EXTENSIONS OF TRUE POSITIVE SKEWNESS FOR UNIMODAL DISTRIBUTIONS

ALEX NEGRÓN, CLARICE PERTEL, AND CHRISTOPHER WANG

ADVISOR: DR. YEVGENIY KOVCHEGOV
OREGON STATE UNIVERSITY

ABSTRACT. In a 2020 paper, Y. Kovchegov introduced the notion of true positive and negative skewness for continuous random variables via Fréchet p-means. In this work, we find novel criteria for true skewness, identify a parameter region for true positive skewness of the Weibull distribution, and formally establish positive skewness of the Lévy distribution for the first time. We discuss the complications of extending the notion of true skewness to discrete random variables and to multivariate settings. Further, several properties of the p-means of random variables are established.

1. INTRODUCTION AND STATEMENT OF RESULTS

A commonly accepted measure for the skewness of a distribution is given by *Pearson's moment coefficient of skewness*, or the standarized third central moment

$$\text{Skew}[X] := E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right],$$

where μ is the mean and σ^2 is the variance. Usually, we say a distribution is positively skewed if $\text{Skew}[X] > 0$ and negatively skewed if $\text{Skew}[X] < 0$. It is also expected that positively skewed distributions satisfy the *mean-median-mode inequalities*

$$\text{mode} < \text{median} < \text{mean}.$$

Similarly, negatively skewed distributions are expected to satisfy

$$\text{mean} < \text{median} < \text{mode}.$$

However, the sign of the moment coefficient of skewness does not always imply the corresponding mean-median-mode inequalities. For instance, the Weibull distribution with shape parameter $3.44 < \beta < 3.60$ has positive moment coefficient of skewness but satisfies the reverse mean-median-mode inequalities, corresponding to negative skew [5]. This discrepancy is important when comparing with other measures of skewness, such as *Pearson's first skewness coefficient*

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

Date: August 14, 2020.

This work was done during the Summer 2021 REU program in Mathematics and Theoretical Computer Science at Oregon State University supported by NSF grant DMS-1757995.

and *Pearson's second skewness coefficient*

$$3 \times \frac{\text{mean} - \text{median}}{\text{standard deviation}}.$$

The direction of skewness can therefore be inconsistent between different skewness measures.

In a recent paper, Y. Kovchegov [8] introduced the notion of *true positive and negative skewness* to unify the previous measures in determining the sign of skewness. It relies on the idea that for skewed distributions, one tail should stochastically dominate the other, resulting in a tail that "spreads shorter" and another that "spreads longer." We use a class of centroids known as Fréchet p-means.

Definition 1.1. For $p \in [1, \infty)$ and random variable X with finite p-th moment, the **Fréchet p-mean**, or simply the **p-mean**, of X is the quantity

$$\nu_p := \arg\min_{a \in \mathbb{R}} E|X - a|^p. \tag{1.1}$$

Occasionally we write $\nu(p)$ to emphasize that ν is a function of p.

The p-mean is uniquely defined for all $p \geq 1$ as $E|X - a|^p$ is a strictly convex function of a. Moreover, ν_p is the unique solution of

$$E[(X - \nu_p)_+^{p-1}] = E[(\nu_p - X)_+^{p-1}], \tag{1.2}$$

which requires only the finiteness of the $(p-1)$-st moment. If X has a density function f, then we can rewrite (1.2) as

$$\int_0^{\nu_p - L} y^{p-1} f(\nu_p - y) dy = \int_0^{R - \nu_p} y^{p-1} f(\nu_p + y) dy \tag{1.3}$$

where X has support on the possibly infinite interval (L, R). Notice that identically distributed random variables have the same p-means and that ν_1 and ν_2 correspond to the median and mean of the distribution, respectively.

Let

$$\mathcal{D}_X := \{p \geq 1 : E|X|^{p-1} < \infty\}$$

be the domain of ν_p for X. If X has a unique mode, then we denote it by ν_0. In this case, we include 0 in the set $\mathcal{D}_X$. We omit the subscript X when the random variable is unambiguous.

Definition 1.2. We say a random variable X (resp. its distribution and density) is **truly positively skewed** if ν_p is a strictly increasing function of p in $\mathcal{D}$, provided $\mathcal{D}$ has non-empty interior. Analogously, X is **truly negatively skewed** if ν_p is a strictly decreasing function of p in $\mathcal{D}$.

Remark 1.3. It may be possible that, for a unimodal distribution, ν_p is strictly increasing only on $\mathcal{D} \setminus \{0\}$ and that there exists $p \in \mathcal{D} \setminus \{0\}$ such that $\nu_p < \nu_0$. Kovchegov [8] differentiates between this case, which is referred to in that work as true positive skewness, and the case where ν_p is increasing on all of $\mathcal{D}$, which is referred to as *true mode positive skewness*. Because we consider mostly unimodal distributions, for simplicity we take true positive skewness to mean true mode positive skewness in the sense of Kovchegov, unless explicitly mentioned otherwise.

If X has unique mode, true positive skewness guarantees $\nu_0 < \nu_1 < \nu_2 < \nu_4$. Thus Pearson's first and second skewness coefficients coincide in sign, and the expected mean-median-mode inequalities are satisfied. The following proposition shows that true positive skewness unifies these measures of skewness with the moment coefficient of skewness in determining the direction of skewness.

Proposition 1.4 (Kovchegov [8], 2020)**.** *For X with finite third moment, $Skew[X] > 0$ if and only if $\nu_2 < \nu_4$.*

An additional advantage of the notion of true positive skewness is that it allows us to characterize the skewness of distributions that have infinite integer moments. Indeed, each of Pearson's skewness coefficients requires at least a finite first moment. This excludes a large class of heavy-tailed distributions from the classical study of skewness. See [11] for a detailed inquiry into heavy-tailed distributions.

Kovchegov proves the true positive skewness of several distributions: exponential, gamma, beta (in certain parameter regions), log-normal, and Pareto. Using criteria established in Kovchegov's paper, we prove true positive skewness of two additional distributions. In particular, the Lévy distribution has undefined moment coefficient of skewness because it has no finite integer moments. Thus, to the authors' knowledge, its positive skewness is formally established for the first time in this work.

Theorem 1.5. *The Lévy distribution is truly positively skewed.*

When considering distribution families, we always assume the location parameter is 0 and the scale parameter is 1, since they do not affect the direction of skewness.

Theorem 1.6. *The Weibull distribution with shape parameter $k > 0$ is truly positively skewed if and only if $0 < k < \frac{1}{1-\log 2}$. Moreover, it is not truly negatively skewed for any $k > 0$.*

It is necessary for practical purposes to develop additional simple criteria for determining the true skewness of distributions. We disprove a conjecture of Kovchegov that true positive skewness is preserved under summation, in both discrete and continuous settings. Products of truly skewed random variables are also considered, and we provide a conjecture on the behavior of true skewness under multiplication. For concave distributions, i.e. having a density function decreasing in its support, we prove the following.

Theorem 1.7. *Let X be a continuous random variable with density function decreasing on its support. If u is convex and strictly increasing, then $u(X)$ is truly positively skewed.*

We say a random variable X (resp. its distribution and density) is **truly non-negatively skewed** if ν_p is a non-decreasing function of p in $\mathcal{D}$. Note that true positive skewness implies true non-negative skewness. This property is closed under weak limits, or convergence in distribution. If X_n converges weakly to X, we write $X_n \Rightarrow X$.

Theorem 1.8. *Suppose $X_n \Rightarrow X$ with $\mathcal{D}_X \subseteq \bigcap_n \mathcal{D}_{X_n}$ and $\{X_n^p\}_{n\in\mathbb{N}}$ uniformly integrable for all $p \in \mathcal{D}$. If the X_n's are truly non-negatively skewed, then so is X.*

A slightly stronger condition yields true positive skewness. Denote by $\nu_n(p)$ the p-mean of X_n.

Corollary 1.9. *Suppose $X_n \Rightarrow X$ with $\mathcal{D}_X \subseteq \bigcap_n \mathcal{D}_{X_n}$ and $\{X_n^p\}_{n\in\mathbb{N}}$ uniformly integrable for all $p \in \mathcal{D}$. If the X_n's are truly positively skewed and there exists $c > 0$ such that $\nu_n'(p) > c$ for all $p \in \mathcal{D}$ and all $n \in \mathbb{N}$, then X is truly positively skewed.*

Additionally, we establish a simple criterion for true positive skewness of random variables supported on the half-line that can be utilized in numerical analyses. In particular, it avoids using information about ν_p, which generally has no closed-form expression, except for ν_0 and ν_1, which are easily computable. In certain cases, the conditions given below can be relaxed.

Theorem 1.10. *Let X have support on $(0,\infty)$ with unimodal density $f \in C^2(0,\infty)$ that is continuous at 0. Suppose f'' has exactly two positive roots θ_1, θ_2 such that $\theta_1 < \nu_0 < \theta_2$, and*

(1) $f'/f > 1/\nu_0$ on $(0,\theta_1)$,
(2) $f'/f > -1/\nu_0$ on (θ_2,∞).

If $\nu_1 > (\nu_0 + \theta_2)/2$, then X is truly positively skewed.

Corollary 1.11. *Let X have support on $(0,\infty)$ with unimodal density $f \in C^2(0,\infty)$ that is continuous at 0. Suppose f'' has exactly one positive root $\theta > \nu_0$. If $\nu_1 > (\nu_0 + \theta)/2$, then X is truly positively skewed.*

The article is structured as follows. In Section 2, we restate and extend several previous results on true skewness, and general properties of ν_p are shown in Section 3. Proofs of the true positive skewness of the Lévy and Weibull distributions are given in Section 4. Section 5 reviews transformations, sums, and products of truly skewed random variables, and we give a proof of Theorem 1.7. In Section 6, we prove Theorem 1.8 and Corollary 1.9. Then, in Section 7, we prove Theorem 1.10 and Corollary 1.11, and we give some examples on how they can be used. A discussion of how true skewness can be extended to the discrete and multivariate settings is given in Section 8. Finally, we suggest a new measure of skewness based on the notion of true positive skewness in Section 9.

2. Preliminary results

In this section, we restate several results in [8] and provide slight extensions of them, which will become necessary in later sections. In general, we use f to denote the density function of a random variable.

Definition 2.1. We say that X **stochastically dominates** Y (resp. their distributions) if the distribution function F of X is dominated by the distribution function G of Y, i.e. if $F(x) \le G(x)$ for all $x \in \mathbb{R}$. We say that the stochastic dominance is *strict* if in addition there exists a point x_0 for which $F(x_0) < G(x_0)$.

We also define a normalizing term H_p as the equal quantities in (1.3), i.e.,

$$H_p := \int_0^{\nu_p - L} y^{p-1} f(\nu_p - y)dy = \int_0^{R-\nu_p} y^{p-1} f(\nu_p + y)dy \qquad (2.1)$$

for $p \in \mathcal{D}$. The first result establishes the relationship between true positive skewness and stochastic dominance of the tails of a distribution.

Theorem 2.2 (Kovchegov [8], 2020)**.** *Let X be a continuous random variable supported on (L,R) with density function f. Fix $p \in \mathcal{D}$. If the distribution with density $\frac{1}{H_p}y^{p-1}f(\nu_p + y)\mathbf{1}_{(0,R-\nu_p)}(y)$ exhibits strict stochastic dominance over the distribution with density $\frac{1}{H_p}y^{p-1}f(\nu_p - y)\mathbf{1}_{(0,\nu_p-L)}(y)$, then ν is increasing at p.*

The crux of the proof of Theorem 2.2 is that ν is increasing if and only if

$$\int_0^{R-\nu_p} y^{p-1}\log y \ f(\nu_p + y)dy - \int_0^{\nu_p - L} y^{p-1}\log y \ f(\nu_p - y)dy > 0. \tag{2.2}$$

Therefore, weaker versions of stochastic ordering between the tails actually suffice for Theorem 2.2, in particular the concave ordering. We refer the reader to [9, 11] for overviews on different stochastic orders.

We heavily utilize the following criterion for true positive skewness.

Lemma 2.3 (Kovchegov [8], 2020)**.** *Let X be a continuous random variable supported on (L,R) with density function f. Fix $p \in \mathcal{D}$. Suppose there exists $c > 0$ such that $f(\nu_p - c) = f(\nu_p + c)$, and $f(\nu_p - x) > f(\nu_p + x)$ for $x \in (0,c)$ while $f(\nu_p - x) < f(\nu_p + x)$ for $x > c$. Suppose also that $\nu_p - L \le R - \nu_p$. Then ν is increasing at p.*

The following is a special case.

Proposition 2.4 (Kovchegov [8], 2020)**.** *If f is strictly decreasing on its support, then X is truly positively skewed.*

The requirement of strict monotonicity in the proposition can be relaxed, which will be necessary when we consider uniform mixtures in a later section.

Proposition 2.5. *If f is non-increasing on its support (L,R), and there exist any two points $y_1, y_2 \in (L,R)$, $y_1 < y_2$, such that $f(y_1) > f(y_2)$, then X is truly positively skewed.*

Proof. Clearly L is finite, otherwise f could not be a non-increasing density function. Notice that $\nu_p < (L+R)/2$ for all $p \in \mathcal{D}$. Otherwise (1.3) fails to hold since $f(\nu_p + y) \le f(\nu_p - y)$ and $R - \nu_p < \nu_p - L$ by assumption.

Now the existence of $y_1 < y_2$ such that $f(y_1) > f(y_2)$ implies that there exists a non-singleton interval in $(0, \nu_p - L)$ on which $f(\nu_p + y) < f(\nu_p - y)$. Then strict stochastic dominance of $\frac{1}{H_p}y^{p-1}f(\nu_p + y)\mathbf{1}_{(0,R-\nu_p)}(y)$ over $\frac{1}{H_p}y^{p-1}f(\nu_p - y)\mathbf{1}_{(0,\nu_p-L)}(y)$ follows by integrating each density, applying monotonicity of the integral, and using equation (2.1). □

The strict inequalities in Lemma 2.3 can be relaxed in a similar manner.

3. Properties of p-means

In this section, we prove several simple but fundamental properties of p-means and their behavior. Here and throughout this work, let ν_p^X denote the p-mean of a random variable X whenever defined. We use ν_p when the random variable in question is unambiguous. The first result bounds ν_p in the support of X.

Proposition 3.1. *Let X be a random variable with support contained in (L,R) for possibly infinite L, R. Then $\nu_p \in (L,R)$ for all $p \in \mathcal{D}$.*

Proof. If $\nu_p \leq L$, then $E[(\nu_p - X)_+^{p-1}] = 0$ while $E[(X - \nu_p)_+^{p-1}] > 0$, contradiction. Similar if $\nu_p \geq R$. □

Another important fact that will frequently be used is that ν_p is a continuously differentiable function for $p > 1$. This was used implicitly in [8], but we provide a proof here for completeness.

Proposition 3.2. *The map $\nu(p) \equiv \nu_p$ is continuously differentiable on $\mathcal{D} \cap (1, \infty)$.*

Proof. Consider the function $\Phi : \mathbb{R} \times \mathcal{D} \to \mathbb{R}$ given by

$$\Phi(a, p) := E[(X - a)_+^{p-1}] - E[(a - X)_+^{p-1}].$$

Let (X, Ω, P) be our probability space. We can rewrite Φ as

$$\Phi(a, p) = \int_\Omega (X - a)^{p-1} \mathbf{1}\{X > a\} dP - \int_\Omega (a - X)^{p-1} \mathbf{1}\{X < a\} dP.$$

Both integrands are continuously differentiable functions of a and p for all $\omega \in \Omega$ and are dominated by an integrable function by the finiteness of $E|X|^p$. Using the Leibniz integral rule, note that

$$\frac{\partial}{\partial a} \Phi(a, p) = -(p - 1) \left(E[(X - a)_+^{p-2}] + E[(a - X)_+^{p-2}] \right) \tag{3.1}$$

is strictly negative and finite for all $a \in \mathbb{R}$ and $p > 1$. The map $p \mapsto (\nu_p, p)$ is the zero level curve of Φ and so $p \mapsto \nu_p$ is countinuously differentiable by the implicit function theorem. □

One must be careful about the domain of p. The expectations in (3.1) are not necessarily finite for $p \leq 1$. Moreover, it is not generally true that ν_p is continuous at $p = 0$ for unimodal distributions.

When investigating specific distribution families, we will assume that the scale and location parameters are 1 and 0 respectively unless otherwise noted. This is justified by the following, which implies that true positive skewness is preserved under positive affine transformations.

Proposition 3.3. *For any $c, s \in \mathbb{R}$ and $p \in \mathcal{D}_X$, $\nu_p^{cX+s} = c\nu_p^X + s$.*

Proof. It is easy to see that $\mathcal{D}_X = \mathcal{D}_{cX+s} =: \mathcal{D}$. We have for all $p \in \mathcal{D}$ that

$$\begin{aligned} E[((cX + s) - (c\nu_p^X + s))_+^{p-1}] = c^{p-1} E[(X - \nu_p^X)_+^{p-1}] = \\ = c^{p-1} E[(\nu_p^X - X)_+^{p-1}] = E[((c\nu_p^X + s) - (cX + s))_+^{p-1}] \end{aligned}$$

and the result follows. □

Next, we consider the asymptotic behavior of ν_p as $p \to \infty$. This requires X to have finite moments of all orders, which clearly holds if X has bounded support. We consider only continuous random variables, but we can obtain similar results in the discrete case.

Proposition 3.4. *Let X be a continuous random variable with support on (L, R) for finite L, R. Then $\nu_p \to (L + R)/2$ as $p \to \infty$.*

Proof. It suffices to show the statement for X with support on $(0,1)$ and $\nu_p \to 1/2$ by Proposition 3.3.

Let $\epsilon > 0$ be given. If $\epsilon \geq 1/2$ then it holds trivially that $\nu_p \in (1/2-\epsilon, 1/2+\epsilon)$. Thus assume $0 < \epsilon < 1/2$ and suppose for the sake of contradiction that there exists a subsequence ν_{p_k} such that $\nu_{p_k} < 1/2 - \epsilon$ for all $k \in \mathbb{N}$. Note that for $p_k > 1$,

$$\begin{aligned} E[(X-\nu_{p_k})^{p_k-1}\mathbf{1}\{X > \nu_{p_k}\}] &> E[(X-1/2+\epsilon)^{p_k-1}\mathbf{1}\{X > 1-\epsilon\}] \\ &> E[2^{-(p_k-1)}\mathbf{1}\{X > 1-\epsilon\}] \\ &= 2^{-(p_k-1)}P(X > 1-\epsilon) \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} E[(\nu_{p_k}-X)^{p_k-1}\mathbf{1}\{X < \nu_{p_k}\}] &< E[(1/2-\epsilon)^{p_k-1}\mathbf{1}\{X < 1/2-\epsilon\}] \\ &= (1/2-\epsilon)^{p_k-1}P(X < 1/2-\epsilon). \end{aligned} \tag{3.3}$$

From (3.2) and (3.3), we have

$$\frac{E[(\nu_{p_k}-X)^{p_k-1}\mathbf{1}\{X < \nu_{p_k}\}]}{E[(X-\nu_{p_k})^{p_k-1}\mathbf{1}\{X > \nu_{p_k}\}]} < \frac{P(X < 1/2-\epsilon)}{P(X > 1-\epsilon)} \cdot (1-2\epsilon)^{p_k-1} \to 0$$

as $k \to \infty$. This contradicts (1.1). Thus every subsequence has only finitely many points that fall below $1/2-\epsilon$. We can similarly show that every subsequence has only finitely many points that fall above $1/2+\epsilon$. Taking $\epsilon \to 0$, it follows that every subsequence converges to $1/2$, which proves the proposition. □

Suppose instead X has infinite support that is bounded below. We have an analogous proposition if the support is instead bounded above.

Proposition 3.5. *Let X be a continuous random variable with support on (L,∞) for finite L. If X has finite moments of all orders, then $\nu_p \to \infty$.*

Proof. We may assume $L = 0$. Suppose for the sake of contradiction that there exists $M > 0$ such that $\nu_p < M$ for all $p \geq 1$. By Proposition 3.3, we may assume $M = 1/2$. Note that $\nu_p - X < 1/2$ on $\{X < \nu_p\}$, so by the bounded convergence theorem $E[(\nu_p - X)_+^{p-1}] \to 0$ as $p \to \infty$. On the other hand,

$$E[(X-\nu_p)_+^{p-1}] \geq E[(X-\nu_p)^{p-1}\mathbf{1}\{X > \nu_p + 1\}] \geq P(X > 3/2) > 0,$$

hence (1.2) fails to hold for large p and we attain the contradiction. □

A consequence of Proposition 3.5 is that no distribution with support on the positive half-line is truly negatively skewed. Similarly, no distribution with support on the negative half-line is truly positively skewed.

4. Examples: Lévy and Weibull distributions

As an example, we demonstrate the true positive skewness of the Lévy distribution using Lemma 2.3.

Proof of Theorem 1.5. The density function of the Lévy distribution over $x \geq 0$ is

$$f(x) = \frac{1}{x^{2/3}\sqrt{2\pi}} e^{-1/(2x)}.$$

The log density ratio is given by

$$\begin{aligned} g_p(x) &:= \log\left(\frac{f(\nu_p - x)}{f(\nu_p + x)}\right) \\ &= \log(f(\nu_p - x)) - \log(f(\nu_p + x)) \\ &= -\frac{2}{3}\log(\nu_p - x) - \frac{1}{2(\nu_p - x)} + \frac{2}{3}\log(\nu_p + x) + \frac{1}{2(\nu_p + x)}, \end{aligned}$$

To find the extrema, we take $g_p'(x)$ and set it to zero:

$$\frac{2}{3(\nu_p - x)} - \frac{1}{2(\nu_p - x)^2} + \frac{2}{3(\nu_p + x)} - \frac{1}{2(\nu_p + x)^2} = 0,$$

Equivalently,

$$4(\nu_p - x)(\nu_p + x)^2 - 3(\nu_p + x)^2 + 4(\nu_p + x)(\nu_p - x)^2 - 3(\nu_p - x)^2 = 0.$$

We can simplify this equation further to obtain a quadratic equation in x,

$$(3 + 4\nu_p)x^2 + 3\nu_p^2 - 4\nu_p^3 = 0.$$

One can verify that the only positive root of this equation is

$$x^* = \sqrt{\frac{4\nu_p^3 - 3\nu_p^2}{3 + 4\nu_p}}.$$

Observe that $\lim_{x \to \nu_p -} g_p(x) = -\infty$ and $g_p(0) = 0$, with the only positive extremum of g_p being x^*. By (1.3), we cannot have $g_p(x) \leq 0$ for all $x > 0$. This implies x^* maximizes g. We have that $g(x)$ is increasing on $[0, x^*)$ and decreasing on (x^*, ν_p). The intermediate value theorem implies the existence of a point $c \in (x^*, \nu_p)$ such that $g(c) = 1$. This satisfies the conditions in Lemma 2.3 and holds for all $p \in \mathcal{D}$, so true positive skewness follows. □

Next, we demonstrate necessary and sufficient conditions on the shape parameter of a Weibull distribution for true positive skewness, again using Lemma 2.3. The Rayleigh distribution, as a special case of Weibull for shape parameter $k = 2$, is truly positively skewed as an immediate consequence.

The density function for shape parameter $k > 0$ is given by

$$f(x) = kx^{k-1}e^{-x^k}\mathbf{1}_{(0,\infty)}(x).$$

Moments of all orders are finite. Formulas for the median and mode are also well-known for $k > 1$:

$$\nu_0 = \left(\frac{k-1}{k}\right)^{1/k}, \tag{4.1}$$

$$\nu_1 = (\log 2)^{1/k}. \tag{4.2}$$

Notice that $\nu_0 < \nu_1$ if and only if $k < (1 - \log 2)^{-1}$. We refer the reader to [10] for general properties of the Weibull distribution.

Proof of Theorem 1.6. The "only if" part follows from (4.1) and (4.2). For $k \leq 1$, f is strictly decreasing, so we are done by Proposition 2.4. Consider $k > 1$.

Let $p \geq 1$ be given and consider the log density ratio

$$g_p(x) := \log\left(\frac{f(\nu_p - x)}{f(\nu_p + x)}\right)$$

defined on $x \in [0, \nu_p)$. Plugging in the Weibull density function for f, we have

$$g_p(x) = (k-1)\log(\nu_p - x) - (k-1)\log(\nu_p + x) - (\nu_p - x)^k + (\nu_p + x)^k;$$
$$g_p'(x) = k(\nu_p - x)^{k-1} + k(\nu_p + x)^{k-1} - \frac{k-1}{\nu_p - x} - \frac{k-1}{\nu_p + x}.$$

Roots of g_p' remain the same after we multiply by $(\nu_p - x)(\nu_p + x)$, so we consider the roots of

$$\begin{aligned} g_p^*(x) &:= (\nu_p - x)(\nu_p + x)g_p'(x) = \\ &= k(\nu_p^2 - x^2)\left[(\nu_p - x)^{k-1} + (\nu_p + x)^{k-1}\right] - 2(k-1)\nu_p. \end{aligned} \tag{4.3}$$

for $x \in [0, \nu_p)$. This domain means the binomial series for $(\nu_p - x)^{k-1}$ and $(\nu_p + x)^{k-1}$ converge, hence

$$(\nu_p - x)^{k-1} + (\nu_p + x)^{k-1} = 2\sum_{n=0}^{\infty}\binom{k-1}{2n}\nu_p^{k-2n-1}x^{2n},$$

where $\binom{a}{b} := \frac{a(a-1)\ldots(a-b+1)}{b!}$ is the generalized binomial coefficient.

Expanding,

$$\begin{aligned} g_p^*(x) &= 2k\sum_{n=0}^{\infty}\binom{k-1}{2n}\nu_p^{k-2n+1}x^{2n} - 2k\sum_{n=1}^{\infty}\binom{k-1}{2n-2}\nu_p^{k-2n+1}x^{2n} - 2(k-1)\nu_p \\ &= 2k\sum_{n=1}^{\infty}\left(\binom{k-1}{2n} - \binom{k-1}{2n-2}\right)\nu_p^{k-2n+1}x^{2n} + 2k\nu_p^{k+1} - 2(k-1)\nu_p. \end{aligned} \tag{4.4}$$

where

$$\binom{k-1}{2n} - \binom{k-1}{2n-2} = \frac{k(k-1)\ldots(k-2n+2)(k+1-4n)}{(2n)!}. \tag{4.5}$$

We analyze the sign changes of the coefficients in (4.4) to determine the number of real positive roots of g_p^* by splitting into several cases for the value of k.

Case 1: $2 < k < 3$. Then $k, k-1, k-2$ are positive and $k-3, \ldots, k-2n+2$ are negative. There are an even number (possibly zero, if $n = 1, 2$) of negative factors in $k(k-1)\ldots(k-2n+2)$, so it is positive. Also note $k+1-4n < 0$ for all $n \geq 1$, hence (4.5) is negative for all $n \geq 1$.

For the series expression $g_p^*(x) = \sum_{m=0}^{\infty} a_m x^m$, we have shown that $a_m < 0$ for non-zero even m and $a_m = 0$ for odd m. By Descartes' rule of signs for infinite series, g_p^* has no positive real roots if $a_0 \leq 0$ and has at most one positive real root if $a_0 > 0$.

Suppose $a_0 \leq 0$ such that g_p^* has no positive real roots. By extension, g_p' has no positive real roots, so g_p has no positive extrema and is strictly monotonic on $[0, \nu_p)$. Since $g_p(0) = 0$

and $\lim_{x\to\nu_p-} g_p(x) = -\infty$, then g_p is strictly negative on $[0, \nu_p)$, hence $f(\nu_p - x) < f(\nu_p + x)$ for all $x \in (0, \nu_p)$. By monotonicity of the integral, this contradicts (1.3). Thus $a_0 > 0$ and g_p^* has exactly one real root $c' \in (0, \nu_p)$.

We know g_p' also has exactly one positive real root since it shares roots with g_p^*. Moreover, $g_p'(0) = a_0 > 0$ and $g_p' \to -\infty$ as $x \to \nu_p$, so $g_p'(x) > 0$ for $x \in (0, c')$ and $g_p'(x) < 0$ for $x \in (c', \nu_p)$. Thus c' is a maximum and the only positive extremum.

Since $g_p(0) = 0$ and $\lim_{x\to\nu_p-} g_p(x) = -\infty$, then $g_p(c') > 0$. The intermediate value theorem implies that g_p has a root $c \in (c', \nu_p)$. Moreover, since c' is the only positive extremum of g_p, then $g_p(x) > 0$ for $x \in (0, c)$ and $g_p(x) < 0$ for $x \in (c, \nu_p)$. This satisfies the conditions of Lemma 2.3.

The above holds for all $p \geq 1$. Moreover, the positivity of

$$a_0 = 2k\nu_p^{k+1} - 2(k-1)\nu_p = 2\nu_p(k\nu_p^k - k + 1)$$

in our above calculations implies $\nu_p > \left(\frac{k-1}{k}\right)^{1/k} = \nu_0$ for all $p \geq 1$, and we are done.

Case 2: $1 < k < 2$. The inequality $k + 1 - 4n < 0$ still holds for all $n \geq 1$, but we now have $k, k-1 > 0$ and $k-2, \dots, k-2n+2 < 0$. If $n > 1$, then there are an odd number of negative factors in $k(k-1)\dots(k-2n+2)$, so the numerator of (4.5) is positive. If $n = 1$, then the numerator is simply $k(k-3) < 0$.

Thus our coefficients in $g_p^*(x) = \sum_{m=0}^{\infty} a_m x^m$ are positive for even $m > 2$ and zero for odd m with $a_2 < 0$. By Descartes' rule of signs, g_p^* has at most two positive real roots if $a_0 > 0$ and at most one if $a_0 \leq 0$.

Suppose $a_0 \leq 0$. If g_p^* has a single positive real root, then $\lim_{x\to\nu_p-} g_p^*(x) > 0$. But by continuity this limit tends to $-2(k-1)\nu_p < 0$. If instead g_p^* has no positive real roots, then by the same argument above we contradict equation (9) in [**?**]. Thus $a_0 > 0$. In this case, if g_p^* has zero or two positive real roots, then again $\lim_{x\to\nu_p-} g_p^*(x) > 0$ and we reach a contradiction. Hence g_p^* has exactly one positive real root. We may now conclude in the same fashion as in Case 1.

Case 3: $3 < k < \frac{1}{1-\log 2}$. We again look at the numerator of (4.5). If $n = 1$, the numerator is $k(k-3) > 0$. If $n = 2$, the numerator is $k(k-1)(k-2)(k-7) < 0$. If $n > 2$, then $k(k-1)\dots(k-2n+2)$ has positive factors $k, \dots, k-3$, and an odd number of negative factors $k-4, \dots, k-2n+2$. Additionally, $k+1-4n < 0$ when $n > 2$, so the numerator of (4.5) is positive.

Our coefficients in $g_p^*(x) = \sum_{m=0}^{\infty} a_m x^m$ are zero for odd m, positive for even $m \geq 6$, negative for $m = 4$, and positive for $m = 2$. This yields two sign changes and hence at most two positive real roots of g_p^* if $a_0 > 0$. Recall that $a_0 > 0$ if and only if $\nu_p > \nu_0$.

Since g_p^* is negative at the right limit of its domain, then it must have exactly one positive real root if $\nu_p > \nu_0$. If this holds for some p_0, then ν_p is increasing at $p = p_0$ by Lemma 2.3. It would then follow that ν_p must be increasing for all $p \geq p_0$ since ν_p is continuous. Since $k < \frac{1}{1-\log 2}$, then $\nu_1 > \nu_0$ by (4.1) and (4.2); hence ν_p is increasing for all $p \geq 1$ and we are done.

Case 4: $k = 2$. We can plug $k = 2$ directly into (4.3) and solve $g_p^* = 0$ to obtain the roots

$$x^2 = \nu_p^2 - \frac{1}{2} = \nu_p^2 - \nu_0^2$$

The argument in Case 1 can be adapted to show that $\nu_p > \nu_0$ for all $p \geq 1$, hence g_p^* has exactly one positive root $\sqrt{\nu_p^2 - \nu_0^2}$. We conclude as in Case 1 that ν_p is increasing for all $p \geq 1$.

Case 5: $k = 3$. Similarly, we plug $k = 3$ into (4.3) and obtain the roots of g_p^*:

$$x^4 = \nu_p^4 - \frac{2}{3}\nu_p = \nu_p(\nu_p^3 - \nu_0^3)$$

One can again check that there is exactly one positive root for any $p \geq 1$. The rest of the argument is the same. ☐

Remark 4.1. Applying Descartes' rule of signs for infinite series is permissible as long as the number of sign changes is finite, see [3]. Moreover, it is generally untrue in the infinite case that the difference between the number of positive roots and the number of sign changes is even.

The parameter region for which the Weibull distribution is truly positively skewed is consistent with previous findings on the Weibull distribution's positive skewness; see [14] for an overview. Indeed, the upper bound $(1 - \log 2)^{-1}$ is derived from requiring the median to be greater than the mode. Our result essentially shows that the median-mode inequality is sufficient to conclude positive skewness for the Weibull distribution under any of Pearson's skewness coefficients.

5. Transformations, Sums, and Products

In this section, we investigate the effect transformations, sums, and products have on truly skewed random variables.

5.1. **Transformations of Truly Skewed Random Variables.**

Before tackling operations on multiple random variables, we begin with transformations of a single random variable. By Proposition 3.3, we already know that positive affine transformations of a truly positively skewed random variable preserve true positive skewness.

Consider a continuous random variable X with density f_X with support $(0, R)$ for possibly infinite R. Let $Y = u(X)$ where u is a measurable, invertible function defined on $(0, R)$. Let $w = u^{-1}$. By a change of variables, the density of Y is

$$f_Y(y) = f_X(w(y)) \cdot |w'(y)| \tag{5.1}$$

We prove the following theorem, again using Lemma 2.3.

Proof of Theorem 1.7. Note that $\mathcal{D}_X = \mathcal{D}_Y = [1, \infty)$ since both X and Y have bounded support. We write $\mathcal{D}$ for this domain. Let f_Y be the density of Y. It suffices to show that

$$\log\left(\frac{f_Y(\nu_p - y)}{f_Y(\nu_p + y)}\right) \tag{5.2}$$

changes sign exactly once on the interval $(0, \infty)$, for all $p \in \mathcal{D}$.

First note that as y increases, $\nu_p - y$ approaches 0. Once $y > \nu_p$, $f_Y(\nu_p - y) = 0$. Thus, on the interval $y \in (\nu_p, \infty)$,

$$\log\left(\frac{f_Y(\nu_p - y)}{f_Y(\nu_p + y)}\right) = -\infty$$

Moreover, we have $\{\nu_p - y\} \in [0, \nu_p)$, and $\{\nu_p + y\} \in [\nu_p, 2\nu_p)$. We expand f_Y:

$$\begin{aligned}\log\left(\frac{f_Y(\nu_p - y)}{f_Y(\nu_p + y)}\right) &= \log\left(\frac{-f_X[w(\nu_p - y)]w'(\nu_p - y)}{-f_X[w(\nu_p + y)]w'(\nu_p + y)}\right) \\ &= \log\left(\frac{f_X[w(\nu_p - y)]}{f_X[w(\nu_p + y)]}\right) + \log\left(\frac{w'(\nu_p - y)}{w'(\nu_p + y)}\right) \quad (5.3)\end{aligned}$$

We show that the left side of (5.3) is positive. The function w is increasing, so

$$w(\nu_p - y) < w(\nu_p + y)$$

Because f_X is strictly decreasing,

$$f_X[w(\nu_p - y)] > f_X[w(\nu_p - y)] \quad (5.4)$$

Now, we show that the right side of (5.3) is positive. Because u is convex and increasing, w is concave, and therefore w' is decreasing. It follows that

$$w'(\nu_p - y) > w'(\nu_p + y)$$

Therefore (5.2) is strictly positive on $(0, \nu_p]$ and strictly negative on (ν_p, ∞). This satisfies Lemma 2.3 for all $p \in \mathcal{D}$, so we obtain true positive skewness. □

We have a similar result if f_X is strictly increasing.

Corollary 5.1. *Let X be a continuous random variable with strictly increasing density function f_X with bounded support. If u is a measurable, strictly decreasing, and convex function on the support of X, then $Y = u(X)$ is truly positively skewed.*

Proof. The proof is analogous to that of Theorem 1.7. □

Example 5.2. Consider exponential random variable X with density function

$$f_X(x) = e^{-x}\mathbf{1}_{(0,\infty)}$$

and $u(x) = x^2$, where $u(x)$ is clearly strictly increasing and convex. By the change of variable formula, random variable $Y = u(X)$ has probability density function

$$f_Y(y) = e^{-\sqrt{y}}\frac{1}{2\sqrt{y}}$$

which is truly positively skewed by Theorem 1.7. It also happens to be decreasing, which we know implies true positive skewness.

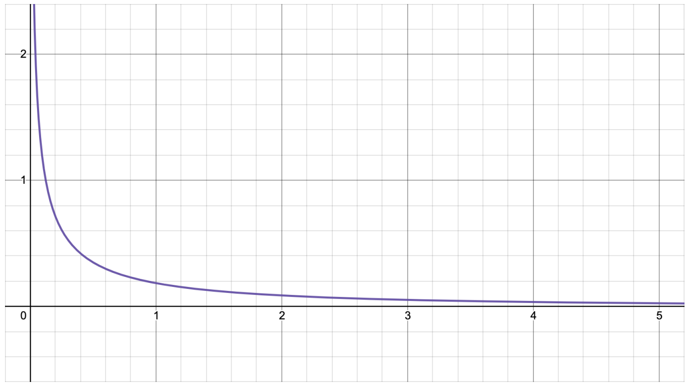


FIGURE 5.1. $f_Y(y) = e^{-\sqrt{y}} \frac{1}{2\sqrt{y}}$

5.2. **Sums of Truly Skewed Random Variables.** Let X and Y be two continuous random variables with density functions f_X and f_Y. The density function of their sum $Z = X + Y$ is given by

$$f_Z(z) = \int f_X(x) f_Y(z - x) dx.$$

One can think about the convolution of two random variables as a representation of the area under the intersection of both probability density functions as one moves over the other (see Figure **??**).

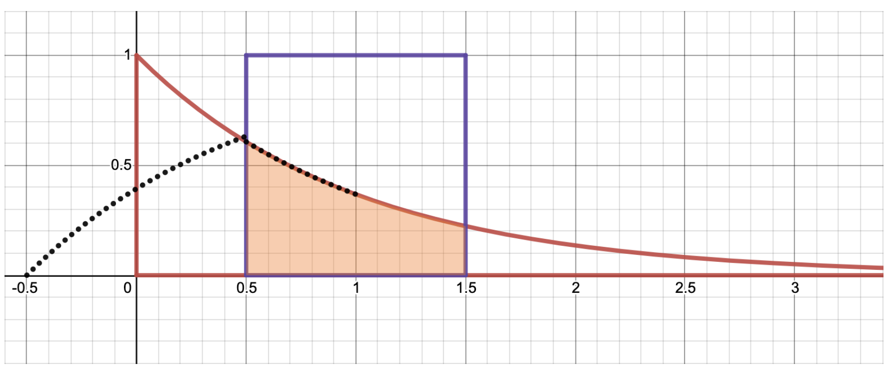


FIGURE 5.2. Convolution

Kovchegov proposes an intuitive conjecture that true positive skewness is preserved under summation.

Conjecture 5.3. *If X and Y are truly positively skewed, then so is $X + Y$.*

However, this statement is false in general. We present two counterexamples, one for discrete random variables and one for continuous.

Example 5.4. Let X and Y be independent Bernoulli random variables with parameter $1/3$. We have shown elsewhere that Bernoulli distributions with parameter less than $1/2$ are truly positively skewed. The sum $Z = X + Y$ is a 2-binomial distribution with parameter 1/3. The median of Z is unique and equal to $[n/3] = 1$ due to Kaas & Buhrman [6, Theorem 1]. However, the mean of Z is $n/3 = 2/3$. Thus $\nu_2 < \nu_1$, so Z is not truly positively skewed.

Example 5.5. Set $\lambda \in (\frac{1}{2}, 1)$ and define the density function

$$f(x) := \begin{cases} \lambda & 0 < x \leq 1 \\ 1-\lambda & 1 \leq x < 2 \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

Let X and Y be independent random variables with density f, so by Proposition 2.5, X and Y are truly positively skewed. However, it is not always true that $X + Y$ is truly positively skewed.

The density function of $Z = X + Y$ is given by the convolution of f with itself. One can check that

$$f_Z(x) := (f * f)(x) = \begin{cases} \lambda^2 x & 0 < x < 1 \\ \lambda(2-3\lambda)x - 2\lambda(1-2\lambda) & 1 \leq x < 2 \\ (1-\lambda)(1-3\lambda)x - 2(1-\lambda)(1-4\lambda) & 2 \leq x < 3 \\ -(1-\lambda)^2 x + 4(1-\lambda)^2 & 3 \leq x < 4 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that f_Z is strictly increasing on $(0,1)$ and strictly decreasing on $(2,4)$ regardless of the value of λ. On $(1,2)$, f_Z is strictly increasing if $\lambda \in (\frac{1}{2}, \frac{2}{3})$ and strictly decreasing if $\lambda \in (\frac{2}{3}, 1)$.

Take $\lambda = \frac{3}{5}$. Since f is continuous, there exists a unique median ν_1 which by solving a straightforward integral equation can be shown to be

$$\nu_1 = \frac{2(2\lambda - 1)}{3\lambda - 2} - \sqrt{\frac{4\lambda^3 - 2\lambda^2 - 3\lambda + 2}{\lambda(3\lambda-2)^2}}$$

when $\lambda > \frac{2}{3}$. For our choice of λ, we have $\nu_1 \approx 1.786$. Recall that ν is increasing at p if and only if (2.2) holds. Simply taking $p = 1$ and using our calculated value for the median, we find that

$$\int_0^{4-\nu_1} \log y \; f_Z(\nu_1 + y) dy - \int_0^{\nu_1} \log y \; f_Z(\nu_1 - y) dy \approx -0.000699,$$

hence ν_p is decreasing at $p = 1$ and so $X + Y$ is not truly positively skewed. Numerical computations of ν_p for small values of p confirm our findings.*

There are certain classes of functions for which true positive skewness is indeed preserved under summation. One simple example is that of decreasing linear densities of the form

$$f(x) = h - \frac{h^2 x}{2}, \; x \in [0, 2/h]. \tag{5.6}$$

for $h > 0$. Visually, these densities are right-facing right triangles with corner at the origin and height h. They are clearly truly positively skewed by Proposition 2.4, and we can prove

*Computations were performed using the software Wolfram Mathematica.

that the sum of any two random variables with densities of this form is also truly positively skewed. The proof is similar to those for the Lévy and Weibull distributions and is rather tedious, so we leave it for Appendix A.

Proposition 5.6. *If X and Y are independent random variables with density functions of the form* (5.6)*, then $X+Y$ is truly positively skewed.*

This suggests the following modified conjecture.

Conjecture 5.7. *If X and Y are independent and have strictly decreasing and continuous density functions, then $X+Y$ is truly positively skewed.*

5.3. **Products of Truly Skewed Random Variables.** Given that sums generally do not preserve true skewness, we next consider products of truly skewed random variables. Let X and Y be two independent random variables with density functions f_X and f_Y. The density function of their product $Z = XY$ is given by

$$f_Z(z) = \int \frac{1}{x} f_X(x) f_Y(z/x) dx \tag{5.7}$$

We consider several examples to motivate the investigation of product distributions.

Example 5.8. Consider independent log-normal random variables X and Y with density functions

$$f_X(x) = f_Y(x) = \frac{1}{2\pi x} \exp\left(-\frac{(\log x)^2}{2}\right) \mathbf{1}_{[0,\infty)}(x)$$

It is a special result of log-normal random variables that the product $Z = XY$ is also log-normally distributed. The log-normal distribution is truly positively skewed [8], hence the product distribution is truly positively skewed.

Example 5.9. Consider independent X and Y uniformly distributed on $(0,1)$. The product $Z = XY$ has density

$$f_Z(z) = \int_z^1 \frac{1}{x} dx = -\log(z) \mathbf{1}_{(0,1)}(z)$$

Since f_Z is decreasing, then Z is truly positively skewed by Proposition 2.4.

Example 5.10. Consider independent X and Y with increasing density functions

$$f_X(x) = \frac{x}{2} \mathbf{1}_{(0,2)}(x), \qquad f_Y(x) = \frac{x^3}{4} \mathbf{1}_{(0,2)}(x)$$

The product $Z = XY$ has density

$$f_Z(z) = \frac{3z}{8} \int_{z/2}^2 dx = \frac{z}{4} - \frac{z^3}{64}$$

for $z \in (0,4)$. One can numerically approximate the median $\nu_1 \approx 2.165$ and the mode $\nu_0 \approx 2.309$. Thus Z is not truly positively skewed.

In Example 5.9, we see that the $\frac{1}{x}$ factor in (5.7) dominates the density functions, which are not increasing. However, in Example 5.10, one of the functions in the integrand is quadratic, which in turn dominates the $\frac{1}{x}$ factor. This motivates the following conjecture.

Conjecture 5.11. *Let X and Y be independent random variables with monotonic density functions f_X and f_Y that have "appropriately" bounded rate of increase. Then XY is truly positively skewed.*

What this "appropriate" bound on the rate of increase needs to be determined.

6. True Skewness under Weak Limits

It is reasonable to conjecture that true skewness is preserved under uniform, if not pointwise, limits of distribution functions since Lemma 2.3 implies that true skewness is, in effect, a feature of a random variable's density function. However, this is not always true. Let $X_n \sim \text{Gamma}(n, \lambda)$ be a sequence of independent gamma random variables with identical second parameter such that each X_n can be expressed as a sum of n i.i.d. exponential random variables. We know from [8] that the X_n's are truly positively skewed, but the central limit theorem implies that their weak limit, after normalization, is Gaussian and thus symmetric.

Therefore, we introduce the notion of *true non-negative skewness* to refer to a random variable whose p-means are non-decreasing, i.e., $d\nu_p/dp \geq 0$, as opposed to the strict increasingness required by true positive skewness. Notice that truly positively skewed as well as symmetric distributions are truly non-negatively skewed. To prove Theorem 1.8, we rely on the following well-known result.

Lemma 6.1 ([2], Theorem 25.12)**.** *If $X_n \Rightarrow X$ and $\{X_n^r\}_{n\in\mathbb{N}}$ is uniformly integrable for some $r > 1$, then $E[X_n^p] \to E[X^p]$ for all $1 \leq p \leq r$. In particular, $\{X_n^p\}_{n\in\mathbb{N}}$ is uniformly integrable.*

By uniform integrability we mean in the typical sense that, for any $\epsilon > 0$, there exists $K > 0$ such that $E[|X_n|\mathbf{1}\{|X_n| \geq K\}] < \epsilon$ for all $n \in \mathbb{N}$. We prove our own simple result.

Lemma 6.2. *If $\{X_n^r\}_{n\in\mathbb{N}}$ is uniformly integrable for some $r > 1$, then for any $a \in \mathbb{R}$, $\{(X_n - a)^r\}_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof. Let $\epsilon > 0$ be given. Choose large $K \geq |a|^r$ such that

$$E[|X_n|^r\mathbf{1}\{|X_n|^r \geq K\}] < \epsilon/2^r$$

for all $n \in \mathbb{N}$. Let $K' = (K^{1/r} + |a|)^r$, so $|X_n - a|^r \geq K'$ implies $|X_n|^r \geq K$ by the triangle inequality. For $|X_n| \geq |a|$, we obtain

$$|X_n - a|^r \leq (|X_n| + |a|)^r \leq 2^r|X_n|^r.$$

Since $K \geq |a|^r$, then $|X_n|^r \geq K \implies |X_n| \geq |a|$, so it follows that

$$E[|X_n - a|^r\mathbf{1}\{|X_n - a|^r \geq K'\}] \leq 2^r E[|X_n|^r\mathbf{1}\{|X_n|^r \geq K\}] < \epsilon$$

for all $n \in \mathbb{N}$. □

The continuous mapping theorem implies that if $X_n \Rightarrow X$, then $(X_n - a)^r \Rightarrow (X - a)^r$, so we arrive at the following corollary via Lemma 6.1 and Lemma 6.2.

Corollary 6.3. *If $X_n \Rightarrow X$ with $\{X_n^r\}_{n\in\mathbb{N}}$ uniformly integrable for some $r > 1$, then*

$$E|X_n - a|^p \to E|X - a|^p$$

for all $1 \leq p \leq r$.

Suppose a_n is a sequence of numbers converging to a. For $\epsilon > 0$, there exists large N such that for all $n \geq N$, we have $||X - a_n|^r - |X - a|^r| < \epsilon/2$ by continuity and hence $|E|X - a_n|^r - E|X - a|^r| < \epsilon/2$ by Jensen's inequality. With Corollary 6.3 we get the next result.

Corollary 6.4. *Let $a_n \to a$ and $X_n \Rightarrow X$ with $\{X_n^r\}_{n\in\mathbb{N}}$ uniformly integrable for some $r > 1$. Then*

$$E|X_n - a_n|^p \to E|X - a|^p$$

for all $1 \leq p \leq r$.

The theorem now follows in typical analytical style.

Proof of Theorem 1.8. For fixed $p \in \mathcal{D}$, define

$$f_n(a) := E|X_n - a|^p,\ a \in \mathbb{R}; \qquad \nu_n := \arg\min_{a\in\mathbb{R}} f_n(a);$$

$$f(a) := E|X - a|^p,\ a \in \mathbb{R}; \qquad \nu := \arg\min_{a\in\mathbb{R}} f(a);$$

so that ν_n and ν are the p-means of X_n and X respectively by Definition 1.1.

Suppose for now that the ν_n's are contained in a compact interval, so every subsequence ν_{n_k} has a limit point ν^*. Since $f_{n_k}(\nu_{n_k}) \leq f_{n_k}(a)$ for all $a \in \mathbb{R}$, then by taking $k \to \infty$ we obtain $f(\nu^*) \leq f(a)$ for all $a \in \mathbb{R}$ via Corollaries 6.3 and 6.4. Then ν^* minimizes f, and since the minimizer of f is unique, we actually have $\nu^* = \nu$. Every subsequence of ν_n converges to ν, hence $\nu_n \to \nu$.

Consider ν_n and ν as functions of p, so ν_n converges pointwise to ν. True non-negative skewness of X_n implies ν_n is non-decreasing, and the pointwise limit of monotone functions is monotone, hence ν is non-decreasing.

It remains to show that the ν_n's are contained in a compact interval. Suppose otherwise, so we have a subsequence $\nu_{n_k} \to \infty$ as $k \to \infty$. (The argument is similar if $\nu_{n_k} \to -\infty$.) Corollary 6.3 gives the pointwise convergence $f_n \to f$, so we fix $\epsilon > 0$ and choose large enough K such that $\nu_{n_k} > \nu$ and $|f_{n_k}(\nu) - f(\nu)| < \epsilon$ for all $k \geq K$. Since f_{n_k} is strictly convex with minimizer ν_{n_k}, then $f_{n_k}(a) < f(\nu) + \epsilon$ for all $a \in [\nu, \nu_{n_k}]$.

Clearly $f \to \infty$ as $a \to \infty$, so choose $x > \nu_{n_K} > \nu$ large enough such that $f(x) > f(\nu) + 2\epsilon$. For any $N > 0$, there exists $k > K$ large enough such that $n_k > N$ and $\nu_{n_k} > x$ such that $f_{n_k}(x) < f(\nu) + \epsilon < f(x) - \epsilon$ as shown above. This contradicts the pointwise convergence $f_{n_k}(x) \to f(x)$, and so concludes the proof. □

If we strengthen the conditions of Theorem 1.8, we can conclude that the limiting distribution is truly positively skewed, which is stated in Corollary 1.9. As expected, this requires each X_n to be truly positively skewed.

Proof of Corollary 1.9. The proof is the same as Theorem 1.8, but also notice that for any $p, q \in \mathcal{D}$, $p < q$, we have $\nu_n(q) - \nu_n(p) > c(q - p)$ for all $n \in \mathbb{N}$ by the mean value theorem. We showed that $\nu_n \to \nu$ pointwise, hence taking limits, we find $\nu(q) - \nu(p) \geq c(q - p) > 0$. It follows easily that for all $p \in \mathcal{D}$,

$$\nu'(p) := \lim_{q\downarrow p} \frac{\nu(q) - \nu(p)}{q - p} \geq c > 0.$$

□

Remark 6.5. The phrasing in Theorem 1.8 that $\{X_n^p\}_{n\in\mathbb{N}}$ be uniformly integrable for all $p \in \mathcal{D}$ matters only when $\mathcal{D}$ is unbounded. If $\mathcal{D}$ is bounded, then we only need to show that $\{X_n^s\}_{n\in\mathbb{N}}$ is uniformly integrable for some s such that $s > p$ for all $p \in \mathcal{D}$.

An application of Theorem 1.8 regards parameter regions of true skewness for distribution families. If we show that a distribution family is truly positively skewed in an open interval in one of its parameters, then we can conclude also that the distribution is truly non-negatively skewed on the closure of this open interval. For example, the theorem applied to the Weibull distribution, using the parameter regions established in Theorem 1.6, implies true non-negative skewness when $k = (1 - \log 2)^{-1}$.

7. Criterion Using Logarithmic Derivatives

Theorem 1.10 establishes a novel criterion for true positive skewness that does not rely on the p-means of a distribution other than its mode and median. It also does not require knowledge of the density f expressed in terms of elementary functions, which has conveniently been provided in each of the specific distributions previously examined. In particular, this theorem may have applications in numerically checking true positive skewness for stable distributions, for which very little descriptive information is known in general, given specific parameter values.

We begin with some notation. Assume X is a continuous unimodal random variable with density f. For $p \in \mathcal{D}$, define

$$h_p(c) := f(\nu_p + c) - f(\nu_p - c), \;\; c \in [0, \nu_p]$$

and

$$\mathcal{S}_p := \{c > 0 : h_p(c) > 0\}.$$

If $\mathcal{S}_p$ is non-empty, then its infimum $c_p := \inf \mathcal{S}_p$ exists and is non-negative. Note that if f is continuous, then so is h_p. Then $\mathcal{S}_p$ is the preimage of an open set under h_p, so $\mathcal{S}_p$ is also open and $c_p \notin \mathcal{S}_p$. Since $h_p(0) = 0$, then continuity implies

$$h_p(c_p) = 0. \tag{7.1}$$

Similarly, if f is differentiable, then so is h_p. Because $h_p(c) \leq 0$ for all $c < c_p$, then

$$h_p'(c_p) \geq 0. \tag{7.2}$$

Remark 7.1. Before we prove Theorem 1.10, we discuss ways in which certain conditions can be relaxed, at the potential cost of practical ease. In particular, the following require one to compute c_1.

(a) The condition $\nu_1 > (\nu_0 + \theta_2)/2$ serves to guarantee that f is convex at $\nu_p + c_p$ for all p, but the required convexity can certainly be achieved for a weaker lower bound on ν_1. Indeed, one can see in the proof below that $\nu_1 + c_1 > \theta_2$ is sufficient; we obtain (7.3) by using Lemma 7.5. Note that this is not *always* a weaker lower bound.

(b) The quantity ν_0 in conditions (1) and (2) can be replaced by $\nu_1 - c_1$.[†] As we show later, $\nu_1 - c_1 < \nu_0$, so actually this replacement creates a stronger condition on the lower bound of f'/f to the left of the mode and a weaker condition on the lower bound of f'/f to the right of the mode. This replacement is useful when the density has a steeper right tail.
(c) Similarly, condition (2) only needs to hold on $(\nu_1 + c_1, \infty)$.

To prove the theorem, we will require several lemmas.

Lemma 7.2. *The density f is convex on $(0, \theta_1) \cup (\theta_2, \infty)$, concave on (θ_1, θ_2), strictly increasing on $(0, \nu_0)$, and strictly decreasing on (ν_0, ∞).*

Proof. Since f is positive on its support, then it must be increasing on $(0, \nu_0)$. If f is concave on $(0, \theta_1)$, then it is convex on (θ_1, θ_2), contradicting the fact that $\nu_0 \in (\theta_1, \theta_2)$. The convexity part of the lemma follows, and since $\theta_2 > \nu_0$ and f is unimodal, then f is decreasing on (ν_0, ∞). Moreover, since $f' > 0$ near 0 and $f' < 0$ near infinity, then f' has an odd number of zeroes. Integrability of f implies $f' \to 0$, hence f' is non-zero everywhere on $(0, \infty)$ except at ν_0, otherwise f would have at least three inflection points. This proves the strictness part. □

Since f is C^2 on $(0, \infty)$, then (7.1) and (7.2) hold. Moreover, h_p is C^2. Let

$$\mathcal{D}^* := \{p \in \mathcal{D} : \nu_p > (\nu_0 + \theta_2)/2\}.^{\ddagger}$$

By assumption $1 \in \mathcal{D}^*$. We show that membership in $\mathcal{D}^*$ is sufficient to make (7.2) a strict inequality.

Lemma 7.3. *$h_p'(c_p) > 0$ for all $p \in \mathcal{D}^*$.*

Proof. First we locate $\nu_p + c_p$ and $\nu_p - c_p$. Note that $f(\nu_p - c)$ is increasing and $f(\nu_p + c)$ is decreasing for $c \in (0, \nu_p - \nu_0)$. The fact that h_p is negative near 0 and $h_p(c_p) = 0$ implies $c_p > \nu_p - \nu_0$. By assumption $\nu_p > (\nu_0 + \theta_2)/2$, so it follows that

$$\nu_p + c_p > \theta_2. \tag{7.3}$$

On the other hand, we easily have $\nu_p - c_p > 0$, otherwise $h_p(c_p) = f(\nu_p + c_p) > 0$. If $\nu_p - c_p > \nu_0$, then $f(\nu_p - c_p) > f(\nu_p + c_p)$ by Lemma 7.2, again contradicting $h_p(c_p) = 0$. Thus

$$0 < \nu_p - c_p < \nu_0. \tag{7.4}$$

Now suppose for the sake of contradiction that $h_p'(c_p) = 0$. Either $\nu_p - c_p \in (0, \theta_1)$ or $\nu_p - c_p \in [\theta_1, \nu_0)$ by (7.4). If the latter holds, then $f''(\nu_p - c_p) < 0$, and (7.3) implies $f''(\nu_p + c_p) > 0$. Note that $h_p''(c_p) = f''(\nu_p + c_p) - f''(\nu_p - c_p)$, hence $h_p''(c_p) > 0$. By continuity of h_p'' and (7.1), h_p is strictly convex and thus positive in a neighborhood of c_p (excluding c_p itself), contradicting the minimality of c_p in $\mathcal{S}_p$.

[†] This makes the proof significantly lengthier; in fact, Lemmas 7.3, 7.4, and 7.5 are otherwise unnecessary. In the proof below we present the most general argument.

[‡] If as in the remark we use the condition $\nu_1 + c_1 > \theta_2$ instead of $\nu_1 > (\nu_0 + \theta_2)/2$, then instead let $\mathcal{D}^* := \{p \in \mathcal{D} : \nu_p \geq \nu_1 + c_1\}$. The following arguments still apply with trivial modifications.

Suppose instead that $\nu_p - c_p \in (0, \theta_1)$. Then from conditions (1) and (2) and equations (7.3) and (7.4), we have the inequalities $\nu_0 f'(\nu_p - c_p) > f(\nu_p - c_p)$ and $\nu_0 f'(\nu_p + c_p) > -f(\nu_p + c_p)$. It follows that $\nu_0 h_p'(c_p) > -h_p(c_p) = 0$ and we obtain a contradiction. □

Next, we prove some properties of c_p.

Lemma 7.4. *The map $p \mapsto c_p$ is continuously differentiable in $\mathcal{D}^*$.*

Proof. Define $\psi : \mathcal{D}^* \times \mathbb{R}_+$ by $\psi(p, c) := h_p(c)$. By Proposition 3.2, ψ is differentiable and has partial derivatives

$$\frac{\partial}{\partial c}\psi(p, c) = f'(\nu_p + c) + f'(\nu_p - c) = h_p'(c),$$

$$\frac{\partial}{\partial p}\psi(p, c) = (f'(\nu_p + c) - f'(\nu_p - c))\frac{d\nu_p}{dp},$$

both of which are jointly continuous in p and c. Thus ψ is continuously differentiable. By Lemma 7.3, $\frac{\partial}{\partial c}\psi(p, c_p) > 0$ for all $p \in \mathcal{D}^*$ and so the continuous differentiability of $p \mapsto c_p$ follows from (7.1) and the implicit function theorem. □

As expected, c_p travels "faster" than ν_p as a function of p.

Lemma 7.5. *For any $p \in \mathcal{D}^*$, $\frac{dc_p}{dp}$ and $\frac{d\nu_p}{dp}$ have the same sign, and $\left|\frac{dc_p}{dp}\right| > \left|\frac{d\nu_p}{dp}\right|$ if they are non-zero.*

Proof. By (7.1), we have $f(\nu_p + c_p) - f(\nu_p - c_p) = 0$. The left side is differentiable in p by Lemma 7.4, so taking derivatives, we obtain

$$f'(\nu_p + c_p)\left(\frac{d\nu_p}{dp} + \frac{dc_p}{dp}\right) - f'(\nu_p - c_p)\left(\frac{d\nu_p}{dp} - \frac{dc_p}{dp}\right) = 0. \tag{7.5}$$

Rearranging yields

$$\frac{dc_p}{dp} = \frac{d\nu_p}{dp}\left(\frac{f'(\nu_p - c_p) - f'(\nu_p + c_p)}{f'(\nu_p - c_p) + f'(\nu_p + c_p)}\right),$$

where the fraction is well-defined with positive denominator by Lemma 7.3. The numerator is positive by Lemma 7.2, (7.3), and (7.4), so $\frac{dc_p}{dp}$ has the same sign as $\frac{d\nu_p}{dp}$. If $\frac{d\nu_p}{dp}, \frac{dc_p}{dp} > 0$, then one can see immediately from (7.5) that $\frac{dc_p}{dp} > \frac{d\nu_p}{dp}$. The reverse also follows. □

Our final lemma concerns a criterion for pointwise increasingness of ν_p.

Lemma 7.6. *Fix $p \in \mathcal{D} \setminus \{0\}$. If*

$$(\nu_p - c_p)\frac{f'(\nu_p + c_p)}{f(\nu_p + c_p)} > -1, \tag{7.6}$$

then ν is increasing at p.

Proof. Since $f(\nu_p + c_p) = f(\nu_p - c_p)$, we may rearrange to obtain

$$-\frac{f(\nu_p - c_p)}{\nu_p - c_p} < f'(\nu_p + c_p).$$

Define the line

$$\ell(x) := -\frac{xf(\nu_p - c_p)}{\nu_p - c_p} + \frac{2\nu_p f(\nu_p - c_p)}{\nu_p - c_p},\ x \geq \nu_p + c_p,$$

and note that

$$\begin{aligned} \ell(\nu_p + c_p) &= f(\nu_p - c_p) = f(\nu_p + c_p), \\ \ell(2\nu_p) &= 0 = f(0). \end{aligned} \tag{7.7}$$

Note that (7.6) implies $f'(\nu_p + c_p) = \ell'(\nu_p + c_p)$, so by convexity of f on (θ_2, ∞), we have $f' > \ell'$ on $(\nu_p + c_p, \infty)$. Via integration we obtain

$$f(\nu_p + c) > \ell(\nu_p + c) \tag{7.8}$$

for $c > c_p$. We now split into two cases to show $f(\nu_p + c) > f(\nu_p - c)$ for $c > c_p$.

Case 1. Suppose $\nu_p - c_p \in (0, \theta_1]$. By (7.7) and the convexity of f on this interval, we have $\ell(x) > f(2\nu_p - x)$ for $x \in (\nu_p + c_p, 2\nu_p)$, making use of the fact that convexity is preserved under reflection and translation. A change of variables gives the inequality $\ell(\nu_p + c) > f(\nu_p - c)$ for $c \in (c_p, \nu_p)$. Combining with (7.8) yields $f(\nu_p + c) > f(\nu_p - c)$ for $c \in (c_p, \nu_p)$. If $c \geq \nu_p$, then $f(\nu_p + c) > 0 = f(\nu_p - c)$ and we are done.

Case 2. Suppose instead $\nu_p - c_p \in (\theta_1, \nu_0)$. If it happens that $\ell(\nu_p + c) > f(\nu_p - c)$ for all $c \in (c_p, \nu_p)$, then the argument in Case 1 applies and we are done. Otherwise, let

$$\tilde{\ell}(x) := -(x - \nu_p - c_p)f'(\nu_p - c_p) + f(\nu_p + c_p),\ x \geq \nu_p + c_p,$$

be the line such that

$$\tilde{\ell}(\nu_p + c_p) = f(\nu_p + c_p)$$

and

$$\tilde{\ell}(\nu_p + c_p + f(\nu_p + c_p)/f'(\nu_p - c_p)) = 0.$$

Lemma 7.3 and convexity imply $f'(\nu_p + c) > \tilde{\ell}'(\nu_p + c)$ and thus $f(\nu_p + c) > \tilde{\ell}(\nu_p + c)$ for all $c > c_p$. By concavity of f near $\nu_p - c_p$, one can easily see that $\tilde{\ell}(\nu_p + c) > f(\nu_p - c)$ for $c \in (c_p, \nu_p - \theta_1]$. It remains to show $\tilde{\ell}(\nu_p + c) > f(\nu_p - c)$ for $c \in (\nu_p - \theta_1, \nu_p)$.

We have by assumption that $f(\nu_p - c) > \ell(\nu_p + c)$ for c in a right neighborhood of c_p. We also have by (7.7) that $f(\nu_p - c_p) = \ell(\nu_p + c_p)$. It follows that

$$-f'(\nu_p - c_p) > \ell'(\nu_p + c_p) = -\frac{f(\nu_p - c_p)}{\nu_p - c_p}.$$

Substituting with $f(\nu_p - c_p)$ with $f(\nu_p + c_p)$ and rearranging yields

$$\nu_p + c_p + \frac{f(\nu_p + c_p)}{f'(\nu_p - c_p)} > 2\nu_p.$$

The left side is precisely the root of $\tilde{\ell}$ whereas the right side is a root of f. We showed previously that $\tilde{\ell}(2\nu_p - \theta_1) > f(\theta_1)$. Convexity of f on $(0, \theta_1)$ implies $\tilde{\ell}(\nu_p + c) > f(\nu_p - c)$ for $c \in (\nu_p - \theta_1, \nu_p)$, and we are done.

We have now shown that, in general, $f(\nu_p + c) > f(\nu_p - c)$ for $c > c_p$. We also know by definition of c_p that $f(\nu_p + c) \leq f(\nu_p - c)$ for $c < c_p$. Thus the conditions of Lemma 2.3 are satisfied and so ν is increasing at the point p. $\square$

We are now ready to prove the main theorem.

Proof of Theorem 1.10. It suffices to show (7.6) for all $p \in \mathcal{D}^*$ and that $\mathcal{D}^* = \mathcal{D} \setminus \{0\}$.

If we have conditions (1) and (2) as they are written, then (7.6) holds for all $p \in \mathcal{D}^*$ immediately by (7.3) and (7.4).

Suppose instead we only have the weaker condition as stated in Remark 7.1(b):

$$f'/f > 1/(\nu_1 - c_1) \text{ on } (0, \theta_1),$$
$$f'/f > -1/(\nu_1 - c_1) \text{ on } (\nu_1 + c_1, \infty).$$

Let $u(p) := \nu_p - c_p$ on $\mathcal{D}^*$, so u is differentiable on its domain. Also note that $u > 0$ by (7.4). Suppose for some $p' \in \mathcal{D}^*$ that $u(p') \leq \nu_1 - c_1$. Then

$$(\nu_{p'} - c_{p'}) \frac{f'(\nu_{p'} + c_{p'})}{f(\nu_{p'} + c_{p'})} > -1,$$

i.e., (7.6) holds for p' and so ν_p is increasing at p'. Lemma 7.5 implies u is decreasing at p'. It follows that for all $p \geq p'$, u is decreasing and so $u(p) \leq \nu_1 - c_1$. In particular, ν_p is increasing for all $p \geq p'$.

By assumption, $1 \in \mathcal{D}^*$ and trivially $u(1) \leq \nu_1 - c_1$. Since $\mathcal{D}^* \in [1, \infty)$, then ν_p is increasing for all $p \in \mathcal{D}^*$. Recall that $p \in \mathcal{D}^*$ if $\nu_p > (\nu_0 + \theta_2)/2$. Since ν_p is increasing at $p = 1$ and for all $p \in \mathcal{D}^*$, then $\nu_p > (\nu_0 + \theta_2)/2$ for all $p \in \mathcal{D} \setminus \{0\}$, hence $\mathcal{D} \setminus \{0\} = \mathcal{D}^*$.[§] □

The proof directly extends to the case where f'' only has a single root $\theta > \nu_0$ by setting $\theta_1 = 0$. In fact, conditions (1) and (2) are not necessary. Indeed, we use them once in the proof of Lemma 7.3 in the case $\nu_p - c_p \in (0, \theta_1)$, but this is no longer relevant if f'' has only one root. The only other time we use the conditions is to prove that (7.6) holds for all $p \in \mathcal{D}^*$. However, if f'' has only one root then (7.6) holds automatically. Note that f is concave and increasing near $\nu_p - c_p$ for all $p \in \mathcal{D}^*$, hence

$$\frac{f(\nu_p - c_p)}{\nu_p - c_p} > f'(\nu_p - c_p).$$

Since $f'(\nu_p + c_p) < 0$, then by substituting $f(\nu_p + c_p) = f(\nu_p - c_p)$, we have

$$(\nu_p - c_p) \frac{f'(\nu_p + c_p)}{f(\nu_p + c_p)} > \frac{f'(\nu_p + c_p)}{f'(\nu_p - c_p)}.$$

The right side dominates -1 as a consequence of Lemma 7.3, so we arrive at (7.6). This proves Corollary 1.11.

We show how Theorem 1.10 and Corollary 1.11 can be used to determine true positive skewness, both numerically and analytically.

Example 7.7 (Beta prime distribution). The beta prime distribution, also known as the inverted beta distribution or the beta distribution of the second kind, is supported on $(0, \infty)$ and, for parameters $\alpha, \beta > 0$, has density function

$$f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha, \beta)}$$

[§] Clearly the argument for $\mathcal{D}^* = \mathcal{D} \setminus \{0\}$ still applies if we use the alternative definition for $\mathcal{D}^*$.

on $(0,\infty)$, where B is the beta function. One can verify analytically that f is unimodal, continuous at 0, twice continuously differentiable, and has two positive inflection points. Take $\alpha = 4$ and $\beta = 5$. We can numerically calculate that

$$\begin{aligned}
\theta_1 &\approx 0.17267,\\
\theta_2 &\approx 0.82733,\\
\nu_0 &= 0.5,\\
\nu_1 &\approx 0.78621,\\
\inf_{x\in(0,\theta_1)} f'(x)/f(x) &\approx 9.69909,\\
\inf_{x\in(\theta_2,\infty)} f'(x)/f(x) &\approx -1.60770.
\end{aligned}$$

We find that the inflection points lie on either side of the mode, that the median lies to the right of the mode, and that the logarithmic derivatives satisfy conditions (1) and (2) of Theorem 1.10. Of course, this is not a formal proof, but we have good reason to believe that the beta prime distribution with parameters $\alpha = 4$, $\beta = 5$ is truly positively skewed. This technique can be applied to many different scenarios.

Example 7.8 (Log-logistic distribution). The log-logistic distribution with shape parameter $\beta > 0$ has density function

$$f(x) = \frac{\beta x^{\beta-1}}{(1+x^\beta)^2},\ x \geq 0$$

If $0 < \beta \leq 1$, then f strictly decreasing and true positive skewness follows from Proposition 2.4. Suppose $\beta > 1$. One can verify that f is unimodal with mode

$$\nu_0 = \left(\frac{\beta-1}{\beta+1}\right)^{1/\beta},$$

median $\nu_1 = 1$, and inflection points

$$\theta^{\pm} = \left(\frac{2\beta^2 - 2 \pm \beta\sqrt{3\beta^2-3}}{\beta^2+3\beta+2}\right)^{1/\beta}.$$

Straightforward computations show that $\theta^- \leq 0$ if and only if $\beta \leq 2$ and $\theta^+ > \nu_0$ if and only if $\beta > 1$. Moreover, $\theta^+ \leq 1$ if and only if $1 \leq \beta \leq 2$. Therefore, the log-logistic distribution is truly positively skewed if $1 < \beta \leq 2$ by Corollary 1.11.

8. True Skewness Beyond Continuous Univariate Distributions

In this section we present preliminary results concerning how one might widen the scope of true skewness theory to include a broader class of random variables. We present some results on true skewness for discrete univariate random variables and discuss the inherent challenges with showing a given discrete random variable is truly skewed. We also present preliminary numerical computations in order to illustrate how true skewness may be interpreted for multivariate distributions.

We first consider the problem of extending true skewness to univariate discrete random variables by way of an example distribution.

8.1. **Bernoulli distribution.** In this case, we can calculate a closed-form expression for ν_p.

Proposition 8.1. *Let* $X \sim \text{Bernoulli}(\lambda)$. *Then*

$$\nu_p = \frac{\left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)}}{1 + \left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)}},$$

and X *is truly positively skewed if and only if* $\lambda < 1/2$.

Proof. We first compute ν_p directly. Recall that ν_p is the unique solution to (1.2) and the probability mass function of X is

$$f(x;\lambda) = \begin{cases} \lambda, & x = 1 \\ 1-\lambda, & x = 0. \end{cases}$$

Equation (1.2) is

$$(1-\nu_p)_+^{p-1}\lambda + (-\nu_p)_+^{p-1}(1-\lambda) = (\nu_p - 1)_+^{p-1}\lambda + (\nu_p)_+^{p-1}(1-\lambda). \tag{2}$$

By Proposition 3.1, $0 < \nu_p < 1$. With these bounds on ν_p, the equation above becomes

$$(1-\nu_p)^{p-1}\lambda = \nu_p^{p-1}(1-\lambda) \iff \left(\frac{\nu_p}{1-\nu_p}\right)^{p-1} = \frac{\lambda}{1-\lambda}.$$

Solving for ν_p, we can isolate

$$\nu_p = \frac{\left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)}}{1 + \left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)}},$$

which has derivative

$$\frac{d\nu_p}{dp} = \frac{-\left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)} \log\left[\frac{\lambda}{1-\lambda}\right]}{(p-1)^2\left(1 + \left(\frac{\lambda}{1-\lambda}\right)^{1/(p-1)}\right)^2}.$$

Because it can be written as a square, the denominator is positive. The numerator is positive if and only if $\log\left(\frac{\lambda}{1-\lambda}\right) < 0$, which is true if and only if $\lambda < \frac{1}{2}$. Therefore $d\nu_p/dp > 0$ if and only if $\lambda < 1/2$. ☐

Remark 8.2. From (1.2), ν_p can be understood as a generalized central moment, which suggests a connection to the moment problem. In particular, it is possible that under certain conditions, the p-means of a random variable uniquely characterize its distribution. Since we have a precise closed-form expression of ν_p for the Bernoulli distribution, we can express the characteristic function of $X \sim \text{Bernoulli}(\lambda)$ as a function of ν_p:

$$\varphi(t) = \lambda\left(\left(\frac{\nu_p}{1-\nu_p}\right)^{p-1} + \left(\frac{\nu_p}{1-\nu_p}\right)^{2p-2} e^{it}\right).$$

Thus, a Bernoulli distribution can be uniquely defined in terms of its p-means. In general, however, it is unclear what connection exists between ν_p and the moments of a given distribution. We refer the reader to [13] for an overview of the moment problem.

To prove a given continuous distribution is truly skewed, a powerful tool is the combination of Theorem 2.2 and Lemma 2.3. We obtain the next result which extends Theorem 2.2 to discrete random variables.

8.2. **A Discrete Analogue for Theorem 2.2.** The main idea of Theorem 2.2 is the following. Given $p \in \mathcal{D}$, reflect the left tail of the distribution about ν_p, shift both the right and left tails backward by ν_p, then show that the (scaled) right tail stochastically dominates the reflected (scaled) left tail. We describe an extension this idea to a discrete distribution.

Let X be a random variable with probability mass function $f(x)$ supported on $\{l, \dots, r\}$, where $l < r$ are integers. Fix $p \in \mathcal{D}$. Once we reflect the left tail about ν_p, the mass falls on points in the set

$$L := \Big\{ \nu_p - \lfloor \nu_p \rfloor + m : m \in \{0, 1, \dots \lfloor \nu_p \rfloor - l\} \Big\}.$$

Similarly, the mass of the right tail falls on points in the set

$$R := \Big\{ \lceil \nu_p \rceil - \nu_p + m : m \in \{0, 1, \dots r - \lceil \nu_p \rceil\} \Big\}.$$

The function representing the reflected/shifted left tail is $f(\nu_p - x)\mathbf{1}_L(x)$ and the function representing the shifted right tail is $f(\nu_p + x)\mathbf{1}_R(x)$. We need to scale these functions by an appropriate factor analogous to (2.1) that turns each into a valid probability mass function. Define

$$J_p := \sum_{y=\nu_p - \lfloor \nu_p \rfloor}^{\nu_p - l} y^{p-1} f(\nu_p - y) = \sum_{y=\lceil \nu_p \rceil - \nu_p}^{r - \nu_p} y^{p-1} f(y + \nu_p).$$

Then we have the following.

Lemma 8.3. *The functions*

$$\frac{1}{J_p} y^{p-1} f(\nu_p - y)\mathbf{1}_L(y) \qquad \textit{and} \qquad \frac{1}{J_p} y^{p-1} f(\nu_p + y)\mathbf{1}_R(y)$$

are probability mass functions.

Proof. We prove

$$\sum_{y=-\infty}^{\infty} \frac{1}{J_p} y^{p-1} f(\nu_p - y)\mathbf{1}_L(y) = 1,$$

and note that the computation for the other function is similar. We have

$$\begin{aligned}
\sum_{y=-\infty}^{\infty} \frac{1}{J_p} y^{p-1} f(\nu_p - y) \mathbf{1}_L(y) &= \sum_{y=\nu_p-\lfloor \nu_p \rfloor}^{\nu_p - l} \frac{1}{J_p} y^{p-1} f(\nu_p - y) \\
&= \sum_{y=\nu_p-\lfloor \nu_p \rfloor}^{\nu_p - l} \left[\frac{1}{\sum_{\tilde{y}=\nu_p-\lfloor \nu_p \rfloor}^{\nu_p - l} \tilde{y}^{p-1} f(\nu_p - \tilde{y})} y^{p-1} f(\nu_p - y) \right] \\
&= \frac{\sum_{y=\nu_p-\lfloor \nu_p \rfloor}^{\nu_p - l} y^{p-1} f(\nu_p - y)}{\sum_{\tilde{y}=\nu_p-\lfloor \nu_p \rfloor}^{\nu_p - l} \tilde{y}^{p-1} f(\nu_p - \tilde{y})} \\
&= 1.
\end{aligned}$$

□

Theorem 8.4. *Let Y be a discrete random variable with probability mass function $f(y)$ supported on $\{l, \ldots, r\}$ with $l < r \in \mathbb{Z}$. Suppose $p \in \mathcal{D}$. If the random variable with probability mass function*

$$\frac{1}{J_p} y^{p-1} f(\nu_p + y) \mathbf{1}_R(y)$$

exhibits strict stochastic dominance over the random variable with probability mass function

$$\frac{1}{J_p} y^{p-1} f(\nu_p - y) \mathbf{1}_L(y),$$

then the function $\nu(p) \equiv \nu_p$ is increasing at p.

Proof. Since ν_p is the unique solution to (1.2), taking $\frac{d}{dp}$ to both sides we have

$$\begin{aligned}
0 &= \frac{d}{dp} \left(E\left[(\nu_p - Y)_+^{p-1} \right] - E\left[(Y - \nu_p)_+^{p-1} \right] \right) \\
&= \frac{d}{dp} \left(\sum_{y=l}^{\lfloor \nu_p \rfloor} (\nu_p - y)^{p-1} f(y) - \sum_{y=\lceil \nu_p \rceil}^{r} (y - \nu_p)^{p-1} f(y) \right) \\
&= \frac{d}{dp} \left(\sum_{y=l}^{\lfloor \nu_p \rfloor} (\nu_p - y)^{p-1} f(y) \right) - \frac{d}{dp} \left(\sum_{y=\lceil \nu_p \rceil}^{r} (y - \nu_p)^{p-1} f(y) \right).
\end{aligned}$$

We handle the derivatives of each of these sums individually. Consider the sum on the left first. Define $g(p, y) = (\nu_p - y)^{p-1} f(y)$. If $\lfloor \nu_p \rfloor = l + k$ for some integer k such that $l + k \leq r - 1$, then

$$\begin{aligned}
\frac{d}{dp}\sum_{y=l}^{\lfloor \nu_p \rfloor} g(p,y) &= \sum_{y=l}^{\lfloor \nu_p \rfloor} g'(p,y) \\
&= \sum_{y=l}^{\lfloor \nu_p \rfloor} \left((p-1)(\nu_p - y)^{p-1}\frac{d\nu_p}{dp} + (\nu_p - y)^{p-1}\log(\nu_p - y) \right) f(y) \\
&= \frac{d\nu_p}{dp}\sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(p-1)(\nu_p - y)^{p-1} f(y)\right] + \sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(\nu_p - y)^{p-1}\log(\nu_p - y) f(y)\right].
\end{aligned}$$

Similarly, for the derivative of the other sum we have

$$\begin{aligned}
\frac{d}{dp}\sum_{y=\lceil \nu_p \rceil}^{r} (y-\nu_p)^{p-1} f(y) &= -\sum_{y=\lceil \nu_p \rceil}^{r} \left((p-1)(y-\nu_p)^{p-2}\frac{d\nu_p}{dp} + (y-\nu_p)^{p-1}\log(y-\nu_p) \right) f(y) \\
&= -\frac{d\nu_p}{dp}\sum_{y=\lceil \nu_p \rceil}^{r}\left[(p-1)(y-\nu_p)^{p-2} f(y)\right] \\
&\quad + \sum_{y=\lceil \nu_p \rceil}^{r}\left[(y-\nu_p)^{p-1}\log(y-\nu_p) f(y)\right].
\end{aligned}$$

Hence

$$\begin{aligned}
0 = \frac{d\nu_p}{dp}&\left(\sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(p-1)(\nu_p - y)^{p-1} f(y)\right] + \sum_{y=\lceil \nu_p \rceil}^{r}\left[(p-1)(y-\nu_p)^{p-2} f(y)\right]\right) \\
&+ \sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(\nu_p - y)^{p-1}\log(\nu_p - y) f(y)\right] - \sum_{y=\lceil \nu_p \rceil}^{r}\left[(y-\nu_p)^{p-1}\log(y-\nu_p) f(y)\right].
\end{aligned}$$

Rearranging,

$$\begin{aligned}
\frac{d\nu_p}{dp} &= \frac{\displaystyle\sum_{y=\lceil \nu_p \rceil}^{r}\left[\log(y-\nu_p)(y-\nu_p)^{p-1} f(y)\right] - \sum_{y=l}^{\lfloor \nu_p \rfloor}\left[\log(\nu_p - y)(\nu_p - y)^{p-1} f(y)\right]}{\displaystyle\sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(p-1)(\nu_p - y)^{p-1} f(y)\right] + \sum_{y=\lceil \nu_p \rceil}^{r}\left[(p-1)(y-\nu_p)^{p-2} f(y)\right]} \\
&= \frac{\displaystyle\sum_{y=\lceil \nu_p \rceil - \nu_p}^{r-\nu_p}\left[\log(y)\; y^{p-1} f(\nu_p + y)\right] - \sum_{y=\nu_p - \lfloor \nu_p \rfloor}^{\nu_p - l}\left[\log(y)\; y^{p-1} f(\nu_p - y)\right]}{\displaystyle\sum_{y=l}^{\lfloor \nu_p \rfloor}\left[(p-1)(\nu_p - y)^{p-1} f(y)\right] + \sum_{y=\lceil \nu_p \rceil}^{r}\left[(p-1)(y-\nu_p)^{p-2} f(y)\right]}.
\end{aligned}$$

The denominator is positive, and the numerator is strictly positive under the strict stochastic dominance assumption of the theorem. More explicitly, let Y_1 denote the random variable with probability mass function $\frac{1}{J_p} y^{p-1} f(\nu_p + y)\mathbf{1}_R(y)$ and let Y_2 denote the random variable with probability mass function $\frac{1}{J_p} y^{p-1} f(\nu_p - y)\mathbf{1}_L(y)$. Since Y_1 strictly stochastically dominates Y_2, we have by a well-known result [9, Theorem 1.2.8] that

$$E\left[h(Y_1)\right] > E\left[h(Y_2)\right]$$

for any increasing function h. In particular,

$$E\left[\log(Y_1)\right] > E\left[\log(Y_2)\right],$$

which implies

$$\frac{1}{J_p} \sum_{y=\lceil \nu_p \rceil - \nu_p}^{r-\nu_p} \log(y)\ y^{p-1} f(\nu_p + y) > \frac{1}{J_p} \sum_{y=\nu_p - \lfloor \nu_p \rfloor}^{\nu_p - l} \log(y)\ y^{p-1} f(\nu_p - y).$$

Therefore $\frac{d\nu_p}{dp} > 0$. □

In general, for a continuous random variable it is difficult to verify the stochastic dominance criterion of Theorem 2.2 directly. This difficulty necessitates Lemma 2.3. Likewise for discrete, the conditions of Theorem 8.4 are difficult to verify. Whereas the main idea of Theorem 2.2 extends neatly to discrete random variables, the main idea of Lemma 2.3 does not, since this lemma depends on the continuity of density functions.

The inherent challenge to build analogous tools for discrete random variables suggests that the true skewness theory developed for continuous random variables may not extend directly to discrete random variables. Despite these challenges, we prove some elementary results about binomial random variables.

Lemma 8.5. *A binomial random variable X with parameter n and λ of the form $\frac{k}{n+1}$ where $k \in \{2, \ldots, \lfloor n/2 \rfloor\}$ is bimodal.*

Proof. See Appendix B. □

Lemma 8.6. *A unimodal binomial random variable X with parameter n and $\frac{1}{n} \leq \lambda < 1/2$, has mode less than or equal to its mean.*

Proof. See Appendix B. □

By working first with decreasing binomial random variables, we hope to prove a result similar to Proposition 2.4 for the discrete cae.

Lemma 8.7. *The probability mass function of a binomial random variable X with parameters $n > 1$ and $\lambda < \frac{1}{n+1}$ is strictly decreasing on its support.*

Proof. We will show that $P(X=k)-P(X=k+1)>0$ for $k\in\{0,\dots,n-1\}$.

$$\begin{aligned}P(X=k)-P(X=k+1) &= \binom{n}{k}\lambda^k(1-\lambda)^{n-k}-\binom{n}{k+1}\lambda^{k+1}(1-\lambda)^{n-k-1}\\ &= \frac{n!\lambda^k(1-\lambda)^{n-k}}{k!(n-k)!}-\frac{n!\lambda^{k+1}(1-\lambda)^{n-k-1}}{(k+1)!(n-k-1)!}\\ &= \frac{n!\lambda^k(1-\lambda)^{n-k-1}}{k!(n-k-1)!}\left[\frac{(1-\lambda)}{(n-k)}-\frac{\lambda}{k+1}\right].\end{aligned}$$

Each term on the left is positive: it remains to show that $\left[\frac{(1-\lambda)}{(n-k)}-\frac{\lambda}{k+1}\right]$ is always positive. We have that $\lambda<1/(n+1)$, so

$$\begin{aligned}\left[\frac{(1-\lambda)}{(n-k)}-\frac{\lambda}{k+1}\right] &< \left[\frac{n}{(n+1)(n-k)}-\frac{1}{(k+1)(n+1)}\right]\\ &= \left[\frac{nk+k}{(n+1)(n-k)(k+1)}\right]>0.\end{aligned}$$

Since $P(X=k)>P(X=k+1)$ for all $k\in\{0,\dots,n-1\}$, then we conclude that the probability mass function of X is strictly decreasing on its support. □

Proposition 8.8. *Let $X\sim$ Binomial(n,λ) and $Y\sim$ Binomial(m,λ) be independent. If the probability mass function of $X+Y$ is strictly decreasing on its support, then the probability mass functions of X and Y are strictly decreasing on their supports, respectively.*

Proof. Let X and Y be Binomial random variables with parameters n and m respectively. Recall that by Lemma 8.7 their convolution, the binomial random variable $X+Y$, is a strictly decreasing distribution if $\lambda<\frac{1}{n+m+1}$.

$$\lambda<\frac{1}{n+m+1}\implies\lambda<\frac{1}{n+1}\text{ and }\lambda<\frac{1}{m+1} \tag{8.1}$$

Thus we have that X and Y are both strictly decreasing. □

However, it is actually not true that a binomial random variable with decreasing probability mass functions is truly positively skewed. We can show this numerically through a graph of ν_p from Mathematica. The following graphical representation of ν_p is for $X\sim$ Binomial$(2,1/3)$ and $p=1.01,1.02,\dots,1.24$.

Remark 8.9. It is worth discussing that the reason why true skewness does not hold in the discrete case is because we can express the discrete random variable X as a continuous random variable with the (generalized) density function

$$f(x)=\sum_{k\in I}\mathbb{P}(X=k)\delta(x-x_k)$$

where $\delta(x-x_k)$ denotes the unit point mass at x_k. This is seen in the following figure.

Because the Dirac delta function can be used to represent discrete distributions, a discrete random variable can be thought of not as unimodal, but as having n modes. Where the disconnect occurs is with our intuition about the "shape" of discrete distributions. It is easy

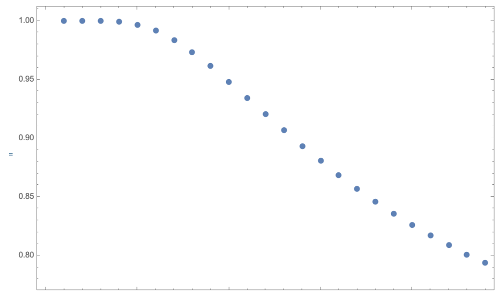
nup.png


Figure 8.3. Values of ν_p for $p \in [1.01, 1.24]$

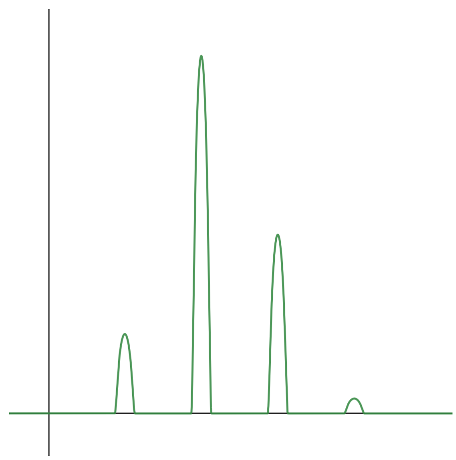
example.png


Figure 8.4. "Binomial" distribution as mixture of bump functions

to imagine a continuous curve mapped over top of the point masses of a discrete distribution. From this perspective, it seems clear that our stochastic dominance tools would apply. However, a more accurate continuous representation of discrete would be to climb and descend each of the point masses. This is the essence of approximation by the Dirac delta function. Stochastic dominance is not an intuitive tool for multi-modal distributions of this form.

8.3. **True skewness of continuous multivariate distributions.** We can extend Definition 1.1 naturally to multivariate distributions as follows.

Let $\mathbf{X} \in \mathbb{R}^k$ be a random vector. The p-mean $\boldsymbol{\nu}_p = \left(\nu_p^{(1)}, \nu_p^{(2)}, \ldots, \nu_p^{(k)}\right)$ of $\mathbf{X}$ is defined

$$\boldsymbol{\nu}_p = \arg\min_{\mathbf{a} \in \mathbb{R}^k} E\|\mathbf{X} - \mathbf{a}\|^p$$

similarly to (1.1), where $\|\cdot\|$ is the usual Euclidean norm.

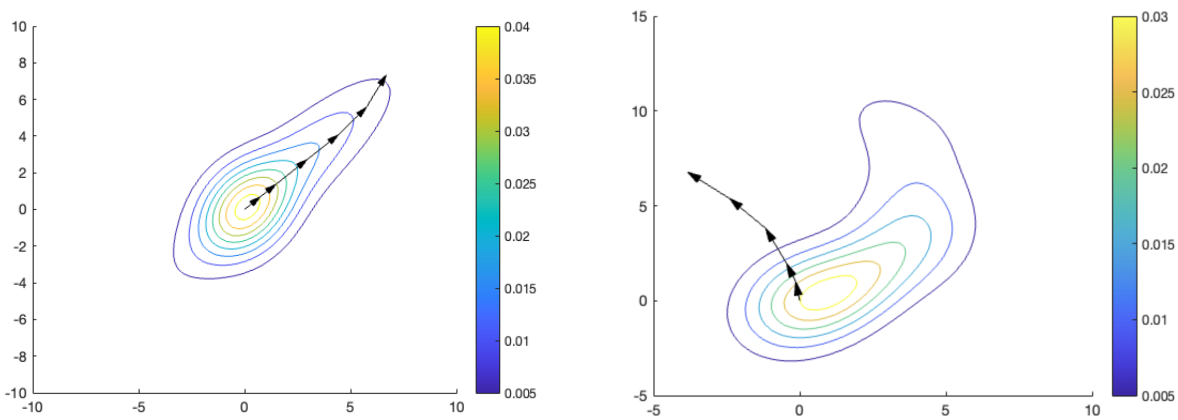
In the univariate setting, true skewness corresponds to the sign of $d\nu_p/dp$. However, it does not make sense to interpret the true skewness of a random vector $\mathbf{X}$ as the "sign" of the vector $d\boldsymbol{\nu}_p/dp$. We adjust our interpretation of true skewness to emphasize the *trajectory* of $d\boldsymbol{\nu}_p/dp$ in $\mathbb{R}^k$.

Let

$$\boldsymbol{\tau}_p := \frac{d\boldsymbol{\nu}_p}{dp} \bigg/ \left\| \frac{d\boldsymbol{\nu}_p}{dp} \right\|$$

and let $\boldsymbol{\tau} := \{\boldsymbol{\tau}_p\}_{p\in\mathcal{D}}$ denote the trajectory of $d\boldsymbol{\nu}_p/dp$ in $\mathbb{R}^k$. Then we say that random vector $\mathbf{X}$ is **truly $\boldsymbol{\tau}$-skewed**, where the trajectory $\boldsymbol{\tau}$ fully characterizes the skewness of $\mathbf{X}$.

To illustrate this interpretation, consider two bivariate Gaussian mixture distributions as shown in Figure 8.5. We numerically computed the trajectories $\boldsymbol{\tau}$ and plotted them over the contour plots of the density functions. These trajectories follow the shape of each distribution. This preliminary numerical demonstration reinforces the idea that the skewness of a distribution is deeply connected to the trajectory $\boldsymbol{\tau}$.



(A) Simple Gaussian mixture (B) "Curving" Gaussian mixture

FIGURE 8.5. Trajectories $\boldsymbol{\tau}$ (black arrows) plotted tip-to-tail superimposed on contour plots of two bivariate Gaussian mixture density functions.

9. METRICS OF SKEWNESS

In addition to developing tools for showing a large class of random variables are truly skewed, we also aim to develop a metric for quantifying the degree to which a distribution is truly skewed. Naturally, this metric will depend on $d\nu_p/dp$. We highlight that this approach is consistent with the classical approach to measuring skewness. Recall Pearson's first and second skewness coefficients,

$$\gamma_1 = \frac{\nu_2 - \nu_0}{(E[(X-\nu_2)^2])^{1/2}} \qquad \text{and} \qquad \gamma_2 = \frac{3(\nu_2 - \nu_1)}{(E[(X-\nu_2)^2])^{1/2}}.$$

Each of these metrics are of the form

$$\gamma = (\text{normalizing coefficient}) \times (\text{difference of } \nu_p).$$

Likewise, in determining true skewness, we are also interested in a "difference" of ν_p, namely $d\nu_p/dp$. It follows that a measure of skewness based on the true skewness theory should be of the form

$$\gamma = (\text{normalizing coefficient}) \times \frac{d\nu_p}{dp}.$$

In line with Arnold & Groeneveld [1], any measure γ ought to obey the following properties:

(I) If X is symmetric, then $\gamma(X) = 0$
(II) For any constants $c > 0$ and $b \in \mathbb{R}$, $\gamma(cX + b) = \gamma(X)$.
(III) $\gamma(-X) = -\gamma(X)$.

To make the notion of true skewness applicable to statistics, we would like our skewness metric to be defined in such a way that admits statistical estimation. Any metric involving $d\nu_p/dp$ explicitly will prevent this. To resolve this issue, we consider a measure based on the *average* of $d\nu_p/dp$ with respect to a smooth, non-negative, decreasing weight function ϕ.

Let $S = \sup \mathcal{D}$. We propose the skewness metric

$$\Gamma(X) = \frac{\displaystyle\int_1^S \frac{d\nu_p}{dp}\phi(p)dp}{|\nu_1 - \nu_0| \displaystyle\int_1^S \phi(p)dp}.$$

Remark 9.1. For a symmetric distribution, ν_p is constant for all p so that $\nu_1 - \nu_0 = 0$. In the case of a symmetric X, we define $\Gamma(X) = 0$.

By considering the average of $d\nu_p/dp$ we can apply integration by parts to the numerator, which gives

$$\Gamma(X) = \frac{\nu_p\phi(p)\Big|_1^S - \displaystyle\int_1^S \nu_p d(\phi(p))}{|\nu_1 - \nu_0| \displaystyle\int_1^S \phi(p)dp}.$$

If $\mathcal{X} = \{x_1, \ldots, x_n\}$ is a given dataset with metric d, we can estimate ν_p via the **p-medioid**

$$\hat{\nu}_p := \arg\min_{a \in \mathcal{X}} \sum_{i=1}^n d(x_i, a)^p.$$

Then a candidate estimator for Γ is

$$\hat{\Gamma}(X) = \frac{\hat{\nu}_p\phi(p)\Big|_1^{S^*} - \displaystyle\int_1^{S^*} \hat{\nu}_p d(\phi(p))}{|\hat{\nu}_1 - \hat{\nu}_0| \displaystyle\int_1^{S^*} \phi(p)dp},$$

where $S^* = \sup\left\{p \geq 1 : \frac{1}{n}\sum_{i=1}^n |x_i|^p < \infty\right\}$. The following proposition demonstrates that Γ is a suitable metric.

Proposition 9.2. *Let X be a continuous univariate random variable with $S = \sup \mathcal{D}$. The skewness metric*

$$\Gamma(X) = \frac{\nu_p \phi(p)\Big|_1^S - \displaystyle\int_1^S \nu_p d(\phi(p))}{|\nu_1 - \nu_0| \displaystyle\int_1^S \phi(p) dp}$$

satisfies the following properties:

(I) If X is symmetric, then $\Gamma(X) = 0$
(II) For any constants $c > 0$ and $b \in \mathbb{R}$, $\Gamma(cX + b) = \Gamma(X)$.
(III) $\Gamma(-X) = -\Gamma(X)$.

Proof. We prove (II) and (III). For simplicity, suppose ϕ is chosen such that it integrates to 1. Suppose $\tilde{\nu}_p$ is the p-mean of $cX + b$ and ν_p is the p-mean of X. By Proposition 3.3,

$$\begin{aligned}
\Gamma(cX+b) &= \frac{\displaystyle\int_1^S \frac{d\tilde{\nu}_p}{dp}\phi(p)dp}{|\tilde{\nu}_1 - \tilde{\nu}_0|} \\
&= \frac{1}{|\tilde{\nu}_1 - \tilde{\nu}_0|}\left[\tilde{\nu}_p \phi(p)\Big|_1^S - \int_1^S \tilde{\nu}_p d(\phi(p))\right] \\
&= \frac{1}{|c\nu_1 - c\nu_0|}\left[\phi(p)(c\nu_p + b)\Big|_1^S - \int_1^S d(\phi(p))(c\nu_p + b)\right] \\
&= \frac{1}{c|\nu_1 - \nu_0|}\left[c(\phi(S)\nu_S - \phi(1)\nu_1) + b(\phi(S) - \phi(1)) - c\int_1^S \nu_p d\phi - \underbrace{b\int_1^S \phi'(p)dp}_{b(\phi(S)-\phi(1))}\right] \\
&= \frac{c}{c|\nu_1 - \nu_0|}\left[\phi(S)\nu_S - \phi(1)\nu_1 - \int_1^S \nu_p d\phi\right] \\
&= \Gamma(X).
\end{aligned}$$

It remains to show (III). Let $\tilde{\nu}_p$ be the p-mean of $-X$. With Proposition 3.3 we have

$$\begin{aligned}
\Gamma(-X) = \frac{\displaystyle\int_1^S \frac{d\tilde{\nu}_p}{dp}\phi(p)dp}{|\tilde{\nu}_1 - \tilde{\nu}_0|} &= \frac{1}{-|\nu_1 - \nu_0|}\left[-\phi(p)\nu_p\Big|_1^S - \int_1^S (-\nu_p) d\phi\right] \\
&= -\left(\frac{1}{|\nu_1 - \nu_0|}\left[\phi(p)\nu_p\Big|_1^S - \int_1^S \nu_p d\phi\right]\right) \\
&= -\Gamma(X).
\end{aligned}$$

□

With regards to choosing the weight function ϕ, note that the classical skewness theory values ν_p only for small values of p, specifically $p = 0, 1, 2$. Intuitively, one might expect ν_p to convey less information about the shape of a distribution for larger values of p. It should be emphasized that "larger values of p" is relative to the distribution in question. For a distribution finite moments of all orders, then $\mathcal{D} = [1, \infty)$, so that p may be taken arbitrarily large in Γ. However, for a distribution such as the Lévy distribution where $\mathcal{D} = [1, 1.5)$, in computing Γ we may only take p as large as 1.5.

Assuming we are dealing with a random variable with finite moments of all orders, the above discussion suggests that we should choose ϕ to be a smooth non-negative, decreasing function of p that takes values close to 1 for small values of p, and values close to 0 for large values of p. Consider a sigmoid function of the form

$$\phi(p; \alpha, \beta, \kappa) = \frac{1}{1 + \exp\{\kappa p^\alpha - \beta\}},$$

where $\alpha, \beta, \kappa > 0$ are parameters we choose to control the shape of ϕ. See Figure 9.6.

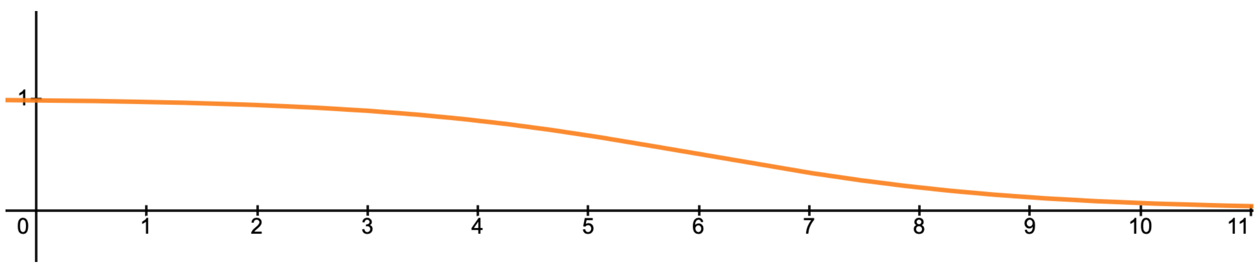


Figure 9.6. $\alpha = 1, \beta = 4.2, \kappa = 0.7$

It is unclear at this point how to appropriately choose α, β, and γ to be consistent and meaningful across all distributions.

Appendix A. Proof of Proposition 5.6

Proof. Let X and Y have density functions f_X and f_Y with corresponding heights h_1 and h_2, respectively. Assume without loss of generality that $h_1 \leq h_2$.

It is straightforward to compute the convolution

$$f_Z(z) = (f_X * f_Y)(z) = \begin{cases} \dfrac{h_1 h_2 z(24 - 6(h_1 + h_2)z + h_1 h_2 z^2)}{24} & z \in \left[0, \frac{2}{h_1}\right] \\ \dfrac{6h_1 h_2 + 2h_2^2 - 3h_1 h_2^2 z}{6h_1} & z \in \left(\frac{2}{h_1}, \frac{2}{h_2}\right] \\ -\dfrac{(-2(h_1 + h_2) + h_1 h_2 z)^3}{24 h_1 h_2} & z \in \left(\frac{2}{h_2}, \frac{2}{h_1} + \frac{2}{h_2}\right] \\ 0 & \text{otherwise.} \end{cases}$$

One can see that f_Z is continuous everywhere and decreasing on $(2/h_1, 2/h_1 + 2/h_2]$ such that a maximum exists in $[0, 2/h_1]$. Taking the second derivative readily tells us that f_Z is strictly concave on $[0, 2/h_1]$ and so has exactly one maximum, which is attained at

$$\nu_0 := \frac{2}{h_1} + \frac{2}{h_2} - 2\sqrt{\frac{1}{h_1^2} + \frac{1}{h_2^2}}\,.$$

Notice $\nu_0 \in [0, 2/h_1]$, which makes the following computation tedious but elementary:

$$\int_0^{\nu_0} (f_Z)(x)dx = \frac{4h_1\sqrt{h_1^2+h_2^2}}{3h_2^2} + \frac{4h_2\sqrt{h_1^2+h_2^2}}{3h_1^2} - \frac{4h_1^2}{3h_2^2} - \frac{4h_2^2}{3h_1^2} - \frac{2}{3}.$$

One can verify that this quantity is strictly less than 1/2 for all $h_1, h_2 > 0$. It follows that the median ν_1 of f_Z satisfies $\nu_1 > \nu_0$.

In the following, we use the term "near" to refer to a small enough neighborhood around a point. If this is nonsensical because a function is only has a left or right limit at that point, then we mean a small enough left or right neighborhood, respectively.

Given $p \geq 1$, suppose $\nu_p > \nu_0$ and consider the function

$$h(y) := f_Z(\nu_p + y) - f_Z(\nu_p - y)$$

defined on $[0, \nu_p]$. Clearly $h(0) = 0$. Since f_Z is unimodal, then it is decreasing at ν_p and so h is negative near 0.

Now notice that $f_Z(\nu_p+y) = 0$ if and only if $y > \frac{2}{h_1} + \frac{2}{h_2} - \nu_p$ and $f_Z(\nu_p - y) = 0$ if and only if $y > \nu_p$. Since f_Z has support on a bounded interval, one can derive the bound $\nu_p < \frac{1}{h_1} + \frac{1}{h_2}$ using (1.3). Since f_Z is strictly positive on its support, then $h(\nu_p) = f_Z(2\nu_p) > 0$. This implies that h has an odd number of zeroes in $(0, \nu_p)$ by the intermediate value theorem.

We wish to show h has exactly one root in $(0, \nu_p)$, so it suffices to show h' has at most one root in this interval by the mean value theorem. Consider $h'(y) = f_Z'(\nu_p + y) + f_Z'(\nu_p - y)$. This is well-defined and continuous for $y \in (0, \nu_p)$, with right and left limits at 0 and ν_p respectively. If $\nu_p \geq 2/h_1$, then both $f_Z'(\nu_p + y)$ and $f_Z'(\nu_p - y)$ are non-decreasing in y, hence h' is non-decreasing. Then h' at most one root and we are done in this case.

Suppose instead $\nu_p \in (\nu_0, 2/h_1)$. Since $\nu_p > \nu_0$ and f_Z is unimodal, then f_Z is decreasing near ν_p and so h' is negative near 0. On the other hand, $\lim_{y\to 0+} f_Z'(0) = h_1 h_2$. Since f_Z is concave on $[0, 2/h_1]$ and convex on $(2/h_1, \infty)$, then f_Z' is minimized at $2/h_1$. Note that

$$f_Z'(2/h_1) = -h_2^2/2 > -h_1 h_2 = -\lim_{y\to 0+} f_Z'(y)$$

which implies that h' is positive near ν_p. By continuity, h' has an odd number of roots in $(0, \nu_p)$.

Now consider $h''(y) = f_Z''(\nu_p + y) - f_Z''(\nu_p - y)$, whenever it is defined, which includes a neighborhood of 0. By concavity of f_Z on $[0, 2/h_1]$, $f_Z'(\nu_p + y) < f_Z'(\nu_p - y)$ near 0. If $h''(y) > 0$ near 0 (say ϵ-close), then the fundamental theorem of calculus gives

$$\int_0^{\epsilon} f_Z''(\nu_p + y)dy > \int_0^{\epsilon} f_Z''(\nu_p - y)dy \implies f_Z'(\nu_p + \epsilon) > -f_Z'(\nu_p - \epsilon).$$

But we showed above that h' is negative near 0, contradiction. Thus h'' is negative near 0.

Note that f''_Z is negative and linear on $(0, 2/h_1)$, zero on $(2/h_1, 2/h_2)$, and postive and linear on $(2/h_2, 2/h_1 + 2/h_2)$. Thus $f''_Z(\nu_p - y)$ and $f''_Z(\nu_p + y)$ intersect at most once, and if they do intersect, it must occur in the interval $(0, 2/h_1 - \nu_p)$. Moreover, because $f''_Z(\nu_p - y)$ is negative for all $y < \nu_p$ and because h'' is negative near 0, then h'' changes sign at most once in $(0, \nu_p)$. It follows that h' has at most two roots in this interval, but because it has an odd number of roots, h' has exactly one root. The same argument shows h has exactly one root in $(0, \nu_p)$.

We have thus shown via h that Lemma 2.3 is satisfied, and so ν_p is increasing when $\nu_p > \nu_0$ for any $p \geq 1$. Since $\nu_1 > \nu_0$, then $\nu_p > \nu_0$ for all $p \geq 1$ by continuity of $p \mapsto \nu_p$. This completes the proof. □

Appendix B. Proof of Results about Binomial

Proof of Lemma 8.5. Recall that the mode of binomial random variable X is $\lfloor (n+1)\lambda \rfloor$ and $\lfloor (n+1)\lambda \rfloor - 1$ if $(n+1)\lambda$ is an integer. Thus the two modes of X are k and $k-1$. Note that

$$k - 1 < \frac{kn}{n+1} < k$$

where $\frac{kn}{n+1}$ is the mean of X. □

Proof of Lemma 8.6. Recall that the mode of a unimodal binomial random variable X is $\lfloor (n+1)\lambda \rfloor$ and that the mean of binomial random variable X is $n\lambda$. There are two cases.

Case 1: λ is of the form $\frac{k}{n}$ where $k \in \{1, \dots \lfloor \frac{n-1}{2} \rfloor\}$. In this case, we have that the mode of X is

$$\lfloor (n+1)\lambda \rfloor = \left\lfloor \frac{k(n+1)}{n} \right\rfloor = k$$

and the mean of X is

$$\lambda n = \frac{k}{n} n = k$$

Therefore the mode is equal to the mean of X (is equal to k).

Case 2: $\frac{1}{n} < \lambda < 1/2$ otherwise.

$$\begin{aligned} \lambda &< 1 \\ n\lambda + \lambda &< n\lambda + 1 \\ \lfloor (n+1)\lambda \rfloor &< n\lambda \end{aligned}$$

where $\lfloor (n+1)\lambda \rfloor$ is the mode of X and $n\lambda$ is the mean. □

References

[1] Arnold, B. C., and Groeneveld, R. A. Measuring skewness with respect to the mode. *The American Statistician 49*, 1 (1995), 34–38.

[2] Billingsley, P. *Probability and Measure*, 3rd ed. Wiley & Sons, 1995.

[3] Curtiss, D. R. Recent extensions of Descartes' rule of signs. *Annals of Mathematics 19*, 4 (1918), 251–278.

[4] Das, S., Mandal, P. K., and Ghosh, D. On homogeneous skewness of unimodal distributions. *Sankhyā: The Indian Journal of Statistics, Series B (2008–)* (2009), 187–205.

[5] Groeneveld, R. A. Skewness for the Weibull family. *Statistica Neerlandica 40*, 3 (1986), 135–140.

[6] Kaas, R., and Buhrman, J. M. Mean, median and mode in binomial distributions. *Statistica Neerlandica 34*, 1 (1980), 13–18.
[7] Kopa, M., and Petrová, B. Strong and weak multivariate first-order stochastic dominance. SSRN 3144058 (2018).
[8] Kovchegov, Y. A new life of Pearson's skewness. arXiv preprint arXiv:2009.10305 (2020).
[9] Müller, A., and Stoyan, D. *Comparison methods for stochastic models and risks.* Wiley, 2002.
[10] Rinne, H. *The Weibull distribution: a handbook.* CRC Press, 2008.
[11] Rojo, J. Heavy-tailed densities. *Wiley Interdisciplinary Reviews: Computational Statistics 5*, 1 (2013), 30–40.
[12] Sato, K.-i. Convolution of unimodal distributions can produce any number of modes. *The Annals of Probability* (1993), 1543–1549.
[13] Schmüdgen, K. *The moment problem*, vol. 9. Springer, 2017.
[14] Versluis, C., and Straetmans, S. Skewness measures for the Weibull distribution. SSRN 2590356 (2015).
[15] Yu, Y. Stochastic ordering of exponential family distributions and their mixtures. *Journal of Applied Probability 46*, 1 (2009), 244–254.

Illinois Institute of Technology
E-mail address: alexnegron18@gmail.com

Cornell University
E-mail address: cep87@cornell.edu

Columbia University
E-mail address: cyw2124@columbia.edu