The Bunching and Monotonicity Properties of Families of Probability Distributions

S. Portnoy[*], N. Torrado[†] and J.J.P. Veerman[‡]

November 20, 2023

Abstract

Measuring the concentration of random variables is a fundamental concept in probability and statistics. Here, we explore a type of concentration measure for continuous random variables with bounded support and use it to provide a notion of stochastic order by concentration. We give an application to the Beta family of distributions, and especially to the one-parameter subfamily with constant mean.

Keywords: Stochastic Order, Concentration, Beta Distribution, Monotonicity

1 Motivation and Introduction

A fundamental focus in probability theory lies in examining the degree of concentration of a random variable. This concept involves investigating how closely or loosely the values of a random variable cluster around a specific value, often its central measure, such as the mean or median. The degree of concentration provides insights into the variability and predictability of the variable's outcomes. Understanding concentration is essential in various fields, including statistics, economics, and risk analysis, as it enables us to assess the reliability and stability of the random variable in question.

There are several ways to measure the concentration of random variables. One of them is through concentration inequalities, which provide valuable bounds on how much random variables

[*]Stat. Depts., Portland State University and Univ. of Illinois; e-mail: sport2@psu.edu

[†]Math. Dept., Universidad Autónoma de Madrid and Instituto de Ciencias Matmáticas, 28049 Madrid; e-mail: nuria.torrado@uam.es; nuria.torrado@icmat.es

[‡]Maseeh Dept. of Math. and Stat., Portland State Univ., Portland, OR, USA; e-mail: veerman@pdx.edu.

deviate from a specific value, typically the expected value. These inequalities are instrumental in helping us to measure the degree of concentration within a given distribution and understand how well the data points cluster around a specific value. They play a crucial role in diverse fields, from probability theory to statistics and machine learning, by allowing us to make informed assessments of the variability and stability of random variables and their associated distributions. In particular, concentration inequalities can be divided into two major groups: those that are distribution-free and those that are dependent on a specific distribution. In the first group, we encounter well-known inequalities such as Chebyshev's inequality, Markov's inequality, Chernoff's inequality, Hoeffding's inequality, and Bernstein's inequality, to name a few. Chebyshev's inequality provides an upper bound for the probability that a random variable deviates by more than a certain number of standard deviations from its mean. Markov's inequality, on the other hand, offers an exonential upper bound and focuses on the probability that a non-negative random variable is greater than or equal to a specific value. In contrast, Chernoff's and Hoeffding's inequalities are designed for independent random variables, aiming to quantify the probability that the sum of these random variables deviates from its mean. Bernstein's inequality serves as a generalization of Hoeffding's inequality and provides bounds for the probability that the sum of independent random variables significantly deviates from its mean.

In the second group, Chvátal's conjecture focussed on the Binomial distribution and has recently attracted the attention of several researchers. Specifically, Chvátal conjectured that for any given n, the probability of a binomial random variable $B(n, m/n)$ with integer expectation m is smallest when m is the integer closest to $2n/3$. [14] showed that this holds when n is large. [24] proved that the sequence $q_k = 1 - p_k = P\{B(n, kn) \geq k + 1\}$ is strictly unimodal with the mode k_0 being the integer closest to $2n/3$ for any fixed $n \geq 2$. [4] established that Chvátal's conjecture is indeed true for every $n \geq 2$. Motivated by these works, [19] studied the cases of Poisson, geometric, and Pascal distributions.

Another way to measure the concentration of a random variable is to examine whether the probability that the random variable is less than or equal to a specific value exhibits a monotonic behavior. Apparently, the first work discussing such monotonicity properties was [11], which provided results for the Chi-square, Fisher-Snedecor F, and Student's t-distributions. Subsequently, [1] explored the case of the incomplete Beta function. A recent work studying the case of the gamma distribution is [22]. Furthermore, the concentration of two random variables can be compared using certain stochastic orders, such as the convex, dispersive, and right spread orders. To explore these and other stochastic orders, refer to Shaked and Shanthikumar's monograph [23]. Some works where characterizations and applications of these orders are studied include the following. [5] characterized the right spread order through the increasing convex order. [21] investigated convex orders for linear combinations of random variables. In [17], they study the convex order between convolution polynomials of finite Borel measures. [6] explored the increasing concave orderings of linear combinations of ordered statistics, applying them to issues related to social well-being. In [7], they examine two categories of optimal insurance decision problems associated with the convex order, where the objective function or the premium valuation is a general function of the expected value, Value-at-Risk, and Average Value-at-Risk of the loss variables.

In this work, we investigate a type of concentration measure for continuous random variables with bounded support within the interval $[0, 1]$ which we refer to as the bunching property. Specifically, we demonstrate that under certain conditions, there exists a unique point x^* around which the distribution is more bunched (or concentrated). Additionally, we study a continuity and monotonicity property of such a point. The structure of this article is as follows. Section 2 presents the definitions that will be employed throughout the manuscript. Section 3 is devoted to the main results for continuous random variables with bounded support within the interval $[0, 1]$. In Section 4, we present an application to Beta distributions.

2 Definitions

In this section, we present a comprehensive set of foundational definitions that contribute to a precise understanding of the key concepts and terminologies that will be employed throughout the article.

One of the main tools we use are the well-known inverse probability transform or the *push-forward* of the measure and the operation it induces on the corresponding density.

Definition 2.1 *Suppose we have a probability measure P on a space $\mathcal{X}$ and a continuous function $y : \mathcal{X} \to \mathcal{Y}$. The push-forward $\widetilde{P}$ of P is a measure on the space $\mathcal{Y}$ defined as follows:*

$$\widetilde{P}(S) := P(y^{-1}(S)),$$

for a (measurable) set $S \subseteq \mathcal{Y}$.

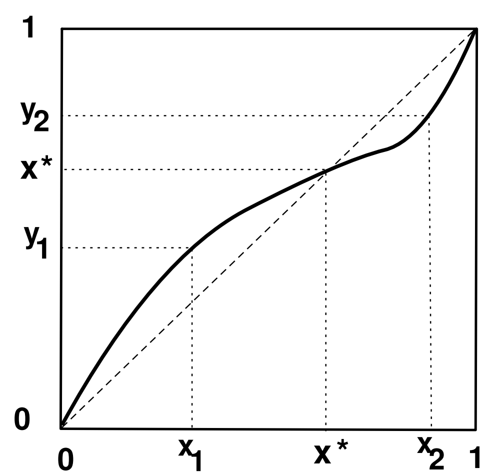

Figure 2.1: *Definition of the pushforward $\widetilde{P}$ by $y : X \to Y$ of a measure P. The measure of $\widetilde{P}([y_1, y_2])$ is set equal to that of $P([x_1, x_2])$ where $y_i := y(x_i)$.*

For the restricted setting: $\mathcal{X} = \mathcal{Y} = [0, 1]$, $y : \mathcal{X} \to \mathcal{Y}$ invertible, and with c.d.f. F (on $\mathcal{X}$), the *push-forward* is just the usual probability transform

$$\widetilde{F}(u) = F(y^{-1}(u)). \tag{2.1}$$

Now let's assume in addition that F and $\widetilde{F}$ and y are continuously differentiable and $y'(x) > 0$. The derivatives of F and $\widetilde{F}$ are probability densities and will be denoted by f and $\widetilde{f}$, respectively, satisfying

$$\widetilde{f}(y) = \frac{f(x)}{y'(x)} \qquad \text{where} \qquad x = y^{-1}(x). \tag{2.2}$$

In this study, we confine our focus to continuous random variables with bounded support within the interval $[0,1]$ such that their probability densities are contained in the class $\mathcal{F}_a$ defined as following.

Definition 2.2 *For a in some interval $A \subset \mathbb{R}_+$, we say f_a (or F_a) is in $\mathcal{F}_a$ if $f_a : [0,1] \to \mathbb{R}$ is a continuous, positive probability density with $f_a(x) > 0$ for all $x \in (0,1)$ and $a \in A$.*

Next, we present a brief review of some notions of stochastic orders (see [23] for an overview of the different notions of ordering).

Definition 2.3 *Let X and Y be univariate random variables with cumulative distribution functions (c.d.f.'s) F and G, survival functions $\bar{F}\,(= 1 - F)$ and $\bar{G}\,(= 1 - G)$.*

i) X is said to be smaller than Y in the usual stochastic order, denoted by $X \leq_{st} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all increasing functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

ii) X is said to be smaller than Y in the convex order, denoted by $X \leq_{cx} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

iii) X is said to be smaller than Y in the concave order, denoted by $X \leq_{cv} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all concave functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

iv) X is said to be smaller than Y in the increasing convex order, denoted by $X \leq_{icx} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all increasing convex functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

v) X is said to be smaller than Y in the increasing concave order, denoted by $X \leq_{icv} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all increasing concave functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

In broad terms, when $X \leq_{cx} Y$ is satisfied, there is a greater propensity for Y to adopt extreme values compared to X. This implies that Y exhibits higher variability than X. Furthermore, it follows that $E[X] \leq E[Y]$ and $Var[X] \leq Var[Y]$, given that $Var[Y] < \infty$. Moreover, $X \leq_{cx} Y$ if, and only if, $X \geq_{cv} Y$, since if ϕ is convex, then $-\phi$ is concave. On the other hand, if $X \leq_{icx} Y$ then X is both smaller and less variable than Y in some stochastic sense. It follows that $X \leq_{icx} Y$ implies $E[X] \leq E[Y]$. It is clear that $X \leq_{st} Y$ implies $X \leq_{icx} Y$ and $X \leq_{cx} Y$ also implies $X \leq_{icx} Y$. In particular, if $E[X] = E[Y]$, then $X \leq_{cx} Y$ if, and only if, $X \leq_{icx} Y$ (see section 3.A in [23]). Both increasing order relations are related by $X \leq_{icv} Y$ if, and only if, $-Y \leq_{icx} -X$. It is worth mentioning that the usual stochastic order is known in economics and finance as the first stochastic dominance (FSD), while the increasing concave order is referred to as the second stochastic dominance (SSD).

Next, we recall a characterization of increasing convex and concave orders based on the number of crossings of distribution or density functions (see [23] or [18]). Let us denote the number of sign changes of a function, g, defined on an interval, I, with

$$S^-(g) = S^-(g(x)) = \sup S^-[g(x_1), \ldots, g(x_n)]$$

where $S^-[y_1, \ldots, y_n]$ is the number of sign changes of the sequence, $y_1, \ldots, y_n$, where the zero terms are omitted, and the supremum is extended over all $x_1 < x_2 < \ldots < x_n$ $(x_i \in I)$, $n < \infty$.

Lemma 2.4 *Let X and Y be univariate random variables with cumulative distribution functions (c.d.f.'s) F and G, density functions f and g, respectively, with finite means.*

i) If $S^-(F - G) \leq 1$ and the sign sequence starts with $-$, then $X \geq_{icv} Y$ if and only if $E(X) \geq E(Y)$, while $Y \geq_{icx} X$ if and only if $E(Y) \geq E(X)$.

ii) Let $S^-(f - g) \leq 2$ and the sign sequence begins with $-$. Then, $X \geq_i cvY$ if and only if $E(X) \geq E(Y)$, while $Y \geq_{icx} X$ if and only if $E(Y) \geq E(X)$.

3 Bunching

We introduce a novel form of stochastic order based on concentration, which we call the "Bunching Property". The main aim is to provide reasonable conditions under which two distributions are ordered according to bunching. We focus on distributions in the smooth family $\mathcal{F}_a$ of Definition 2 above, and provide the following developments.

3.1 Basic results

Proposition 3.1 *Let $f_{a_1}, f_{a_2} \in \mathcal{F}_a$. Then, for all a_1 and a_2 in A, there is a unique diffeomorphism $y : (0, 1) \to (0, 1)$ such that*

$$f_{a_2}(y(x))y'(x) = f_{a_1}(x) \, .$$

Furthermore, y can be extended to a continuous function from $[0, 1]$ to itself and $y(0) = 0$ and $y(1) = 1$.

Proof From the definition (2.1) of the push-forward, we see that y is defined as the solution of

$$H(x, y) := \int_0^y f_{a_2}(u)\, du - \int_0^x f_{a_1}(v)\, dv = F_{a_2}(y) - F_{a_1}(x) = 0 \, .$$

Since both densities are positive, it follows that $y(0) = 0$ and $y(1) = 1$. Now for every given x, by continuity (and positivity of both densities), we see that there must be a unique y. Furthermore, since both $\partial_x H$ and $\partial_y H$ are non-zero in $(0,1)^2$, a straightforward application of the implicit function theorem, tells us that there are differentiable functions $y(x)$ and $x(y)$ so that $H(x,y(x)) = H(x(y),y) = 0$. Naturally, these are inverses of one another, and so $y(x)$ is a diffeomorphism on $(0,1)$. Differentiating $H(x,y(x))$ with respect to x gives the well-known formula in the proposition. ■

For the next result, we first need a simple convexity lemma.

Lemma 3.2 *Let $g : \mathbb{R} \to \mathbb{R}$ be twice differentiable on some interval I. If $x_1 < x_2 < x_3$ and*

$$\frac{g(x_3) - g(x_2)}{x_3 - x_2} \leq \frac{g(x_2) - g(x_1)}{x_2 - x_1},$$

then there is a point $x \in I$ where $g''(x) \leq 0$.

Proof The proof is elementary and consists of applying the mean value theorem repeatedly. ■

Lemma 3.3 *Let f_{a_1} and f_{a_2} two twice differentiable densities belong to $\mathcal{F}_a$ with $a_2 > a_1$ and a push-forward y as in Proposition 3.1. Suppose they satisfy the additional requirements that $\lim_{x \searrow 0} y'(x)$ and $\lim_{x \nearrow 1} y'(x)$ are greater than one (or tend to infinity) and that the second derivative of $f_{a_1}(x)/f_{a_2}(x)$ is strictly positive on $(0,1)$. Then:*

i) $y(x) - x = 0$ *if and only if* $x \in \{0, x^*, 1\}$ *and*

ii) $\forall\, x \in (0, x^*) \,:\, y(x) - x > 0 \quad$ and $\quad \forall\, x \in (x^*, 1) \,:\, y(x) - x < 0$ *and*

iii) $f_{a_1}(x^*) \leq f_{a_2}(x^*)$.

Proof It is sufficient to prove that y has a unique fixed point (i.e. $y(x) = x$) in $(0,1)$. We follow standard procedure and rewrite the differential equation in Proposition 3.1 as a 2-dimensional autonomous system on $(0,1) \times (0,1)$ with reparametrized time:

$$\begin{cases} \dot{x} &= f_{a_2}(y) \\ \dot{y} &= f_{a_1}(x) \end{cases} \tag{3.1}$$

The RHS of this system is continuously differentiable in $(\epsilon, 1-\epsilon) \times (\epsilon, 1-\epsilon)$ for any positive ϵ. We know from Proposition 3.1 that there is a unique solution $(x, y(x))$ such that $y(0) = 0$ and $y(1) = 1$. By hypothesis, for x close to zero, $y(x) - x > 0$, and so $y(x)$ starts out above the diagonal. Similarly, for x close to 1, $y(x) - x < 0$ by the boundary conditions $\lim_{x \nearrow 1} y'(x)$ and $y(1) = 1$.

We now study the intersections of $\gamma(x) := (x, y(x))$ with the diagonal. The tangent of the vector field *along the diagonal* (x,x) in the RHS of (3.1) is $t(x) := \frac{f_{a_1}(x)}{f_{a_2}(x)}$. By hypothesis we have

$$t''(x) > 0\,. \tag{3.2}$$

Set $f(x) := y(x) - x$. Now let us suppose there are three distinct points $0 < x_1 < x_2 < x_3 < 1$ in the open interval $(0,1)$ where $f(x_i) = 0$. Then by Lemma 3.2, there is a point v where $t''(v) \leq 0$, which contradicts (3.2).

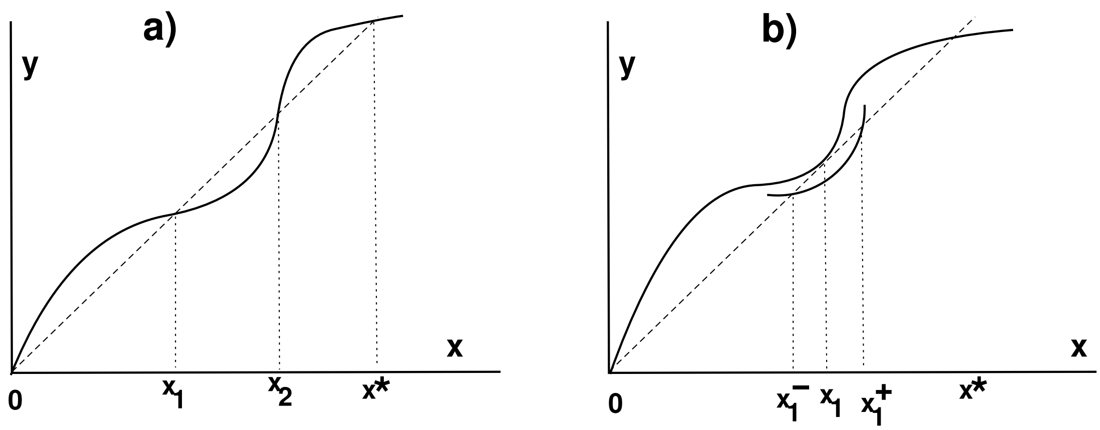

Figure 3.1: *A non-simple intersection of $y(x)$ and the diagonal at x_1. A local twice continuously differentiable nearby solution resolves this into two simple intersections.*

The conclusion of this reasoning is that there are at most two distinct points where $y(x) = x$. Since near $x = 0$, γ is above the diagonal, and near $x = 1$ below it, there must be (by the intermediate value theorem) at least one crossing at x^* from above to below the diagonal.

The only possibilities this leaves is either a unique crossing at x^* or else a crossing at x^* plus a point x_1 where γ is tangent to the diagonal. We have to rule out the latter. So suppose this is the case and assume $x_1 < x^*$ as in Figure 3.1a. Now solutions to the ODE of (3.1) are unique and cannot cross [3, 13]. Furthermore, nearby solutions approximate one another. So a solution that starts at (x_1^-, x_1^+) very close to γ as illustrated Figure 3.1b, must cross the diagonal again at $x_1^+ > x_1$ but very close to it. This again gives three distinct points $0 < x_1^- < x_1^+ < x^* < 1$ such that, respectively,

$$t(x_1^-) \leq 1\,, \quad t(x_1^+) \geq 1\,, \quad t(x^*) \leq 1\,.$$

Lemma 3.2 yields a value x^{**} where $t''(x^{**}) \leq 0$. Thus the intersection of γ with the diagonal in $(0, 1)$ is unique and this gives us items i and ii of the proposition. Item (iii) follows from the fact that $y(x)$ is differentiable and crosses the diagonal in the downward direction. ■

Next, we present a formal statement of the Bunching Property.

Theorem 3.4 (The Bunching Property) *Fix $0 < a_1 < a_2$ and let $F_{a_1}, F_{a_2} \in \mathcal{F}_a$ such that they satisfy the conditions of Lemma 3.3. There is a unique $x^* \in (0, 1)$ so that for all x_1 and x_2 with $0 < x_1 < x^* < x_2 < 1$, we have $F_{a_1}(x_1) > F_{a_2}(x_1)$ and $1 - F_{a_1}(x_2) > 1 - F_{a_2}(x_2)$. Thus, we may say that F_{a_2} is more* bunched *around x^* than is F_{a_1}.*

Remark 3.5 Observe that, from Theorem 3.4, we get $S^-(F_{a_2} - F_{a_1}) = 1$ and the sign sequence starts with $-$. Therefore, if X_{a_1} and X_{a_2} be two random variables with distribution functions F_{a_1} and F_{a_2}, respectively, from Lemma 2.4(i), we obtain that $X_{a_2} \geq_{icv} X_{a_1}$ if and only if $E(X_{a_2}) \geq E(X_{a_1})$, while $X_{a_1} \geq_{icx} X_{a_2}$ if and only if $E(X_{a_1}) \geq E(X_{a_2})$. □

3.2 The Continuity Property

In this subsection, we present a continuity property for the point x^* where one distribution is more bunched than the other.

Lemma 3.6 *Let f and g be two probability densities that are positive on $(0,1)$. Set $t(x) := \frac{f(x)}{g(x)}$ and suppose that t is positive and strictly convex with*

$$\lim_{x\to 0} t(x) = \lim_{x\to 1} t(x) = \infty\,.$$

Then $f(x) = g(x)$ in exactly two distinct points in $(0,1)$.

Proof By the strict convexity, it is clear that $t(x)$ intersects the line $y = 1$ either twice or not at all But in the latter case, $f(x)$ is strictly larger than $g(x)$, which conflicts with the fact that both are probability densities (integrating to 1). ■

Proposition 3.7 *Let F_n and G_n be c.d.f.'s with densities f_n and g_n continuous in n and satisfying Lemma 3.6. Assume the conditions for Bunching (Theorem 3.4) and let x^* be the unique solution of*

$$F_n(x) - G_n(x) = \int_0^x (f_n(t) - g_n(t))\, dt = 0\,.$$

Then x^* depends continuously on n.

Proof We first show $f_n(x^*) \neq g_n(x^*)$. Suppose otherwise. Then there must be a $0 < x_- < x^*$ such that $f_n(x_-) = g_n(x_-)$, for otherwise one density would dominates the other, and so the integrals could not be equal. Similarly, since $1 - F_n(x^*) = 1 - G_n(x^*)$, the densities must be equal at some point $x_+ \in (x^*, 1)$. This would provide three points where the densities are equal, contradicting Lemma 3.6. Thus $f_n(x^*) \neq g_n(x^*)$. Now (by continuity of f_n and g_n in n), a small change in n will cause a small change in the integral:

$$\int_0^{x^*} (f_{n'}(t) - g_{n'}(t))\, dt = \delta\,.$$

Since near x^* and for n' close enough to n, the measures are not equal, a small adjustment in x^* will bring the integral back to zero. ■

3.3 A Monotonicity Property

To provide the main result of this subsection, consider a family of distribution functions $\{F_{a;n,m}\} \subset \mathcal{F}$ (see Definition 2.2) depending on two additional real parameters: $n \geq m > 0$.

Condition 3.8 *Given $0 < a_1 < a_2$ and $n \geq m > 0$. Let $F_a(x)$ and its derivative $f_a(x)$ be c.d.f.'s and densities as described above. Furthermore, for $x \in (0,\,1)$:*

i: If $n = m$, $f_a(x)$ is symmetric under $x \leftrightarrow 1 - x$ and

ii: $\partial_n f_a(x)$ is negative and increasing on $(0, 1)$ and

iii: $\partial_a \partial_n f_a(x) < 0$.

Lemma 3.9 *(See also Proposition 3.7.) Assume the hypotheses for Bunching (Theorem 3.4) and for Lemma 3.6, Let x^* in $(0, 1)$ be the unique point where $y(x^*) = x^*$. Then*

$$f_{a_2}(x^*) > f_{a_1}(x^*)\,.$$

Proof. At the point where $y(x^*) = x^*$, we have $F_{a_2}(x^*) - F_{a_1}(x^*) = 0$. By Lemma 3.3, $f_{a_1}(x^*) \leq f_{a_2}(x^*)$ and so we only have to rule out the possibility that they are equal. By Lemma 3.6, $f_{a_2}(x) = f_{a_1}(x)$ in exactly two distinct points in $(0, 1)$, say x_1 and x_2 as in Figure 3.2. But since the f_{a_i} are densities (i.e. they integrate to 1), it impossible for $F_{a_2}(x)$ and $F_{a_1}(x)$ to be equal at either x_1 or x_2. ■

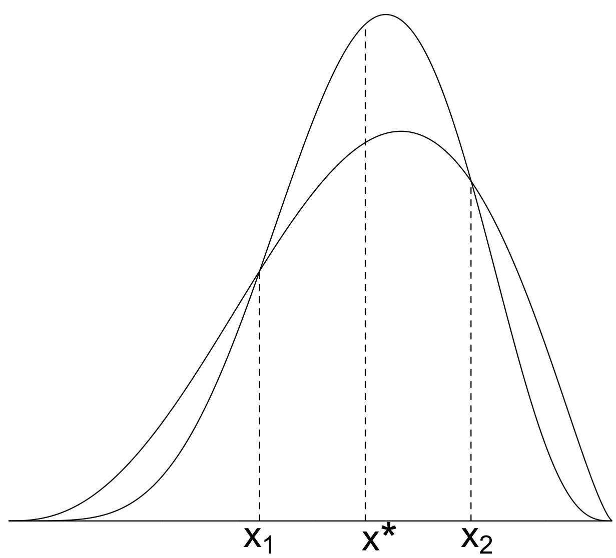

Figure 3.2: *Two Beta densities ordered by Bunching. Under conditions, if $a_2 > a_1 > 0$, f_{a_1} and f_{a_2} intersect in exactly two points $\{x_1,\, x_2\} \subset (0, 1)$. At x^*, we have that $\int_0^{x^*} f_{a_1}\,dt = \int_0^{x^*} f_{a_2}\,dt$.*

In what follows, we will always assume that $a_2 > a_1 > 0$ and $n > m > 0$. From the previous section, we know that there is unique solution for x of

$$F_{a_2}(x) - F_{a_1}(x) = 0\,.$$

We will track this solution as function of n and x, holding fixed all the other parameters. To facilitate this, we define

$$J(n,x) := F_{a_2}(x) - F_{a_1}(x)\,.$$

Proposition 3.10 *Assume Condition 3.8 and the hypotheses for Bunching (Theorem 3.4) and for Lemma 3.6. Fix $0 < a_1 < a_2$ and $n \geq m > 0$. We vary n and hold m and the a_i constant. For $n = m$, the locus $x^*(n)$ of the unique zero of $J(n,x)$ equals 1/2. We have that x^* is a differentiable function satisfying $x^*(m) = 1/2$ and $\partial_n x^* > 0$.*

Proof. By the symmetry of f (see Condition 3.8, (ii)) we have that if $n = m$, then $x(n) = 1/2$. We also have

$$\frac{d}{dn} J(n,x) = \partial_x J \, \frac{dx}{dn} + \partial_n J = 0\,. \tag{3.3}$$

Now,

$$\partial_x J(n,x) = f_{a_2}(x) - f_{a_1}(x)\,.$$

By Lemma 3.9, have $f_{a_2}(x^*) > f_{a_1}(x^*)$ or

$$\partial_x J > 0\,.$$

The implicit function theorem now says that near this point there is differentiable function h such that $J(n, h(n)) = 0$. By Condition 3.8(iii),we also have

$$\partial_n J = \int_0^x \left(\partial_n f_{a_2}(t) - \partial_n f_{a_1}(t)\right)\, dt < 0\,. \tag{3.4}$$

Together with (3.3), this establishes that the function $x^*(n)$ is differentiable and strictly increasing, as required. ■

4 Applications to Beta distributions

In this section, we present an application of the main results to a beta distribution defined in the interval $[0,1]$. To do this, let us recall the definition of this distribution. Here we denote the density function of a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ as

$$f_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{with} \qquad B(\alpha,\beta) = \int_0^1 f_{\alpha,\beta}(x)\, dx.$$

In what follows, we will focus on Beta distributions with parameters na and ma for fixed n and m with $a > 0$. That is, the density and distribution functions of a random variable $X \sim Beta(na, ma)$ are denoted:

$$p_a(x) = f_{na,ma}(x) \quad \text{and} \quad F_a(x) = \int_0^x p_a(t)\,dt \ , \tag{4.1}$$

respectively. Note that the notation suppresses dependence on n and m; but when needed we may write $p_{a;n,m}(x)$ and $F_{a;n,m}(x)$. From (4.1), we see that

$$E[X] = \frac{n}{n+m} \quad \text{and} \quad V[X] = \frac{nm}{(n+m)^2(na+ma+1)}.$$

So the mean is constant and the variance decreases with a. This would seem to imply that the distribution becomes more and more concentrated around the mean.

Next, we prove that the Beta distributions of equation (4.1) satisfy the requirements of Proposition 3.1 and Lemma 3.3. First, we need a little lemma.

Lemma 4.1 *Let $t(x) = x^{-p}(1-x)^{-q}$. If p and q are both positive, then t is strictly convex on $[0, 1]$.*

Proof. A tedious, but straightforward, computation gives

$$t''(x) = \frac{qx^2 + p(1-x)^2 + (qx - p(1-x))^2}{x^{2+p}(1-x)^{2+q}} \ ,$$

and the result follows immediately. ∎

Remark. It is easy to see that the product of two monotone increasing (or decreasing) convex functions is again convex. In our case, t is the product of one increasing and one decreasing convex function. It is by no means obvious the the product should be convex. In fact, $p = q = -1$ gives a counter-example.

Lemma 4.2 *Let $\{p_a\}$ be the family of Beta densities (see equation 4.1) and fix $a_2 > a_1$. Define $t(x) := \frac{p_{a_1}(x)}{p_{a_2}(x)}$. Then for y in Proposition 3.1*

$$\begin{aligned} i: &\quad p_a(x) > 0 \ \text{ for } \ x \in (0,1) \\ ii: &\quad \lim_{x \searrow 0} y'(x) = \lim_{x \nearrow 1} y'(x) = \infty \\ iii: &\quad t''(x) > 0 \ \text{ for } \ x \in (0,1) \\ iv: &\quad \lim_{x \searrow 0} t(x) = \lim_{x \nearrow 1} t(x) = \infty\,. \end{aligned}$$

Proof. The first statement is obvious.

The approximate solution for x and y very close to zero can be found by neglecting the $(1-x)$ terms in the integration (since they are going to be very close to 1). So

$$G(x,y) \approx \frac{\int_0^y u^{na_2-1}\,du}{B(na_2,ma_2)} - \frac{\int_0^x v^{na_1-1}\,dy}{B(na_1,ma_1)} = 0 \quad \Longleftrightarrow \quad \frac{y^{na_2}}{na_2B_2} \approx \frac{x^{na_1}}{na_1B_1}\,.$$

(Here we abbreviated $B(na_i,ma_i)$ as B_i.) This gives

$$y \approx \left(\frac{a_2B_2}{a_1B_1}\right)^{1/na_2} x^{a_1/a_2}\,.$$

The first limit follows from $a_2 > a_1$. The second limit can be evaluated in the same way by changing variables $\widetilde{x} = 1-x$ and $\widetilde{y} = 1-y$. This proves the second statement.

Note that $t(x) = Kx^{-n(a_2-a_1)}(1-x)^{-m(a_2-a_1)}$. The third statement follows from Lemma 4.1. The fourth statement follows directly from the expression for $t(x)$ just given. ■

Theorem 4.3 *Let $\{F_a(x)\}$ denote the Beta distributions in* (4.1)*, and fix $0 < a_1 < a_2$ and $n \geq m > 0$. There is a unique $x^* \in (0,1)$ so that for all x_1 and x_2 with $0 < x_1 < x^* < x_2 < 1$, we have $F_{a_1}(x) > F_{a_2}(x)$ and $1 - F_{a_1}(x) > 1 - F_{a_2}(x)$. Thus, we may say that F_{a_2} is more* bunched *around x^* than is F_{a_1}.*

Proof. The situation is exactly as sketched in Figure 2.1 with p_{a_2} on the vertical axis being the push-forward by $x \to y(x)$ of p_{a_1}. In a case like this, the bunching property described in Theorem 3.4 holds. This is most easily seen by noting that the theorem implies that for $y_1 \in (0,x^*)$:

$$F_{a_2}(y_1) = F_{a_1}(x_1) < F_{a_1}(y_1) \qquad \text{where} \qquad y = y(x).$$

The latter inequality holds because we also know that $0 < x < y(x) < x^*$ and $p_{a_1}(x) > 0$ on $[x, y(x)]$. Similarly, one shows that for $x^* < y_2 < 1$, $1 - F_{a_2}(y_2) < 1 - F_{a_1}(y_2)$.

■

For the family of Beta distributions, the following result is known (see, e.g., [2] or [20]).

Proposition 4.4 *Let $X \sim Beta(\alpha_1,\beta_1)$ and $Y \sim Beta(\alpha_2,\beta_2)$, then $Y \leq_{st} X$ if, and only if, $\alpha_1 \geq \alpha_2$ and $\beta_1 \leq \beta_2$.*

In our case, $\alpha_i = na_i$ and $\beta_i = ma_i$ for $i = 1,2$ and $n, m > 0$ two real numbers fixed. Therefore, it is evident that $Y \nleq_{st} X$ and $Y \ngeq_{st} X$. Note that, from Theorem 4.3, we have $S^-(F_{a_2} - F_{a_1}) = 1$ and the sign sequence starts with $-$. Therefore, from Lemma 2.4(i), we have the following result.

Corollary 1 *Let $X_{a_i} \sim Beta(na_i, ma_i)$ for $i = 1, 2$ and $n, m > 0$. Then, $X_{a_2} \geq_{icv} X_{a_1}$ for $0 < a_1 < a_2$.*

It is worth mentioning that the previous corollary coincides with Theorem 2 in [8], as X_{a_i} satisfies the conditions of that result for $i = 1, 2$. Therefore, Theorem 2 in [8] can be viewed as a special case of our Theorem 3.4 for Beta distributions (under certain conditions).

Moreover, since the Beta distributions satisfy all our hypotheses (see Lemma 4.1), the continuity property holds for them.

Proposition 4.5 *Let $\{F_a(x)\}$ denote the beta distributions in (4.1), and fix $0 < a_1 < a_2$ and $n \geq m > 0$. The location of x^* is a continuous function of n.*

5 Conclusions and Motivating Problem

We have introduced a novel form of stochastic order based on concentration. Basically, we say that one distribution is more bunched that another if there is a point x^* such that interval probabilities both to the left and right of x^* are smaller (for the more-bunched distribution). We have provided sufficient conditions for the existence of x^* and give conditions for some continuity and monotonicity properties.

While these results are of general interest, it may be noted that there was a specific motivating problem: a colleague, Subhash Kochar, asked if we could show that $P_a(\frac{1}{2})$ is monotonically decreasing in $a > 0$ if $n > m$ for the restricted Beta subfamily. Apparently, the earliest paper discussing this monotonicity result is a Clemson University Technical Report, Alam [1], which motivates the problem in terms of Ranking and Selection probabilities. The problem also arises in comparing Gamma distributions in reliability theory: if U and V have Gamma distributions with mean parameters $n\alpha$ and $m\alpha$ respectively, then $\Pr\{U < V\} = P_\alpha(\frac{1}{2})$. This suggested a possible connection to stochastic dominance within the restricted beta family.

Specifically, if one could prove that $x^*(n) > \frac{1}{2}$ for $n > m$ in this restricted beta family, then monotonicity of $P_a(\frac{1}{2})$ would follow immediately from Bunching for the family (Theorem 4.3). Unfortunately, it has not been possible to show that Condition 3.8 holds for the Beta subfamily. It is possible to show that an alternative sufficient condition would be $\partial_n \partial_a F_{a;n,m}(x) > 0$ for $x = x*$ and $n > m$, but this also has not been shown for the Beta subfamily, and so the desired monotonicity of $P_a(\frac{1}{2})$ remains conjectural.

Finally, on generalizing these ideas: clearly our results on Bunching do not depend on the specific domain interval, but do require a one-dimensional (smooth) family in order to apply the "push-forward" probability transform. In multivariate situations, this can often be replaced by measure transport. See [12] for a summery of some recent statistical work using this concept. We conjecture that it is possible to develop reasonable conditions for smooth multivariate distributions under which there is a unique point such that the probabilities are monotonically decreasing for all closed convex sets not containing the point. Such case would clearly provide a definition for multivariate bunching.

Funding

N. Torrado states that this manuscript is part of the project TED2021-129813A-I00 and the grant CEX2019-000904-S. She thanks the support of MCIN/AEI/10.13039/501100011033 and the European Union “NextGenerationEU”/PRTR.

References

[1] Alam, K. (1979). A note on the incomplete beta function, *Clemson University Technical Report no. 316*.

[2] Arab, I., Oliveira, P.E., and Wiklund, T. (2021). Convex transform order of Beta distributions with some consequences, *Statistica Neerlandica, 75*, 238-256.

[3] V. I. Arnold, *Ordinary Differential Equations*, 3rd Edition, Springer-Verlag, Berlin, 1992.

[4] Barabesi, L., Pratelli, L., and Rigo, P. (2023). On the Chvátal–Janson conjecture. *Statistics & Probability Letters, 194*, 109744.

[5] Belzunce, F. (1999). On a characterization of right spread order by the increasing convex order. *Statistics & Probability Letters, 45(2)*, 103-110.

[6] Castaño-Martínez, A., Pigueiras, G., Psarrakos, G., & Sordo, M. A. (2020). Increasing concave orderings of linear combinations of order statistics with applications to social welfare.*Metrika, 83*, 699-712.

[7] Cheung, K.C., Chong, W.F., & Yam, S.C.P. (2015). Convex ordering for insurance preferences.*Insurance: Mathematics and Economics, 64*, 409-416.

[8] Braouezec, Y., & Cagnol, J. (2023). A lattice approach to the Beta distribution induced by stochastic dominance: Theory and applications. *Journal of the Operational Research Society, 74(6)*, 1424-1442.

[9] Fresen, D. (2022). Optimal tail comparison under convex majorization, *arXiv preprint arXiv:2207.01872*.

[10] Finner H. and Roters, M. (1997). Log-concavity and inequalities for chi-square, f, and beta distributions with applications in multiple comparisons, *Statistica Sinica, 7*, 771–787.

[11] Ghosh, B. K. (1973). Some monotonicity theorems for χ^2, F and t distributions with applications. *Journal of the Royal Statistical Society: Series B (Methodological), 35(3)*, 480-492.

[12] Hallin, Marc. (2022). Measure transportation and statistical decision theory. *Annual Review of Statistics and Its Application, 9*, 401-424.

[13] M. W. Hirsch, S. Smale, R. L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, 3rd Edition, Elsevier, Amsterdam, 2013.

[14] Janson, S. (2021). On the probability that a binomial variable is at most its expectation. *Statistics & Probability Letters, 171*, 109020.

[15] Jimenex, H., Naszódi, M, and Villa, R. (2014). Push forward measures and concentration phenomena, *Math. Nachrichten, 287*, 585-594.

[16] Jones, M.C. (2002). The complementary beta distribution. *Journal of Statistical Planning and Inference 104* 329–337.

[17] Komisarski, A., & Rajba, T. (2019). Convex order for convolution polynomials of Borel measures. *Journal of Mathematical Analysis and Applications, 478(1)*, 182-194.

[18] Lando, T., & Bertoli-Barsotti, L. (2020). Stochastic dominance relations for generalised parametric distributions obtained through composition. *METRON, 78(3)*, 297-311.

[19] Li, F. B., Xu, K., & Hu, Z. C. (2023). A study on the Poisson, geometric and Pascal distributions motivated by Chvátal's conjecture. *Statistics & Probability Letters*, 109871.

[20] Lisek, B. (1978). Comparability of special distributions. Series Statistics, 9, 587–598.

[21] Ma, C. (2000). Convex orders for linear combinations of random variables. *Journal of Statistical Planning and Inference, 84(1-2)*, 11-25.

[22] Pinelis, I. (2021). Monotonicity properties of the gamma family of distributions. *Statistics & Probability Letters, 171*, 109027.

[23] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. New York: Springer.

[24] Sun, P. (2021). Strictly unimodality of the probability that the binomial distribution is more than its expectation. *Discrete Applied Mathematics, 301*, 1-5.