the previous result implies monotonicity properties for the probability of exceeding the mode or anti-mode for Beta distributions. For that, we must restrict ourselves to parameters a and b such that $\mathrm{Beta}(a, b)$ actually has a unique mode or anti-mode, that is to say, when $a, b > 1$ or $a, b < 1$, respectively. In either case the mode or anti-mode, respectively, is $(a-1)/(a+b-2)$.

Corollary 16. *For $a, b > 0$ let $X_{a,b} \sim \mathrm{Beta}(a, b)$.*

If $a, b > 1$ let $\mathrm{mode}(X_{a,b}) = (a-1)/(a+b-2)$, then the mapping $(a, b) \mapsto \mathbb{P}(X_{a,b} > \mathrm{mode}(X_{a,b}))$ is decreasing in a and increasing in b.

If $a, b < 1$ let $\text{anti-mode}(X_{a,b}) = (a-1)/(a+b-2)$, then the mapping $(a, b) \mapsto \mathbb{P}(X_{a,b} > \text{anti-mode}(X_{a,b}))$ is increasing in a and decreasing in b.

Recall that $B \sim \mathrm{Bin}(n, p)$ if $\mathbb{P}(B = k) = \binom{n}{k} p^k (1-p)^{n-k}$ or equivalently if it is a sum of n independent and equally distributed indicators that are equal to 1 with probability p. Using a link between the Beta and the Binomial distributions allows to prove some monotonicity properties for the probabilities that a Binomial variables exceeds certain values close to its mean. As noted in the Introduction, the quantity $\mathbb{P}(B_{n,p} \leq np)$, where $B_{n,p} \sim \text{Bin}(n,p)$ has garnered some interest recently. The map $p \mapsto \mathbb{P}(B_{n,p} \leq np)$ is not monotone even when restricting to $p = 0, \frac{1}{n}, \ldots, 1 - \frac{1}{n}, 1$, where np is an integer. Using our results we prove that slightly changing np renders monotonicity.

Corollary 17. *For $n \in \mathbb{N}$ and for each $p \in [0,1]$ let $B_{n,p} \sim \text{Bin}(n,p)$. The map $p \mapsto \mathbb{P}(B_{n,p} > np - p)$ is increasing for $p = 1/(n-1), \ldots, (n-2)/(n-1)$, and the map $p \mapsto \mathbb{P}(B_{n,p} > np - (1-p))$ is decreasing for $p = 1/(n+1), \ldots, n/(n+1)$.*

Proof: For each $a, b > 0$ let $X_{a,b} \sim \text{Beta}(a,b)$. It is well known that $\mathbb{P}(X_{k+1,n-k} \geq p) = \mathbb{P}(B_{n,p} \leq k)$, for $k = 0, \ldots, n$. The equality can for example be established by repeated integration by parts. As the distribution of $X_{k+1,n-k}$ has mean $(k+1)/(n+1)$ and mode $k/(n-1)$, it follows from Corollaries 13 and 16, that $k \mapsto \mathbb{P}(B_{n,\frac{k+1}{n+1}} \geq k)$ is decreasing and $k \mapsto \mathbb{P}(B_{n,\frac{k}{n-1}} \geq k)$ is increasing. Reparameterising in terms of p yields $k = np + p - 1$ and $k = np - p$, so the result follows. ∎

4.2. (Anti)mode-median-mean inequalities. If $X_{a,b} \sim \text{Beta}(a,b)$ then the random variable $1 - X_{a,b}$ is distributed according to $\text{Beta}(b,a)$. As the convex transform order is invariant with respect to translations, Theorem 9 implies that when $a \leq b$ we have that $-X_{a,b} \leq_c X_{a,b}$. Since the convex transform order orders only the underlying distribution the following definition due to van Zwet [23] is justified.

Definition 18 (Positive/negative skew)**.** Let P be a probability distribution and $X \sim P$ a random variable with distribution P. We say that P is *positively skewed* if $-X \leq_c X$ and that P is *negatively skewed* if $X \leq_c -X$.

Thus, according to this definition, the Beta distributions have positive skew when $a \leq b$ and negative skew when $a \geq b$.

As noted by van Zwet [23] Definition 18 provides an intuitive condition for inequalities between the mode, median and mean to hold. We give an alternative proof of this fact, based on the results in the previous section. This alternative proof yields a similar inequality for the anti-mode.

Theorem 19. *Let P be a positively skewed distribution.*

If P is unimodal with mode m_0, then there exists a median m_1 of P such that $m_0 \leq m_1$.

If P has finite mean m_2, then there exists a median m_1 of P such that $m_1 \leq m_2$.

If P is uniantimodal with anti-mode m_3, then there exists a median m_1 of P such that $m_1 \leq m_3$.

Proof: We prove only the first statement as the remaining ones are proved analogously. Let X be a random variable with distribution P and m_0 the mode of P. Then $m_1 = \sup\{m \mid \mathbb{P}(X \leq m) \leq 1/2\}$ is a median of P. Since P is positively skewed it follows by Corollary 15 that $\mathbb{P}(X \leq m_0) \leq \mathbb{P}(-X \leq -m_0)$. Moreover, $\mathbb{P}(-X \leq -m_0) = 1 - \mathbb{P}(X \leq m_0)$, so that $\mathbb{P}(X \leq m_0) \leq 1/2$. Therefore $m_0 \leq m_1$.

For the second statement apply Corollary 12 instead of Corollary 15. ■

Having a median lying between the mode and mean is usually called satisfying the *mode-median-mean inequality.* Analogously we will say that a distribution satisfies the *median-anti-mode* inequality if it has a median smaller than its anti-mode.

As already noted before, when $a \leq b$, the distribution $\mathrm{Beta}(a, b)$ is positively skewed. The following slight generalisation of the known result concerning the ordering of the mode, median, and mean of the Beta distribution is now immediate.

Corollary 20. *If $1 \leq a \leq b$ then $\mathrm{Beta}(a, b)$ satisfies the mode-median-mean inequality. If $a \leq b \leq 1$ then $\mathrm{Beta}(a, b)$ satisfies the median-mean and median-anti-mode inequalities.*

5. Proofs

This section collects all the proofs related to establishing Theorems 7, 8 and 9, stated in Section 3. To improve the readability, results relevant to each Theorem are presented in a separate subsection.

Most of the proofs rely on keeping track of sign changes of various functions. Throughout $S(x \in I \mapsto f(x)) = S(x \mapsto f(x)) = S(f(x)) = S(f) \in \mathcal{S} = \{\mathtt{0}, \mathtt{-}, \mathtt{+}, \mathtt{-+}, \mathtt{+-}, \dots\}$ denotes the sequence of signs of a function $f\colon I \to \mathbb{R}$. Formal definitions, notation, and standard results concerning sign patterns can be found in Appendix A.

The following technical lemma summarises the basic strategy used throughout the proofs of the main results in the upcoming section

Lemma 21. *For $a, b, a', b', c > 0$ and $d < 1$ denote by F and G the distribution functions of $\mathrm{Beta}(a, b)$ and $\mathrm{Beta}(a', b')$ and $\ell(x) = cx + d$.*

Then for $I = \{x \in [0,1] \mid 0 < \ell(x) < 1\} = (\max(0, -\frac{d}{c}), \min(1, \frac{1-d}{c}))$ *one has*

$$\begin{aligned} S(x \in [0,1] \mapsto F(x) - G(\ell(x))) &= S(x \in I \mapsto F(x) - G(\ell(x))) && (2)\\ &\le \sigma_1 \cdot S(x \in I \mapsto p_1(x)) && (3)\\ &\le \sigma_1 \cdot S(x \in I \mapsto p_2(x)) && (4)\\ &\le \sigma_1 \cdot \sigma_2 \cdot S(x \in I \mapsto p_3(x)) && (5)\\ &\le \sigma_1 \cdot \sigma_2 \cdot S(x \in I \mapsto p_4(x)), && (6) \end{aligned}$$

where

$$\sigma_1 = \mathrm{Sign}(-d), \qquad \sigma_2 = \begin{cases} \mathrm{Sign}(a' - a), & \textit{if } d = 0, a' \neq a, \\ \mathrm{Sign}(1 - a), & \textit{if } d > 0, a \neq 1, \\ \mathrm{Sign}(a' - 1), & \textit{if } d < 0, a' \neq 1, \\ \texttt{0}, \texttt{-}, \textit{ or } \texttt{+}, & \textit{otherwise}, \end{cases}$$

and

$$\begin{aligned} p_1(x) &= \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} - c\frac{\ell(x)^{a'-1}(1-\ell(x))^{b'-1}}{\mathrm{B}(a',b')}, \\ p_2(x) &= (a-1)\log(x) + (b-1)\log(1-x) \\ &\quad - (a'-1)\log(\ell(x)) - (b'-1)\log(1-\ell(x)) + C, \\ p_3(x) &= \frac{a-1}{x} - \frac{b-1}{1-x} - \frac{c(a'-1)}{\ell(x)} + \frac{c(b-1)}{1-\ell(x)}, \\ p_4(x) &= c_3x^3 + c_2x^2 + c_1x + c_0, \end{aligned}$$

for $c_3 = (a-a'+b-b')c^2$, $c_2 = -(a-a'+1-b')c^2 - (a-a'+b-1)c(1-d) - (b'-b+1-a)cd$, $c_1 = (a-a')c(1-d) + (a-b')c(-d) + (a+b-2)(1-d)(-d)$, $c_0 = -(a-1)(d-1)d$, *and* $C = \log\frac{\mathrm{B}(a',b')}{c\,\mathrm{B}(a,b)}$.

Proof: Write $[0,1] = J \cup I \cup J'$ where $J = [0, \max(0, -\frac{d}{c})]$ and $J' = [\min(1, \frac{1-d}{c}), 1]$. Then $S(x \in [0,1] \mapsto F(x) - G(\ell(x))) = S(x \in J \mapsto F(x) - G(\ell(x))) \cdot S(x \in I \mapsto F(x) - G(\ell(x))) \cdot S(x \in J' \mapsto F(x) - G(\ell(x)))$. By construction the first and third terms are just a single sign that coincides with the first and final sign of $S(x \in I \mapsto F(x) - G(\ell(x)))$ and can hence be dropped. This proves (2).

Assertion (3) is now immediate from Propositions 39 and 40 and (4) follows by taking logarithms of both terms and simplifying.

Moreover, (5) follows by another application of Propositions 39 and 40 and (6) follows by multiplication with $x(1-x)\ell(x)(1-\ell(x))$ which is positive on I by definition. ■

5.1. Stochastic dominance ordering. Before actually proving Theorem 7, we shall prove that the stochastic dominance is a *necessary* condition for ordering compactly supported distributions with respect to the star-shape transform or the convex transform orders. Although the result concerning the stochastic dominance is well established, we present a proof using sign patterns.

A first result concerns a simple relation between the star-shaped transform ordering and the stochastic dominance order.

Proposition 22. *Let $X \sim P$ and $Y \sim Q$ be random variables with distributions P and Q supported on $[0,1]$. Then $X \leq_* Y$ implies $Y \leq_{\mathrm{st}} X$.*

Proof: Let F and G be the distribution functions of X and Y, respectively. As $G^{-1}(F(x))/x$ is increasing, it follows that $G^{-1}(F(x))/x \leq G^{-1}(F(1)) = 1$, thus $G^{-1}(F(x)) \leq x$ and $G(x) \geq F(x)$, meaning $Y \leq_{\mathrm{st}} X$. ■

Since the convex transform order implies the star-shape transform order, the following is immediate.

Corollary 23. *Let $X \sim P$ and $Y \sim Q$ be random variables with distributions P and Q supported on $[0,1]$. Then $X \leq_c Y$ implies $Y \leq_{\mathrm{st}} X$.*

In the above statement the use of the unit interval is for notational convenience. Using invariance under orientation preserving affine transformations the statement generalises to distributions on any bounded interval.

Using the above we may now establish *necessary* conditions for one Beta distribution to be smaller than another according to convex- or star-shaped transform orders. We do this by characterising when one is smaller than the other according to stochastic dominance.

The proof is elementary, but since it illustrates well the style of the upcoming proofs we formulate it in terms of an analysis of sign patterns.

Proof of Theorem 7: Let F, G, f, and g be the distribution and density functions of $\mathrm{Beta}(a,b)$ and $\mathrm{Beta}(a',b')$. Denote $H(x) = F(x) - G(x)$. We need to prove that $S(H) = $ - if and only if $a \geq a'$ and $b \leq b'$. We have

$$S(H'(x)) = S\left(\frac{x^{a'-1}(1-x)^{b'-1}}{\mathrm{B}(a,b)}\left(x^{a-a'}(1-x)^{b-b'} - \frac{\mathrm{B}(a,b)}{\mathrm{B}(a',b')}\right)\right)$$
$$= S\left(x^{a-a'}(1-x)^{b-b'} - \frac{\mathrm{B}(a,b)}{\mathrm{B}(a',b')}\right).$$

Since the case $a = a'$ and $b = b'$ is trivial, we may assume H is not constant 0 and so, since $H(0) = H(1) = 0$, that neither $S(H') = $ + nor $S(H') = $ -.

If $a \geq a'$ and $b \leq b'$, with at least one strict, we have $S(H') \leq$ -+ since $x^{a-a'}(1-x)^{b-b'}$ is increasing. Only $S(H') = $ -+ is possible so Propositions 39 and 40 imply $- \cdots - = S(H) \leq$ -+ with $S(H) = $ - the only option.

Assume now that $b > b'$ and $a > a'$. Clearly $S(H') = - \cdots -$ and since $x^{a-a'}(1-x)^{b-b'}$ is unimodal either $S(H') = $ - or $S(H') = $ -+-. Only $S(H') = $ -+- is possible, so Proposition 40 implies that $S(H) = - \cdots + \neq$ -.

Using that $\mathrm{Beta}(a,b) \leq_{\mathrm{st}} \mathrm{Beta}(a',b')$ if and only if $\mathrm{Beta}(b',a') \leq_{\mathrm{st}} \mathrm{Beta}(a,b)$ and that $\leq_{\mathrm{st}}$ is a partial order covers the remaining cases. ■

5.2. Star-shape ordering. We now prove Theorem 8, showing that, apart reversing the order direction, we find the same parameter characterisations as for the stochastic dominance.

Proof of Theorem 8: The necessity follows from Proposition 22 and Theorem 7. As for the sufficiency, it is enough to prove the statement when $a > a'$, $b = b'$ and when $a = a'$, $b < b'$. The general statement then follows by transitivity since then $\mathrm{Beta}(a,b) \leq_* \mathrm{Beta}(a,b') \leq_* \mathrm{Beta}(a',b')$. Moreover, we may assume that either $b \leq b' \leq 1$ or $1 \leq b \leq b'$. Since the remaining case, $b \leq 1 \leq b'$, follows again by transitivity.

Let F and G be the distribution functions of $\mathrm{Beta}(a,b)$ and $\mathrm{Beta}(a',b')$, respectively, with f and g the corresponding density functions as in (1). By Proposition 38 we need to prove that for every $c > 0$

$$S(x \in [0,1] \mapsto G^{-1}(F(x)) - cx) = S(F(x) - G(cx)) \leq \text{-+}. \tag{7}$$

As the assumptions on the parameters are the same as in Theorem 7, it follows that $G^{-1}(F(x)) \leq x$, meaning (7) is trivially satisfied when $c \geq 1$. Moreover, both G^{-1} and F are increasing, so (7) is again trivial for $c \leq 0$. It is therefore enough to consider $c \in (0,1)$.

The conclusion follows by analysing three different cases.

Case 1. $b = b' \leq 1$, $a > a'$**:** Using (3) from Lemma 21 with $d = 0$ gives

$$\begin{aligned} S(G^{-1}(F(x)) - cx) &\leq S\left(\frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} - \frac{c^{a'}x^{a'-1}(1-cx)^{b-1}}{\mathrm{B}(a',b)}\right) \\ &= S\left(x^{a-a'}\left(\frac{1-cx}{1-x}\right)^{1-b} - c^{a'}\frac{\mathrm{B}(a,b)}{\mathrm{B}(a',b)}\right) \\ &\leq \texttt{-+}, \end{aligned}$$

since $\frac{1-cx}{1-x}$ is increasing for $a \geq a'$ and $b \leq 1$.

Case 2. $b = b' \geq 1$, $a > a'$**:** Applying Lemma 21 with $d = 0$ we have $c_3 = (a-a')c^2 > 0$, $c_2 = -(b-1)c(1-c) - (a-a')c(1+c)$, $c_1 = (a-a')c > 0$, $c_0 = 0$, $\sigma_1 = \texttt{0}$, and $\sigma_2 = \texttt{+}$, meaning (6) gives

$$S(F(x) - G(cx)) \leq \texttt{+} \cdot S(c_3x^3 + c_2x^2 + c_1x) = \texttt{+} \cdot S(c_3x^2 + c_2x + c_1).$$

Since $c_3 > 0$ we have $S(x \in [0,1] \mapsto c_2x^2 + c_1x + c_0) \leq \texttt{+-+}$. But $c_1 > 0$ and $c_3 + c_2 + c_1 = -(b-1)c(1-c) < 0$ meaning we must have $S(x \in [0,1] \mapsto c_2x^2 + c_1x + c_0) = \texttt{+-}$. Hence $S(F(x) - G(cx)) \leq \texttt{-+-}$. But since $F(1) - G(c) = 1 - G(c) > 0$ we have $S(F(x) - G(cx)) \leq \texttt{-+}$.

Case 3. $a = a'$, $b < b'$**:** Applying Lemma 21 with $d = 0$ we have $c_3 = -(b'-b)c^2 < 0$, $c_2 = (1-b)c - (1-b')c^2$, $c_1 = 0$, $c_0 = 0$, $\sigma_1 = \texttt{0}$, and $\sigma_2 \in \{\texttt{0}, \texttt{-}, \texttt{+}\}$, meaning (6) gives

$$S(F(x) - G(cx)) \leq \sigma_2 \cdot S(c_3x^3 + c_2x^2) = \sigma_2 \cdot S(c_3x + c_2) \leq \sigma_2 \cdot \texttt{+-}.$$

In any case $S(F(x) - G(cx)) \leq \texttt{-+-}$ no matter the value of σ_2. But $F(1) - G(c) > 0$, so we must have $S(F(x) - G(cx)) \leq \texttt{-+}$.

This concludes the proof. ■

As will become apparent in the next section, this characterisation of the star-shape transform ordering is an essential first step towards proving the corresponding statement for the convex transform order.

5.3. Convex transform ordering. To characterise how the Beta distributions are ordered according to the convex transform order we will apply a strategy similar to the one used in previous sections. According to Proposition 38 we need to prove that for $a \geq a'$ and $b \leq b'$ the distribution functions F and G of $\mathrm{Beta}(a,b)$ and $\mathrm{Beta}(a',b')$, respectively, satisfy

$$S(x \in [0,1] \mapsto F(x) - G(\ell(x))) \leq \texttt{+-+}, \tag{8}$$

for every affine function ℓ with positive slope.

First we need an auxiliary result, generalising Theorem 6.1 in [4], which corresponds to taking $x_0 = y_0 = 0 = \inf I$ in the statement below.

Proposition 24. *Let $f\colon I \mapsto \mathbb{R}$ where I is an interval. If for some $x_0 \leq \inf I$ and y_0 it holds that $S(x \in I \mapsto f(x) - \ell(x)) \leq$ -+ for all affine functions ℓ such that $\ell(x_0) = y_0$ then $S(x \in I \mapsto f(x) - \tilde{\ell}(x)) \leq$ -+ for all affine functions $\tilde{\ell}$ such that $\tilde{\ell}(x_0) \geq y_0$.*

The analogous conclusion holds considering the sign pattern +- and taking $x_0 \geq \sup I$ satisfying $\tilde{\ell}(x_0) \leq y_0$.

Proof: The proof is not too difficult. The main idea is given graphically in Figure 1. ■

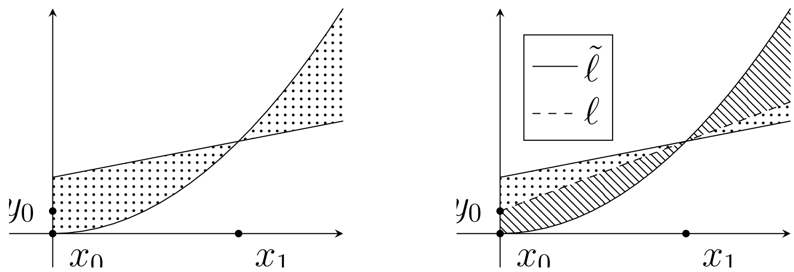

FIGURE 1. Main idea of proof of Proposition 24.

The above statement may be combined with the characterisation of how Beta distributions are ordered according to the star-shaped transform order that was established in the previous section. Doing so allows us to immediately take care of a number of affine ℓ in (8).

Corollary 25. *Let F and G be the distribution functions of* $\mathrm{Beta}(a, b)$ *and* $\mathrm{Beta}(a', b')$*, respectively, and assume that $a \geq a'$ and $b \leq b'$. If ℓ is any affine function satisfying $\ell(0) \geq 0$ or $\ell(1) \notin (0, 1)$ then* (8) *is satisfied.*

Proof: We analyze three different cases.

- **Case 1.** $\ell(0) \geq 0$**:** According to Theorem 8 and Proposition 38, we have that, for any $c \in \mathbb{R}$, $S(F(x) - G(cx)) \leq$ -+. Taking into account Proposition 24, this implies $S(F(x) - G(\ell(x))) \leq$ -+ $\leq$ +-+.
- **Case 2.** $\ell(1) \leq 0$**:** In this case ℓ is always negative, and the result is immediate.

Case 3. $\ell(1) \geq 0$**:** It enough to prove that $S(F(x) - G(c(x-1)+1)) \leq$ +- for every $c \in \mathbb{R}$. Indeed, once this proved, the conclusion follows using Proposition 24 again.

Note that $F_-(x) = 1 - F(1-x)$ is the distribution function of $\text{Beta}(b, a)$ and $G_-(x) = 1 - G(1-x)$ is the distribution function of $\text{Beta}(b', a')$. The characterisation of star-shape transform order proved in Theorem 8 together with Proposition 38, means that $S(1 - G(1 - x) - 1 + F(1 - c'x)) \leq$ -+, for every $c' \in \mathbb{R}$. For any $c \in \mathbb{R}$ we may apply this to $c' = 1/c$, which gives

$$\begin{aligned} S(F(x) - G(c(x-1)+1)) &= S(1 - G(c(x-1)+1) - 1 + F(x)) \\ &= \mathbf{rev}\, S(1 - G(1-x) - 1 + F(1 - x/c)) \\ &\leq \mathbf{rev}(\texttt{-+}) = \texttt{+-}. \end{aligned}$$

■

The proof of Theorem 9, establishing the convex transform ordering within the Beta family is achieved through the analysis of several partial cases. For improved readability we will be presenting these in several lemmas.

Lemma 26. *Let $X \sim \text{Beta}(a, b)$ and $Y \sim \text{Beta}(1, b)$, with $a \geq 1$ and $b > 0$. Then $X \leq_c Y$.*

Proof: Let F and G be the distribution functions of the $\text{Beta}(a, b)$ and $\text{Beta}(1, b)$ distributions, respectively, and f and g their densities. Taking into account Proposition 38 and Corollary 25, we need to show that, for every affine function $\ell(x) = cx + d$ satisfying $\ell(0) = d < 0$ and $\ell(1) = c + d \in (0, 1)$ one has that (8) is satisfied. We need to separate the arguments into three cases.

Case 1. $b = 1$**:** In this case the statement follows directly from the convexity of $F(x) = x^a$ and that $G(x) = x$.

Case 2. $b \in (0, 1)$**:** Applying Lemma 21 we have $I = (-d/x, 1)$, $\sigma_1 = \texttt{+}$, and $\sigma_2 \in \{\texttt{0}, \texttt{-}, \texttt{+}\}$, meaning (5) gives

$$S(F(x) - G(cx)) \leq \texttt{+} \cdot \sigma_2 \cdot S\left(\frac{a-1}{x} - \frac{b-1}{1-x} + \frac{c(b-1)}{1-\ell(x)}\right).$$

But since for $x \in I$

$$\frac{a-1}{x} - \frac{b-1}{1-x} + \frac{c(b-1)}{1-\ell(x)} = \frac{a-1}{x} + \frac{(1-b)(1-(c+d))}{(1-x)(1-\ell(x))} > 0,$$

we have $S(F(x) - G(\ell(x))) \leq \texttt{+} \cdot \sigma_2 \cdot \texttt{+} \leq \texttt{+-+}$ no matter the value of σ_2.

Case 3. $b > 1$**:** Applying Lemma 21 we have $\sigma_1 = \mathtt{+}$ and $\sigma_2 \in \{\mathtt{0}, \mathtt{-}, \mathtt{+}\}$, meaning (6) gives

$$\begin{aligned} S(F(x) - G(cx)) &\le \mathtt{+} \cdot \sigma_2 \cdot S(c_3x^3 + c_2x^2 + c_1x + c_0) \\ &= \mathtt{+} \cdot \sigma_2 \cdot S(l(x)(c_2'x^2 + c_1'x + c_0')) \\ &= \mathtt{+} \cdot \sigma_2 \cdot S(c_2'x^2 + c_1'x + c_0') \end{aligned}$$

where $c_2' = (a-1)c$, $c_1' = -(a-b)c - (a+b-2)(1-d)$, and $c_0' = (a-1)(1-d)$. Since $c_2' > 0$ we have $S(c_2'x^2 + c_1'x + c_0') \le \mathtt{+-+}$. On the other hand, $c_2' + c_1' + c_0' = -(b-1)(1-(c+d)) < 0$, hence $S(c_2'x^2 + c_1'x + c_0') \le \mathtt{+-}$. Combining these inequalities yields

$$S(F(x) - G(\ell(x))) \le \mathtt{+}\sigma_2\mathtt{+-} \le \mathtt{+-+-}.$$

Finally, as $F(1) - G(\ell(1)) > 0$, it follows that $S(F(x) - G(\ell(x))) \le \mathtt{+-+}$. So, taking into account Proposition 38, the proof is concluded. ∎

The second lemma is similar in that but covers the case where $a \ge 1$.

Lemma 27. *Let* $X \sim \mathrm{Beta}(1, b)$ *and* $Y \sim \mathrm{Beta}(a, b)$ *with* $a \le 1$ *and* $b > 0$. *Then* $X \le_c Y$.

Proof: Let F and G represent the distribution functions of $\mathrm{Beta}(1, b)$ and $\mathrm{Beta}(a, b)$, Note that the meaning of the symbols F and G are interchanged relative to their use in the proof of Lemma 26. Taking into account Proposition 38 and Corollary 25, we need to show that (8) holds for $\ell(x) = cx + d$ such that $\ell(0) = d < 0$ and $\ell(1) = c + d \in (0, 1)$. This is equivalent to $S(G(x) - F(\ell^{-1}(x))) \le \mathtt{-+-}$, where $\ell^{-1}(x) = (x-d)/c$ satisfies $\ell^{-1}(0) \in (0, 1)$ and $\ell^{-1}(1) > 1$.

Reversing the roles of F and G the proof is now analogous to that of Lemma 26 except that $a < 1$ and we wish to establish $S(x \in I \mapsto G(x) - F(\ell^*(x))) \le \mathtt{-+-}$ for $\ell^*(x) = c^*x + d^*$ with $\ell^*(0) = d^* \in (0, 1)$ and $\ell^*(1) = c^* + d^* > 1$ on the interval $I = (0, (1-d^*)/c^*)$. ∎

Comparing the distributions for more general pairs of parameters a and a' requires separate analyses depending on whether $b > 1$ or $b \in (0, 1)$.

Lemma 28. *Let* $X \sim \mathrm{Beta}(a, b)$ *and* $Y \sim \mathrm{Beta}(a', b)$ *with* $a > a'$ *and* $b > 1$. *Then* $X \le_c Y$.

Proof: Let F and G be the distribution functions of $\mathrm{Beta}(a, b)$ and $\mathrm{Beta}(a', b)$. By Proposition 38 and Corollary 25, it is enough to prove that (8) holds when $\ell(x) = cx + d$ is such that $\ell(0) = d < 0$, $\ell(1) = c + d \in (0, 1)$.

Applying Lemma 21 we have $c_3 = (a - a')c^2$, $\sigma_1 = \texttt{+}$, and $\sigma_2 \in \{\texttt{0}, \texttt{-}, \texttt{+}\}$, meaning (6) gives

$$S(F(x) - G(\ell(x))) \leq \texttt{+} \cdot \sigma_2 \cdot S(c_3x^3 + c_2x^2 + c_1x^1 + c_0).$$

Since $c_3 > 0$ we have $S(c_3x^3+c_2x^2+c_1x^1+c_0) \leq \texttt{-+-+}$. But $c_3+c_2+c_1+c_0 = -(b-1)(c+d)(1-(c+d)) < 0$ so $S(c_3x^3 + c_2x^2 + c_1x^1 + c_0) \leq \texttt{-+-}$.

Combining the sign pattern inequalities, we have derived that

$$S(x \in I \mapsto F(x) - G(\ell(x))) \leq \texttt{+} \cdot \sigma_2 \cdot \texttt{-+-} = \texttt{+-+-},$$

regardless of the value of σ_2. Finally $F(1) - G(\ell(1)) > 0$ so we conclude that $S(x \in I \mapsto F(x) - G(\ell(x))) \leq \texttt{+-+}$, hence proving the result. ■

Lemma 29. *Let* $X \sim \mathrm{Beta}(a, b)$ *and* $Y \sim \mathrm{Beta}(a', b)$ *with* $0 < b \leq 1$ *and either* $1 > a > a' > 0$ *or* $a > a' > 1$. *Then* $X \leq_c Y$.

Proof: We may assume, without loss of generality, that $a - a' < 1$. Indeed, if $a - a' \geq 1$, one may choose for sufficiently large N a sequence $a_0 = a, a_1, \ldots, a_N = a'$ such that $a_{i_1} - a_i < 1$ for all $i = 1, \ldots, N$ and apply transitivity to conclude $\mathrm{Beta}(a_0, b) \leq_c \cdots \leq_c \mathrm{Beta}(a_N, b)$. Let F and G be the distribution functions of $\mathrm{Beta}(a, b)$ and $\mathrm{Beta}(a', b)$, respectively. Based on Proposition 38 and Corollary 25, it is enough to prove that (8) holds for every $\ell(x) = cx + d$ such that $\ell(0) = d < 0$ and $\ell(1) = c + d \in (0, 1)$. Using Lemma 21 we have for $I = (-d/c, 1)$ that

$$\begin{aligned} &S(x \in [0,1] \mapsto F(x) - G(\ell(x))) \\ &\quad \leq \texttt{+} \cdot S\left(x \in I \mapsto \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} - \frac{\ell(x)^{a'-1}(1-\ell(x))^{b-1}}{\mathrm{B}(a',b)}\right) \\ &\quad = \texttt{+} \cdot S\left(x \in I \mapsto \frac{x^{a-1}}{\ell(x)^{a'-1}} - \frac{c\,\mathrm{B}(a,b)}{\mathrm{B}(a',b')}\Big(\frac{1-\ell(x)}{1-x}\Big)^{b-1}\right). \end{aligned} \tag{9}$$

For convenience define $C = c\,\mathrm{B}(a, b)/\mathrm{B}(a', b') > 0$, $q_1(x) = x^{a-1}/\ell(x)^{a'-1}$, $q_2(x) = (1 - x)/(1 - \ell(x))$, and $q(x) = q_1(x) - Cq_2(x)^{1-b}$. Restricting to I we have that q_2 is decreasing and concave. Hence, as $b \leq 1$, it follows that $x \in I \mapsto -Cq_2(x)^{1-b}$ is non-decreasing and convex.

A simple computation yields $q_1'(x) = ((a - a')cx + (a - 1)d)/(x^{2-a}\ell(x)^{a'})$ which has unique root at $x_0 = -(a - 1)d/((a - a')c)$. Letting $c_2^* = (a -$

$a')(a-a'-1)c^2$, $c_1^* = 2(a-a'-1)(a-1)cd$, $c_0^* = (a-1)(a-2)d^2$, and $p(x) = c_2^* x^2 + c_1^* x + c_0^*$ we have after a another straight forward computation $q_1''(x) = p(x)/(x^{3-a}\ell(x)^{a'+1})$.

Case 1. $1 > a > a'$**:** In this case $(a-a')cx + (a-1)d > 0$ for $x > 0$, which implies that q_1 is increasing on I. Therefore $x \in I \mapsto q_1(x) - Cq_2(x)^{1-b}$ is increasing, meaning $S(x \in I \mapsto q_1(x) - Cq_2(x)^{1-b}) \leq$ -+. Plugged into (9) this gives $S(F(x) - G(\ell(x))) \leq$ +-+.

Case 2. $a > a' > 1$**:** A direct verification shows that $x_0 \in I$ so that $I_1 = (-d/c, x_0]$ and $I_2 = (x_0, 1)$ are well defined and non-empty. Since $I = I_1 \cup I_2$ and $I_1 < I_2$

$$S(x \in I \mapsto q(x)) = S(x \in I_1 \mapsto q(x)) \cdot S(x \in I_2 \mapsto q(x)).$$

Sign pattern in I_1**:** As $c_2^* = (a-a')(a-a'-1)c^2 < 0$ it follows that $S(x \in I \mapsto q_1''(x)) = S(x \in I \mapsto p(x)) \leq$ -+-. But $p(-d/c) = (a'-1)a'd^2 > 0$ and $p(x_0) = \frac{(a-1)(a'-1)d^2}{a-a'} > 0$ so $S(x \in I \mapsto p(x)) =$ +.

This implies that q_1 is convex in I_1. As we have proved the convexity of $-Cq_2(x)^{1-b}$ in I, it follows that $q(x) = q_1(x) - Cq_2(x)^{1-b}$ is convex in I_1. According to Proposition 37 it follows that

$$S(x \in I_1 \mapsto q(x)) \leq \text{+-+}.$$

Sign pattern in I_2**:** Noting that $S(x \in I \mapsto q_1'(x)) = S((a-a')cx + (a-1)d) \leq$ -+ and $q_1'(x_0) = 0$, it follows that q_1' is positive in I_2. Thus q_1 is increasing in I_2. We have proved above that $x \in I \mapsto -Cq_2(x)^{1-b}$ is increasing, so $q(x)$ is increasing in the interval I_2. Therefore

$$S(x \in I_2 \mapsto q(x)) \leq \text{-+}.$$

If $q(x_0) < 0$ then $S(x \in I_1 \mapsto q(x)) \leq$ +-. If $q(x_0) \geq 0$ then $S(x \in I_2 \mapsto q(x)) =$ + since q is increasing on I_2. In either case $S(x \in I_1 \mapsto q(x)) \cdot S(x \in I_2 \mapsto q(x)) \leq$ +-+.

Putting the above into (9) we have

$$S(x \in I \mapsto F(x) - G(\ell(x))) \leq \text{+} \cdot \text{+-+} = \text{+-+}$$

as required. ■

We now state, without proof, a straightforward result, helpful for the conclusion of the final characterisation within the Beta family.

Proposition 30. *Let $X \sim P$ and $Y \sim Q$ be random variables with some distributions P and Q, then $X \leq_c Y$ if and only if $1 - Y \leq_c 1 - X$.*

We now have all the necessary ingredients to prove the main theorem.

Proof of Theorem 9: The necessity is a direct consequence of Theorem 7. The sufficiency follows from Lemmas 26, 27, 28, 29, and the transitivity of the convex transform order. First note that we obtain

$$\mathrm{Beta}(a, b) \leq_c \mathrm{Beta}(a', b), \tag{10}$$

when $a = a'$ (trivial), $b > 1$ (use Lemma 28), $b \leq 1$ and either $1 > a > a'$ or $a > a' > 1$ (use Lemma 29). The order relation (10) also holds if $b \leq 1$ and $a > 1 > a'$ by combining Lemmas 26 and 27, since then $\mathrm{Beta}(a, b) \leq_c \mathrm{Beta}(1, b) \leq_c \mathrm{Beta}(a', b)$. Using this and Proposition 30 we also have $\mathrm{Beta}(a', b) \leq_c \mathrm{Beta}(a', b')$, concluding the proof. ■

References

[1] Abadir, K. M. (2005). The Mean-Median-Mode Inequality: Counterexamples, Econom. Th. 21, 477–482.

[2] Arab, I. and Oliveira, P. E. (2019). Iterated Failure Rate Monotonicity And Ordering Relations Within Gamma And Weibull Distributions. Probab. Eng. Inform. Sc. 33, 64–80. doi: 10.1017/S0269964817000481.

[3] Arab, I. and Oliveira, P. E. (2018). Iterated Failure Rate Monotonicity And Ordering Relations Within Gamma And Weibull Distributions – Corrigendum. Probab. Eng. Inform. Sc. 32, 640–641. doi: 10.1017/S0269964818000372.

[4] Arab, I., Hadjikyriakou, M., and Oliveira, P. E. (2020). Failure rate properties of parallel systems, Adv. App. Probab. 52 (to appear).

[5] Becchetti, L., Clementi, A. E., Natale, E., Pasquale, F., Silvestri, R., and Trevisan, L. (2017). Simple dynamics for plurality consensus. Distrib. Comput. 30, 293—306.

[6] Barlow, R. E., Marshall, A. W., and Proschan, F. (1969). Some inequalities for starshaped and convex functions. Pacific J. Math. 29, 19—42.

[7] Cortes, C., Mansour, Y., and Mohri, M. (2010). Learning Bounds for Importance Weighting. MIT Press, Vancouver, Canada.

[8] Doerr, B. (2018). An elementary analysis of the probability that a binomial random variable exceeds its expectation. Statist. Probab Letters 139, 67–74.

[9] Greenberg, S., Mohri, M. (2014). Tight lower bound on the probability of a binomial exceeding its expectation. Statist. Probab. Lett. 86, 91—98.

[10] Jacobson, N. (1985). Basic Algebra I. W. H. Freeman and Company, San Francisco.

[11] Jeon, J., Kochar, S. and Park, C. G. (2006). Dispersive ordering - Some applications and examples. Statist. Papers 47, 227–247.

[12] Karppa, M., Kaski, P., and Kohonen, J. (2016). A faster subquadratic algorithm for finding outlier correlations. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 1288—1305.

[13] Marshall, A. W. and Olkin, I. (2007). *Life Distributions.* Springer, New York.

[14] Mitzenmacher, M. and Morgan, T. (2017). Reconciling graphs and sets of sets (2017). CoRR, abs/1707.05867.
[15] Nanda, A. K., Hazra, N. K., Al-Mutairi, D. K. and Ghitany, M. E. (2017). On some generalized ageing orderings. Commun. Stat. Theory Methods 46, 5273–5291. doi: 10.1080/03610926.2015.1100738.
[16] Oja, H. (1981). On Location, Scale, Skewness and Kurtosis of Univariate Distributions. Scand. J. Statist. 8, 154–168.
[17] Pelekis, Ch. (2016). A lower bound on binomial tails: an approach via tail conditional expectations. ArXiv e-prints, 1609.06651.
[18] Pelekis, Ch., and Ramon, J. (2016). A lower bound on the probability that a binomial random variable is exceeding its mean. Statist. Probab. Lett. 119, 305-–309.
[19] Shaked, S. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer, New York.
[20] Wiklund, T.(2018). The deficiency introduced by resampling. Math. Methods Statist. 27, 145—161.
[21] Zheng, S., Mogusu, E., Veeranki, S. P., Quinn, M.m and Cao, Y. (2017).The relationship between the mean, median, and mode with grouped data. Commun. Stat. Theory Methods 46, 4285–4295. doi: 10.1080/03610926.2015.1081948.
[22] van Zwet, W. R. (1964). Convex transformations of random variables. MC Tracts 7, Amsterdam.
[23] van Zwet, W. R. (1979). Mean, median, mode II. Statistica Neerlandica 33, 1–5.

Appendix A. An algebra for sign variation

The main tool of all proofs concerning the ordering within the Beta family is the study of sign patterns of functions. While such techniques have a long tradition in probability theory, for our purposes it turns out to be computationally convenient to give a presentation slightly more algebraic as compared to what appears to be the convention, using a suitable monoid (see, for example, Jacobson [10]).

Definition 31. Let $(\mathcal{S}, \cdot) = \langle \mathtt{+}, \mathtt{-} \mid \mathtt{+} \cdot \mathtt{+} = \mathtt{+}, \mathtt{-} \cdot \mathtt{-} = \mathtt{-} \rangle$ be the monoid generated by two idempotent elements $\mathtt{+}$ and $\mathtt{-}$ and with unit $\mathtt{0}$.

We shall call elements of $\mathcal{S}$ *sign patterns*. When unambiguous we will denote products $\sigma \cdot \sigma'$ by simply juxtaposing the factors as in $\sigma\sigma'$, so that $\mathcal{S} = \{\mathtt{0}, \mathtt{+}, \mathtt{-}, \mathtt{+-}, \mathtt{-+}, \mathtt{+-+}, \dots\}$.

For any $\sigma = \sigma_1 \cdots \sigma_n \in \mathcal{S}$ where $\sigma_1, \dots, \sigma_n \in \{\mathtt{+}, \mathtt{-}\}$ let $\mathbf{rev}\, \sigma = \sigma_n \cdots \sigma_1$ be the sign pattern given by reversing the order of signs and let

$$\overline{\sigma} = \overline{\sigma_0} \cdot \dots \cdot \overline{\sigma_n}, \qquad \text{where} \qquad \overline{\sigma_i} = \begin{cases} \mathtt{+} & \sigma_i = \mathtt{-}, \\ \mathtt{-} & \sigma_i = \mathtt{+}, \end{cases}$$

denote the sign pattern given by flipping the signs. Note in particular that $\overline{\mathtt{0}} = \mathtt{0}$.

Sign patterns have a natural order structure.

Definition 32. Given $\sigma, \sigma' \in \mathcal{S}$ we say that $\sigma \leq \sigma'$ if $\sigma' = \pi \cdot \sigma \cdot \pi'$ for some $\pi, \pi' \in \mathcal{S}$. In other words, if σ is a substring of σ'.

Proposition 33. $(\mathcal{S}, \cdot, \leq)$ *is a partially ordered monoid in the sense that* $(\mathcal{S}, \leq)$ *is a partially ordered set and if* $\sigma, \sigma' \in \mathcal{S}$ *are such that* $\sigma \leq \sigma'$ *then for any* $\pi, \pi' \in \mathcal{S}$ *one has* $\pi \cdot \sigma \cdot \pi' \leq \pi \cdot \sigma' \cdot \pi'$*.*

We can now describe the sign variations of a function in terms of the simple sign function.

Definition 34 (Sign function)**.** The sign function $\mathrm{Sign}\colon \mathbb{R} \to \mathcal{S}$ is defined by $\mathrm{Sign}(x) = \texttt{+}$ if $x > 0$, $\mathrm{Sign}(x) = \texttt{0}$ if $x = 0$, and $\mathrm{Sign}(x) = \texttt{-}$ if $x < 0$.

Definition 35 (Sign patterns and finite sign variation)**.** Given $I \subseteq \mathbb{R}$, we say that a function $f\colon I \mapsto \mathbb{R}$ is of *finite sign variation* if the set

$$\{\mathrm{Sign}(f(x_1)) \cdot \mathrm{Sign}(f(x_2)) \cdot \cdots \cdot \mathrm{Sign}(f(x_n)) \mid n \in \mathbb{N}, x_1 \leq \cdots \leq x_n \in I\}$$

has a (unique) maximal element in $\mathcal{S}$. This maximal element is then denoted by $S(x \in I \mapsto f(x))$ and called the *sign pattern* of f.

When unambiguous, we will abbreviate $S(x \in I \mapsto f(x)) = S(x \mapsto f(x)) = S(f(x)) = S(f)$ and write for readability $\overline{S}(f) = \overline{S(f)}$.

The proposition below gives some standard rules of calculation for sign patterns which are straightforward to prove and used without explicit mention throughout the proofs.

Proposition 36. *Let* $I \subset \mathbb{R}$ *and* $f, g\colon I \to \mathbb{R}$ *be such that* f *and* $f - g$ *are of finite sign variation.*

1. *For any* $J \subset I$ *one has* $S(x \in J \mapsto f(x)) \leq S(x \in I \mapsto f(x))$*.*
2. *For any* $J \leq K$ *such that* $I = J \cup K$ *one has* $S(x \in I \mapsto f(x)) = S(x \in J \mapsto f(x)) \cdot S(x \in K \mapsto f(x))$*.*
3. *For any positive* $h\colon I \to \mathbb{R}$ *one has* $S(f(x)) = S(f(x)h(x))$*.*
4. *For* $J \subset \mathbb{R}$ *and* $\eta\colon J \to I$ *increasing (or decreasing) one has* $S(x \in I \mapsto f(x)) = S(x \in J \mapsto f(\eta(x)))$ *(respectively* $= \boldsymbol{rev}\, S(x \in J \mapsto f(\eta(x)))$*).*
5. *For* $J \subset f(I) \cup g(I)$ *and* $\eta\colon J \to \mathbb{R}$ *increasing (or decreasing) one has* $S(f(x) - g(x)) = S(\eta(f(x)) - \eta(g(x)))$ *(respectively* $= \overline{S}(\eta(f(x)) - \eta(g(x)))$*).*

Sign patterns provide a useful tool for establishing convexity or star-shapedness of functions (see for example Lemma 11 and Theorem 20 in Arab and Oliveira [2]).

Proposition 37. *A continuous function f is convex (respectively, star-shaped) if and only if $S(f(x)-\ell(x)) \leq$ +-+ (respectively, $S(f(x)-\ell(x)) \leq$ -+), for all affine functions ℓ (respectively, for all affine functions ℓ vanishing at 0).*

Applied to the convex (IFR) and star-shape transform (IFRA) orders, these characterisations translate into the following equivalent conditions for being ordered.

Proposition 38. *Let X and Y be random variables with distributions given by distribution functions F and G, respectively. Then $X \leq_c Y$ (respectively $X \leq_* Y$) if and only if $S(F(x)-G(\ell(x))) \leq$ +-+ (resp., $S(F(x)-G(\ell(x))) \leq$ -+) for every affine function ℓ (resp., for every affine function ℓ vanishing at 0).*

The following slight generalisation of a well known relationship between the sign pattern of a differentiable function and the sign pattern of its derivative is also used throughout our proofs.

Proposition 39. *Let $f\colon I \mapsto \mathbb{R}$ be continuously differentiable with finite sign pattern $S(x \in I \to f(x)) = \sigma \cdots$, then $S(x \in I \to f(x)) \leq \sigma \cdot S(x \in I \to f'(x))$.*

Proof: Let $S(x \in I \to f(x)) = \sigma_0\sigma_1 \cdots \sigma_n$. Therefore there exists a sequence $x_0 < x_1 < \cdots < x_n$ with $\mathrm{Sign}(f(x_i)) = \sigma_i$. By the mean value theorem there exist $y_1, \ldots, y_n$ such that $f'(y_i) = \frac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}}$. Since, in particular, $\mathrm{Sign}(f'(y_i)) = \sigma_i$, we have that $\sigma_1 \cdots \sigma_n \leq S(x \in I \to f'(x))$. ■

If in the statement of Proposition 39 the initial sign of $S(x \in I \to f'(x))$ is the same as σ the inequality becomes $S(x \in I \to f(x)) \leq S(x \in I \to f'(x))$. This becomes particularly useful in combination with the following, elementary, proposition.

Proposition 40. *For $b > a$ let $f\colon [a,b] \mapsto \mathbb{R}$ be a continuously differentiable function with finite sign patterns $S(x \in I \to f(x)) = \sigma \cdots \sigma'$ and $S(x \in I \to f'(x))) = \tau \cdots \tau'$. If $f(a) = 0$ then $\sigma = \tau$ and if $f(b) = 0$ then $\sigma' = \overline{\tau}'$.*

The interval $[a, b]$ *may be replaced by* $(a, b]$, $[a, b)$ *or* (a, b) *if the conditions* $f(a) = 0$ *and* $f(b) = 0$ *are replaced by* $\lim_{x \to a+} f(x) = 0$ *or* $\lim_{x \to b-} f(x) = 0$, *as appropriate.*

Idir Arab
University of Coimbra, CMUC, Department of Mathematics, Portugal
E-mail address: idir.bhh@gmail.com

Paulo Eduardo Oliveira
University of Coimbra, CMUC, Department of Mathematics, Portugal
E-mail address: paulo@mat.uc.pt

Tilo Wiklund
Department of Mathematics, Uppsala University, Sweden
E-mail address: tilo.wiklund@math.uu.se