BIAS AND DIVISION IN THE FREE WORLD

LARRY GOLDSTEIN[1] AND TODD KEMP[2]

Abstract. Sampling bias is a foundational concept in statistics; associated bias transforms, such as size bias, have come to play important roles in probability theory of late. The first author and G. Reinert introduced *zero bias*, a transform whose unique fixed point is the normal distribution; it has become a standard tool in Stein's method and Gaussian approximation. Very recently, connections between zero bias and the class of infinitely divisible distributions have been found.

In this paper, we develop a free probabilistic analog of the zero bias transform, proving its existence and regularity. The *free zero bias* has the semicircle law (free probability's central limit distribution) as its unique fixed point. We offer a construction of the free zero bias that mirrors a classical one incorporating *square bias* with a mollifier, and in the process develop a surprisingly new class of distributional operations through their Cauchy transforms.

We then explore connections between the free zero bias, and size bias, with the class of *freely* infinitely divisible distributions. We develop a new self-contained treatment of the subject, together with a new characterization of free infinite divisibility using bias transforms. We also develop a parallel treatment of *positively* freely infinitely divisible distributions, characterized by a different from of the Lévy–Khintchine formula which we use to both give several new descriptions of such distributions and furnish new examples using these bias methods.

Contents

1. Introduction 2
1.1. (Free) Zero Bias and (Free) Stein Kernels 10
2. A Free Probability Primer 11
2.1. Noncommutative Probability Spaces 12
2.2. Free Independence 12
2.3. Cauchy Transform 13
2.4. The Voiculescu Transform and the R-transform 15
2.5. Free Poisson Distributions 17
2.6. Subordination and Free Conditional Expectation 19
2.7. Free Infinite Divisibility 20
2.8. Positive Free Infinite Divisibility 22
3. Existence and properties of the Free Zero Bias and other Transformations 23
3.1. Geometric Mean Transformation 23
3.2. The Size Bias s, Square Bias $\square$ and El Gordo $\flat$ Transformations 24
3.3. Construction of the Free Zero Bias $\circ$ 29
3.4. The 'Replace One' Property 32
3.5. A Transform that Fixes the Free Poisson Distribution 35

[2] supported in part of NSF Grants DMS-2055340 and DMS-2400246.

3.6. Free Zero Bias Transform Examples 36
4. Regularity of the Free Zero Bias 38
4.1. Absolute Continuity of the Free Zero Bias 38
4.2. Support of μ° for Finitely-Supported μ 40
5. Infinite Divisibility 44
5.1. New Formulations of Free Infinite Divisibility 45
5.2. Free Compound Poisson Distributions Revisited 48
5.3. Free Infinite Divisibility 52
5.4. The Compactly-Supported Case 59
6. New Examples of freely infinitely divisible Distributions 60
6.1. Using the Free Zero Bias 60
6.2. Using the Size Bias 66
Acknowledgments 69
References 70

1. Introduction

Distributional 'bias' transformations and their fixed points have been important tools in multiple contexts throughout probability, statistics, and beyond. Well known is the one that samples quantities proportional to their size, the **size bias** (1.1). As common sampling methods may suffer inadvertently from its effects, it was likely first regarded only as a statistical nuisance. Indeed, schemes for determining the sizes of new oil reserves where fields are explored uniformly at random, or surveys which hope to determine family sizes by asking participants how many siblings they have, will result in biased estimates of the desired quantities. Later work revealed that size bias can play more positive roles: it is responsible for waiting-time paradoxes, and has applications to tightness, analysis of the lognormal distribution, Skorohod embedding, number theory, concentration inequalities, Stein's method, and infinite divisibility; see [4]. Presently, we focus on the final topic, which we explore using size bias and related transforms like square bias, zero bias, and new bias transforms (introduced below) possessing notable unique distributional fixed points such as the normal or the semi-circle law.

The size bias transformation $X \mapsto X^s$ with domain $\mathcal{D}_m^+$ (the set of Borel probability measures on $[0,\infty)$ with mean $E[X] = m > 0$) is characterized by its distribution function

$$P(X^s \le x) = \frac{E[X\mathbf{1}(X \le x)]}{E[X]}. \tag{1.1}$$

Equivalently, X^s can be characterized by expectations over a function class by

$$E[Xf(X)] = mE[f(X^s)] \quad \text{for all } f \in C_b^\infty(\mathbb{R}), \tag{1.2}$$

where $C_b^\infty(\mathbb{R})$ is the class of all bounded infinitely differentiable functions. Size biasing has proved especially useful in connection with the Poisson distribution, see [17], [5] and [3]. In particular, the mapping $X \mapsto X^s - 1$ on $\mathcal{D}_m^+$ has the Poisson distribution with mean m as its unique fixed point. Similarly, on the domain $\mathcal{D}_{m,\sigma^2}$ of Borel probability measures on $\mathbb{R}$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$, we

define the **square bias** transformation $X \mapsto X^{\square}$ via

$$P(X^{\square} \leq x) = \frac{E[X^2 \mathbf{1}(X \leq x)]}{E[X^2]}, \tag{1.3}$$

which is likewise equivalent to

$$E[X^2 f(X)] = E[X^2]E[f(X^{\square})] \quad \text{for all } f \in C_b^{\infty}(\mathbb{R}). \tag{1.4}$$

The **zero bias** transform was introduced in [24] as a component of Stein's method [18, 43] for normal approximation. Given a random variable X with law $\mathcal{L}(X) = \mu \in \mathcal{D}_{0,\sigma^2}$, there exists a unique distribution μ^* defined by the property that, for a random variable X^* with law $\mathcal{L}(X^*) = \mu^*$,

$$E[Xf(X)] = \sigma^2 E[f'(X^*)] \quad \text{for all } f \in C_b^{\infty}(\mathbb{R}). \tag{1.5}$$

The associated mapping $\mu \mapsto \mu^*$ with domain $\mathcal{D}_{0,\sigma^2}$ has the Gaussian $\mathcal{N}(0, \sigma^2)$ as its unique fixed point (i.e. reproducing the usual Gaussian integration by parts formula that motivates the Stein equation). Despite a characterizing equation less intuitive than (1.2) for size biasing, there is a simple probabilistic construction of X^*: first take its square bias $X^{\square}$, and then multiply it by an independent Unif$[0, 1]$ random variable U; that is,

$$X^* \stackrel{\mathrm{d}}{=} UX^{\square}. \tag{1.6}$$

The proof can be found in [25, Theorem 2.1].

This and related transformations have found uses in normal and other distributional approximations [18] and e.g. [17], concentration inequalities [14], [16], [22], [31], and data science [19], [34], [35], [29], [20].

In this paper we provide free probability analogs of these results and the existence of the associated transformations, starting with the zero bias, which we call the **free zero bias** $\mu \mapsto \mu^{\circ}$ (or correspondingly for random variables $X \mapsto X^{\circ}$), also defined on the set $\mathcal{D}_{0,\sigma^2}$. We can extend the transformation to variables X with non-zero mean m, as is in Theorem 5.1, by

$$X^{\circ} = (X - m)^{\circ} + m, \tag{1.7}$$

that is, by applying the transformation to the centered variable $X - m$, and then adding m back in.

To define the putative free zero bias X° of a random variable X, we rely on a core plank of "free harmonic analysis": replacing the derivative $f' = f^{(1)}$ by the difference quotient $f^{[1]}$:

$$f^{[1]}(x, y) = \frac{f(x) - f(y)}{x - y}.$$

The difference quotient is a function of two variables, and we will evaluate it on *two i.i.d. copies* $X^{\circ} \stackrel{\mathrm{d}}{=} Y^{\circ}$, giving the following characterizing equation, analogous to (1.5):

$$E[Xf(X)] = \sigma^2 E[f^{[1]}(X^{\circ}, Y^{\circ})] \quad \text{for all } f \in C_b^{\infty}(\mathbb{R}). \tag{1.8}$$

It is easy to verify that X° in (1.7) satisfies (1.8) when X on the left hand side is replaced by $X - m$, with similar remarks applying to X^* and (1.5).

The difference quotient $f^{[1]}$ appears in this form often in literature on perturbation of eigenvalues, cf. [32], and noncommutative function theory, cf. [41, Proposition 4.3.1], [30, Example 6.5], owing to its relationship to ordinary (Fréchet) derivatives of functional calculus. In free probability, a more commonly used object is the

cyclic derivative ∂f, which is a noncommutative version of the difference quotient. For a polynomial f in one variable, ∂f is a noncommutative polynomial in two variables; $f \mapsto \partial f$ is linear, and for the monomial $p_n(X) = X^n$

$$(\partial p_n)(X,Y) = \sum_{k=1}^{n} X^{k-1}Y^{n-k}.$$

If X and Y commute, this coincides with the difference quotient $p_n^{[1]}(X,Y)$.

The characterizing equation (1.8) could be formulated using the cyclic derivative instead: we again ask for a distribution commonly shared by X° and Y°, now assumed to be *freely* independent (see Definition 2.1), satisfying

$$E[Xf(X)] = \sigma^2 E[\partial f(X^\circ, Y^\circ)] \quad \text{for all polynomials } f. \tag{1.9}$$

Because the cyclic derivative introduces no terms that alternate more than once between X and Y, 1.8 and 1.9 yield exactly the same moment recursion relations when applied to polynomial functions f, and thus the free zero bias X° defined by (1.8) also satisfies (1.9); the reverse is also true, at least for distributions determined by their moments. We work with the formulation in (1.8) as the classical independence there affords a larger robust toolset to study the free zero bias.

Our first main theorem, Theorem 1.1, asserts that the free zero bias distribution putatively defined by (1.8) or (1.9) does indeed exist and is unique; the theorem also collects all of its main regularity properties. Going forward, let $\rightharpoonup$ denote weak convergence of a sequence of probability measures.

Theorem 1.1. *Given any $\sigma^2 > 0$ and probability measure μ in $\mathcal{D}_{0,\sigma^2}$, there is a unique probability measure μ° on $\mathbb{R}$ that satisfies Equation (1.8) with $X \stackrel{\mathrm{d}}{=} \mu$ and $X^\circ \stackrel{\mathrm{d}}{=} \mu^\circ$. This measure is called the* **free zero bias***; we also refer to any random variable X° with distribution μ° as a* **free zero bias** *of X.*

The free zero bias has the following properties:

a) The unique fixed point of $\mathcal{D}_{0,\sigma^2} \ni \mu \mapsto \mu^\circ$ is the semicircle law of mean 0 and variance σ^2, with density $\frac{1}{2\pi\sigma^2}\sqrt{(4\sigma^2 - x^2)_+}$.

b) For any mean zero μ with finite, non-zero variance, the free zero bias μ° is absolutely continuous with respect to Lebesgue measure, and for all $a \le b$

$$\mu^\circ([a,b])^2 \le \left(\frac{b-a}{\sigma^2}\right) E[|X|].$$

c) The support of μ° is contained in the convex hull of the support of μ.

d) If $\mu_n \rightharpoonup \mu$ and $\int t^2 \mu_n(dt) \to \int t^2 \mu(dt)$, then $\mu_n^\circ \rightharpoonup \mu^\circ$.

The existence claim of Theorem 1.1 and part a) follow from a more constructive definition of the free zero bias in Definition 3.10 which mirrors the construction of the zero bias in (1.6), and is shown in Lemma 3.15 to be equivalent to (1.8); parts b) and c) are shown in Theorem 4.1, and part d) is Lemma 3.12(c).

It is convenient to write the difference quotient in the form

$$f^{[1]}(x,y) = \frac{f(x) - f(y)}{x - y} = E[f'(Ux + (1-U)y)] \tag{1.10}$$

where U is a Uniform random variable on $[0,1]$; if, for a smooth f, we choose to define $f^{[1]}$ on the diagonal as $f^{[1]}(x,x) = f'(x)$, this identity holds for all x, y including $x = y$. A priori, Equation (1.8) requires the two variable function $f^{[1]}$ to

be defined on the diagonal, and so we may canonically make the above choice when $f \in C^1$. That being said: due to Theorem 1.1(b), X° always has an absolutely continuous distribution, and as such, the diagonal values of $f^{[1]}$ do not affect the value of $E[f^{[1]}(X^\circ, Y^\circ)]$.

Remark 1.2. *Theorem 1.1(c) indicates that the free zero bias cannot "expand" the support of the measure, which is also true for the (classical) zero bias. However, the two differ on this point: the support of the classical zero bias μ^* is always* equal *to the convex hull of the support of μ, cf. [24], while strict containment is possible for the free zero bias μ° whose support need not be connected (see Example 4.2, Theorem 4.5 and (4.9)).*

Remark 1.3. *Taking test functions $f \in C_b^\infty(\mathbb{R})$ and comparing (1.5) and (1.8) in light of (1.10), we see the following relationship between the (classical) zero bias and the free zero bias:*

$$X^* \stackrel{\mathrm{d}}{=} UX^\circ + (1-U)Y^\circ \tag{1.11}$$

where U is a uniform random variable on $[0,1]$ (classically) independent from X° and Y°. This equivalence is proved carefully as (3.19) in Lemma 3.15 and will be useful below in Sections 3 and 4.

A number of benefits reaped by the classical zero and size bias transformations are consequences of a key 'replace one property' that allows one to construct the zero bias, or size bias, of a sum of independent variables by replacing a single summand. We derive the free analogs in Theorems 3.18 and 3.21. Though these forms are somewhat more complicated than their classical counterparts, they still expresses the (free) zero bias and size bias of a sum in terms of the (free) zero bias and size bias of a single summand.

As a main application of our results on and around the free zero bias, we give a new approach to, and provide new results about, **free infinite divisibility**. Recall that a random variable X is (classically) infinitely divisible if, for each $n \geq 1$, X can be represented as an independent sum

$$X \stackrel{\mathrm{d}}{=} X_{1,n} + \cdots + X_{n,n} \quad \text{where } \{X_{j,n}\}_{1 \leq j \leq n} \text{ are i.i.d.} \tag{1.12}$$

In the 1930s Kolmogorov, and then independently Lévy and Khintchine, proved the following: an L^2 random variable X with mean m and variance σ^2 is infinitely divisible if and only if there exist a probability measure ν on $\mathbb{R}$, known as the **Lévy–Khintchine measure**, such that the characteristic function $\psi_X(\xi) = E[\exp(i\xi X)]$ of X satisfies

$$\log \psi_X(\xi) = im\xi + \sigma^2 \int_{\mathbb{R}} \frac{\exp(i\xi t) - 1 - i\xi t}{t^2}\, \nu(dx), \qquad t \in \mathbb{R}. \tag{1.13}$$

(Here the integrand is taken to equal its limit value $-\frac{1}{2}\xi^2$ at $t = 0$; this is relevant if ν has mass at 0, which happens, for example, when X is normally distributed.) The measure ν is unique when $\sigma^2 > 0$. These results were later extended beyond the L^2 setting to all infinitely divisible distributions; see Section 2.7.

The (classical) zero bias and size bias transformations have deep connections to infinitely divisible distributions. [2, Proposition 3.8], [36, Lemma 4.2(b), eq. (4.3)], [28], [44], and [4] give, via the 'replace one property', insightful characterizations of infinitely divisible probability distributions on the real line that shed light on the

meaning of the Lévy–Khintchine measure. Notably, in [28] it is shown that (1.13) holds if and only if on a possibly enlarged probability space there exists a uniform $U \stackrel{\mathrm{d}}{=} \mathrm{Unif}[0,1]$ random variable and Y_0 such that

$$X^* \stackrel{\mathrm{d}}{=} X + UY_0 \qquad \text{with } U,\ X,\ Y_0 \text{ independent,} \tag{1.14}$$

and in this case, the distribution of Y_0 in (1.14) is precisely the Lévy–Khintchine measure ν in (1.13). The work [44] yields an analogous result in the positive L^1 category, showing that $\mathcal{L}(X) \in \mathcal{D}_m^+$ is infinitely divisible if and only if its size bias X^s satisfies

$$X^s \stackrel{\mathrm{d}}{=} X + Y_+ \qquad \text{with } X \text{ and } Y_+ \text{ independent.} \tag{1.15}$$

In this paper, we prove free probability analogs of (1.14) and (1.15), giving new characterizations of the free probability version of infinite divisibility and, in fact, shedding new light on both the classical and free Lévy–Khintchine formulas.

To properly state these results, we briefly introduce the distributional transforms needed. The characteristic function ψ_X encodes the distribution of X, and also encodes independence particularly well in terms of the cumulant-generating function $\log \psi_X$: X, Y are independent iff $\log \psi_{X+Y} = \log \psi_X + \log \psi_Y$ (i.e. the cumulant-generating function linearizes independence). In free probability, instead of ψ_X it is more convenient to use the **Cauchy transform**

$$G_X(z) = E\left[\frac{1}{z - X}\right], \quad z \in \mathbb{C}_+$$

where $\mathbb{C}_+$ denotes the upper half-plane $\{z \in \mathbb{C} \colon \operatorname{Im} z > 0\}$. A Cauchy transform G_X takes values in the lower half-plane $\mathbb{C}_- = -\mathbb{C}_+$; as such, since 0 is not in its range, the **reciprocal Cauchy transform** $F_X = 1/G_X$ is well-defined, and will also be used throughout this paper.

The 'free' in free probability refers to an alternative notion of independence, modeled on free groups, which makes sense for a wider class of random variables that do not necessarily commute under multiplication (for example high-dimensional random matrices); see Definition 2.1. The analog of the cumulant-generating function which linearizes **free independence** is called the **R-transform** R_X, defined in an open domain in $\mathbb{C}_+$; its relation (2.4) to G_X is not as simple as the exponential relation between ψ_X and $\log \psi_X$, but the two functions determine each other in an often computationally feasible way, and $R_{X+Y} = R_X + R_Y$ precisely when X, Y are freely independent. Corresponding to the reciprocal Cauchy transform, it is often useful to use a "reflected" version of R_X, the **Voiculescu transform** $\varphi_X(z) = R_X(1/z)$. Section 2.4 below describes all these transforms and their domains in precise detail.

Free infinite divisibility is defined as in (1.12) with (classical) independence of the summands $X_{j,n}$ replaced by free independence; see Section 2.7. In [8], Bercovici and Voiculescu proved that an L^2 random variable X with mean m and variance σ^2 is freely infinitely divisible if and only if its R-transform is of the form

$$R_X(z) = m + \sigma^2 \int_{\mathbb{R}} \frac{z}{1 - tz}\, \nu(dt) \tag{1.16}$$

for some probability measure ν on $\mathbb{R}$. Equation (1.16) is a free analog of the Lévy–Khintchine formula (1.13). It is no accident that we use ν to denote the measure here as we did in (1.16). The triple (m, σ^2, ν) characterizes all L^2 classically infinitely divisible distributions and likewise characterizes all L^2 freely infinitely

divisible distributions. This sets up an important bijection, known as the *Bercovici–Pata bijection* cf. [7], between the classical and free cases, which we describe in Section 2.7. (Moreover: as in the classical case, the setup can be modified to capture all freely infinitely divisible distributions, not only L^2 ones; we explain this in Section 2.7.)

The equivalence between the free Lévy–Khintchine formula (1.16) (or more generally (2.14)) and free infinite divisibility was a landmark result of Bercovici–Voiculescu in [8, 10]. Theorem 1.6 below gives a different approach to this equivalence along with a new characterization — the free analog (1.18) of (1.14) — using the free zero bias. What's more, in Theorem 1.8, we provide a completely new parallel development and new equivalent characterization of *positively* freely infinitely divisible distributions (see Definition 1.7) in terms of a different Lévy–Khintchine type formula, using the size bias.

In order to state these theorems, we need the following result which introduces (for our purposes) a free analog of multiplying a random variable by an independent uniform random variable.

Lemma 1.4. *Let μ be any probability measure on $\mathbb{R}$. There is a unique probability measure $\mu^\flat$ whose Cauchy transform satisfies $G_{\mu^\flat}(z) = -\sqrt{G_\mu(z)/z}$.*

If W has distribution $\mathcal{L}(W) = \mu$, we abuse notation as usual and let $W^\flat$ denote a random variable with distribution $\mathcal{L}(W^\flat) = \mu^\flat$. Lemma 1.4 is restated and proved as Lemma 3.4, which follows from the curious fact (Lemma 3.2) that the geometric mean of two Cauchy transforms of probability measures is again the Cauchy transform of some probability measure; this result may be of independent interest.

Remark 1.5. *We view the operation $W \mapsto W^\flat$ as a free version of the operation $Y \mapsto UY$ of multiplying by an independent* Unif$[0,1]$ *random variable because of the form of Theorem 1.6(c) as it relates to (1.14); see also our construction (3.12) of the free zero bias compared to its classical counterpart (1.6). To date, the authors have not been able to realize $W \mapsto W^\flat$ in more probabilistic terms (i.e. as a direct operation on random variables, or on von Neumann algebras) rather than on distributional transforms; this is a good question for future thought.*

The following main theorem introduces a new take, and a totally new equivalent condition, on infinite divisibility in the free world.

Theorem 1.6. *Let X be a random variable with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$. The following are equivalent.*

a) X is freely infinitely divisible.

b) There is a random variable V_0 so that the Voiculescu transform φ_X has the form

$$\varphi_X(z) = m + \sigma^2 G_{V_0}(z). \tag{1.17}$$

c) The free zero bias of X satisfies

$$F_{X^\circ}(z) = F_{V_0^\flat}(F_X(z)), \tag{1.18}$$

for some random variable V_0.

In the case that these equivalent statements hold true, the random variables V_0 in b) and c) have the same law, which is the weak limit as $n \to \infty$ of the square biases $X_{1,n}^\square$ (cf. (1.3)) of the summand distributions where $X = X_{1,n} + \cdots + X_{n,n}$ for freely independent identically distributed $\{X_{j,n}\}_{1 \le j \le n}$.

The equivalence of a) and b) was proved in [8], and the characterization of the free Lévy measure as a limit of square biases (without using that language) was proved in [7, Theorem 3.3(2)]. Our contribution, other than providing probabilistic arguments rather than purely complex analytic ones, is the third leg c) of the equivalence, which crucially gives the free Lévy–Khintchine measure $\mathcal{L}(V_0)$ new meaning, relating to the free zero bias. Our proof, which is the subject of Section 5, proceeds by first treating compound free Poisson distributions (cf. Section 2.5), and then extending results broadly by showing all freely infinitely divisible distributions are well-controlled limits of compound free Poissons. (Indeed, we prove that compound free Poisson distributions are precisely those for which the random variable V_0 in Theorem 1.6 satisfies $E[V_0^{-2}] < \infty$; see Theorem 5.8.) Our proof is fully self-contained in the case of compact support (Section 5.4), relying weakly on the results of [10, 8] for requisite tightness in the general case.

We also move beyond the L^2 category and treat L^1 freely infinitely divisible distributions that are positive in a strong sense.

Definition 1.7. *We say X is* **positively freely infinitely divisible**, *or free regular as in* [1], *when for all $n \geq 1$*

$$X \overset{\mathrm{d}}{=} X_{1,n} + \cdots + X_{n,n} \tag{1.19}$$

where $\{X_{j,n}\}_{1 \leq j \leq n}$ are freely independent, identically distributed, and $X_{j,n} \geq 0$.

The key defining feature here is that the summands $X_{j,n}$ are all non-negative. For classically infinitely divisible random variables that are non-negative, this holds automatically (owing to the linearity of classical convolution); in the free world, it is a strong constraint: for example shifted semicircular distributions can be positive and are freely infinitely divisible, but are never positive*ly* freely infinitely divisible; see Section 2.7.

Theorem 1.8. *Let X be a non-negative random variable with mean $m > 0$. The following are equivalent.*

a) X is positively freely infinitely divisible.

b) There is a random variable V_+ so that the Voiculescu transform φ_X has the form

$$\varphi_X(z) = mzG_{V_+}(z). \tag{1.20}$$

c) The size bias of X satisfies

$$F_{X^s}(z) = F_{V_+}(F_X(z)). \tag{1.21}$$

In the case that these equivalent statements hold true, the random variables V_+ in b) and c) have the same law, which is the weak limit as $n \to \infty$ of the size biases $X_{1,n}^s$ (cf. (1.1)) of the summand distributions where $X = X_{1,n} + \cdots + X_{n,n}$ for freely independent identically distributed $\{X_{j,n}\}_{1 \leq j \leq n}$.

Both conditions b) and c) are entirely new to the theory of free infinite divisibility; moreover, the 'positive Lévy–Khintchine formula' of (1.20) has no classical equivalent that the authors are aware of.

One more advantage of our approach: it yields the previously unknown result that that *every* probability measure is the free Lévy–Khintchine measure for a two-parameter family of L^2 freely infinitely divisible distributions, indexed by their mean and variance; also, every positively-supported probability measure yields a

one-parameter family of positively freely infinitely divisible distributions, indexed by their mean.

Proposition 1.9. *Let ν be a probability measure on $\mathbb{R}$, and let $m \in \mathbb{R}$ and $\sigma^2 \geq 0$.*
(1) There is a unique L^2 freely infinitely divisible random variable X with mean m and variance σ^2 such that $\varphi_X(z) = m + \sigma^2 G_\nu(z)$; i.e. such that Theorem 1.6(b) holds with $V_0 \stackrel{\mathrm{d}}{=} \nu$.
(2) If $\operatorname{supp}\nu \subseteq [0,\infty)$ and $m > 0$, there is a unique L^1 positively freely infinitely divisible random variable X with mean m such that $\varphi_X(z) = mzG_\nu(z)$; i.e. such that Theorem 1.8(b) holds with $V_+ \stackrel{\mathrm{d}}{=} \nu$.

The proof can be found in Propositions 5.12 and 5.14.

In the special case that a positively freely infinitely divisible random variable is not just L^1 but L^2, it possess both types of free Lévy–Khintchine measures, i.e. both V_0 and V_+. They are nicely related, which also gives a new elegant characterization of L^2 positively infinitely divisible distributions (and their translates).

Theorem 1.10. *Let X be a freely infinitely divisible L^2 random variable. Assume X is bounded below. The following are equivalent:*

(1) There is some $\alpha > 0$ so that $X + \alpha$ is a positively freely infinitely divisible random variable.
(2) The random variable V_0 corresponding to X in Theorem 1.6(b) is non-negative and has a finite negative moment: $V_0 \geq 0$ and $E[V_0^{-1}] < \infty$.

When these equivalent conditions hold, if X' is any positively freely infinitely divisible translate of X, and V_+ is the random variable corresponding to X' in Theorem 1.8(b), then

$$(V_+)^s = V_0 \quad \text{and} \quad E[V_+] = \frac{\mathrm{Var}[X']}{E[X']}.$$

The proof of this final main theorem appears towards the end of Section 5.3.

Remark 1.11. *For X and Y two freely independent, positively freely infinitely divisible L^1 random variables with means m, m' respectively, by the additivity of the Voiculescu transformation and Theorem 1.8(b),*

$$\varphi_{X+Y}(z) = mzG_{V_+(X)}(z) + m'zG_{V_+(Y)}(z) = (m+m')zG_{V_+(X+Y)}(z)$$

where

$$V_+(X+Y) = \frac{m}{m+m'}V_+(X) + \frac{m'}{m+m'}V_+(Y). \tag{1.22}$$

That is, the free Lévy–Khintchine measure V_+ of the sum is the mixture of the measures of the summands, taken proportional to the means.

Similarly when X and Y are freely independent, freely infinitely divisible L^2 random variables we have via Theorem 1.6(b) that

$$\varphi_{X+Y}(z) = m + m' + (\sigma^2 + \tau^2)G_{V_0(X+Y)}(z)$$

where σ^2, τ^2 are the respective variances of X and Y, and

$$V_0(X+Y) = \frac{\sigma^2}{\sigma^2+\tau^2}V_0(X) + \frac{\tau^2}{\sigma^2+\tau^2}V_0(Y). \tag{1.23}$$

That is, in the L^2 case the free Lévy–Khintchine measure V_0 of the sum is the mixture taken proportional to the variances.

By the relation $\varphi_{\alpha X}(z) = \alpha\varphi(z/\alpha)$ one can verify that non-zero scalings in the L^2 case, and positive scalings in the L^1 case, both yield the same scaling of V_0 and V_+ respectively. However, the two cases differ widely for translations, as can be seen by specializing the relations above to the case where Y is a point mass.

In particular, for a given infinitely divisible L^2 random variable X, its corresponding V_0 in Theorem 1.6(b) is invariant under translations of X; the free Lévy–Khintchine formula (1.17) only changes by translating the isolated mean term.

On the other hand, as Theorem 1.10 involves translations which leave V_0 invariant but do change V_+, one may wonder about the validity of the final claim that $(V_+)^s = V_0$. This apparent problem is resolved by specializing (1.22) to the case $\mathcal{L}(Y) = \delta_{m'}$, yielding

$$V_+(X + m') = \frac{m}{m + m'} V_+(X) + \frac{m'}{m + m'} \delta_0, \tag{1.24}$$

which shows that translations affect V_+ only by mixing with a point mass at 0, and such mixtures do not change the size bias of V_+, cf. Lemma 3.9(b).

More generally, we may move directly from (1.22) to (1.23) using (3.11), and in the reverse direction taking the inverse size bias of both its sides, and also using the final claim of Theorem 1.10 which provides the factor $E[V_+] = \sigma^2/m$ that converts a mixture by the mean to one taken by the variance.

This work is organized as follows. The remainder of this Introduction, Section 1.1, discusses the connection between the zero bias transform and Stein kernels in both the classical and free probability contexts. Section 2 is a fairly self-contained primer on free probability, in particular as it relates to analytic distributional transforms, conditioning, and infinite divisibility. In Section 3 we prove the existence of the free zero bias and related distributional transforms, and develop their basic properties and relations to each other. In particular, Theorems 3.18 and 3.21 provide free analogs of the classical key 'replace one properties' for the free zero and size bias transformations. In Section 4 we show that, like its classical cousin, the free version zero bias has an absolutely continuous distribution, completing the proof of Theorem 1.1. and explore how disconnected its support can be with a full treatment of free zero bias transforms of finitely-supported distributions. We then turn, in Section 5, to infinitely divisible random variables, and prove Theorems 1.6, 1.8, and their consequences discussed above. Finally, Section 6 uses the results of Theorems 1.6 and 1.8 to compute (quite effectively) the distributions of several new freely infinitely divisible distributions, including a new two-parameter family of positively freely infinitely divisible distributions.

1.1. **(Free) Zero Bias and (Free) Stein Kernels.** Our work is related to that of [13] and [21], where Stein kernels, familiar in classical probability [12, 15], are shown to also exist in free probability. Classically, we say that a random variable $A_*(X)$ is a Stein kernel for X if

$$E[Xf(X)] = E[A_*(X)f'(X)] \quad \text{for all } f \in C_b^\infty(\mathbb{R}). \tag{1.25}$$

We note that (1.25) is similar to (1.5) in that it modifies Stein's original identity for the normal by changing some feature on the right hand side, for (1.5) the distribution of X, and for (1.25) the multiplier of the derivative of f. If the law $\mathcal{L}(X^*)$ is absolutely continuous with respect to $\mathcal{L}(X)$ with Radon-Nikodym derivative D_*,

then from (1.5) and the change-of-variables theorem,

$$E[Xf(X)] = \sigma^2 E[f'(X^*)] = E\left[\sigma^2 D_*(X) f'(X)\right].$$

In particular, under the stated absolute continuity requirement, $A_*(X) = \sigma^2 D_*(X)$ provides a Stein kernel.

The notion of a Stein kernel in free probability was first introduced in [21], and used to obtain a free counterpart of the HSI inequality, connecting entropy, Stein discepancy and free Fisher information, as well as a rate of convergence in the entropic free central limit theorem. The work [13] showed that free Stein kernels, unlike in the classical case, always exist; moreover, it showed that free Stein kernels are never unique by providing two constructions that lead to distinct results, one inspired by an unpublished note of Giovanni Peccati and Roland Speicher in the case where the potential is $V(x) = n^{-1}\sum_{i=1}^n x_i^2$ and where $n = 1$, and a second one constructed via the Reisz representation theorem.

For a single random variable X, we say that $A_0(X, Y)$ is a Free Stein kernel for X when (consistent with our choice above to replace derivative with difference quotient (1.10))

$$E[Xf(X)] = E[A_0(X, Y) f^{[1]}(X, Y)] \quad \text{for all } f \in C_b^\infty(\mathbb{R}), \tag{1.26}$$

where Y is a (classically) independent copy of X. A relation similar to the one in the classical case holds between free Stein kernels and the free zero bias. Indeed, when the distribution $\mathcal{L}(X^\circ)$ of X° is absolutely continuous with respect to $\mathcal{L}(X)$ with Radon–Nikodym derivative D_0, from (1.8) and the change-of-variables theorem we have

$$E[Xf(X)] = \sigma^2 E[f^{[1]}(X^\circ, Y^\circ)] = \sigma^2 E[D_0(X) D_0(Y) f^{[1]}(X, Y)]. \tag{1.27}$$

Comparing with (1.26), we see, as in the classical case under a similar absolute continuity assumption, that free zero bias provides a free Stein kernel, here given by $A_0(X, Y) = D_0(X)D_0(Y)$, which has the form of a product of some function of X multiplied by the same function of Y.

In the setting of (1.26), the first of the two distinct Stein kernels discussed in [13] corresponds to $A_0(x, y) = (x-y)^2/2$, for which (1.26) is nearly immediately verified to hold upon applying the definition of the difference quotient $f^{[1]}$ and the fact that $E[(X - Y)(f(X) - f(Y))] = 2E[Xf(X)]$. This kernel must be distinct from the one in (1.27) generated by the free zero bias, as $(x - y)^2$ cannot be written in the necessary product form. The construction of the second kernel in [13] requires that X satisfies the free Poincaré inequality. However, no such requirement is needed for the kernel for X in (1.27). In other words: the free zerio bias affords yet a third distinct construction of a free Stein kernel.

2. A Free Probability Primer

Free probability is a noncommutative probability theory introduced by Voiculescu in the late 1980s and early 1990s, initially with the intent to import core ideas from probability, statistics, and information theory to solve structure and isomorphism problems in operator algebras. Through its intimate connection with random matrix theory, it has grown over the last three decades into a flourishing field of study on its own. The reader may wish to consult [40, 38] for detailed treatises on free probability.

2.1. **Noncommutative Probability Spaces.** The arena is a *noncommutative probability space* $(\mathcal{A}, E)$, which replaces the algebra of bounded random variables and the expectation functional on them in the classical case. $\mathcal{A}$ is a unital $*$-algebra, like the algebra of $n \times n$ complex matrices for example, where the adjoint $*$ operation is conjugate transpose. (Note: to avoid notational conflict with the (classical) zero bias of (1.5), for the rest of this section we use $\dagger$ to refer to the adjoint operation rather than $*$.) $E\colon \mathcal{A} \to \mathbb{C}$ is a *state*: a linear functional satisfying $E(1) = 1$ that is *positive* in the sense that, for any $X \in \mathcal{A}$ other than 0, $E(X^\dagger X) > 0$. As the notation suggests, E should be thought of as *expectation*; indeed, if $\mathcal{A}$ is the algebra of bounded random variables on some probability space, then *the* expectation is the standard choice for E. Some authors prefer to use a letter like τ or φ for the state functional in free probability; we will use E throughout.

For some results (which we will need), more assumptions are required. A W^*-*probability space* is a noncommutative probability space $(\mathcal{A}, E)$ where $\mathcal{A}$ is a von Neumann algebra (a weakly closed subalgebra of the algebra of bounded operators on a Hilbert space), and E is *tracial*: $E(X_1X_2) = E(X_2X_1)$ for all $X_1, X_2 \in \mathcal{A}$, and also continuous in the weak* topology (identifying the algebra with its predual). This latter continuity condition, sometimes called *normal continuity*, is designed to make a version of the Dominated Convergence Theorem hold for E; in particular, an equivalent statement is that, for any bounded increasing net of positive operators $\{X_\alpha\}$, $\sup_\alpha E[X_\alpha] = E[\sup_\alpha X_\alpha]$. We will make these assumptions going forward, for convenience. Typical examples are $n \times n$ matrices equipped with the normalized trace; any classical probability space; and various standard type II_1 von Neumann algebras. As with classical probability, one doesn't typically work in some particular explicit (noncommutative) probability space; the random variables (i.e. elements of $\mathcal{A}$) are assumed to "live" on a space with enough structure to make the usual results work, and that typically means a W^* probability space.

2.2. **Free Independence.** The *free* in free probability refers to a noncommutative "independence" notion for random variables. Algebraically, statistical independence of random variables (corresponding to joint laws being product measures) corresponds to tensor product; in free probability, the independence rule is instead modeled on *free* product from group theory. Ultimately, it is best to express it as a moment-factoring condition. For classical bounded random variables, independence can be stated combinatorially as the exepctation factoring property $E(X_1^n X_2^m) = E(X_1^n)E(X_2^m)$ for all $n, m \in \mathbb{N}$. This equality will also hold for freely independent random variables, but it is not sufficient: when random variable multiplication is noncommutative, this rule doesn't handle expressions like $E(X_1X_2X_1X_2)$. To handle such expressions, freeness is most easily stated in terms of subalgebras.

Definition 2.1. *Let $(\mathcal{A}, E)$ be a noncommutative probability space. A collection $\mathcal{A}_1, \ldots, \mathcal{A}_d$ of unital $*$-subalgebras of $\mathcal{A}$ is called* **freely independent** *if the following holds true: given indices $i_1, i_2, \ldots, i_n$ that are consecutively distinct (i.e. $i_1 \neq i_2$, $i_2 \neq i_3$, etc.) and random variables $X_k \in \mathcal{A}_{i_k}$ that are centered (i.e. $E(X_k) = 0$), the product $X_1X_2 \cdots X_k$ is also centered, that is, $E[X_1X_2 \cdots X_k] = 0$.*

We say random variables $X_1, \ldots, X_d$ are freely independent if the $*$-subalgebras they generate are freely independent.

For our purposes, it is only necessary to deal with *selfadjoint* variables $X = X^\dagger$. For a selfadjoint X, the $*$-algebra it generates is just the set of all polynomial functions of it, $\{p(X) \colon p \text{ is a polynomial of one variable}\}$. For such variables, then, free independence can be expressed a bit more directly: selfadjoint $X_1, \ldots, X_d$ are freely independent if, given any consecutively distinct indices $i_1 \neq i_2 \neq i_3 \neq \cdots \neq i_n$, and polynomials $p_1, \ldots, p_n$ such that $E[p_1(X_{i_1})] = \cdots = E[p_n(X_{i_n})] = 0$, it follows that $E[p_1(X_{i_1}) \cdots p_n(X_{i_n})] = 0$. By using the usual centering trick, applying this property to variables $X_i^m - E[X_i^m]$, one learns that free random variables satisfy the usual factoring property $E[X_1^n X_2^m] = E[X_1^n]E[X_2^m]$, as well as a litany of more complex noncommutative factoring properties, for example $E[X_1X_2X_1X_2] = E[X_1^2]E[X_2]^2 + E[X_1]^2E[X_2^2] - E[X_1]^2E[X_2]^2$. In general, if random variables are freely independent, any mixed moment can be expressed (as a polynomial function) in terms of the variables' individual moments.

Another benefit of selfadjoint variables is that, like classical random variables, they have probability distributions. Associated to each selfadjoint X in a W^* probability space is a unique probability measure μ_X on $\mathbb{R}$ sharing the same moments:

$$\int_{\mathbb{R}} x^n \, \mu_X(dx) = E[X^n], \qquad n \in \mathbb{N}.$$

Mirroring the classical setting, we denote $\mu_X = \mathcal{L}(X)$ as the **law** of X. Existence follows from the spectral theorem for selfadjoint operators: associated to X is a projection-valued Borel measure Π^X on $\mathbb{R}$ such that $X = \int_{\mathbb{R}} x \, \Pi^X(dx)$, and then functional calculus of X is given by $f(X) = \int_{\mathbb{R}} f(x) \, \Pi^X(dx)$ (which is consistent with analytic functional calculus when f is analytic, e.g. a polynomial). Composing with the state E on both sides yields $E[f(X)] = \int_{\mathbb{R}} f(x) \, E \circ \Pi^X(dx)$, showing that $\mu_X = E \circ \Pi^X$ is the desired law.

For now, we restrict to the W^* probability space context, meaning that the random variables are bounded operators on a Hilbert space. This implies that their distributions are compactly-supported (we will discuss removing this requirement shortly), and hence uniquely determined by their moments, which establishes uniqueness of the law. Moreover, any compactly-supported probability distribution μ arises this way (just take any classical random variable with law μ). A deeper result is that any collection of compactly-supported probability measures can be coupled together as the laws of freely independent random variables on a common space (see Definition 2.1). This leads to a new operation on such measures, **free convolution** $\boxplus$. The measure $\mu \boxplus \nu$ is defined to be the law of $X_1 + X_2$ where X_1 and X_2 are freely independent with laws μ and ν respectively. To see this is well defined, we look at moments. The moments of $X_1 + X_2$ are all linear combinations of mixed moments in X_1 and X_2, which in turn factor in terms of moments of X_1 and X_2 separately. Ergo, the moments of $X_1 + X_2$ are completely determined by the moments of X_1 and X_2 separately. Unlike the classical case, there is no simple integral formula for $\mu \boxplus \nu$ in terms of μ and ν; but there is an approach using analytic transforms which we now describe.

2.3. **Cauchy Transform.** As mentioned on page 6, for a probability measure μ on $\mathbb{R}$, its **Cauchy transform** G_μ is the analytic function defined on the complement of the support of μ by

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - t} \, \mu(dt). \tag{2.1}$$

If X is a selfajoint random variable, then from the definition (above) of its law, we have

$$G_{\mu_X}(z) = E[(z-X)^{-1}]$$

where the function $f_z(x) = (z-x)^{-1}$ is real analytic on $\mathbb{R}$ for any $z \in \mathbb{C}_+$. This reflects the change-of-variables formula in classical probability. When convenient, we may write this as $G_X = G_{\mu_X}$.

Any Cauchy transform is an analytic map from the upper half plane $\mathbb{C}_+ = \{x+iy \colon x \in \mathbb{R}, y > 0\}$ into the lower half plane $\mathbb{C}_- = -\mathbb{C}_+$. (Since μ is a real measure, the domain of G_μ can really be taken as $\mathbb{C} \setminus \mathbb{R}$, and $G_\mu(\bar{z}) = \overline{G_\mu(z)}$, hence it also maps $\mathbb{C}_-$ into $\mathbb{C}_+$.) The Cauchy transform also satisfies $iy\, G_\mu(iy) \to 1$ as $y \uparrow \infty$, and these conditions characterize Cauchy transforms.

Proposition 2.2. *Let $G \colon \mathbb{C}_+ \to \mathbb{C}_-$ be a holomorphic function with the property that $\lim_{y\to\infty} iy\, G(iy) = 1$. Then there is a unique probability measure μ on $\mathbb{R}$ such that $G = G_\mu$.*

More generally, the asymptotic $\frac{1}{z}$ behavior of Cauchy transforms holds along any path with $|z| \to \infty$ that stays in a cone $y > m|x|$, $m > 0$ in the upper half plane.

In the case that a random variable X has an exponential moment $E[\exp(\alpha|X|)] < \infty$ for some $\alpha > 0$, the Cauchy transform G_X has a Laurent series expansion

$$G_X(z) = E\left[\frac{1}{z-X}\right] = \frac{1}{z} E\left[\frac{1}{1-X/z}\right] = \sum_{n\geq 0} E[X^n] z^{-(n+1)} \tag{2.2}$$

which converges locally uniformly for $|z|$ sufficiently large. (Indeed, the ordinary moment generating function $M_X(z) = \sum_n E[X^n] z^n$ is given by $M_X(z) = \frac{1}{z} G_X(\frac{1}{z})$ in this case, and has positive radius of convergence.)

The Cauchy transform is well-defined even if X has no moments at all; (2.2) suggests that information about moments can be derived from certain limits or derivatives of the Cauchy transform. We provide one simple such result here which will be useful in Section 5.3.

Lemma 2.3. *Let $X \geq 0$ be a non-negative random variable. Suppose that the limit*

$$\lim_{y\to\infty} iy\,(iy\, G_X(iy) - 1) = r$$

exists and is in $\mathbb{R}$. Then $X \in L^1$ and $E[X] = r$.

Proof. For any $z \in \mathbb{C}_+$, note that

$$zG_X(z) - 1 = zE\left[\frac{1}{z-X}\right] - 1 = E\left[\frac{z}{z-X}\right] - E\left[\frac{z-X}{z-X}\right] = E\left[\frac{X}{z-X}\right].$$

Hence

$$z(zG_X(z) - 1) = E\left[\frac{zX}{z-X}\right] = E\left[\frac{X}{1-X/z}\right].$$

Taking $z = iy$ we have

$$iy(iy\, G_X(iy) - 1) = E\left[\frac{X}{1-X/iy}\right] = E\left[\frac{X(1+X/iy)}{1+X^2/y^2}\right].$$

Since both $\frac{X}{1+X^2/y^2}$ and $\frac{X^2}{y(1+X^2/y^2)}$ are bounded hence L^1, this decomposes into its real and imaginary parts as

$$iy(iy\, G_X(iy) - 1) = E\left[\frac{X}{1+X^2/y^2}\right] - iE\left[\frac{X^2}{y(1+X^2/y^2)}\right].$$

By assumption, the limit as $y \to \infty$ is equal to $r \in \mathbb{R}$, and so it follows that

$$\lim_{y\to\infty} E\left[\frac{X}{1+X^2/y^2}\right] = r \quad \text{and} \quad \lim_{y\to\infty} E\left[\frac{X^2}{y(1+X^2/y^2)}\right] = 0.$$

Since $X \geq 0$, $\frac{X}{1+X^2/t^2}$ increases to X as $t \uparrow \infty$; thus, the result follows from the Monotone Convergence Theorem. □

The Cauchy transform of a probability measure μ completely characterizes μ, and moreover this relationship is robust under weak limits. The following is a complex analysis exercise; a detailed proof can be found in [33, Section 8].

Proposition 2.4. *Let μ be a probability measure on $\mathbb{R}$, with Cauchy transform G_μ. The* **Stieltjes inversion formula** *recovers μ from G_μ as the weak limit $\mu_\epsilon \rightharpoonup \mu$ as $\epsilon \downarrow 0$, where μ_ϵ has real analytic density ϱ_ϵ*

$$\varrho_\epsilon(x) = -\frac{1}{\pi}\operatorname{Im} G_\mu(x+i\epsilon).$$

Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures on $\mathbb{R}$.

1. *If $\mu_n \rightharpoonup \mu$ for some probability measure μ, then G_{μ_n} converges to G_μ uniformly on compact subsets of $\mathbb{C}_+$.*
2. *Conversely, if G_{μ_n} converges pointwise to a function G that is analytic on $\mathbb{C}_+$, then $G = G_\mu$ for some finite measure μ with $\mu(\mathbb{R}) \leq 1$, and $\{\mu_n\}_{n=1}^\infty$ converges vaguely to μ.*

Example 2.5. *Let X be a Cauchy distributed random variable, with density $\varrho_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Then its Cauchy transform is $G_X(z) = \frac{1}{z+i}$. This can be computed from the definition using contour integrals; it also follows directly from Proposition 2.4. Indeed, it is straightforward to check that the function $G(z) = \frac{1}{z+i}$ is a Cauchy transform, since $zG(z) \to 1$ as $|z| \to \infty$ and $G(\mathbb{C}_+) \subseteq \mathbb{C}_-$. From the Stieljtes inversion formula, the measure whose Cauchy transform is G has density*

$$-\lim_{\epsilon\downarrow 0}\frac{1}{\pi}\operatorname{Im} G(x+i\epsilon) = -\lim_{\epsilon\downarrow 0}\frac{1}{\pi}\operatorname{Im}\left(\frac{1}{x+i\epsilon+i}\right) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Thus, the unique measure whose Cauchy transform is $G(z) = \frac{1}{z+i}$ is the Cauchy distribution.

This gives the odd association that the Cauchy distribution behaves like a point mass at $-i$. Indeed, it points out that the injectivity $\mu \mapsto G_\mu$ only holds for probability measures on $\mathbb{R}$, not generally for probability measures on $\mathbb{C}$.

Remark 2.6. *Example 2.5 shows that the limit in Lemma 2.3 can exist (as a complex number) even if $X \notin L^1$. Indeed, if X has the Cauchy distribution,*

$$z(zG_X(z)-1) = z\left(z\cdot\frac{1}{z+i}-1\right) = \frac{-i}{1+z/i}$$

and so the limit as $|z| \to \infty$ exists and is equal to $-i$. The Cauchy distribution does not have a finite moment, however.

2.4. **The Voiculescu Transform and the R-transform.** The domain of invertibility of a Cauchy transform will be important in our analysis. For $\alpha, \beta > 0$, the **truncated cone** $\Gamma_{\alpha,\beta}$ (also known as a **Stolz angle**) is

$$\Gamma_{\alpha,\beta} = \{z = x+iy \in \mathbb{C}_+ : \alpha y > |x|, |z| > \beta\}.$$

Then $zG_\mu(z) \to 1$ for any path $z \in \Gamma_{\alpha,\beta}$ with $|z| \to \infty$. It is often convenient to work with the **reciprocal Cauchy transform** mentioned above: $F_\mu = 1/G_\mu$; which is an analytic self-map of the upper half plane $\mathbb{C}_+$ satisfying $F_\mu(z) \sim z$ for z large in some $\Gamma_{\alpha,\beta}$. This asymptotic behavior suggests that F_μ (and hence G_μ) is an invertible function for sufficiently large α, β, and this is indeed true.

Proposition 2.7 (Proposition 5.4 in [8]). *If μ is any probability measure on $\mathbb{R}$, and $\alpha > 0$, then for any $\epsilon \in (0, \alpha)$, there exists some $\beta > 0$ such that F_μ is univalent in $\Gamma_{\alpha,\beta}$ (i.e. is one-to-one and has an analytic inverse), with range $F_\mu(\Gamma_{\alpha,\beta}) \supseteq \Gamma_{\alpha-\epsilon,\beta(1+\epsilon)}$.*

That is: on a sufficiently truncated cone, F_μ has an analytic inverse whose domain includes a truncated cone arbitrarily close to the domain of F_μ.

On an appropriate truncated cone, F_μ^{-1} is an analytic map, and given that $F_\mu(z) \sim z$, the same holds true for F_μ^{-1}. The **Voiculescu transform** φ_μ, introduced on page 6, is defined, a priori on some truncated cone $\Gamma_{\alpha,\beta}$, by

$$\varphi_\mu(z) = F_\mu^{-1}(z) - z \quad \text{for } z \in \Gamma_{\alpha,\beta}. \tag{2.3}$$

From Proposition 2.7, the domain of φ_μ can be taken as a truncated cone $\Gamma_{\alpha,\beta_\alpha}$ for any $\alpha > 0$ (given large enough β_α); and the conjugate symmetry implies the vertical reflection in the lower half plane can be included as well. So φ_μ is defined on the union over $\alpha > 0$ of such truncated cones.

The Voiculescu transform satisfies $\varphi_\mu(z)/z \to 0$ as $z \to \infty$ in its domain. Moreover, this convergence (and any truncated cone in the domain) is robust under weak convergence of measures.

Proposition 2.8 (Proposition 5.7 in [8]). *A sequence $\{\mu_n\}_{n=1}^\infty$ of probability measures on $\mathbb{R}$ converges weakly $\mu_n \rightharpoonup \mu$ to some probability measure μ if and only if there are $\alpha, \beta > 0$ such that φ_{μ_n} all converge uniformly on compact subsets of $\Gamma_{\alpha,\beta}$ to a function φ and $\varphi_{\mu_n}(z) = o(z)$ uniformly in n as $|z| \to \infty$, $z \in \Gamma_{\alpha,\beta}$. In this case, $\varphi = \varphi_\mu$ in $\Gamma_{\alpha,\beta}$.*

The **R-transform** R_μ is defined in term of the Voiculescu transform as $R_\mu(z) = \varphi_\mu(1/z)$; from (2.3), the domain of R_μ contains the reciprocal image of the above-described region, which is an open disk centered at the origin with narrow regions along the real axis possibly removed. Equation (2.3) defining the Voiculescu transform φ_μ on a precise domain can be rephrased as an implicit equation for φ_μ in terms of the reciprocal Cauchy transform G_μ; correspondingly, R_μ satisfies an implicit equation in terms of the Cauchy transform G_μ:

$$F_\mu(\varphi_\mu(z) + z) = z, \qquad G_X\left(R_X(z) + 1/z\right) = z. \tag{2.4}$$

The key property of the Voiculescu transform and the R-transform is that they linearize free independence:

$$\varphi_{\mu \boxplus \nu}(z) = \varphi_\mu(z) + \varphi_\nu(z), \qquad R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z)$$

for all z in appropriate domains where the transforms make sense. This fact was proved for compactly-supported measures, the only kind for which the combinatorial definition (Definition 2.1) of freeness makes sense a priori, in [45]. Later it was shown [8, 37] that for *any* probability measures μ, ν on $\mathbb{R}$, there is a unique probability measure υ such that $\varphi_\mu + \varphi_\nu = \varphi_\upsilon$ in a truncated cone; this allows the extension

of free convolution to all probability measures, defining $\mu \boxplus \nu = \upsilon$. To be clear: $\mu \boxplus \nu$ is defined by computing φ_μ and φ_ν (each determined completely by its measure, through F_μ and F_ν), summing, and then computing $F^{-1}_{\mu\boxplus\nu}(z) = \varphi_\mu(z) + \varphi_\nu(z) + z$ on an appropriate truncated cone. By Proposition 2.7 and the above-mentioned papers, the inverse of this function coincides with reciprocal Cauchy transform F_υ of a unique probability measure υ on a closely related truncated cone. This function has an analytic continuation to all of $\mathbb{C}_+$, and then (taking reciprocal) the measure $\upsilon = \mu \boxplus \nu$ can be recovered via the Stieltjes inversion formula.

If μ is compactly-supported, the domain of analyticity of the R-transform R_μ always includes an open disk centered at the origin. Its Taylor coefficients at the origin are polynomial functions of the moments of μ, called the **free cumulants**; indeed, the first three are the classical cumulants (mean, variance, skewness), while they disagree from the fourth on. The R-transform is the free analog of the cumulant-generating function (i.e. the log-characteristic function). The relationship between R_μ and G_μ (2.4) is more complicated than the simple exponential relationship between the characteristic function ψ_μ and the cumulant-generating function $\log \psi_\mu$, but it is still possible in many cases to use this relationship for exact computation.

Remark 2.9. *As with the Cauchy transform and its reciprocal, we abuse notation with the Voiculescu transform and R-transform and write $\varphi_X = \varphi_{\mu_X}$ and $R_X = R_{\mu_X}$, for a (potentially unbounded) selfadjoint random variable X. This gives the concise statement that two random variables X_1, X_2 are freely independent if and only if $R_{X_1+X_2} = R_{X_1} + R_{X_2}$.*

Like the Cauchy transform (Proposition 2.4), analytic regularity of the R-transform shows that free convolution is robust under weak limits.

Proposition 2.10 (Proposition 5.7 in [8])**.** *Let μ_n, μ, ν_n, ν be probability measures on $\mathbb{R}$, and suppose $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Then $\mu_n \boxplus \nu_n \rightharpoonup \mu \boxplus \nu$.*

2.5. **Free Poisson Distributions.** One of the core laws in classical probability is the Poisson distribution: $\mathrm{Poiss}(\lambda) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \delta_n$. It arises in many contexts parallel to the way the normal distribution does, since it is a universal scaling limit (in a different scaling regime from the normal). In simplest terms, the Poisson limit theorem states that $\mathrm{Poiss}(\lambda)$ approximates an i.i.d. sum of n Bernoulli trials with success rate λ/n; it is the weak limit

$$\mathrm{Poiss}(\lambda) = \lim_{n\to\infty} \left(\left(1 - \frac{\lambda}{n}\right) \delta_0 + \frac{\lambda}{n} \delta_1 \right)^{*n} \tag{2.5}$$

(where, in (2.5), the $*$ in the exponent refers to convolution). More generally, *compound Poisson* distributions are achieved by replacing δ_1 in (2.5) with any probability measure υ, thence denoted the *jump distribution*.

Free Poisson distributions are defined using the same limit procedure as (2.5) but with free convolution.

Definition 2.11. *Let $\lambda > 0$ and let υ be a probability distribution on $\mathbb{R}$. The* **compound free Poisson** *distribution $\mathrm{FP}(\lambda, \upsilon)$ is defined to be the following weak limit of free convolution powers of mixtures of δ_0 and υ:*

$$\mathrm{FP}(\lambda, \upsilon) := \lim_{n\to\infty} \left(\left(1 - \frac{\lambda}{n}\right) \delta_0 + \frac{\lambda}{n} \upsilon \right)^{\boxplus n}.$$

The parameters λ and υ are called the rate *and* jump distribution *as in the classical case (although there are no jumps in the free case).*

The R-transform can be used to characterize $\mathrm{FP}(\lambda, \upsilon)$; as shown in [39, Proposition 12.15],

$$R_{\mathrm{FP}(\lambda,\upsilon)}(z) = \lambda \int_{\mathbb{R}} \frac{t}{1-tz}\, \upsilon(dt). \tag{2.6}$$

This shows that $\mathrm{FP}(\lambda, \nu)$ has as many finite moments as υ does. Compound free Poisson distributions will play an important role in Section 5.

If $M_j \overset{\mathrm{d}}{=} \mathrm{FP}(\lambda_j, \upsilon_j)$ are freely independent for $1 \leq j \leq n$, it follows easily from Definition 2.11 and R-transform calculations that the sum $M_1 + \cdots + M_n$ is distributed as $\mathrm{FP}(\lambda, \upsilon)$ where $\lambda = \sum_{j=1}^n \lambda_j$ and ν is the mixture $\nu = \sum_{j=1}^n \frac{\lambda_j}{\lambda} \upsilon_j$. (Nearly the same calculation, using characteristic functions, shows that sums of independent compound free Poisson random variables are compound free Poisson with the sum of the rates and the appropriate mixture of the jump distributions). Since every measure is approximated by a mixture of finitely-supported measures, in principle any compound free Poisson can be built as a linear combination of free Poisson distributions are given by the two parameter family

$$\mathrm{FP}(\lambda, \alpha) := \mathrm{FP}(\lambda, \delta_\alpha) = \lim_{n\to\infty} \left(\left(1 - \frac{\lambda}{n}\right) \delta_0 + \frac{\lambda}{n} \delta_\alpha \right)^{\boxplus n}.$$

(One could similarly define a two parameter family of classical Poisson distributions, but a simple calculation shows that $\mathrm{Poiss}(\lambda, \delta_\alpha)$ is just the dilation of $\mathrm{Poiss}(\lambda)$ by α. The same does not hold true for free Poissons.) Equation 2.6 shows that

$$R_{\mathrm{FP}(\lambda,\alpha)}(z) = \frac{\lambda\alpha}{1-\alpha z}. \tag{2.7}$$

From here it follows that the mean of $\mathrm{FP}(\lambda, \alpha)$ is $\lambda\alpha$, and the variance is $\lambda\alpha^2$. (In general, the free cumulants of $\mathrm{FP}(\lambda, \nu)$ are λ times the moments of ν – this is the power series interpretation of (2.6).)

From the R-transform (2.7), we can explicit computation of the Cauchy transform of $\mathrm{FP}(\lambda, \alpha)$. Indeed, from (2.4), we have that $G_{\mathrm{FP}(\lambda,\alpha)}$ is the function inverse of $\frac{\lambda\alpha}{1-\alpha z} + \frac{1}{z}$; in other words

$$\frac{\lambda\alpha}{1 - \alpha G_{\mathrm{FP}(\lambda,\alpha)}(z)} + \frac{1}{G_{\mathrm{FP}(\lambda,\alpha)}(z)} = z. \tag{2.8}$$

This is a quadratic equation, whose explicit solution (with the correct asymptotic and mapping behavior for a Cauchy transform) is

$$G_{\mathrm{FP}(\lambda,\alpha)}(z) = \frac{z + (1-\lambda)\alpha - \sqrt{(z-(1+\lambda)\alpha)^2 - 4\alpha^2\lambda}}{2\alpha z}. \tag{2.9}$$

Here the square root has a branch cut along $[0, \infty)$, i.e. $\sqrt{re^{i\theta}} = r^{1/2}e^{i\theta/2}$ with $\theta \in [0, 2\pi)$. Using the Stiletjes inversion formula, this yields a law with a continuous compactly supported density $\varrho_{\lambda,\alpha}$ in the case that $\lambda \geq 1$:

$$\varrho_{\lambda,\alpha}(x) = \frac{\sqrt{4\lambda\alpha^2 - (x - \alpha(1+\lambda))^2}}{2\pi\alpha x}, \qquad \alpha(1-\sqrt{\lambda})^2 \leq x \leq \alpha(1+\sqrt{\lambda})^2.$$

When $\lambda < 1$, this density has total mass λ, and here $\mathrm{FP}(\lambda, \alpha)$ is a mixture of this density and a point mass at 0 of weight $1 - \lambda$. (This follows from the elementary calculation $G_{\mathrm{FP}(\lambda,\alpha)}(z) = G_{\mathrm{FP}(1/\lambda,\alpha)}(z/\lambda) + (1-\lambda)\frac{1}{z}$.)

It is also worth noting that, in the special case $\alpha = 1$, this law coincides with the **Marchenko–Pastur** distribution of aspect ratio λ; i.e. it is the large-n limit of the density of singular values of an $m(n) \times n$ matrix all of whose entries are i.i.d. with mean 0 and variance $1/n$, where $m(n)/n \to \lambda$ as $n \to \infty$.

2.6. **Subordination and Free Conditional Expectation.** The R-transform and Voiculescu transform of $\mu \boxplus \nu$ are easily described in terms of those of μ and ν separately, but the implicit relationship between these and the Cauchy transform makes it harder to compute directly with $G_{\mu\boxplus\nu}$. There is a function relating them, known as the **subordinator**.

Proposition 2.12 (Proposition 4.4 in [46]; Theorem 3.1 in [11]; Proposition 7.3 in [9])**.** *Let μ, ν be probability measures on $\mathbb{R}$. There is a unique analytic function $\omega_{\mu,\nu} \colon \mathbb{C}_+ \to \mathbb{C}_+$ that satisfies*

$$G_{\mu\boxplus\nu}(z) = G_\mu(\omega_{\mu,\nu}(z)) \quad \textit{or equivalently} \quad F_{\mu\boxplus\nu}(z) = F_\mu(\omega_{\mu,\nu}(z)), \quad z \in \mathbb{C}_+.$$

Moreover

$$F_{\mu\boxplus\nu}(z) = F_\mu(\omega_{\mu,\nu}(z)) = F_\nu(\omega_{\nu,\mu}(z)) = \omega_{\mu,\nu}(z) + \omega_{\nu,\mu}(z) - z.$$

If $V \overset{\mathrm{d}}{=} \mu$, $W \overset{\mathrm{d}}{=} \nu$ are freely independent random variables, we denote $\omega_{V,W} = \omega_{\mu,\nu}$.

In principal, these properties follow simply from the defining relation of the subordinator, which can informally be written as $\omega_{\mu,\nu} = F_\mu^{-1} \circ F_{\mu\boxplus\nu}$. Using Proposition 2.7, this equality rigorously defines $\omega_{\mu,\nu}$ on an appropriate truncated cone, but the key to the use of subordinators is their analyticity on the whole upper half plane.

Like Cauchy transforms (and other analytic transforms defined above), subordinators are robust under weak convergence. This fact is apparently folklore in the literature; we provide a simple proof outline below.

Proposition 2.13. *If $\{\mu_n\}_{n=1}^\infty$ and $\{\nu_n\}_{n=1}^\infty$ are sequences of probability measures with weak limits $\mu_n \rightharpoonup \mu$, $\nu_n \rightharpoonup \nu$, then $\omega_{\mu_n,\nu_n} \to \omega_{\mu,\nu}$ uniformly on compact subsets of $\mathbb{C}_+$.*

Proof. By Proposition 2.8 there is a fixed truncated cone $\Gamma_{\alpha,\beta}$ on which φ_{μ_n} converges uniformly on compact subsets to φ_μ. By Propositions 2.4 and 2.10, $F_{\mu_n\boxplus\nu_n}$ converges to $F_{\mu\boxplus\nu}$ uniformly on compact subsets of $\mathbb{C}_+$. What's more, $\varphi_{\mu_n\boxplus\nu_n}$ converges uniformly on compact subsets of another fixed truncated cone $\Gamma_{\alpha',\beta'}$. Note that the intersection $\Gamma_{\alpha,\beta} \cap \Gamma_{\alpha',\beta'}$ is another truncated cone. Let V be an open subset with compact closure $\overline{V} \subset \Gamma_{\alpha,\beta} \cap \Gamma_{\alpha',\beta'}$; thus $\varphi_{\mu_n\boxplus\nu_n}$ converges uniformly on V.

Appealing to the definition (2.3), $\varphi_{\mu_n\boxplus\nu_n}(z) = F^{-1}_{\mu_n\boxplus\nu_n}(z) - z$, and thus the preimage of V under $F_{\mu_n\boxplus\nu_n}$ is $F^{-1}_{\mu_n\boxplus\nu_n}(V) = \{\varphi_{\mu_n\boxplus\nu_n}(v) + v \colon v \in V\}$. As $\varphi_{\mu_n\boxplus\nu_n}$ converges uniformly, there is thus a fixed non-empty open set $W \subseteq \mathbb{C}_+$ with $F^{-1}_{\mu_n\boxplus\nu_n}(V) \supseteq W$ for all n. Hence, $F_{\mu_n\boxplus\nu_n}(W) \subseteq V \subseteq \Gamma_{\alpha,\beta}$.

Applying (2.3) once more,

$$\omega_{\mu_n,\nu_n} = F_{\mu_n}^{-1} \circ F_{\mu_n\boxplus\nu_n} = \varphi_{\mu_n} \circ F_{\mu_n\boxplus\nu_n} + F_{\mu_n\boxplus\nu_n}$$

is defined and analytic on W since $F_{\mu_n\boxplus\nu_n}$ maps W into $\Gamma_{\alpha,\beta}$. The composition thus converges to $\varphi_\mu \circ F_{\mu\boxplus\nu} + F_{\mu\boxplus\nu} = \omega_{\mu,\nu}$ uniformly on compact subsets of the open subset W of the upper half plane. Since ω_{μ_n,ν_n} and $\omega_{\mu,\nu}$ are all analytic in the full upper half plane, the proposition now follows from Montel's theorem. □

Our main use of subordinators is for computations of **conditional expectation** in a free probabilistic context. If $(\mathcal{A}, E)$ is a W^*-probability space, and $\mathcal{B} \subseteq \mathcal{A}$ is a von Neumann subalgebra, then there is a "conditioning" map from $\mathcal{A}$ to $\mathcal{B}$. As in the classical case, it can be constructed as an L^2-projection: since the starting point is the algebra $\mathcal{A}$ (analogous to L^∞) which is contained in $L^2(\mathcal{A}, E)$, the orthogonal projection maps $\mathcal{A}$ into $L^2(\mathcal{B}, E)$; in point of fact, the image of the projection of elements in the algebra $\mathcal{A}$ is contained in the algebra $\mathcal{B} \subset L^2(\mathcal{B}, E)$. (This requires E to be a faithful normal trace.) The projection therefore defines a linear map $E[\,\cdot\,|\mathcal{B}]\colon \mathcal{A} \to \mathcal{B}$ which extends to a contraction $L^p(\mathcal{A}, E) \to L^p(\mathcal{B}, E)$ for all $p \geq 1$, which satisfies the following two key properties:

(1) If $X_1, X_2 \in \mathcal{B}$ and $Y \in \mathcal{A}$, then $E[X_1 Y X_2|\mathcal{B}] = X_1\, E[Y|\mathcal{B}]\, X_2$.
(2) If $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq A$, then for any $X \in \mathcal{A}$, $E[E[X|\mathcal{B}_2]|\mathcal{B}_1] = E[X|\mathcal{B}_1]$.

If X is freely independent from $\mathcal{B}$, then $E[X|\mathcal{B}] = E[X]$ is constant; this mirrors the same result which holds classically if X is independent from $\mathcal{B}$. However, things diverge significantly between the classical and free worlds, as we'll see shortly.

If X is a random variable in $(\mathcal{A}, E)$, and if $\mathcal{B} = W^*(X)$ is the von Neumann subalgebra generated by X (in the selfadjoint case this can be interpreted as the set of all measurable bounded functions f of X, $f(X)$), then we typically denote $E[\,\cdot\,|\mathcal{B}] = E[\,\cdot\,|X]$. This is relevant for the present work in the case of computing $E[f(X+Y)|X]$ where f is a smooth bounded test function and X, Y are freely independent. If X, Y were *classically* independent, this is simply equal to $g(X)$ where for $t \in \mathbb{R}$ $g(t) = E[f(t+Y)]$; but that is not true in the free world. Instead, we have the following important result.

Proposition 2.14 (Theorem 3.1 in [11])**.** *Let X and Y be freely independent self-adjoint random variables with laws μ and ν, and let $\omega_{\mu,\nu}$ denote the subordinator given by Proposition 2.12. Let $k_{\mu,\nu}$ be the Feller–Markov kernel whose Cauchy transform is given by*

$$\int_{\mathbb{R}} \frac{k_{\mu,\nu}(t, du)}{z - u} = \frac{1}{\omega_{\mu,\nu}(z) - t}, \quad t \in \mathbb{R}. \tag{2.10}$$

Then for any bounded Borel measurable function f,

$$E[f(X+Y)|X] = f_{\mu,\nu}(X) \quad \text{where} \quad f_{\mu,\nu}(t) = \int f(u)\, k_{\mu,\nu}(t, du). \tag{2.11}$$

In particular: fixing $z \in \mathbb{C}_+$, the function $f_z(t) = 1/(z-t)$ is bounded and continuous; (2.10) and (2.11) together then say

$$E[(z - X - Y)^{-1}|X] = (\omega_{\mu,\nu}(z) - X)^{-1}. \tag{2.12}$$

2.7. **Free Infinite Divisibility.** As introduced already in the Introduction, a probability measure μ on $\mathbb{R}$ is called **freely infinitely divisible** if, for each $n \in \mathbb{N}$, there is a probability measure μ_n such that $\mu = \mu_n \boxplus \cdots \boxplus \mu_n = \mu_n^{\boxplus n}$. In terms of random variables: (the law of) X is infinitely divisible if, for each n, there are freely independent identically distributed random variables $\{X_{j,n}\}_{1 \leq j \leq n}$ so that $X = X_{1,n} + \cdots + X_{n,n}$. A standard triangular arrays argument, analogous to the classical approach, shows that it is equivalent to ask only that such $X_{j,n}$ exist where the sum $X_{1,n} + \cdots + X_{n,n}$ *converges weakly* to X. Prominent examples of freely independent laws include semicircular distributions (since the sum of two free semicricular random variables is semicircular), and the compound free Poisson

distributions $\mathrm{FP}(\lambda,\nu)$ of Definition 2.11 (this follows directly from their definition: the $X_{j,n}$ can be taken to have distribution $(1-\frac{\lambda}{n})\delta_0+\frac{\lambda}{n}\nu$).

Characterizing laws in terms of their R-transform, μ is freely infinitely divisible if, for each $n\in\mathbb{N}$, $\frac{1}{n}R_\mu$ is the R-transform of a probability measure. By taking convolution powers, this means that tR_μ has the form R_{μ_t} for some probability measure μ_t for any rational $t\geq 0$; a continuity result [8, Lemma 5.9] then shows that this holds for all real $t\geq 0$ as well. Thus, freely infinitely divisible measures correspond to **free convolution semigroups** μ_t satisfying $\mu_{t+s}=\mu_t\boxplus\mu_s$ for $s,t\geq 0$, and $\mu_0=\delta_0$ (the original measure is then identified as $\mu=\mu_1$). A major result identifying such measures/semigroups is [8, Theorem 5.10], which states that a measure μ is freely infinitely divisible if and only if φ_μ (and hence R_μ) has an analytic continuation to all of $\mathbb{C}_+$ (which may take values in $\mathbb{C}_-\cup\mathbb{R}$). That same theorem also proves the free Lévy–Khintchine formulas (1.16) and (2.14).

One of the major theorems in free probability, connecting it to classical probability, is the *Bercovici–Pata bijection* Λ introduced in [7]. Denote by $\mathrm{ID}_*(\mathbb{R})$ the (classically) infinitely divisible distributions on $\mathbb{R}$, and by $\mathrm{ID}_\boxplus(\mathbb{R})$ the freely infinitely divisible distributions on $\mathbb{R}$. If $\mu\in\mathrm{ID}_*$ and has 2 finite moments, the Lévy–Khintchine formula (1.13) yields a probability measure ν which, together with the mean m and variance σ^2, characterizes μ. The triple (m,σ^2,ν) is *also* associated to a unique probability measure $\Lambda(\mu)$ with 2 finite moments, via (1.16). The mapping $\mu\mapsto\Lambda(\mu)$ sets up a bijection between $\mathrm{ID}_*(\mathbb{R})\cap L^2$ and $\mathrm{ID}_\boxplus(\mathbb{R})\cap L^2$. In fact, it extends all the way to a bijection $\Lambda\colon\mathrm{ID}_*(\mathbb{R})\to\mathrm{ID}_\boxplus(\mathbb{R})$.

Theorem 2.15. *Let μ be a probability measure on $\mathbb{R}$.*

a) $\mu\in\mathrm{ID}_*(\mathbb{R})$ if and only if there is a constant $\gamma\in\mathbb{R}$ and a finite positive Borel measure ς on $\mathbb{R}$ so that the characteristic function $\psi_\mu(\xi)=\int e^{i\xi t}\mu(dt)$ satisfies

$$\log\psi_\mu(\xi)=i\gamma\xi+\int_{\mathbb{R}}\frac{\exp(i\xi t)-1-i\xi t}{t^2}\,(t^2+1)\,\varsigma(dt). \tag{2.13}$$

b) $\mu\in\mathrm{ID}_\boxplus(\mathbb{R})$ if and only if there is a constant $\gamma\in\mathbb{R}$ and a finite positive Borel measure ς on $\mathbb{R}$ so that the R-transform R_μ satisfies

$$R_\mu(z)=\gamma+\int_{\mathbb{R}}\frac{z+t}{1-tz}\,\varsigma(dt). \tag{2.14}$$

c) Given $\mu\in\mathrm{ID}_*(\mathbb{R})$, let (γ,ς) be the constant and finite measure associated to μ by (2.13). There is a unique probability measure $\Lambda(\mu)\in\mathrm{ID}_\boxplus(\mathbb{R})$ with this same pair (γ,ς) characterizing $R_{\Lambda(\mu)}$ in (2.14). The map $\Lambda\colon\mathrm{ID}_*(\mathbb{R})\to\mathrm{ID}_\boxplus(\mathbb{R})$ is a bijection, and satisfies $\Lambda(\mu_1*\mu_2)=\Lambda(\mu_1)\boxplus\Lambda(\mu_2)$.

Item a) is classical, proved for example in the textbook [23, Section 18]. The free version in item b) was proved in [8].

Remark 2.16. *Noting that $\frac{z}{1-tz}\cdot\frac{t}{t^2+1}=\frac{z+t}{1-tz}+\frac{t}{t^2+1}$, in the case that ς has a finte moment, (2.14) can be rewritten as*

$$R_\mu(z)=\gamma'+\int_{\mathbb{R}}\frac{z}{1-tz}(t^2+1)\,\varsigma(dt) \tag{2.15}$$

where $\gamma'=\gamma+\int t\,\varsigma(dt)$. Further, if ς has two finite moments, then $(t^2+1)\,\varsigma(dt)$ is a finite measure, and expanding the right-hand side of (2.15) as a power series in z shows that its mass is $\sigma^2=\mathrm{Var}(\mu)$. Thus, letting $\nu(dt)=\frac{1}{\sigma^2}(t^2+1)\,\varsigma(dt)$, we see that (2.15) recovers (1.16), where $\gamma'=m$. An analogous analysis comparing (2.13)

and (1.13) *shows that* Λ *defined in Theorem 2.15(c) indeed maps* $\mathrm{ID}_*(\mathbb{R}) \cap L^2$ *onto* $\mathrm{ID}_{\boxplus}(\mathbb{R}) \cap L^2$ *bijectively as described above.*

Remark 2.17. *From the convolution interchange property of Theorem 2.15(c) and from direct power-series expansions, one sees that* $\Lambda(\mu)$ *always has the same number of finite moments as* μ*, and up to that order, the classical cumulants of* μ *are equal to the free cumulants of* $\Lambda(\mu)$*.*

For example: the normal distribution $\mathcal{N}(m, \sigma^2)$ has Lévy measure $\nu = \delta_0$; the image $\Lambda(\mathcal{N}(m, \sigma^2))$ under the Bercovici–Pata bijection is then the law whose R-transform satisfies $R(z) = m + \sigma^2 z$, which is the semicircular law $\mathcal{S}(m, \sigma^2)$ with density

$$\frac{1}{2\pi\sigma}\sqrt{4\sigma^2 - (x-m)^2}\,\mathbf{1}_{|x-m|\le 2\sigma}. \tag{2.16}$$

Similarly, if $\mu = \mathrm{Poiss}(\lambda, \upsilon)$ is a compound Poisson distribution, then $\Lambda(\mu) = \mathrm{FP}(\lambda, \upsilon)$ is the compound free Poisson with the same rate λ and the same jump distribution υ.

2.8. **Positive Free Infinite Divisibility.** Classically, infinite divisibility and positivity intertwine well. If $\mu \in \mathrm{ID}_*(\mathbb{R}_+)$,i.e. μ is infinitely divisible and is supported on $[0, \infty)$, then all convolution roots of μ are also supported on $[0, \infty)$. Indeed, if $X = X_{1,n} + \cdots + X_{n,n}$ are i.i.d. and $P(X_{j,n} < 0) = p > 0$ then $P(X < 0) \ge P(X_{j,n} < 0 \text{ for } 1 \le j \le n) = p^n > 0$. This elementary argument fails in the free world, however. Take, for example, any semicircular distribution $\mathcal{S}(m, \sigma^2)$ with $m \ge 2\sigma$; from (2.16) this is supported in $[0, \infty)$. However, the nth $\boxplus$-root is $\mathcal{S}(m/n, \sigma^2/n)$, and no matter how large $m - 2\sigma > 0$ is, $\frac{m}{n} - 2\frac{\sigma}{\sqrt{n}}$ is negative for all large n. That is, referring to Definition 1.7, this example shows that there are measures that are positive and freely infinitely divisible but not positive*ly* freely infinitely divisible.

We denote the class of positively freely infinitely divisible measures by $\mathrm{ID}^+_{\boxplus}(\mathbb{R}_+)$. This notion was introduced first in [42] and explored further in [1], where such measures were called *free regular* and the class of them was denoted $I^{\boxplus}_{r+}$ rather than our $\mathrm{ID}^+_{\boxplus}(\mathbb{R}_+)$. The authors of [1] were interested in these measures because (as they show) they are precisely the laws of all free Lévy processes with positive increments.

The following theorem is proved as part of [1, Theorem 4.2].

Theorem 2.18. *Let* μ *be a probability measure on* $\mathbb{R}$*. Then* μ *is positively freely infinitely divisible if and only if there is a measure* $\nu \in \mathrm{ID}_*(\mathbb{R}_+)$ *with* $\Lambda(\eta) = \mu$*. I.e.* $\mathrm{ID}^+_{\boxplus}(\mathbb{R}_+) = \Lambda(\mathrm{ID}_*(\mathbb{R}_+))$*. Moreover, in this case, the Lévy–Khintchine measure* ς *for* μ*, cf.* (2.14)*, is supported on* $[0, \infty)$*.*

It is straightforward to see, using the homomorphism property of the Bercovici–Pata bijection Λ in Theorem 2.15(c), and the fact the classically infinitely divisible distributions supported on $[0, \infty)$ have all convolution roots supported on $[0, \infty)$, that $\Lambda(\mathrm{ID}_*(\mathbb{R}_+)) \subseteq \mathrm{ID}^+_{\boxplus}(\mathbb{R}_+)$; the reverse containment is one of the main theorems of [1].

In the L^2 setting most relevant to this paper, the free Lévy–Khintchine measure ν of (1.16) and Theorem 1.6(b) is related to ς by $\nu(dt) = \frac{1}{\sigma^2}(t^2+1)\varsigma(dt)$ where σ^2 is the variance of μ, cf. Remark 2.16. Hence, we have the following important corollary.

Corollary 2.19. *Let X be an L^2 positively freely infinitely divisible random variable. Then its Voiculescu transform φ_X has the form*

$$\varphi_X(z) = m + \sigma^2 G_{V_0}(z)$$

where $m = E[X]$, $\sigma^2 = \mathrm{Var}[X]$, and $V_0 \geq 0$; i.e. the free Lévy–Khintchine measure is supported on $[0, \infty)$.

3. Existence and properties of the Free Zero Bias and other Transformations

In Section 3.1 we first prove the existence of a distributional transformation based on the fact that the "geometric mean" of two Cauchy transforms of probability measures is again the Cauchy transform of a probability measure, cf. Lemma 3.2. In Section 3.2 we use a special case of Lemma 3.2 to produce a new transformation we call the *El Gordo transform*; we also review the square bias transformation and provide some of its properties. Combining these transformations in Section 3.3 then yields the construction of the free zero bias putatively defined by (1.8), and the proof that its unique fixed point is the semi-circle distributions. Finally, in Section 3.6 we consider a few examples.

3.1. Geometric Mean Transformation.

Complex square roots will come up frequently in this section and beyond, so we begin by setting notation.

Notation 3.1. *For $\zeta \in \mathbb{C}$, the (principal)* **square root** *$\sqrt{\zeta}$ is defined uniquely in polar coordinates: writing $\zeta = re^{i\theta}$ for $r \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{\zeta} = r^{1/2} e^{i\theta/2}$. The function $\zeta \mapsto \sqrt{\zeta}$ is holomorphic on $\mathbb{C} \setminus [0, \infty)$, with image contained in $\mathbb{C}_+$.*

Note that if $\zeta \in \mathbb{C}$ and $y > 0$ then $\sqrt{y\zeta} = \sqrt{y}\sqrt{\zeta}$; but if $\zeta, \xi \in \mathbb{C}_-$ then $\sqrt{\zeta\xi} = -\sqrt{\zeta}\sqrt{\xi}$. Similarly, if $z \in \mathbb{C} \setminus [0, \infty)$ then $\sqrt{1/z} = -1/\sqrt{z}$. Finally: note that if $\zeta, \xi \in \mathbb{C}_-$, then $\zeta\xi \in \mathbb{C} \setminus [0, \infty)$.

With the square root in hand, the following elementary lemma yields a (surprisingly) new operation on probability distributions.

Lemma 3.2. *Given any two random variables X and Y, there is a unique probability measure (which we denote as the law of a new random variable $X \flat Y$) whose Cauchy transform satisfies*

$$G_{X\flat Y}(z) = -\sqrt{G_X(z) G_Y(z)} = \sqrt{G_X(z)}\sqrt{G_Y(z)}. \tag{3.1}$$

I.e. $G_{X\flat Y}$ is the "geometric mean" of the Cauchy transforms of X and Y. The map $(\mathcal{L}(X), \mathcal{L}(Y)) \to \mathcal{L}(X \flat Y)$ is bi-continuous in the sense that $X_n \flat Y_n \rightharpoonup X \flat Y$ whenever $X_n \rightharpoonup X$ and $Y_n \rightharpoonup Y$, as $n \to \infty$.

Proof. G_X and G_Y are analytic maps from $\mathbb{C}_+$ to $\mathbb{C}_-$, hence for any $z \in \mathbb{C}_+$ the product $z \mapsto G_X(z) G_Y(z)$ defines an analytic function taking values in $\mathbb{C} \setminus \mathbb{R}$. Hence $G_{X\flat Y}(z) = -\sqrt{G_X(z) G_Y(z)}$ is analytic, and takes values in the lower half plane $\mathbb{C}_-$. That it is equal to $\sqrt{G_X(z)}\sqrt{G_Y(z)}$ follows from the second paragraph in Notation 3.1.

Hence, to show that $G_{X\flat Y}$ is the Cauchy transform of a probability measure, it suffices to verify that $iyG_{X\flat Y}(iy) \to 1$ as $y \uparrow \infty$. For this, we appeal to the second representation in (3.1). Since G_X and G_Y are Cauchy transforms of probability measures, $iyG_X(iy) \to 1$ and $iyG_Y(iy) \to 1$, and therefore $yG_X(iy) \to -i$ and $yG_Y(iy) \to -i$ as $y \uparrow \infty$. Since $y \mapsto yG_X(iy)$ is continuous and takes values in

$\mathbb{C}\setminus[0,\infty)$, the function $y\mapsto\sqrt{yG_X(iy)}$ is also continuous, and hence $\sqrt{yG_X(iy)}\to\sqrt{-i}=e^{\frac{3\pi}{4}i}$; similarly $\sqrt{yG_Y(iy)}\to e^{\frac{3\pi}{4}i}$. Hence

$$yG_{X\flat Y}(iy)=\sqrt{yG_X(iy)}\sqrt{yG_Y(iy)}\to(e^{\frac{3\pi}{4}i})^2=-i\quad\text{as } y\uparrow\infty$$

and we conclude that $iyG_{X\flat Y}(iy)\to i(-i)=1$, showing that $G_{X\flat Y}$ is the Cauchy transform of a probability measure as claimed, and uniqueness follows from Stieltjes inversion (cf. Proposition 2.4.

For the continuity claim: if $X_n\rightharpoonup X$ and $Y_n\rightharpoonup Y$ then $G_{X_n}\to G_X$ and $G_{Y_n}\to G_Y$ uniformly on compact subsets of $\mathbb{C}_+$ by Proposition 2.4(1). Since G_{X_n} and G_{Y_n} take values in $\mathbb{C}_-$ where the square root is continuous, it then follows that $G_{X_n\flat Y_n}=\sqrt{G_{X_n}}\sqrt{G_{Y_n}}$ converges uniformly on compact subsets to $\sqrt{G_X}\sqrt{G_Y}=G_{X\flat Y}$. The convergence $X_n\flat Y_n\rightharpoonup X\flat Y$ now follows from Proposition 2.4(2). □

Example 3.3. *a) For any random variable X,*

$$G_{X\flat X}(z)=\sqrt{G_X(z)}\sqrt{G_X(z)}=G_X(z)$$

i.e. $X\flat X\overset{d}{=}X$.
b) Consider the case that X and Y are constants; e.g. $X=1$ and $Y=-1$. Then

$$G_{X\flat Y}(z)=-\sqrt{G_X(z)G_Y(z)}=-\sqrt{\frac{1}{z-1}\frac{1}{z+1}}=\frac{1}{\sqrt{z^2-1}}.\tag{3.2}$$

(The last equality comes from Notation 3.1, since $z^2-1\in\mathbb{C}\setminus[0,\infty)$ when $z\in\mathbb{C}_+$.) Now employing the Stieltjes inversion formula (Proposition 2.4), the law $X\flat Y$ can be recovered, and possesses a density:

$$\varrho_{X\flat Y}(x)=-\frac{1}{\pi}\lim_{y\downarrow0}\operatorname{Im}G_\nu(x+iy)=\frac{1}{\pi\sqrt{1-x^2}}\mathbf{1}_{[-1,1]}(x).\tag{3.3}$$

That is to say: the law of $X\flat Y$ in (3.3) is the arcsine distribution on $[-1,1]$. More generally, letting X and Y be point masses at $a<b$ respectively, scaling and translating appropriately we obtain

$$G_{a\flat b}(z)=\frac{1}{\sqrt{(z-a)(z-b)}}\quad\text{and}\quad\varrho_{a\flat b}(x)=\frac{1}{\pi\sqrt{(x-a)(b-x)}}\mathbf{1}_{[a,b]}(x)$$

which is the arcsine distribution on $[a,b]$.
c) We can combine any random variable X with a point mass a this way, producing a new distribution $\mathcal{L}(a\flat X)$ whose Cauchy transform is $G_{a\flat X}(z)=-\sqrt{G_X(z)/(z-a)}$. This construction, with $a=0$, will be very important through the remainder of the paper, so we address it in the next result.

3.2. **The Size Bias s, Square Bias $\square$ and El Gordo $\flat$ Transformations.** In this section, we introduce two old (size bias $X\mapsto X^s$, square bias $X\mapsto X^\square$) and one new (El Gordo $X\mapsto X^\flat$) transformation on probability measures, focusing on their actions on Cauchy transforms. We begin with El Gordo.

Lemma 3.4. *For any probability measure $\mathcal{L}(X)$,*

$$G_{X^\flat}(z)=-\sqrt{\frac{1}{z}G_X(z)}\tag{3.4}$$

is the Cauchy transform of a probability distribution. We call $\mathcal{L}(X^\flat)$ the **El Gordo transform** *of $\mathcal{L}(X)$. This transformation on the set of all probability measures is continuous in the topology of convergence in distribution, has point mass at zero as*

its unique fixed point, satisfies $(\alpha X)^\flat \stackrel{\mathrm{d}}{=} \alpha X^\flat$ *for all* $\alpha \in \mathbb{R}$*, and is injective but not surjective.*

Proof. As $G_{\delta_0}(z) = 1/z$, we find that $X^\flat \stackrel{\mathrm{d}}{=} X \flat \delta_0$, and the existence and continuity properties follow by Lemma 3.2. That δ_0 is the map's unique fixed point follows immediately by solving the equality $G_X(z) = -\sqrt{G_X(z)/z}$, implying either $G_X(z) = 1/z$ or $G_X \equiv 0$ which is not the Cauchy transform of a probability measure.

Next, if $\alpha = 0$ we have that $(\alpha X)^\flat \stackrel{\mathrm{d}}{=} \alpha X^\flat$ as the left hand side is zero with probability 1 since the El Gordo transform takes δ_0 to δ_0. When $\alpha \neq 0$ then

$$\begin{aligned} G^2_{(\alpha X)^\flat}(z) = \frac{1}{z} E\left[\frac{1}{z - \alpha X}\right] = \frac{1}{\alpha z} E\left[\frac{1}{z/\alpha - X}\right] = \frac{1}{\alpha z} G_X\left(\frac{z}{\alpha}\right) \\ = \frac{1}{\alpha^2}\left(\frac{\alpha}{z} G_X\left(\frac{z}{\alpha}\right)\right) = \frac{1}{\alpha^2} G^2_{X^\flat}\left(\frac{z}{\alpha}\right) = G^2_{\alpha X^\flat}(z). \end{aligned}$$

Taking square roots gives $G_{(\alpha X)^\flat} = G_{\alpha X^\flat}$, and this implies $(\alpha X)^\flat \stackrel{\mathrm{d}}{=} \alpha X^\flat$ by the Stieltjes inversion formula, cf. Proposition 2.4.

To show that the El Gordo transform is injective, when the distributions of $X^\flat$ and $Y^\flat$ are equal, then

$$-\sqrt{\frac{1}{z} G_X(z)} = -\sqrt{\frac{1}{z} G_Y(z)} \quad \text{for all } z \in \mathbb{C}_+,$$

and squaring both sized and multiplying by z implies $G_X = G_Y$, and yielding $X \stackrel{\mathrm{d}}{=} Y$ by the Stieltjes inversion formula.

To show that the map is not surjective, we show that no point mass, other than zero, is in the range of the El Gordo transform. Indeed, if $\mathcal{L}(Y^\flat) = \delta_p$, meaning that $-\sqrt{G_Y(z)/z} = G_{Y^\flat}(z) = G_{\delta_p}(z) = (z-p)^{-1}$, it follows that

$$G_Y(z) = z G^2_{Y^\flat}(z) = \frac{z}{(z-p)^2} = \frac{z(\overline{z} - p)^2}{|z-p|^4}.$$

Note, however, that $\operatorname{Im} G_Y(z)$ is then a positive multiple of

$$\begin{aligned} \operatorname{Im}\left(z(\bar{z} - p)^2\right) = \operatorname{Im}\left(z\bar{z}^2 - 2pz\bar{z} + p^2 z\right) = |z|^2 \operatorname{Im} \bar{z} + p^2 \operatorname{Im} z \\ = (p^2 - |z|^2) \operatorname{Im} z. \end{aligned}$$

Hence, unless $p = 0$, the above function maps the half-disk of radius p in the upper half plane back into the upper half plane, and so the putative G_Y is not the Cauchy transform of a probability measure. □

We now recall the square bias transform given in (1.3), and make the convention that if $E[X^2] = 0$ then $\mathcal{L}(X^\square) = \delta_0$, a point mass at zero. In Theorem 5.8 and Proposition 5.12 we will also have need to consider the following pseudo-inverse of the square bias transforms.

Definition 3.5. *If* $\mathcal{L}(X) = \delta_0$*, let* $\mathcal{L}(X^{-\square}) = \delta_0$*. If* $E[X^{-2}] < \infty$ *then let the distribution* $\mathcal{L}(X^{-\square})$ *have Radon-Nikodym derivative with respect to that of* X *given by*

$$\frac{dP^{-\square}}{dP} = \frac{X^{-2}}{E[X^{-2}]}.$$

The next lemma collects many useful properties of the square bias.

Lemma 3.6. *Let X and Y be L^2 random variables, and let Z be any random variable.*

a) Let $\alpha \in [0,1]$ and

$$\mathcal{L}(W) = \alpha\mathcal{L}(X) + (1-\alpha)\mathcal{L}(Y).$$

Then $E[W^2] = \alpha E[X^2] + (1-\alpha)E[Y^2]$, and when $E[W^2] > 0$ then

$$\mathcal{L}(W^{\square}) = \frac{\alpha E[X^2]}{E[W^2]}\mathcal{L}(X^{\square}) + \frac{(1-\alpha)E[Y^2]}{E[W^2]}\mathcal{L}(Y^{\square}). \tag{3.5}$$

b) When $E[X^2] > 0$, X and Y satisfy $E[X^2] \geq E[Y^2]$ and $\alpha = E[Y^2]/E[X^2]$, then

$$\mathcal{L}(X^{\square}) = \mathcal{L}(Y^{\square}) \text{ if and only if } \mathcal{L}(Y) = (1-\alpha)\delta_0 + \alpha\mathcal{L}(X).$$

c) If X_n is a sequence of square integrable random variables such that $X_n \rightharpoonup X$ and $E[X_n^2] \to E[X^2] \in (0,\infty)$, then $X_n^{\square} \rightharpoonup X^{\square}$. The restriction that the second moment of the limit be positive is necessary.
d) When $E[X^2] > 0$ the Cauchy transform of $X^{\square}$ is given by

$$G_{X^{\square}}(z) = \frac{1}{E[X^2]}\left(z^2 G_X(z) - E[X] - z\right).$$

e) For all $\alpha \in \mathbb{R}$ we have $(\alpha X^{\square}) \overset{\mathrm{d}}{=} \alpha X^{\square}$.
f) If $E[Z^{-2}] < \infty$ then

$$E[(Z^{-\square})^2] = \frac{1}{E[Z^{-2}]} \quad \text{and} \quad \mathcal{L}((Z^{-\square})^{\square}) = \mathcal{L}(Z).$$

Proof. a) The expression given for $E[W^2]$ is clear from the representation of W as a mixture. Under the assumption $E[W^2] > 0$, for any bounded continuous f identity (1.3) yields

$$E[f(W^{\square})] = \frac{E[W^2 f(W)]}{E[W^2]} = \frac{\alpha E[X^2 f(X)] + (1-\alpha)E[Y^2 f(Y)]}{E[W^2]}. \tag{3.6}$$

As at least one of $\alpha E[X^2], (1-\alpha)E[Y^2]$ must be positive, without loss generality, by relabelling X and Y if necessary, we may assume $E[X^2]$ and α are both positive. If $E[Y^2] = 0$ then $E[W^2] = \alpha E[X^2]$ and the second term in the numerator of (3.6) is zero, producing $\mathcal{L}(W^{\square}) = \mathcal{L}(X^{\square})$, in agreement with (3.5), whose second term likewise vanishes.
When $E[Y^2] > 0$, then continuing from (3.6) we obtain the claim by

$$\begin{aligned} E[f(W^{\square})] &= \frac{1}{E[W^2]}\left(\alpha E[X^2]\frac{E[X^2 f(X^2)]}{E[X^2]} + (1-\alpha)E[Y^2]\frac{E[Y^2 f(Y^2)]}{E[Y^2]}\right) \\ &= \frac{1}{E[W^2]}\left(\alpha E[X^2]E[f(X^{\square})] + (1-\alpha)E[Y^2]E[f(Y^{\square})]\right). \end{aligned} \tag{3.7}$$

b) First suppose that $\mathcal{L}(X^{\square}) = \mathcal{L}(Y^{\square})$. Note that this common distribution cannot be point mass at zero, being excluded by the condition $E[X^2] > 0$.

Letting A be a measurable subset of $\mathbb{R}$ not containing zero and $\mathcal{L}(Z) = \delta_0$, we obtain

$$\begin{aligned} P(Y \in A) &= E[\mathbb{1}(Y \in A)] = E[Y^{-2}Y^2\mathbb{1}(Y \in A)] \\ &= E[Y^2]E[(Y^{\square})^{-2}\mathbb{1}(Y^{\square} \in A)] = \alpha E[X^2]E[(X^{\square})^{-2}\mathbb{1}(X^{\square} \in A)] \\ &= \alpha E[X^{-2}X^2\mathbb{1}(X \in A)] = \alpha P(X \in A) \\ &= (1-\alpha)P(Z \in A) + \alpha P(X \in A). \quad (3.8) \end{aligned}$$

Specializing to $A = \mathbb{R} \setminus \{0\}$ we obtain

$$P(Y \neq 0) = (1-\alpha)P(Z \neq 0) + \alpha P(X \neq 0) = \alpha P(X \neq 0)$$
$$\text{or} \quad P(Y = 0) = (1-\alpha)P(Z = 0) + \alpha P(X = 0),$$

thus showing the left and right hand sides of (3.8) are equal for all measurable $A \subseteq \mathbb{R}$. The other direction of the claim follows from part a).

c) As $E[X_n]$ converges to a positive limit, for all n greater than some n_0 the second moments of X_n are strictly positive. In particular, for any bounded continuous function f,

$$\lim_{n\to\infty} E[f(X_n^{\square})] = \lim_{n\to\infty} \frac{E[X_n^2 f(X_n)]}{E[X_n^2]} = \frac{E[X^2 f(X)]}{E[X^2]} = E[f(X^{\square})],$$

where for taking the limit in the numerator we use that $\{X_n^2, n \geq n_0\}$ is uniformly integrable and hence so is $\{X_n^2 f(X_n), n \geq n_0\}$, by the boundedness of f.

If X_n takes on the value $n^{1/4}$ with probability $1/n$, and is otherwise zero, then $X_n \rightharpoonup X$ where $\mathcal{L}(X) = \delta_0$ and $E[X_n^2] = \sqrt{n}/n \to 0 = E[X^2]$, but, by part a) $X_n^{\square}$ takes the value $n^{1/4}$ with probability one, and hence does not converge in distribution to zero, showing the necessity of the positivity condition.

d) Applying the definition of the Cauchy transform and of $\mathcal{L}(X^{\square})$, and then adding and subtracting z^2 yields

$$\begin{aligned} G_{X^{\square}}(z) &= E\left[\frac{1}{z - X^{\square}}\right] = \frac{1}{E[X^2]} E\left[\frac{X^2}{z - X}\right] \\ &= \frac{1}{E[X^2]} E\left[\frac{z^2 + X^2 - z^2}{z - X}\right] = \frac{1}{E[X^2]} E\left[\frac{z^2 + (X-z)(X+z)}{z - X}\right] \\ &= \frac{1}{E[X^2]} E\left[\frac{z^2}{z - X} - X - z\right], \end{aligned}$$

yielding the claim.

e) When $\alpha E[X^2] = 0$, then either $\alpha = 0$ or $X \stackrel{\mathrm{d}}{=} \delta_0$, and in both cases both sides of the claimed identity are equal to point mass at zero. Otherwise, for a given bounded measurable function $f(\cdot)$, letting $g(\cdot) = f(\alpha\cdot)$, we have

$$\begin{aligned} E[f((\alpha X)^{\square})] = \frac{E[(\alpha X)^2 f(\alpha X)]}{E[(\alpha X)^2]} &= \frac{E[X^2 f(\alpha X)]}{E[X^2]} \\ &= \frac{E[X^2 g(X)]}{E[X^2]} = E[g(X^{\square})] = E[f(\alpha X^{\square})]. \end{aligned}$$

f) Applying Definition 3.5 for $f(z) = xz2$ yields

$$E[(Z^{-\square})^2] = \frac{1}{E[Z^{-2}]},$$

and now with this identity, and also the help of Definition 3.5, for any bounded continuous f we obtain

$$E[f(Z^{-\square})^{\square})] = \frac{E[(Z^{-\square})^2 f(Z^{-\square})]}{E[(Z^{-\square})^2]}$$
$$= E[Z^{-2}]E[(Z^{-\square})^2 f(X^{-\square})] = E[Z^{-2}]\frac{E[Z^{-2}Z^2 f(Z)]}{E[Z^{-2}]} = E[f(X)].$$

□

Finally, we summarize the definition and a few important properties of the well-known size bias transformation.

Definition 3.7. *For $\mathcal{L}(X)$ the law of a non-negative random variable X with positive finite mean, let $\mathcal{L}(X^s)$, the size bias transform of $\mathcal{L}(X)$, have Radon-Nikodym derivative with respect to that of X given by*

$$\frac{dP^s}{dP} = \frac{X}{E[X]}. \tag{3.9}$$

Similar to Definition 3.5, we have

Definition 3.8. *For X non-negative and $E[X^{-1}] < \infty$, let the distribution $\mathcal{L}(X^{-s})$ have Radon-Nikodym derivative with respect to that of X given by*

$$\frac{dP^{-s}}{dP} = \frac{X^{-1}}{E[X^{-1}]}. \tag{3.10}$$

The following facts regarding the size bias transformation mirror those about the square bias transform in Lemma 3.6.

Lemma 3.9. *Let X and Y be non-negative L^1 random variables, and let Z be a non-negative random variable.*

a) Let $\alpha \in [0,1]$ and

$$\mathcal{L}(W) = \alpha\mathcal{L}(X) + (1-\alpha)\mathcal{L}(Y).$$

Then $E[W] = \alpha E[X] + (1-\alpha)E[Y]$, and when $E[W] > 0$,

$$\mathcal{L}(W^s) = \frac{\alpha E[X]}{E[W]}\mathcal{L}(X^s) + \frac{(1-\alpha)E[Y]}{E[W]}\mathcal{L}(Y^s). \tag{3.11}$$

b) When $E[X] > 0$, X and Y satisfy $E[X] \geq E[Y]$ and $\alpha = E[Y]/E[X]$, then

$$\mathcal{L}(X^s) = \mathcal{L}(Y^s) \text{ if and only if } \mathcal{L}(Y) = (1-\alpha)\delta_0 + \alpha\mathcal{L}(X).$$

c) If X_n is a sequence of square integrable random variables such that $X_n \rightharpoonup X$ and $E[X_n] \to E[X] \in (0,\infty)$, then $X_n^s \rightharpoonup X^s$. The restriction that the first moment of the limit be positive is necessary.

d) When $E[X] > 0$ the Cauchy transform of X^s is given by

$$G_{X^s}(z) = \frac{1}{E[X]}\left(zG_X(z) - 1\right).$$

e) For all $\alpha > 0$ we have $(\alpha X^s) \overset{\mathrm{d}}{=} \alpha X^s$.

f) If $E[Z^{-1}] < \infty$ then

$$E[Z^{-s}] = \frac{1}{E[Z^{-1}]} \quad \text{and} \quad \mathcal{L}((Z^{-s})^s) = \mathcal{L}(Z).$$

g) If $Z = Y^s$, then $E[Z^{-1}] = 1/E[Y] < \infty$.

Proof. Properties a) through f) are exact analogs of the properties of square bias, and the proofs of those properties in Lemma 3.6 work mutatis mutandis here. We therefore provide a proof only of g).

First note that by Lemma 3.9 b) we may assume without loss of generality that Y has no mass at zero, and in particular that $P(Yf(Y) = 1) = 1$ for $f(y) = 1/y$. Now noting that by standard convergence arguments (1.2) can be extended to all functions f for which the left hand side exists we obtain

$$1 = E[Yf(Y)] = E[Y]E[f(Z)] \quad \text{yielding} \quad E[Z^{-1}] = \frac{1}{E[Y]} < \infty.$$

□

3.3. **Construction of the Free Zero Bias $\circ$.** We now arrive at the main construction: the free zero bias. To motivate it, we return for a moment to the classical zero bias X^* which is defined by the functional equation (1.5), and has the probabilistic construction (1.6) whose proof can be found in [25, Theorem 2.1]. This construction appears in (12) of that work when specializing to the case $P(x) = x$ and $m = 0$, upon noting that the random variable Y whose distribution is given in (11) has the $X^\square$ distribution by virtue of the fact that $Q(x)$ defined in (10) is equal to $P(x)$ in this case.

One may wonder if some version of this statement, with independence replaced by free independence and uniform replaced by some other distribution perhaps, might characterize the free zero bias of (1.8). While nothing of this nature has presented itself, the following definition does have the same flavor, and (as we see shortly) works.

Definition 3.10. *If X is an L^2 random variable with $E[X] = 0$, the* free zero bias *of X, denoted X°, is defined (by its law) as follows:*

$$X^\circ \stackrel{\mathrm{d}}{=} (X^\square)^\flat. \tag{3.12}$$

That is: the law $\mathcal{L}(X^\circ)$ is defined by its Cauchy transform, which satisfies

$$G_{X^\circ}(z) = G_{(X^\square)^\flat}(z), \qquad z \in \mathbb{C}_+$$

following Lemmas 3.4 and 3.6(d) defining the $\flat$ and $\square$ transforms in terms of Cauchy transforms.

Note: the free zero bias, like the square bias and El Gordo transforms it is composed of, acts only on probability measures; throughout this paper, we have it act on random variables by acting on their laws, and abuse notation writing (as above) X° rather than $\mathcal{L}(X)^\circ$.

Remark 3.11. *We see now that to compute either the classical or free zero bias of a mean zero X the first step is to compute $X^\square$, followed by multiplying by an independent $U \stackrel{\mathrm{d}}{=} \mathrm{Unif}[0,1]$ in the classical case. That multiplication in the free case is replaced by taking the El Gordo transformation in* (3.12).

Here are several important properties of the free zero bias, including a concrete formula for its Cauchy transform.

Lemma 3.12. *The free zero bias defined above satisfies the following properties.*

a) $X^\circ \stackrel{\mathrm{d}}{=} Y^\circ$ if and only if $X^\square \stackrel{\mathrm{d}}{=} Y^\square$.

b) For all $\alpha \in \mathbb{R}$ we have $(\alpha X)^\circ \stackrel{\mathrm{d}}{=} \alpha X^\circ$.
c) If $X_n \rightharpoonup X$ and $E[X_n^2] \to E[X^2] > 0$, then $X_n^\circ \rightharpoonup X^\circ$.
d) The Cauchy transform of X° satisfies

$$G_{X^\circ}(z) = -\sqrt{\frac{zG_X(z) - 1}{\mathrm{Var}(X)}}. \tag{3.13}$$

Proof. For a), if $X^\square, Y^\square$ have the same law then the same holds for X°, Y° via Definition 3.10. Conversely, as the El Gordo transform is injective by Lemma 3.2, whenever X°, Y° are equally distributed the same is true for $X^\square, Y^\square$. Next, for b), note that by Lemma 3.4 and e) of Lemma 3.6, for all $\alpha \in \mathbb{R}$,

$$(\alpha X)^\circ = ((\alpha X)^\square)^\flat = (\alpha X^\square)^\flat = \alpha (X^\square)^\flat = \alpha X^\circ.$$

For the next claim, part c) of Lemma 3.6 shows that the assumed convergence conditions on X imply that $X_n^\square \rightharpoonup X^\square$, so the claim follows by the distributional continuity of the El Gordo transform provided by Lemma 3.2.

Writing out the composition in (3.12) using the definition (3.4) and then applying d) of Lemma 3.6(d) in the case $E[X] = 0$ yields

$$G_{(X^\square)^\flat}(z) = -\sqrt{\frac{1}{z} G_{X^\square}(z)} = -\sqrt{\frac{1}{zE[X^2]} \left(z^2 G_X(z) - z\right)}$$

which immediately simplifies to (3.13). □

Remark 3.13. *From Lemma 3.12(b), if X is symmetric (i.e. has the same distribution as $-X$) then so is X°: $-X^\circ \stackrel{\mathrm{d}}{=} (-X)^\circ \stackrel{\mathrm{d}}{=} X^\circ$. (The same argument shows this result holds for the classical zero bias.) We will see this observation witnessed in several examples below.*

Remark 3.14. *Neither the square bias or El Gordo transforms require centered random variables, and indeed Definition 3.10 makes perfect sense for all L^2 random variables, centered or not. All the properties in Lemma 3.12 parts a), b), and c) hold true if $E[X] \neq 0$, while the reader can readily verify that (3.13) is replaced more generally with*

$$G_{(X^\square)^\flat}(z) = -\sqrt{\frac{1}{E[X^2]} \left(zG_X(z) - \frac{E[X]}{z} - 1 \right)}. \tag{3.14}$$

Nevertheless, as mentioned in (1.7), where X has mean $m \in \mathbb{R}$, we define

$$X^\circ \stackrel{\mathrm{d}}{=} (X - m)^\circ + m, \tag{3.15}$$

noting that this definition is in agreement with the previous one in the case $m = 0$. Extending the definition in this way preserves the fixed point property of the transformation, as explored in Example 3.25(b).

However, the mean 0 condition is required for the defining functional equation to hold, in both the classical and free cases. In comparison to the classical case: if $U \stackrel{\mathrm{d}}{=} \mathrm{Unif}[0,1]$ is independent from $X^\square$ then $UX^\square$ is a random variable, but it only coincides with X^* when $E[X] = 0$.

We now show the main reason why the free zero bias transformation is of interest. We will find it convenient to work with resolvent functions $f_z\colon \mathbb{R}\to\mathbb{C}$

$$f_z(u)=\frac{1}{z-u},\ z\in\mathbb{C}_+ \quad \text{which satisfy} \quad f_z^{[1]}(u,v)=\frac{1}{(z-u)(z-v)}. \tag{3.16}$$

These functions are bounded and smooth on $\mathbb{R}$.

We recall that the semi-circle law S_{σ^2} with mean zero and variance σ^2 has Cauchy transform

$$G_{S_{\sigma^2}}(z)=\frac{z-\sqrt{z^2-4\sigma^2}}{2\sigma^2}. \tag{3.17}$$

Lemma 3.15. *For any $\sigma^2>0$ the mapping $\mathcal{L}(X)\to\mathcal{L}(X^\circ)$ with domain $\mathcal{D}_{0,\sigma^2}$ has the semi-circle law with variance σ^2 as its unique fixed point, and for all $X\in\mathcal{D}_{0,\sigma^2}$*

$$E[Xf(X)]=\sigma^2E[f^{[1]}(X^\circ,Y^\circ)] \tag{3.18}$$

for all Lipshitz functions f, and

$$X^*\stackrel{\mathrm{d}}{=}UY^\circ+(1-U)X^\circ \tag{3.19}$$

where Y°, X° and $U\stackrel{\mathrm{d}}{=}\mathcal{U}[0,1]$ are independent, and $Y^\circ\stackrel{\mathrm{d}}{=}X^\circ$.

Proof. Setting $X^\circ\stackrel{\mathrm{d}}{=}X$ in (3.13) of Lemma 3.12 yields

$$\sigma^2G_X^2(z)-zG_X(z)+1=0,$$

implying that $G_X(z)=G_{S_{\sigma^2}}(z)$ via (3.17). For the second claim, letting f_z be a resolvent function for a given $z\in\mathbb{C}_+$, and using (3.13), we have

$$E[Xf_z(X)]=E\left[\frac{X}{z-X}\right]=E\left[\frac{z+X-z}{z-X}\right]=zG_X(z)-1=\sigma^2G_{X^\circ}^2(z). \tag{3.20}$$

Now, from (3.16),

$$E[f_z^{[1]}(X^\circ,Y^\circ)]=E\left[\frac{1}{(z-X^\circ)(z-Y^\circ)}\right]=E\left[\frac{1}{z-X^\circ}\right]E\left[\frac{1}{z-Y^\circ}\right]=G_{X^\circ}^2(z).$$

This identity, along with (3.20), shows that (3.18) holds for all $f_z, z\in\mathbb{C}_+$.

Now, for all such f_z, rewriting $E[f_z^{[1]}(X^\circ,Y^\circ)]$ in terms of the derivative of f_z and an independent $U\stackrel{\mathrm{d}}{=}\mathcal{U}[0,1]$, we obtain

$$\begin{aligned}\sigma^2E[f_z'(X^*)]&=\sigma^2E[Xf_z(X)]=\sigma^2E[f_z^{[1]}(X^\circ,Y^\circ)]\\&=\sigma^2E\left[\frac{f_z(Y^\circ)-f_z(X^\circ)}{Y^\circ-X^\circ}\right]=\sigma^2E[f_z'(UY^\circ+(1-U)X^\circ)].\end{aligned}$$

Letting $\Upsilon=UY^\circ+(1-U)X^\circ$, we have therefore shown that $E[f_z'(X^*)]=E[f_z'(\Upsilon)]$ for all $z\in\mathbb{C}_+$. Since $f_z'(u)=1/(z-u)^2=-\frac{d}{dz}f_z(u)$, we have

$$E[f_z'(\Upsilon)]=-E[\frac{d}{dz}f_z(\Upsilon)]=-\frac{d}{dz}E[f_z(\Upsilon)]=-\frac{d}{dz}G_\Upsilon(z)$$

where the second equality (differentiating under the integral) follows because $(u,z)\mapsto f_z(u)$ and $(u,z)\mapsto f_z'(u)$ are continuous on a neighborhood of each point $(u,z)\in\mathbb{R}\times\mathbb{C}_+$. Similarly, $E[f_z'(X^*)]=-\frac{d}{dz}G_{X^*}(z)$. Hence, we've established that $\frac{d}{dz}G_\Upsilon(z)=\frac{d}{dz}G_{X^*}(z)$ for all $z\in\mathbb{C}_+$, which means G_Υ and G_{X^*} differ by a constant, which must be 0 since both Cauchy transforms are $O(1/z)$ near ∞. Hence

$G_\Upsilon = G_{X^*}$, and by the Stieltjes inversion formula, it follows that $\Upsilon \stackrel{\mathrm{d}}{=} X^*$, establishing (3.19).

Now, via (3.19), for all Lipschitz f we have

$$E[Xf(X)] = \sigma^2 E[f'(X^*)] = \sigma^2 E[f'(UY^\circ + (1-U)X^\circ)]$$
$$= \sigma^2 E\left[\frac{f(Y^\circ) - f(X^\circ)}{Y^\circ - X^\circ}\right] = \sigma^2 E[f^{[1]}(X^\circ, Y^\circ)],$$

as desired.

□

Remark 3.16. *As stated, the final equality in the proof of Lemma 3.15 a priori requires $f^{[1]}$ to be defined as f' on the diagonal. We will see in Theorem 4.1 below that the law of X° is always absolutely continuous, and so it is not actually necessary to define $f^{[1]}$ on the diagonal when taking expectations of $f^{[1]}(X^\circ, Y^\circ)$.*

3.4. **The 'Replace One' Property.** In the classical world, much of the success that the zero bias transformation enjoys is due to what can be called 'the replace one property'. The following lemma gives its free analog; we include the classical one for the sake of comparison. For a clearer parallel, we extend the scope of the classical zero bias to general mean random variables, as in (3.15).

We begin with the following fact which is useful in both the size bias and free zero bias cases below:

Lemma 3.17. *When $S_n = X_1 + \ldots + X_n$, a sum of freely independent variables, and f_z is a resolvant function, then*

$$E[S_n f_z(S_n)] = \sum_{i=1}^n E[X_i f_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i)],$$

where $\omega_{\mu_i,\nu_{n\neg i}}$ is a subordinator as in Proposition 2.12.

Proof. Applying (free) conditioning as in Section 2), (2.12) for the third equality, and the definition of f_ω in the last, we obtain

$$E[S_n f_z(S_n)] = \sum_{i=1}^n E[X_i f_z(S_n)] = \sum_{i=1}^n E[X_i E[f_z(S_n)|X_i]]$$
$$= \sum_{i=1}^n E\left[\frac{X_i}{\omega_{\mu_i,\nu_{n\neg i}}(z) - X_i}\right] = \sum_{i=1}^n E[X_i f_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i)].$$

□

In what follows, we adopt the definition of the classical X^* for non-mean zero random variables just as was done in (1.7) for the free case.

Theorem 3.18. *Consider variables $X_1, \ldots, X_n$ with means $m_1, \ldots, m_n$ and finite, non-zero variances $\sigma_1^2, \ldots, \sigma_n^2$, and sum $S_n = X_1 + \cdots + X_n$ where $\sigma^2 = \mathrm{Var}(S_n)$. For $i = 1, \ldots, n$, let I be a random index with distribution $P(I = i) = \sigma_i^2/\sigma^2$.*

When the summand variables are classically independent, and I is independent of $\{X_i, X_i^*, i = 1, \ldots, n\}$ where X_i^* is independent of $X_j, j \neq i$ and has the X_i zero-bias distribution, then S_n^* given by $S_n - X_I + X_I^*$ has the S_n-zero bias distribution, and the characteristic function of S_n^* satisfies

$$\psi_{S_n^*}(t) = E[\psi_{X_I^*}(t)\psi_{S_n - X_I}(t)]. \tag{3.21}$$

When the summand variables are freely independent, then the Cauchy transform of S_n° satisfies

$$G^2_{S_n^\circ}(z) = E[G^2_{X_I^\circ}(\omega_{X_I, S_n - X_I}(z))]. \tag{3.22}$$

When the summands are identically distributed as X, these relations respectively specialize to

$$\psi_{S_n^*}(t) = \psi_{X^*}(t)\psi_{S_{n-1}}(t) \quad \text{and} \quad G_{S_n^\circ}(z) = G_{X^\circ}(\omega_{X_n, S_n - X_n}(z)). \tag{3.23}$$

Remark 3.19. *As a reminder: $\omega_{V,W}$ is the subordinator function introduced in Proposition 2.12 by the relation $G_{V+W}(z) = G_V(\omega_{V,W}(z))$ for $z \in \mathbb{C}_+$ when V, W are freely independent. Here $V = X_n$ and $W = S_n - X_n = X_1 + \cdots + X_{n-1}$ are indeed freely independent, since all the X_k are free from each other. Subordinators connect Cauchy transforms of free sums, in an analogous (but more complicated) manner to the simple multiplicative relationship between characteristic functions of independent sums. It is therefore natural for them to appear in the free analog of "the replace one property".*

Remark 3.20. *The classical "replace one property" for, say, i.i.d. variables does* not *hold for freely independent random variables; that is to say, in general*

$$S_n^\circ \overset{\mathrm{d}}{\neq} S_n - X_n + X_n^\circ$$

where X_n° is freely independent from $X_1, \ldots, X_{n-1}$. Indeed: from Proposition 2.12,

$$G_{S_n - X_n + X_n^\circ}(z) = G_{X_n^\circ}(\omega_{X_n^\circ, S_n - X_n}(z)).$$

From Theorem 3.18, this would equal $G_{S_n^\circ}(z)$ if and only if $G_{X_n^\circ}(\omega_{X_n^\circ, S_n - X_n}(z)) = G_{X_n^\circ}(\omega_{X_n, S_n - X_n}(z))$, and (from the univalence of $G_{X_n^\circ}$ and subordinators near ∞) this would imply that $X_n^\circ \overset{\mathrm{d}}{=} X_n$. Thus, from Lemma 3.15, the only *case in which $S_n^\circ \overset{\mathrm{d}}{=} S_n - X_n + X_n^\circ$ is when the summands X_n are all semicircular — in which case the equation reduces to the tautology $S_n \overset{\mathrm{d}}{=} S_n$.*

Proof of Theorem 3.18: For a proof of (3.21) in the mean zero case, see [24]. Using that result in the third equality, and with $m = E[S_n]$, in the general mean case we have

$$\begin{aligned}
\psi_{S^*}(t) = \psi_{(S-m)^*+m}(t) &= e^{itm}\psi_{(S-m)^*}(t) \\
&= e^{itm} E\left[E[\psi_{(X_I - m_I)^*}(t)\psi_{(S_n - m) - (X_I - m_I)}(t)|I]\right] \\
&= E\left[E[\psi_{(X_I - m_I)^* + m_I}(t)\psi_{(S_n - m) + m - [(X_I - m_I) + m_I]}(t)|I]\right] \\
&= E[\psi_{X_I^*}(t)\psi_{S_n - X_I}(t)].
\end{aligned}$$

For the free case, this same type of reasoning, and the simply proved fact that

$$\omega_{\mu,\nu}(z - (b + a)) = \omega_{\mu+a, \nu+b}(z) - a$$

allows us to reduce to the mean zero case. Next, using (3.16), for any random variable V with independent copy W, and any $z \in \mathbb{C}_+$ we have

$$\begin{aligned}
E[f_z^{[1]}(V, W)] &= E[(z - V)^{-1}(z - W)^{-1}] \\
&= E[(z - V)^{-1}]E[(z - W)^{-1}] = G_V^2(z).
\end{aligned} \tag{3.24}$$

Now, let μ_i and $\nu_{n\neg i}$ be the probability distributions of X_i and $S_n - X_i$ respectively. Using (3.24) with $V = S_n^\circ$, T_n° an independent copy of S_n°, (3.18) of Lemma 3.15, and the definition of S_n, we have

$$\sigma^2 G^2_{S_n^\circ}(z) = \sigma^2 E[f_z^{[1]}(S_n^\circ, T_n^\circ)] = E[S_n f_z(S_n)] = \sum_{i=1}^n E[X_i f_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i)],$$

invoking Lemma 3.17 for the final equality.

Applying again the functional equation (3.18), the summands in the last expression are equal to

$$\sigma_i^2 E[f^{[1]}_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i^\circ, Y_i^\circ)] = \sigma_i^2 G^2_{X_i^\circ}(\omega_{\mu_i,\nu_{n\neg i}}(z))$$

with the equality again via (3.24). Summing up, we have therefore shown that

$$\sigma^2 G^2_{S_n^\circ}(z) = \sum_{i=1}^n \sigma_i^2 G^2_{X_i^\circ}(\omega_{\mu_i,\nu_{n\neg i}}(z))$$

and dividing through by σ^2 yields (3.22). The final statement of the theorem, in the identical distribution case, follows immediately as in this case $\sigma^2 = n\sigma_i^2$ and $\omega_{\mu_i,\nu_{n\neg i}} = \omega_{\mu_n,\nu_{n\neg n}}$ for all $i = 1, \ldots, n$. □

We now give the parallel result for the size bias.

Theorem 3.21. *Consider non-negative variables $X_1, \ldots, X_n$ with finite non-zero means $m_1, \ldots, m_n$ and sum $S_n = X_1 + \cdots + X_n$ with $m = E[S_n]$. For $i = 1, \ldots, n$, let I be a random index with distribution $P(I = i) = m_i/m$. When the summand variables are classically independent, and I is independent of $\{X_i, X_i^s, i = 1, \ldots, n\}$ where X_i^s is independent of $X_j, j \neq i$ and has the X_i size-bias distribution, then S_n^s given by $S_n - X_I + X_I^s$ has the S_n-zero bias distribution, and the characteristic function of S_n^s satisfies*

$$\psi_{S_n^s}(t) = E[\psi_{X_I^s}(t)\psi_{S_n - X_I}(t)]. \tag{3.25}$$

When the summand variables are freely independent, then the Cauchy transform of S_n^s satisfies

$$G_{S_n^s}(z) = E[G_{X_I^s}(\omega_{X_I, S_n - X_I}(z))].$$

When the summands are identically distributed as X, these relations respectively specialize to

$$\psi_{S_n^s}(t) = \psi_{X^s}(t)\psi_{S_{n-1}}(t) \quad \text{and} \quad G_{S_n^s}(z) = G_{X^s}(\omega_{X_n, S_n - X_n}(z)). \tag{3.26}$$

Proof. For the classical result (3.25) see, for example, [27] or [4]. Next,

$$\begin{aligned} mG_{S_n^s}(z) = mE[f_z(S_n^s)] = E[S_n f_z(S_n)] &= \sum_{i=1}^n E[X_i f_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i)] \\ &= \sum_{i=1}^n m_i E[f_{\omega_{\mu_i,\nu_{n\neg i}}(z)}(X_i^s)] = m\sum_{i=1}^n \frac{m_i}{m} G_{X_i^s}(\omega_{\mu_i,\nu_{n\neg i}}(z)), \end{aligned}$$

where we have used Lemma 3.17 for the third equality, and the size bias property (1.4) for the fourth; the proof now follows as for Theorem 3.18. □

Remark 3.22. *Mirroring Remark 3.20, it follows now that freely independent sums do* not *satisfy the classical "replace one" property for size bias. Indeed: if it were true that $S_n^s \stackrel{\mathrm{d}}{=} S_n - X_n + X_n^s$, it would then follow that $G_{S_n^s}(z) =$*

$G_{X_n^s}(\omega_{X_n^s, S_n - X_n}(z))$; comparing to (3.26), we see this is only possible if $X_n^s \stackrel{\mathrm{d}}{=} X_n$, meaning that all the summands are constant.

3.5. **A Transform that Fixes the Free Poisson Distribution.** The (classical) zero bias arises as a tool for Gaussian approximation in Stein's method. As noted just following its definition (1.5), the normal distribution $X \stackrel{\mathrm{d}}{=} \mathcal{N}(0, \sigma^2)$ is the unique fixed point of the zero bias transform $X \mapsto X^*$; correspondingly, certain metric differences between X and X^* yield quantitative bounds on distances between $\mathcal{L}(X)$ and $\mathcal{N}(0, \sigma^2)$.

Stein's method is not solely restricted to the realm of normal approximation, however. Along similar lines, the (shifted) size bias plays the analogous role in the world of *Poisson approximation*: Poiss(λ) is the unique fixed point of the transform $X \mapsto X^s - 1$ on the domain $\mathcal{D}_\lambda^+$ (positively supported probability distributions with mean $\lambda > 0$). Accordingly, functionals of the difference $X - X^s + 1$ can be used to construct bounds on the total variation distance between $\mathcal{L}(X)$ and Poiss(λ); see [17], [5] and, for example, [26], where in the first two references the use of size bias is implicit.

The main result of this section is the following identification of a transform (involving both the size bias and El Gordo transforms) for which the free Poisson law $\mathrm{FP}(\lambda, \alpha)$ of Section 2.5 is the unique fixed point in $\mathcal{D}_{\alpha\lambda}^+$.

Theorem 3.23. *On the domain $\mathcal{D}_\beta^+$, the transform*

$$\mathcal{L}(X) \mapsto \mathcal{L}(X^{\bullet,\lambda}) := \left(\lambda \mathcal{L}(X^s) + (1-\lambda)\mathcal{L}(X)\right)^\flat, \quad \lambda \in (0, 1],$$

has the free Poisson distribution $\mathrm{FP}(\lambda, \beta/\lambda)$ as its unique fixed point.

Proof. If $\mathcal{L}(X)$ is a fixed point, then $\mathcal{L}(X) = \left(\lambda \mathcal{L}(X^s) + (1-\lambda)\mathcal{L}(X)\right)^\flat$, which means the two sides have the same (squared) Cauchy transform:

$$G_X^2(z) = G_{X^{\bullet,\lambda}}^2(z).$$

From the definition of $\flat$, and the fact that Cauchy transforms of mixtures are given by the corresponding convex combinations of Cauchy transforms, this yields

$$G_X(z)^2 = \frac{1}{z}\left[\lambda G_{X^s}(z) + (1-\lambda) G_X(z)\right].$$

For notational convenience, set $\alpha = \beta/\lambda$, so that $\beta = \alpha\lambda$. On $\mathcal{D}_\beta^+ = \mathcal{D}_{\alpha\lambda}^+$, the definition of the size bias then gives

$$z G_X^2(z) = \lambda \cdot \frac{1}{\alpha\lambda}(z G_X(z) - 1) + (1-\lambda) G_X(z)$$

which is the quadratic equation

$$\alpha z G_X^2(z) = z G_X(z) - 1 + \alpha(1-\lambda) G_X(z) = (z + \alpha(1-\lambda)) G_X(z) - 1.$$

The quadratic formula then gives the solution

$$G_X(z) = \frac{z + \alpha(1-\lambda) - \sqrt{(z+\alpha(1-\lambda))^2 - 4\alpha z}}{2\alpha z}.$$

Finally, a calculation shows that the quantity inside the square root satisfies

$$(z + \alpha(1-\lambda))^2 - 4\alpha z = (z - \alpha(1+\lambda))^2 - 4\alpha^2\lambda$$

hence showing that $G_X = G_{\mathrm{FP}(\lambda,\alpha)}$. This shows that $\mathrm{FP}(\lambda, \alpha)$ is the only possible fixed point; reading the proof backwards shows that if $X \stackrel{\mathrm{d}}{=} \mathrm{FP}(\lambda, \alpha)$ then $G_X^2 =$

$G^2_{X^{\bullet,\lambda}}$. Thus the two Cauchy transforms must be equal (they must have the same sign in the square root, owing to the asymptotic and mapping constraints), and it follows that $\mathcal{L}(X) = \mathcal{L}(X^{\bullet,\lambda})$ as desired. □

Remark 3.24. *The mixture defining $X^{\bullet,\lambda}$ only makes sense if $0 \leq \lambda \leq 1$. For $\lambda > 1$, the free Poisson is a rescaling of the $1/\lambda$ case with a point mass removed (see the discussion following (2.9)), and so one could in principle write down a modified version of $X \mapsto X^{\bullet,\lambda}$ that applies in this setting (involving adding and removed a point mass).*

It is also worth noting that Theorem 3.23 shows that the $\flat$ of a distribution need not be absolutely continuous: $\mathrm{FP}(\lambda, \alpha)$ has a point mass for $0 < \lambda < 1$, and it is the $\flat$ of another distribution (a mixture of itself and its size bias, as the theorem shows). In Section 4 we will show that this cannot happen if we also square bias: that is, the free zero bias $X^\circ = (X^\square)^\flat$ is always absolutely continuous.

3.6. **Free Zero Bias Transform Examples.**

Example 3.25. *a) We compute $\mathcal{L}((X^\square)^\flat)$ for two distributions that do not have mean zero, and compare it to (3.15) that defines the free zero bias in the general mean case in order to demonstrate that the results of these two transforms, equal when the mean is zero, are not necessarily equal in general.*
First consider the semi-circle distribution S with mean zero and variance one, and let $X = S + m$ for $m \in \mathbb{R}$. Using (3.17) with $S \overset{\mathrm{d}}{=} S_{0,1}$ to obtain

$$G_X(z) = G_{S+m}(z) = G_S(z-m) = \frac{z - m - \sqrt{(z-m)^2 - 4}}{2},$$

identity (3.14) yields

$$\begin{aligned} G^2_{(X^\square)^\flat}(z) &= \frac{1}{m^2+1}\left(zG_S(z-m) - \frac{m}{z} - 1\right) \\ &= \frac{1}{m^2+1}\left(\frac{1}{2}z^2 - m\left(\frac{z}{2} + \frac{1}{z}\right) - 1 - \frac{z}{2}\sqrt{(z-m)^2-4}\right). \end{aligned}$$

Using (3.15) in the second equality, and that $S^\circ \overset{\mathrm{d}}{=} S$ in the third, we obtain

$$X^\circ = (S+m)^\circ = S^\circ + m = S + m = X,$$

thus confirming that X is a fixed point, and hence

$$G^2_{X^\circ}(z) = G^2_{S+m}(z) = \left(\frac{z - m - \sqrt{(z-m)^2-4}}{2}\right)^2.$$

We see that $G_{(X^\square)^\flat}$ and G_{X° cannot be not equal, the former not possessing a limit as $y \downarrow 0$ when evaluated at $z = iy$, unlike the latter.
For our second such example consider mixtures of point masses at zero and at some non-zero value a,

$$\mathcal{L}(X) = \alpha\delta_a + (1-\alpha)\delta_0 \quad \text{for } \alpha \in (0,1].$$

Then $E[X] = \alpha a$, and shifting the distribution to have mean zero we have

$$\mathcal{L}(X - \alpha a) = \alpha\delta_{(1-\alpha)a} + (1-\alpha)\delta_{-\alpha a}.$$

We have $E(X-\alpha a)^2 = \alpha(1-\alpha)a^2$, *and hence*

$$P(X^{\square} = (1-\alpha)a) = \frac{((1-\alpha)a)^2 P(X=(1-\alpha)a)}{\alpha(1-\alpha)a^2} = (1-\alpha),$$

thus implying

$$\mathcal{L}((X-\alpha a)^{\square}) = (1-\alpha)\delta_{(1-\alpha)a} + \alpha\delta_{-\alpha a}.$$

Therefore

$$G_{(X-\alpha a)^{\square}}(z) = \frac{1-\alpha}{z-(1-\alpha)a} + \frac{\alpha}{z+\alpha a} = \frac{z}{(z+\alpha a)(z-(1-\alpha)a)}.$$

Now taking the El Gordo transformation produces

$$G_{((X-\alpha a)^{\square})^{\flat}}(z) = -\sqrt{\frac{1}{(z+\alpha a)(z-(1-\alpha)a)}}$$

and adding the mean αa *back in we obtain*

$$G_{((X-\alpha a)^{\square})^{\flat}+\alpha a}(z) = G_{((X-\alpha a)^{\square})^{\flat}}(z-\alpha a) = \frac{1}{\sqrt{z(z-a)}}, \tag{3.27}$$

corresponding to the arcsine distribution on $[0,a]$.
Alternatively, computing $(X^{\square})^{\flat}$ *as in Definition 3.10, we first have* $\mathcal{L}(X^{\square}) = \delta_a$ *from b) of Lemma 3.6, and from (3.4) we find*

$$G_{(X^{\square})^{\flat}}(z) = \frac{1}{\sqrt{z(z-a)}},$$

agreeing with (3.27).
b) For the mean zero, variance ab mixture

$$\mathcal{L}(X) = \frac{b}{b+a}\delta_{-a} + \frac{a}{b+a}\delta_b$$

of point masses at $-a, b$ *for positive* a, b, *by a) of Lemma 3.6 we have*

$$\mathcal{L}(X^{\square}) = \frac{a^2}{ab}\frac{b}{b+a}\delta_{-a} + \frac{b^2}{ab}\frac{a}{b+a}\delta_b = \frac{a}{b+a}\delta_{-a} + \frac{b}{b+a}\delta_b.$$

Hence, using Definition 3.10 for the second equality,

$$G_{X^{\square}}(z) = \frac{z}{(z+a)(z-b)} \quad \text{and} \quad G_{X^{\circ}}(z) = \frac{1}{\sqrt{(z+a)(z-b)}}.$$

In particular, when X *is Rademacher,*

$$G_{X^{\circ}}(z) = \frac{1}{\sqrt{z^2-1}}.$$

Recalling Remark 3.11 that contrasts the constructions of the classical zero bias with the free one, note that U, *used in the classical construction, has the* $\mathrm{Beta}(1,1)$ *distribution, while the standard arcsine, of which all examples here are scaled and shifted versions via Example 3.3, has the* Beta *distribution.*
c) Taking one further iterate of the free zero bias distribution of the Rademacher in b), the free zero bias distribution of the arcsine law on $[-1,1]$, *having mean zero and variance* $1/2$, *has Cauchy transform*

$$G_{(X^{\circ})^{\circ}}(z) = -\sqrt{2}\sqrt{\frac{z}{\sqrt{z^2-1}} - 1}.$$

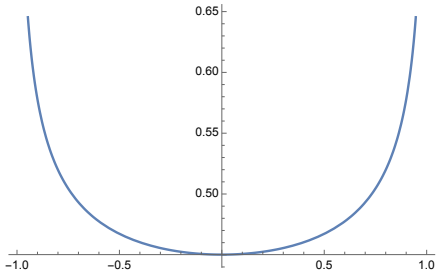

FIGURE 1. The density (3.28) of the iterated free zero bias of the Rademacher distribution, i.e. the free zero bias of the arcsine law.

Using Stieltjes inversion (Proposition 2.4), we evaluate $z = x+i\epsilon$ *and let* $\epsilon \downarrow 0$. *Note that* $z^2-1 = x^2-1-\epsilon^2+2i\epsilon x$ *and this number is in the* lower *half plane when* $x < 0$*; thus, its square root has* negative *real part, in particular* $\sqrt{z^2-1} \to -\sqrt{x^2-1}$ *for* $x < 0$ *as* $\epsilon \downarrow 0$. *Similarly* $\sqrt{z^2-1} \to \sqrt{x^2-1}$ *when* $x > 0$. *Now evaluating, we see that* $G_{(X^\circ)^\circ}(x+i\epsilon)$ *converges as* $\epsilon \downarrow 0$ *to real values when* $|x| > 1$*, meaning that such points are outside the support of* $(X^\circ)^\circ$. *Meanwhile, in the interval* $[-1,1]$*, the formula for the imaginary part of the complex square root yields the density*

$$\varrho_{(X^\circ)^\circ}(x) = \frac{1}{\pi}\sqrt{1+\frac{1}{\sqrt{1-x^2}}}\mathbf{1}_{x\in[-1,1]}. \tag{3.28}$$

4. Regularity of the Free Zero Bias

In this section we prove that the free zero bias μ° of any distribution μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, with support contained in the convex hull of the support of μ. This behavior is similar to, but not exactly the same as, that of classical zero bias μ^*, which is also absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, but whose support is always *equal* to the convex hull of the support of μ, cf. [24]. By comparison, μ° may have support *strictly* contained in the support of μ, cf. Example 4.2.

Section 4.1 proves the absolute continuity of the free zero bias, including quantitative bounds, and Section 4.2 studies the support of μ° for finitely-supported μ, showing in particular that it is always disconnected when the finite support has size three or more.

4.1. Absolute Continuity of the Free Zero Bias.

The following result provides quantitative bounds that were previously unknown even in the classical case; the the remaining claims regarding μ^* are included for comparison and completeness, and can be found in [24] and [18].

Theorem 4.1. *Let* μ^* *and* μ° *respectively be the classical and free zero bias of a mean zero probability measure* μ *with finite, non-zero variance* σ^2. *Then*

$$\mu^*([a,b]) \leq \left(\frac{b-a}{\sigma^2}\right)E[|X|] \quad \textit{and} \quad \mu^\circ([a,b])^2 \leq \left(\frac{b-a}{\sigma^2}\right)E[|X|], \tag{4.1}$$

and both μ^* *and* μ° *are absolutely continuous with respect to Lebesgue measure. It holds that* $\operatorname{supp}(\mu^*) = \operatorname{co}(\operatorname{supp}\mu)$*, and* $\operatorname{supp}(\mu^\circ) \subseteq \operatorname{co}(\operatorname{supp}\mu)$.

Proof. Let X have law μ, and consider the case where $\mathrm{Var}(X) = 1$. We first show that μ° has no point masses. For the sake of contradiction, assume that X° takes on the value $a \in \mathbb{R}$ with non-zero probability p_a. Then, in view of (1.11), from

$$X^* = UX^\circ + (1-U)Y^\circ$$

where Y° is an independent copy of X°, we see that $X^* = a$ with probability at least p_a^2, in contradiction to the absolute continuity of the X^* distribution, see [24]. Recalling $g^{[1]}(x,y) = (g(y)-g(x))/(y-x)$, as $Y^\circ - X^\circ$ has a continuous distribution, we have that $X^\circ = Y^\circ$ with probability zero, hence the value along the diagonal $x = y$ of double integrals such as (4.3) below, with respect to the product measure $\mu^\circ \times \mu^\circ$, is zero.

With $a < b$ arbitrary, take $g(t) = (t-a)_+ \wedge (b-a)$. Then

$$\begin{aligned} E[Xg(X)] &= E[X(X-a)\mathbf{1}(a < X \le b)] + (b-a)E[X\mathbf{1}(X > b)] \\ &\le (b-a)E[|X|\mathbf{1}(a < X \le b)] + (b-a)E(|X|\mathbf{1}(X > b)] \\ &= (b-a)E[|X|\mathbf{1}(X > a)] \le (b-a)E[|X|]. \end{aligned} \tag{4.2}$$

Using that $g'(t) = \mathbf{1}_{[a,b]}$ and (1.5), which also holds for Lipshitz functions [18], we find via (4.2) that

$$\mu^*([a,b]) = E[g'(X^*)] = E[Xg(X)] \le (b-a)E[|X|],$$

thus showing the first claim of (4.1) when $\sigma^2 = 1$.

For the second claim, dividing up the plane according to Figure 2, we first consider the (upper $y = x$ diagonal) collection of cases, the remaining cases across the diagonal given by interchanging x and y.

$$g^{[1]}(x,y) = \begin{cases} 1 & a \le y \le b, a \le x \le b \\ \frac{y-a}{y-x} & a \le y \le b, x < a \\ \frac{b-x}{y-x} & y > b, a \le x \le b \\ \frac{b-a}{y-x} & y > b, x < a \\ 0 & y < a, x < a \quad \text{or} \quad y > b, x > b. \end{cases}$$

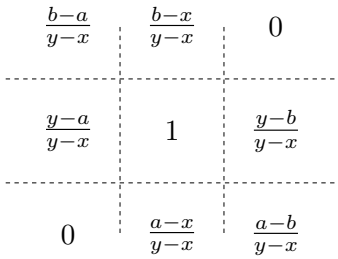

FIGURE 2. Evaluation of $\partial g(x,y)$ on the partition of the plane induced by the horizontal and vertical lines $y = a, y = b$ and $x = a, x = b$ for $a < b$.

Using symmetry when interchanging x and y, we obtain

$$E[g^{[1]}(X^\circ, Y^\circ)] = \mu^\circ([a,b])^2 + 2\int_{(-\infty,a)}\int_{[a,b]} \frac{y-a}{y-x} d\mu^\circ(y) d\mu^\circ(x)$$
$$+ 2\int_{[a,b]}\int_{(b,\infty)} \frac{b-x}{y-x} d\mu^\circ(y) d\mu^\circ(x) + 2\int_{(-\infty,a)}\int_{(b,\infty)} \frac{b-a}{y-x} d\mu^\circ(y) d\mu^\circ(x). \quad (4.3)$$

As the integrands are all non-negative, using (4.2) for the final inequality we obtain

$$\mu^\circ([a,b])^2 \le E[g^{[1]}(X^\circ, Y^\circ)] = E[Xg(X)] \le (b-a)E[|X|],$$

showing (4.1) when $\sigma^2 = 1$.

Now turning to the support, letting $g(t) = (t-a)_+$, for $x < y$ we have

$$g^{[1]}(x,y) = \begin{cases} \frac{y-a}{y-x} & y > a, x < a \\ 1 & y > a, x \ge a \\ 0 & y < a, x < a, \end{cases}$$

and hence

$$E[X(X-a)\mathbf{1}(X>a)] = E[Xg(X)]$$
$$= E[g^{[1]}(X^\circ, Y^\circ)] = 2\int_{-\infty}^{a}\int_{a}^{\infty} \frac{y-a}{y-x} d\mu^\circ(y) d\mu^\circ(x) + \mu^\circ((a,\infty))^2.$$

As $y > a$ in the first integrand, the first integral is non-negative and hence the right hand side is non-negative.

For any a such that $P(X \ge a) = 0$ the left hand side above is zero, and hence both terms on the right are zero; in particular $P(X^\circ > a) = \mu^0((a,\infty)) = 0$. Now let a be such that $P(X \le a) = 0$, or, equivalently, that $P(-X \ge -a) = 0$. Using the scaling property b) of Lemma 3.12 with $\alpha = -1$ yields $(-X)^\circ \stackrel{\mathrm{d}}{=} -X^\circ$, hence $0 = P((-X)^\circ \ge -a) = P(-X^\circ \ge -a) = P(X^\circ \le a) = \mu^\circ((-\infty, a]))$, showing that the support of μ° is contained in $[a,b]$.

For X having measure μ with arbitrary variance $\sigma^2 > 0$, applying the case shown for the unit variance variable $Y = X/\sigma$ and using that $X^* \stackrel{\mathrm{d}}{=} \sigma Y^*$ yields

$$\mu^*([a,b]) = P(X^* \in [a,b]) = P(\sigma Y^* \in [a,b]) = P(Y^* \in [a/\sigma, b/\sigma])$$
$$\le \left(\frac{b-a}{\sigma}\right) E|Y| = \left(\frac{b-a}{\sigma^2}\right) E|X|.$$

The argument for general σ in the free zero bias case (4.1) is identical.

The absolute continuity claims now follows, and the final one, regarding the support of μ° follows from the unit variance case, and that for any $\sigma > 0$ we have $\mathrm{supp}(X^\circ) \subseteq \mathrm{co}(\mathrm{supp} X)$ if and only if $\mathrm{supp}((\sigma X)^\circ) \subseteq \mathrm{co}(\mathrm{supp}(\sigma X))$. □

4.2. **Support of μ° for Finitely-Supported μ.** The (classical) zero bias μ^* of a probability measure is always supported exactly on the convex hull of $\mathrm{supp}\,\mu$, cf. [24]. In this section we show that this fails to be true in the free case, where $\mathrm{supp}\,\mu^\circ$ can be a strict subset of the convex hull of $\mathrm{supp}\,\mu$. In fact, we show that when μ is finitely supported the support of μ° is a finite union of disjoint closed intervals.

We begin with the following illustrative example.

Example 4.2. *Consider the three-point centered random variable*

$$X \stackrel{\mathrm{d}}{=} \frac{3}{5}\delta_{-1} + \frac{1}{5}\delta_1 + \frac{1}{5}\delta_2 \quad \textit{with} \quad \mathrm{Var}(X) = \frac{8}{5} =: \sigma^2. \tag{4.4}$$

The free zero bias X° has Cauchy transform satisfying

$$\sigma^2 G_{X^\circ}^2(z) = zG_X(z) - 1 = E\left[\frac{z}{z-X}\right] - 1 = E\left[\frac{X}{z-X}\right].$$

Hence

$$\frac{8}{5}G_{X^\circ}^2(z) = \frac{3}{5}\frac{-1}{z+1} + \frac{1}{5}\frac{1}{z-1} + \frac{1}{5}\frac{2}{z-2} = \frac{8z-10}{5(z^2-1)(z-2)}$$

and from (3.13) *it follows that*

$$G_{X^\circ}(z) = -\sqrt{\frac{z-5/4}{(z^2-1)(z-2)}}.$$

The rational function inside the square root is continuous on $\mathbb{C}\setminus\{-1,1,2\}$, and hence at all other points $x \in \mathbb{R}$,

$$-\frac{1}{\pi}\lim_{y\downarrow 0}\mathrm{Im}\, G_{X^\circ}(x+iy) = \frac{1}{\pi}\mathrm{Im}\sqrt{\frac{x-5/4}{(x^2-1)(x-2)}}.$$

The rational expression takes negative values on the disconnected set $(-1,1)\sqcup(5/4,2)$, and non-negative values elsewhere. Thus, with the possible exception of point masses at ± 1 and 2, the density of X° is

$$\varrho_{X^\circ}(x) = \frac{1}{\pi}\sqrt{\frac{x-5/4}{(x^2-1)(2-x)}}\mathbb{1}_{\{x\in[-1,1]\sqcup[5/4,2]\}}.$$

With the aid of Mathematica (or a sufficiently intricate trigonometric substitution) one can check that ϱ_{X° is genuinely a probability density (total mass 1), and hence there are no point masses. Thus, X° has a continuous density (with three poles) on a **disconnected** *support set.*

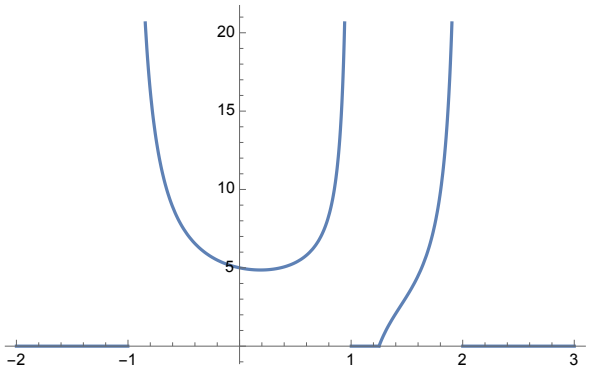

FIGURE 3. The density of the free zero bias of the three point distribution in (4.4).

Example 4.3. *The phenomenon demonstrated in Example 4.2 above and Theorem 4.5 below is not restricted to the finite-support case. Indeed, one can mimic the calculations in Example 4.2 with "bumps" of support instead of point masses. For example: let X have the arcsine law on $[0,1]$ (with density $1/\pi\sqrt{x(1-x)}$), and let μ mixture the mixture of the laws of $X-1$, $X+1$, and $X+3$ with coefficients $\frac{5}{6}$, $\frac{1}{12}$, and $\frac{1}{12}$ (which is centered). Then the Cauchy transform of μ is*

$$\begin{aligned} G_\mu(z) &= \frac{5}{6}G_X(z+1) + \frac{1}{12}G_X(z-1) + \frac{1}{12}G_X(z-3) \\ &= \frac{5}{6}\frac{1}{\sqrt{z(z+1)}} + \frac{1}{12}\frac{1}{\sqrt{(z-1)(z-2)}} + \frac{1}{12}\frac{1}{\sqrt{(z-3)(z-4)}} \end{aligned}$$

A laborious but elementary calculation provides an explicit radical-rational formula for the square $G^2_{\mu^\circ}$ of the Cauchy transform of the free zero bias μ°, and a simple analysis like above then shows that the support of the law μ° includes the intervals $[-1,0]$ and $[1,2]$, but does not include any points in $(0,1)$.

Theorem 4.5 below shows that this case is 'canonical' for finitely supported X. Toward that goal, we begin with the following lemma which we will then apply to the square bias (en route to the free zero bias).

Lemma 4.4. *Let $X \stackrel{\mathrm{d}}{=} \mu := \sum_{i=1}^n p_i\delta_{s_i}$ with $s_1 < \cdots < s_n$, $p_i > 0$, and $\sum_{i=1}^n p_i = 1$. The Cauchy transform*

$$G_\mu(z) = \sum_{i=1}^n \frac{p_i}{z - s_i} \tag{4.5}$$

can be written as the rational function $G_\mu(z) = Q(z)/P(z)$, where

$$P(z) = \prod_{i=1}^n (z - s_i), \quad Q(z) = \sum_{i=1}^n p_i Q_i(z) \quad \text{with} \quad Q_i(z) = \prod_{1\le j\le n, j\neq i} (z - s_j). \tag{4.6}$$

The polynomials P and Q are both monic, with respective degrees of n, and $n-1$. The polynomial Q has exactly $n-1$ real roots $r_j \in (s_j, s_{j+1}), 1, \ldots, n-1$.

When $E[X^{-1}] = 0$ there exists $i^* \in \{1, \ldots, n-1\}$ such that $0 \in (s_{i^*}, s_{i^*+1})$, and

$$\{x \in \mathbb{R} : G_{\mu^\flat}(x) \le 0\} = \bigcup_{1\le j\le i^*-1} [s_j, r_j] \cup [s_{i^*}, s_{i^*+1}] \cup \bigcup_{i^*+1\le j\le n-1} [r_j, s_{j+1}].$$

Proof. The representation of G_μ as a rational function with denominator and numerator as in (4.6) is immediate, as is the fact that for all $i = 1, \ldots, n$, P and Q_i are monic polynomials of respective degrees n and $n-1$. In particular, the coefficient of x^{n-1} of Q is given by $\sum_{i=1}^n p_i = 1$.

Now, for any $s \in \mathbb{R}$,

$$\lim_{x\uparrow s} \frac{1}{x-s} = -\infty \quad \text{and} \quad \lim_{x\downarrow s} \frac{1}{x-s} = \infty, \tag{4.7}$$

and so we see that

$$\lim_{x\uparrow s_{j+1}} G_\mu(x) = -\infty \quad \text{and} \quad \lim_{x\downarrow s_j} G_\mu(x) = \infty \quad \text{for } j = 1, \ldots, n-1.$$

As $Q(x)$ is continuous in $I_j = (s_j, s_{j+1}), j = 1, \ldots, n-1$, it must have least one root in this interval. As Q has degree $n-1$ and therefore at most this many distinct

real roots, each of the $n-1$ intervals $I_j, j = 1, \ldots, n-1$ must contain exactly one real root, r_j.

For the next claim, when $E[X^{-1}] = 0$ the random variable X must have both positive and negative support, and so $s_1 < 0 < s_n$. Hence, as all support points are non-zero there exists exactly one index $i^* \in \{1, \ldots, n-1\}$ such that $0 \in (s_{i^*}, s_{i^*+1})$. Since

$$0 = -E[X^{-1}] = G_\mu(0) \quad \text{and} \quad Q(0) = G_\mu(0)P(0) \quad \text{as} \quad P(0) = (-1)^n \prod_{i=1}^{n} s_i \neq 0,$$

we find that Q has a root at zero, that is, that $r_{i^*} = 0$, allowing us to write

$$Q(x) = xT(x) \quad \text{where} \quad T(x) = \prod_{j:1\leq j\leq n-1, j\neq i^*} (x - r_j).$$

Cancelling the factor of z in (3.4) one obtains

$$G^2_{\mu^\flat}(z) = \frac{T(z)}{P(z)}, \tag{4.8}$$

where T is a monic polynomial in z of degree $n-2$, with real coefficients and having no root in the interval I_{i^*}.

As n and $n-2$ share the same parity and P and T are monic of those degrees respectively, the limits of $G^2_{\mu^\flat}(x)$ at both plus and minus infinity are 1, and we find that the sign of (4.8) must be positive on $(-\infty, s_1)$ and (s_n, ∞).

Assume that for some $1 \leq j \leq i^* - 1$ the sign of (4.8) is positive in some non trivial open interval with right endpoint s_j. At s_j, via (4.7), the sign of a multiplicative factor of (4.8) changes, thus flipping the sign of (4.8) to negative, and remaining negative until encountering the root r_j. At this (single) root the sign flips to positive, and remains so until the end of the interval at the pole s_{j+1}.

As the premise is true for $j = 1$, it remains true for all $j \leq i^* - 1$, and the sign of (4.8) is negative in the intervals (s_j, r_j) and positive in the interval (r_j, s_{j+1}). As the sign, via the same reasoning, changes to negative at s_{i^*} and the root previously in I_{i^*} has been cancelled, the sign of (4.8) is negative throughout the interval (s_{i^*}, s_{i^*+1}).

Arguing under the premise that the sign of (4.8) is negative in some non trivial open interval with right endpoint s_j, the same inductive argument shows that the sign of (4.8) is negative in the intervals (r_j, s_{j+1}) for $i^* + 1 \leq j \leq n-1$, completing the proof of the lemma. □

We now move to the consideration of the support of the free zero bias of a finitely supported measure.

Theorem 4.5. *Let X have mean zero and variance $\sigma^2 \in (0, \infty)$, and be supported on a finite set, and let $S = \text{supp}(X) \setminus \{0\} = \{s_1, \ldots, s_n\}$ with $s_1 < \ldots < s_n$ and $P(X = s_j) = p_j, 1 \leq j \leq n$. Then*

$$G^2_{X^\circ}(z) = \frac{1}{z\sigma^2} \sum_{j=1}^{n} \frac{s_i^2 p_i}{z - s_j},$$

and letting

$$Q(z) = \sum_{i=1}^{n} \frac{s_i^2 p_i}{\sigma^2} Q_i(z) \quad \textit{where} \quad Q_i(z) = \prod_{1\leq j\leq n, j\neq i} (z - s_j),$$

then the polynomial $Q(z)$ has $n-1$ real roots $r_1 < \cdots < r_{n-1}$, there exists $i^* \in \{1,\dots,n-1\}$ such that $0 \in (s_{i^*}, s_{i^*+1})$,

$$\operatorname{supp}(X^\circ) = \bigcup_{1\le j\le i^*-1} [s_j, r_j] \cup [s_{i^*}, s_{i^*+1}] \cup \bigcup_{i^*+1\le j\le n-1} [r_j, s_{j+1}], \tag{4.9}$$

and the density ϱ_{X° of X° is given by

$$\varrho_{X^\circ}(x) = \frac{1}{\pi}\sqrt{\max\{0, -G^2_{X^\circ}(x)\}}.$$

Proof. Identity (1.3) yields

$$P(X^\square = s_i) = \frac{s_i^2 p_i}{\sigma^2} \quad \text{for} \quad s_i \in \operatorname{Sup}(S),$$

and as these probabilities sum to one we obtain $\operatorname{supp}(X^\square) = S$. In addition, applying Definition 3.5,

$$E[(X^\square)^{-1}] = \frac{E[(X^{-1})X^2]}{E[X^2]} = \frac{E[X]}{E[X^2]} = 0.$$

The results on Q and (4.9) now follow by applying Definition 3.10, which yields $X^\circ \stackrel{\mathrm{d}}{=} (X^\square)^\flat$, and the final part of Lemma 4.4 with $\mathcal{L}(X)$ replaced by $\mathcal{L}(X^\square)$, and p_i in (4.5) replaced by $s_i^2 p_i/\sigma^2, s_i, i = 1,\dots,n$.

The density formula follows from the Stieltjes inversion formula (Proposition 2.4), noting that in the case at hand taking limits of $z = x + i\epsilon$ as $\epsilon \downarrow 0$ yields the real valued function $G^2_{X^\circ}(x)$ for all $x \notin \{s_1,\dots,s_n\}$, and hence the square root of $G_{X^\circ}(x)$ will be imaginary if and only if $G^2_{X^\circ}(x) < 0$. □

5. Infinite Divisibility

This section is devoted to the proofs of Theorems 1.6 and 1.8, which connect the free zero bias and the size bias transformations to free infinite divisibility; i.e. to random variables $X \stackrel{\mathrm{d}}{=} X_{1,n}+\cdots+X_{n,n}$ for any $n \in \mathbb{N}$, where $\{X_{j,n}\}_{1\le j\le n}$ are freely independent and identically distributed. The seminal papers [8, 10] by Voiculescu and Bercovici explore and characterize the distributions of such random variables in terms of an analogue (1.16) to the classical Lévy–Khintchine representation. In their complex analytic approach, compactly-supported laws are treated in [10], and extended to all freely infinitely divisible laws (with no moment assumptions) in [8]. Our contributions below begin by giving a new intuitive and probabilistic proof of the free Lévy–Khintchine formula in the L^2 case, via a new Cauchy transform identity using the free zero bias.

In addition, our new approach shows that *any* finite measure ρ corresponds, through (1.16), to a unique (up to a choice of mean) L^2 freely infinitely divisible random variable X whose variance is the mass of ρ, and that the probability measure $\rho/\rho(\mathbb{R})$ has a probabilistic interpretation connecting X and its free zero bias X°: this probability measure is the law of the random variable V_0 in (5.1)-(5.3). This free Lévy law has an immediate probabilistic interpretation: it is the weak limit of the square biases $X^\square_{1,n}$ of the nth root distributions of X (ergo the L^2 requirement for this approach).

In parallel, we give a new characterization of all *positively* freely infinitely divisible L^1 random variables (meaning that all free convolution roots are positively-supported) and we prove a completely new positive Lévy–Khintchine type formula

(1.20) for these laws. The measure in this positive formulation is the law of a different random variable V_+ connected to X via the size bias, and also having an appealing probabilistic construction: V_+ is the weak limit of the size biases $X_{1,n}^s$ of the nth root distributions of X (ergo the L^1 assumption here). What's more, at the intersection of the two cases (positively freely infinitely divisible L^2 random variables), V_+ is the size bias of V_0, and this is characterized by V_0 having a negative moment, as shown in Theorem 1.10.

We begin in Section 5.1, which is encompassed largely by Theorems 5.1 and 5.3. These theorems prove the "symbolic equivalence" expressed in Theorems 1.6 and 1.8 — in the first case, the equivalence of (5.1)–(5.3) in the first case and of (5.5)–(5.7) in the second case, *on some truncated domains* — with the majority of the remainder of the section devoted to extending the relations to the full upper half plane. To that end, Section 5.2 focuses on compound free Poisson distributions (cf. Section 2.5) and shows that are rich enough to produce all the laws $\mathcal{L}(V_0)$ for which $E[V_0^{-2}] < \infty$ in the first case, and all the laws $V_+ \geq 0$ for which $E[V_+^{-1}] < \infty$ in the second.

Section 5.3 begins, in parallel between the V_0 and V_+ cases, by removing the negative moment constraints with an approximations schemes (therefore escaping the compound free Poisson world) in Proposition 5.12, and then completes the pieces of the proof of Theorems 1.6 and 1.8 in Theorems 5.15 and 5.17. It then combines these results in the overlapping case to Prove Theorem 1.10. Section 5.4 gives an alternate approach under the stronger assumption of bounded random variables; this approach is more self-contained, and also introduces a result, Lemma 5.19, on the uniformity of the support of free convolution roots that is of independent interest. Throughout the section, and at its conclusion, are several examples demonstrating the computational effectiveness of our new perspective on free infinite divisibility.

5.1. **New Formulations of Free Infinite Divisibility.** This section is devoted to preliminary versions of our main Theorems 1.6 and 1.8, in the forms of Theorem 5.1 and 5.3 below. Theorem 5.1 shows that Bercovici and Voiculescu's free Lévy–Khintchine formula in this L^2 setting is equivalent to our new formulation (5.1) relating the reciprocal Cauchy transform of the free zero bias X° of the freely infinitely divisible random variable X to that of the El Gordo transform $V_0^\flat$ of its free Lévy–Khintchine measure $\mathcal{L}(V_0)$. But the equivalence here is only proved for these analytic transforms on certain truncated open domains of the complex upper half-plane; further tools later in this section are needed to extend these results to their full strength. Meanwhile, Theorem 5.3 provides a completely new kind of free Lévy–Khintchine formula (5.6) for L^1 *positively* freely infinitely divisible random variables, and proves an equivalent formulation (5.5) relating the reciprocal Cauchy transform of the size bias of X to that of its (new) free Lévy–Khintchine measure $\mathcal{L}(V_+)$ (cf. Theorem 1.8(b)), again with formulas holding for now only on certain truncated domains, with extension to the full upper half-plane proved in following sections.

Theorem 5.1. *Let X be a mean m random variable with variance $\sigma^2 \in (0, \infty)$, and V_0 a random variable. The following are equivalent.*

$$F_{X^\circ}(w) = F_{V_0^\flat}(F_X(w)) \tag{5.1}$$

for $w \in \mathbb{C}_+$.

$$R_X(z) = m + \sigma^2 E\left[\frac{z}{1 - zV_0}\right] \tag{5.2}$$

for z in a reciprocal truncated cone.

$$G_{V_0}(z) = (\varphi_X(z) - m)/\sigma^2, \tag{5.3}$$

for z in a truncated cone.

If these equivalent conditions hold for both the pairs (X, V_0) and (X, V_0') then $\mathcal{L}(V_0) = \mathcal{L}(V_0')$, and if for both pairs (X, V_0) and (X', V_0) then, assuming $\sigma_X^2 \geq \sigma_{X'}^2$ without loss of generality,

$$X \overset{\mathrm{d}}{=} (X' - m_{X'})^{\boxplus t} + m_X \quad \textit{where} \quad t = \frac{\sigma_X^2}{\sigma_{X'}^2},$$

where, for any W, $W^{\boxplus t}$ is the t-fold free convolution of W.

Proof. To begin, note that $\varphi_X(z) = R_X(1/z)$ from the definitions (2.4) and (2.3) of the R-transform and the Voiculescu transform. Hence, if (5.3) holds we have

$$\begin{aligned} R_X(z) = \varphi_X(1/z) &= m + \sigma^2 G_{V_0}(1/z) \\ &= m + \sigma^2 E\left[\frac{1}{1/z - V_0}\right] = m + \sigma^2 E\left[\frac{z}{1 - zV_0}\right], \end{aligned}$$

yielding (5.2). Similarly if (5.2) holds then

$$(\varphi_X(z) - m)/\sigma^2 = (R_X(1/z) - m)/\sigma^2 = E\left[\frac{1/z}{1 - V_0/z}\right] = G_{V_0}(z).$$

To show the equivalence of (5.1) and (5.3) we may reduce to the case $m = 0$. Indeed, if we know that (5.1) holds for mean zero variables then

$$F_{X^\circ}(z) = F_{(X-m)^\circ + m}(z) = F_{(X-m)^\circ}(z - m) = F_{V_0^\flat}(F_{X-m}(z - m)) = F_{V_0^\flat}(F_X(z)),$$

and if (5.3) holds for mean zero variables, then since $\varphi_X(z) = m + \varphi_{X-m}(z)$, we find

$$(\varphi_X(z) - m)^2/\sigma^2 = \varphi_{X-m}(z)^2/\sigma^2 = G_{V_0}(z).$$

Hence we may assume $m = 0$ without loss of generality.

Now, suppose that (5.3) holds in some truncated cone $\Gamma_{\alpha,\beta}$ in the domain of φ_X (ergo the reciprocal Cauchy transform F_X has an analytic inverse F_X^{-1} there, cf. Proposition 2.7); under the assumption $m = 0$ this is the statement $\varphi_X(z) = \sigma^2 G_{V_0}(z)$. From the definition (2.3), $\varphi_X(z) = F_X^{-1}(z) - z$ and hence $F_X^{-1}(z) - z = \sigma^2 G_{V_0}(z)$ on $\Gamma_{\alpha,\beta}$. Letting $w = F_X^{-1}(z)$ for $z \in \Gamma_{\alpha,\beta}$, we have $F_X(w) = z$, and so

$$w - F_X(w) = \sigma^2 G_{V_0}(F_X(w)). \tag{5.4}$$

Using (3.13) for the first inequality and then the result of multiplying (5.4) through by $G_X(w)/\sigma^2$ and noting $F_X = 1/G_X$, and finally applying definition (3.4) for the last, we have

$$\begin{aligned} G_{X^\circ}^2(w) = \frac{wG_X(w) - 1}{\sigma^2} &= G_X(w) G_{V_0}(F_X(w)) \\ &= \frac{1}{F_X(w)} G_{V_0}(F_X(w)) = G_{V_0^\flat}^2(F_X(w)). \end{aligned}$$

The equality $G^2_{X^\circ}(w) = G^2_{V_0^\flat}(F_X(w))$ implies that $G_{X^\circ}(w) = G_{V_0^\flat}(F_X(w))$ (since the negative of a Cauchy transform is not the Cauchy transform of a positive measure). Taking reciprocals verifies (5.1) for $w \in F_X^{-1}(\Gamma_{\alpha,\beta})$, which is open (since $\Gamma_{\alpha,\beta}$ is open and F_X is continuous). Thus, the two analytic functions F_{X° and $F_{V_0^\flat} \circ F_X$ are equal on an open set, and hence open on their full domain $\mathbb{C}_+$ as claimed. The reverse implication (5.1) $\implies$ (5.3) is proved in exactly the same manner in reverse.

Turning to the second half of the theorem: When any of these equivalent identities is satisfied for X then V_0 is uniquely determined by its Voiculescu transform in (5.3), which depends only on the distribution of X, which itself encodes the mean. When both X and X' satisfy any of the three, and therefore all of the given identities for the same V_0, by (5.2) we must have

$$(\varphi_X(z) - m_X)/\sigma_X^2 = (\varphi_{X'}(z) - m_{X'})/\sigma_{X'}^2,$$

and hence

$$R_X(z) = t(R_{X'}(z) - m_{X'}) + m_X,$$

thus yielding the final claim. □

Example 5.2. *When X is a distribution with zero mean and variance $\sigma^2 > 0$, we know by Lemma 3.15 that the unique solution to $X^\circ \overset{\mathrm{d}}{=} X$ is the semicircle distribution with variance σ^2. In addition, this distribution must also satisfy (5.1) with $G_{V_0^\flat}(z) = 1/z$, hence, by Lemma 3.4, with $\mathcal{L}(V_0) = \delta_0$. In this case the Free–Lévy measure in (5.2) is given by point mass at zero, implying the well-known fact that the R-transform of a semicircle is given by $R_{S_{\sigma^2}}(z) = \sigma^2 z$. In particular $R_{S_{\sigma^2}}(z)$ is an R-transform for all $\sigma^2 \geq 0$, and thus, as for all $n \geq 1$ we have $R_{S_{\sigma^2}}(z) = R_{S_{\sigma^2/n}}(z) + \cdots + R_{S_{\sigma^2/n}}(z)$, the distribution of S_{σ^2} is infinitely divisible.*

In the other direction, given the infinity divisibility of the semi-circle distributions and their corresponding R-transforms $R_S(z) = \sigma^2 z$, one can find their Lévy–Khintchine measures by (5.2). Indeed, from there we obtain

$$G_{V_0}(z) = \varphi_S(z)/\sigma^2 = R_S(1/z)/\sigma^2 = 1/z,$$

demonstrating that $V_0 \overset{\mathrm{d}}{=} \delta_0$.

We now present an analogous result for non-negative random variables, and size biasing, under the weaker condition of the existence of a first moment.

Theorem 5.3. *Let X be a non-negative, mean $m \in (0, \infty)$ random variable, and V_+ a random variable. The following are equivalent.*

$$F_{X^s}(w) = F_{V_+}(F_X(w)) \tag{5.5}$$

For $w \in \mathbb{C}_+$,

$$R_X(z) = mE\left[\frac{1}{1 - zV_+}\right] \tag{5.6}$$

for z in a reciprocal truncated cone, and

$$G_{V_+}(z) = \varphi_X(z)/mz \tag{5.7}$$

for z in a truncated cone.

If these equivalent conditions hold for both the pairs (X, V_+) *and* (X, V_+') *then* $\mathcal{L}(V_+) = \mathcal{L}(V_+')$*, and if for both pairs* (X, V_+) *and* (X', V_+) *then, assuming* $m_X \geq m_{X'}$ *without loss of generality,*

$$X \stackrel{\mathrm{d}}{=} (X')^{\boxplus t} \quad \text{where} \quad t = \frac{m_X}{m_{X'}},$$

where, for any W*,* $W^{\boxplus t}$ *is the* t*-fold free convolution of* W*.*

Remark 5.4. *As we will see below in Proposition 5.14, when the equivalent conditions of this theorem hold (on larger domains), the random variable* V_+ *must be be* ≥ 0*.*

Proof. The equivalence between (5.6) and (5.7) follows from a simple calculation noting that $\varphi_X(z) = R_X(1/z)$, just as in the proof of Theorem 5.1.

To prove the equivalence of (5.7) and (5.5) on appropriate domains, we follow the same outline as in Theorem 5.1 (although with fewer calculations needed in this case). On a truncated cone where $\varphi_X(z)$ is defined, it is equal to $F_X^{-1}(z) - z$. Hence, on the image domain under F_X (which, according to Proposition 2.7, contains another truncated cone), setting $z = F_X(w)$, (5.5) is equivalent to

$$mF_X(w)G_{V_+}(F_X(w)) = w - F_X(w).$$

Multiplying both sides by $G_X(w)/m = 1/mF_X(w)$ (which is never 0) yields

$$G_{V_+}(F_X(w)) = \frac{wG_X(w) - 1}{m}$$

and the right-hand-side is equal to the Cauchy transform of the size bias $G_{X^s}(w)$, cf. Lemma 3.9(d). Taking reciprocals yields (5.5). Every step was reversible, and so the two are in fact equivalent. □

5.2. **Free Compound Poisson Distributions Revisited.** Free Poisson distributions $\mathrm{FP}(\lambda, \nu)$ for any $\lambda > 0$ and any probability measure ν on $\mathbb{R}$ were introduced in Section 2.5. They are preeminent examples of freely infinitely divisible distributions, as they are defined as weak limits of f.i.d. triangular arrays, cf. Definition 2.11. Equation 2.6 gives a general expression for the R-transform of any (compound) free Poisson; this formula is left as an exercise in [39]. It will be derived below as a consequence of our new perspective, with more insight into the nature of the "jump distribution". Indeed, let U be a random variable with $U \stackrel{\mathrm{d}}{=} \nu$; 2.6 then reads

$$R_{\mathrm{FP}(\lambda,\nu)}(z) = \lambda E\left[\frac{U}{1 - zU}\right]. \tag{5.8}$$

Note: the free infinite divisibility of $\mathrm{FP}(\lambda, \nu)$ follows immediately from (5.8): since the nth $\boxplus$-root of a distribution μ (should it exist) has R-transform $\frac{1}{n}R_\mu$, (5.8) shows that $\mathrm{FP}(\lambda/n, \nu)$ is the nth $\boxplus$-root of $\mathrm{FP}(\lambda, \nu)$ for each n.

Suppose that $U \in L^2$. Noting that $\frac{U}{1-zU} = U + \frac{zU^2}{1-zU}$, it then follows that

$$R_{\mathrm{FP}(\lambda,\nu)}(z) = \lambda E\left[U + \frac{zU^2}{1 - zU}\right] = \lambda E[U] + \lambda E[U^2] E\left[\frac{z}{1 - zU^\square}\right] \tag{5.9}$$

where the last equality is an application of the characterizing equation (1.4)for the square bias with test function $f(u) = \frac{z}{1-zu}$. Moreover, since the R-transform records the mean and variance as its orders 0 and 1 coefficients, (2.6) shows that the mean of $\mathrm{FP}(\lambda, \nu)$ is $\lambda E[U]$ and the variance of $\mathrm{FP}(\lambda, \nu)$ is $\lambda E[U^2]$. Note that,

for any random variable X and any $m \in \mathbb{R}$, $R_{X-m}(z) = R_X(z) + m$. As such, (5.9) shows the following.

Lemma 5.5. *For every L^2 random variable U, and every $m \in \mathbb{R}$ and $\sigma^2 \geq 0$, there is a translated compound free Poisson random variable X with mean m and variance σ^2 satisfying*

$$R_X(z) = m + \sigma^2 E\left[\frac{z}{1 - zU^{\square}}\right]. \tag{5.10}$$

Similarly, suppose instead that $U \in L^1$, $U \geq 0$, and $E[U] > 0$. Then the size bias U^s is well-defined by (1.2), and applying it with the test function $f(u) = \frac{1}{1-zu}$ yields

$$R_{\mathrm{FP}(\lambda,\nu)}(z) = \lambda E\left[\frac{U}{1 - zU}\right] = \lambda E[U] E\left[\frac{1}{1 - zU^s}\right]. \tag{5.11}$$

This yields the following result, complementing Lemma 5.5.

Lemma 5.6. *For every L^1 random variable $U \geq 0$ with $E[U] > 0$, and every $m \in \mathbb{R}$, there is a compound Poisson random variable $X \geq 0$ with mean m satisfying*

$$R_X(z) = mE\left[\frac{1}{1 - zU^s}\right]. \tag{5.12}$$

Remark 5.7. *The fact that the random variable X in Lemma 5.5 is non-negative follows by definition: here $X \stackrel{\mathrm{d}}{=} \mathrm{FP}(\lambda, \nu)$ where $\lambda = E[U]/m$ and $\nu = \mathcal{L}(U)$. Since ν is positively supported, all the summand distributions $(1 - \frac{\lambda}{n})\delta_0 + \frac{\lambda}{n}\nu$ in Definition 2.11 are positively supported, and hence so are their free convolution powers and the weak limit X.*

Now we turn Lemmas 5.5 and 5.6 around, thinking of $U^{\square}$ or U^s as the basic inputs to Equations (5.10) and (5.12) instead of U. Indeed, from Lemma 3.6(f), a random variable V is the square bias of some other random variable U, $V = U^{\square}$, if and only if $E[V^{-2}] < \infty$; similarly, from Lemma 3.9 parts f) and g), a random variable $V \geq 0$ is the size bias of some other L^1 random variable $U \geq 0$, $V = U^s$, if and only if $E[V^{-1}] < \infty$. This argument gives helpful intuition for the following theorem, which is the main result of this section.

Theorem 5.8. *Let V_0 be any random variable satisfying $E[V_0^{-2}] < \infty$, and let $m \in \mathbb{R}$ and $\sigma^2 \geq 0$. There is a (unique in law) freely infinitely divisible random variable X whose R-transform is*

$$R_X(z) = R_{(m,\sigma^2,V_0)}(z) := m + \sigma^2 E\left[\frac{z}{1 - zV_0}\right]. \tag{5.13}$$

Similarly, let V_+ be any non-negative random variable satisfying $E[V_+^{-1}] < \infty$. There is a (unique in law) positively freely infinitely divisible random variable X whose R-transform is

$$R_X(z) = R_{(m,V_+)}(z) := mE\left[\frac{1}{1 - zV_+}\right]. \tag{5.14}$$

(A reminder to the reader: *positively freely infinitely divisible* measures are characterized by having all convolution roots positively supported, cf. Definition 1.7.)

Although Theorem 5.8 essentially follows from the preceding lemmas, we will provide a different proof below which derives explicitly from the triangular array

construction of compound free Poisson distributions, and thus identifies the jump distribution $\mathrm{FP}(\lambda,\nu)$ probabilistically in terms of the free summands. The reader could then view our approach below, together with the calculations in (5.9) and (5.11) (read backwards) as alternate proofs of the characterizing equation (2.6) of the R-transform of $\mathrm{FP}(\lambda,\nu)$, at least in the L^2 or L^1 non-negative cases.

To proceed, we need several building blocks which hold in the general context of a sequence $\{X_n\}_{n\geq n_0}$, for some $n_0 \geq 1$, of row sums of triangular arrays:

$$X_n \stackrel{\mathrm{d}}{=} X_{1,n} + \cdots + X_{n,n} \quad \text{where } \{X_{j,n}\}_{1\leq j\leq n} \text{ are f.i.d.} \tag{5.15}$$

(Here and in the sequel "f.i.d." is short for *freely independent and identically distributed*.) The first lemma, though simple, will be used multiple times below.

Lemma 5.9. *Let* (5.15) *hold.*

(1) If $\lim_{n\to\infty} \mathrm{Var}(X_n)/n = 0$ *in* (5.15) *then* $\mathrm{Var}(X_{n,n}) \to 0$*, and if in addition* $E[X_n]/n \to 0$ *then* $X_{n,n} \rightharpoonup 0$.

(2) If the summands $X_{j,n}$ *are non-negative for* $n \geq n_0$ *and* $E[X_n]/n \to 0$ *then* $X_{n,n} \rightharpoonup 0$.

Proof. For (1), by the free independence and identical distribution assumption we have

$$\limsup_{n\to\infty} \mathrm{Var}(X_{n,n}) = \limsup_{n\to\infty} \mathrm{Var}(X_n)/n = 0,$$

and hence $X_{n,n} = (X_{n,n} - E[X_{n,n}]) + E[X_{n,n}] \to_p 0$ by Chebyshev's inequality and the fact that $E[X_{n,n}] = E[X_n]/n$, and the claim follows as convergence in probability and in distribution are equivalent when the limit is constant. For (2), by Markov's inequality, for $\epsilon > 0$ we have

$$P(X_{n,n} > \epsilon) \leq \frac{E[X_{n,n}]}{\epsilon} = E[X_n]/\epsilon n \to 0,$$

thus showing $X_{n,n}$ converges to zero in probability to zero, and hence also in distribution, as the limit is constant. □

Letting $X_n^{(n)} = X_n - X_{n,n}$, the free "replace one property" Theorem 3.18, notably in the form (3.23), says

$$G_{X_n^\circ}(z) = G_{X_{n,n}^\circ}(\omega_{X_{n,n},X_n^{(n)}}(z)). \tag{5.16}$$

Lemma 5.10. *Let* $\{X_n\}_{n\geq n_0}$ *be given as in* (5.15).

(1) If the summands $X_{j,n}$ *have finite variance, and along some subsequence* $\{n_k\}_{k\geq 1}$ *we have a random variable* V_0 *such that* $X_{n_k,n_k}^{\square} \rightharpoonup V_0$, $X_{n_k} \rightharpoonup X$, *and* $\mathrm{Var}(X_{n_k}) \to \mathrm{Var}(X)$, *then* (5.1)–(5.3) *hold for the pair* X, V_0.

(2) If the summands $X_{j,n}$ *are non-negative with finite strictly positive mean, and along some subsequence* $\{n_k\}_{k\geq 1}$ *we have a random variable* V_+ *such that* $X_{n_k,n_k}^{s} \rightharpoonup V_+$, $X_{n_k} \rightharpoonup X$, *and* $E[X_{n_k}] \to E[X] > 0$, *then* (5.5)–(5.7) *hold for the pair* X, V_+.

Proof. For convenience, we drop the subscript k from the subsequence. We prove (1) here; the proof of (2) is parallel, using (2) of Lemma 5.9(2).

As $\mathrm{Var}(X_n) \to \mathrm{Var}(X)$, the sequence $X_n, n \geq \lambda$ is uniformly integrable, and $E[X_n] \to E[X]$. Hence (1) of Lemma 5.9 obtains on the subsequence, and by the definition $X_n^{(n)} = X_n - X_{n,n}$ we have $X_n - X_n^{(n)} = X_{n,n} \rightharpoonup 0$.

As $X_n \rightharpoonup X$, we conclude that $X_n^{(n)} \rightharpoonup X$. Therefore, by Proposition 2.13, it follows that $\omega_{X_{n,n},X_n^{(n)}}(z) \to \omega_{0,X}(z)$ uniformly on compact subsets of $\mathbb{C}_+$ as $n \to \infty$. Since 0 is freely independent from any random variable X, by definition $\omega_{0,X}$ satisfies $G_X(z) = G_{0+X}(z) = G_0(\omega_{0,X}(z))$, and $G_0(z) = 1/z$, so we conclude that

$$\omega_{X_{n,n},X_n^{(n)}}(z) \to 1/G_X(z) = F_X(z)$$

uniformly on compact subsets of $\mathbb{C}_+$ as $n \to \infty$.

As $X_n^\circ \rightharpoonup X^\circ$ and $X_{n,n}^\square \rightharpoonup V_0$ by Lemma 3.12 and the hypotheses, we have $(W_n^\square)^\flat \rightharpoonup V_0^\flat$ by Lemma 3.4, and taking the limit in (5.16) along the (uniformly converging) subsequence yields (5.1), and (5.2) by the equivalence provided by Theorem 5.1. □

We now prove Theorem 5.8.

Proof of Theorem 5.8: We begin with the first conclusion, (5.13). By assumption $E[V_0^{-2}] < \infty$, and so V_0 is not identically 0; therefore $E[V_0^2] > 0$. From Lemma 3.6(f), we may take $U := V_0^{-\square}$, in which case $U^\square \stackrel{\mathrm{d}}{=} V_0$ and $E[U^2] = 1/E[V_0^2] > 0$.

Let $\lambda = \sigma^2/E[U^2]$, and for each n let $\{X_{j,n}\}_{1\le j\le n}$ be freely independent random variables each with law $(1-\frac{\lambda}{n})\delta_0 + \frac{\lambda}{n}\mathcal{L}(U)$, and define $X_n = X_{1,n} + \cdots + X_{n,n}$. Immediate calculation verifies that

$$E[X_{n,n}^k] = \frac{\lambda}{n}E[U^k], \qquad k \ge 1 \tag{5.17}$$

and hence, from the free independence of the $X_{j,n}$, we have

$$\mathrm{Var}(X_n) = \lambda E[U^2] - \frac{\lambda^2}{n}E[U]^2 = \sigma^2 - \frac{\sigma^4 E[U]^2}{n\,E[U^2]^2}. \tag{5.18}$$

If these variables were *classically* independent, it is known (from antiquity) that X_n converges in law (and all moments) to a compound Poisson with rate λ and jump distribution $\mathcal{L}(U)$. It therefore follows from [7, Theorem 3.4] that, in this free setting, $X_n \rightharpoonup M$ for some random variable M, and all moments of X_n converge to the corresponding moments of M. In particular, from (5.18) it follows that $\mathrm{Var}(M) = \sigma^2$.

Finally, from Lemma 3.6(b) we have that $X_{n,n}^\square \stackrel{\mathrm{d}}{=} U^\square = V_0$ for each $n \ge 1$, and hence this sequence certainly converges in distribution to V_0. We have thus confirmed all of the hypotheses of Lemma 5.10(1), and we may conclude that

$$R_M(z) = m' + \sigma^2 E\left[\frac{z}{1-zV_0}\right]$$

where $m' = \lambda E[U] = \sigma^2 E[U]/E[U^2]$. Note that $R_{M-m'}(z) = R_M(z) + m'$. Hence, with $M' = M - m'$ we have

$$R_{M'}(z) = \sigma^2 E\left[\frac{z}{1-zV_0}\right].$$

It follows that M' is freely infinitely divisible: as $\sigma^2 > 0$ was chosen at the beginning, we may now construct M_n' as above from V_0 and σ^2/n, and conclude that $R_{M_n'}(z) = \frac{1}{n}R_{M'}(z)$; this means that $\mathcal{L}(M_n')^{\boxplus n} = \mathcal{L}(M')$.

Finally, take $X = M' + m$. This random variable is also freely infinitely divisible, and satisfies (5.13), as desired.

The second claim, (5.14), can be shown in a parallel way, so we only provide the highlights. The random variable V_+ must be non-trivial, and by Lemma 3.9(f) $U := V_+^{-s}$ is L^1 and ≥ 0, in which case $U^s \stackrel{\mathrm{d}}{=} V_+$ and $E[U] = 1/E[V_+] > 0$. Let $\lambda = m/E[U]$, and for each n let $\{X_{j,n}\}_{1\leq j\leq n}$ and X_n be defined as for the previous case. Identity (5.17) continues to hold and we may again conclude that $X_n \rightharpoonup M$ for some M with mean m. From Lemma 3.9(b) we have that $X_{n,n}^s \stackrel{\mathrm{d}}{=} U^s = V_+$ for each $n \geq 1$, and have thus confirmed all of the hypotheses of Lemma 5.10(2), and we may conclude that

$$R_M(z) = mE\left[\frac{1}{1 - zV_+}\right]$$

where $m = \lambda E[U]$. It follows that M is freely infinitely divisible as we can follow this same construction with m replaced by m/n, similar to the case above.

□

Remark 5.11. *The following observation was used in the above proof, but is worth commenting on separately. Given any $R(z)$ of the form in (5.19), replacing m and σ^2 by m/t and σ^2/t respectively for any $t > 0$ yields an R transform via Proposition 5.12. Hence any such R transform corresponds to a variable for which a possibly non-integer t^{th} free convolution root exists. In particular, such random variables are always freely infinitely divisible.*

5.3. **Free Infinite Divisibility.** We now move beyond the compound free Poisson setting, which (as shown in the last section) corresponds to the case that the freely infinitely divisible random variable X has a negative second moment. We begin with Proposition 5.12 and 5.14 which show that the putative Lévy–Khintchine R-transform formulas (5.19) and (5.22) really do define R-transforms of freely infinitely divisible random variables in the L^2 or positive L^1 settings. This first step is accomplished by removing the restriction that $E[V_0^{-2}] < \infty$ in (5.13) in Theorem 5.8 by adapting the argument showing Lemma 1.1.3 in [28] that proves an analog of this result in the classical case by constructing a sequence of variables approximating V_0 whose inverse second moment exists.

Following that step, we prove Theorems 5.15 and 5.17, the converse direction, completing the proof of our main result: any freely infinitely divisible (L^2 or positive L^1) random variable X gives rise to another random variable (V_0 or V_+) satisfying the conditions of Theorem 5.1 or 5.3, arising as the limit of the square bias transforms of the free convolution roots of X.

Proposition 5.12. *For any random variable V_0, $m \in \mathbb{R}$ and $\sigma^2 \in [0, \infty)$, the function*

$$R(z) = m + \sigma^2 E\left[\frac{z}{1 - zV_0}\right] \tag{5.19}$$

is an R-transform that corresponds to a freely infinitely divisible X with mean m and variance σ^2.

Proof. As in the proof of Theorem 5.8, using the translation property of R-transforms when adding constants, it suffices to prove the result for $m = 0$.

If $\sigma^2 = 0$ then $R(z) = 0$, which is the R transform of δ_0, and the result holds trivially. Now take $\sigma^2 > 0$. When $\mathcal{L}(V_0) = \delta_0$, in view of Example 5.2, $R(z)$ is the R transform of the semi-circle distribution, which is shown there to be freely infinitely divisible.

Hence, we restrict to the case where $\sigma^2 > 0$ and $\mathcal{L}(V_0) \neq \delta_0$. For $k \geq 1$ let $D_k = \mathbb{R} \setminus [-c/k, c/k]$ where we take $c > 0$ small enough so that $P(V_0 \in D_1) > 0$. Set $V_{0,k} := \mathcal{L}(V_0 | V_0 \in D_k)$; in other words, the law of $V_{0,k}$ is determined by

$$E[f(V_{0,k})] = \frac{E[f(V_0)\mathbf{1}_{D_k}]}{P(V_0 \in D_k)}.$$

Similarly, define $\mathcal{L}(V_{0,\neg 0}) = \mathcal{L}(V_0 | V_0 \neq 0)$; the restrictions above and choice of c makes all of these well defined laws, and $V_{0,k} \rightharpoonup V_{0,\neg 0}$.

Note that $|V_{0,k}| \geq c/k$ with probability 1, and hence $E[V_{0,k}^{-2}] < \infty$. Let $\mathcal{L}(U_k) = \mathcal{L}(V_{0,k}^{-\boxed{}})$ as in Definition 3.5, so that $\mathcal{L}(U_k^{\boxed{}}) = \mathcal{L}(V_{0,k})$ by Lemma 3.6(f).

By Lemma 5.5, for any $\sigma^2 \geq 0$ there is a compound free Poisson random variable X_k whose R-transform satisfies

$$R_{X_k}(z) = \sigma^2 E\left[\frac{z}{1 - zU_k^{\boxed{}}}\right] = \sigma^2 E\left[\frac{z}{1 - zV_{0,k}}\right]. \tag{5.20}$$

Hence, for any $k \in \mathbb{N}$ and $\sigma^2 \geq 0$,

$$\varphi^{k,\sigma^2}(z) := \varphi_{X_k}(z) = R_{X_k}(1/z) = \sigma^2 E\left[\frac{1}{z - V_{0,k}}\right] = \sigma^2 G_{V_{0,k}}(z)$$

is a Voiculescu transform, whose domain manifestly has an analytic continuation to the whole complex plane, as it coincides on a truncated cone (an open set) with a scaled Cauchy transform. As $V_{0,k} \rightharpoonup V_{0,\neg 0}$ as $k \to \infty$, by Proposition 2.4, $G_{V_{0,k}}$ converges to $G_{V_{0,\neg 0}}$ uniformly on compact subsets of $\mathbb{C}_+$, and hence φ^{k,σ^2} converges uniformly on compact subsets of $\mathbb{C}_+$ to the analytic function $\varphi = \sigma^2 G_{V_{0,\neg 0}}$. What's more, $|\varphi^{k,\sigma^2}(z)| = \sigma^2 |G_{V_{0,k}}(z)| \leq \sigma^2/|\mathrm{Im}(z)|$ (the last inequality comes from the fact that $|G(z)| \leq 1/|\mathrm{Im}(z)|$ for any Cauchy transform G, as is easily verified using the triangle inequality in the definition (2.1)), and hence $\varphi^{k,\sigma^2}(z) = o(z)$ uniformly in k. Therefore, it follows from Proposition 2.8 that the limit φ is the Voiculescu transform of some probability measure, and so $R(z) = \varphi(1/z)$ is the R-transform of that probability measure.

Letting $\gamma = P(V_0 \neq 0)$ we have

$$\mathcal{L}(V_0) = \gamma \mathcal{L}(V_{0,\neg 0}) + (1 - \gamma)\delta_0,$$

and we may thus express

$$R(z) := \sigma^2 E\left[\frac{z}{1 - zV_0}\right] = \gamma \sigma^2 E\left[\frac{z}{1 - zV_{0,\neg 0}}\right] + (1 - \gamma)\sigma^2 z \tag{5.21}$$

as the sum of two R transforms, and thus conclude that $R(z)$ is an R transform for all V_0 and $\sigma^2 \geq 0$. The mean and variance can be read off directly from (5.19).

The fact that $R = R_X$ for a freely infinitely divisible random variable X now follows as in Remark 5.11. □

Remark 5.13. *Note that $R(z) = (1-\gamma)\sigma^2 z$ is the R-transform of a centered semicircular random variable. From (5.21) we see directly that the infinitely divisible distribution constructed from the arbitrary V_0 is the free convolution of a semicircle with a limit of compound free Poisson distributions. With the converse, Theorem 5.15, to Proposition prop:ForAnyV in hand, we see that* every *L^2 freely infinitely divisible distribution has such a decomposition.*

We similarly have the parallel result for the case of non-negative random variables with finite first moments, handled using size biasing.

Proposition 5.14. *For any non-negative random variable V_+ and $m \geq 0$ the function*

$$R(z) = mE\left[\frac{1}{1 - zV_+}\right] \tag{5.22}$$

is an R-transform that corresponds to a positively freely infinitely divisible X with mean m.

Proof. We may assume that $V_+ \neq \delta_0$ and $m \geq 0$, since in these cases we have the R-transform of a non-negative constant random variable which is evidently positively freely infinitely divisible. Take $D_k = [c/k, \infty)$ for $c > 0$ small enough that $P(V_+ \in D_1) > 0$. Let $\mathcal{L}(V_{+,k}) = \mathcal{L}(V_+|V_+ \in D_k)$, $\mathcal{L}(V_{+,\neg 0}) = \mathcal{L}(V_+|V_+ \neq 0)$ so that $V_{+,k} \rightharpoonup V_{+,\neg 0}$. Noting that $V_{+,k}$ satisfies the conditions of Definition 3.8, let $\mathcal{L}(U_k) = \mathcal{L}(V_{+,k}^{-s})$.

Applying Lemma 5.6 we obtain, for any $m > 0$, a compound free Poisson random variable Y_k whose R-transform satisfies

$$R_{Y_k}(z) = mE\left[\frac{1}{1 - zV_{+,k}}\right]$$

Since $V_{+,k}$ is supported in $[0, \infty)$ (in fact in $[c/k, \infty)$), it follows from Remark 5.7 that X_k is positively freely infinitely divisible. Now following exactly the same argument as the proof of Theorem 5.12 following (5.20), we conclude, since $V_{+,k} \rightharpoonup V_{+,\neg 0}$, that $R(z) = mE[\frac{1}{1-zV_{+,\neg 0}}]$ is the R-transform of a random variable Y which, as the weak limit of positively freely infinitely divisible random variables Y_k is itself positively freely infinitely divisible (as the reader may quickly verify).

Letting $\gamma = P(V_+ \neq 0)$ we have

$$\mathcal{L}(V_+) = \gamma \mathcal{L}(V_{+,\neg 0}) + (1 - \gamma)\delta_0,$$

and therefore we can express

$$R(z) := mE\left[\frac{1}{1 - zV_+}\right] = \gamma m E\left[\frac{1}{1 - zV_{+,\neg 0}}\right] + (1 - \gamma)m$$

as the sum of R transforms of two positively freely infinitely divisible random variables, which therefore yields the desires conclusion that $R = R_X$ for a positively freely infinitely divisible random variable X, whose mean is $R_X(0) = m$. □

Theorem 5.15. *Let X be a square integrable freely infinitely divisible random variable with mean 0 and variance $\sigma^2 \in (0, \infty)$. For $n \geq 1$, let $\mathcal{L}(X_{n,n})$ be the distribution of the n freely independent, identically distributed summands satisfying $\mathcal{L}(X_{n,n})^{\boxplus n} = \mathcal{L}(X)$. Then the weak limit V_0 of $X_{n,n}^{\square}$ exists and uniquely satisfies (5.3).*

Proof. Recall that $X_n^{(n)} = X - X_{n,n}$. The "replace one" property, Theorem 3.18, in this case says (cf. (5.16))

$$G_{X^\circ}(z) = G_{X_{n,n}^\circ}(\omega_{X_{n,n}, X_n^{(n)}}(z)).$$

For readability, set $\omega_n(z) = \omega_{X_{n,n}, X_n^{(n)}}(z)$. Using (3.4) defining $\flat$ and (3.13) defining $\circ$, and squaring both sides, this yields

$$\frac{1}{\sigma^2}(zG_X(z) - 1) = \frac{1}{\omega_n(z)} G_{X_{n,n}^{\square}}(\omega_n(z))$$

or equivalently

$$G_{X_{n,n}^{\square}}(\omega_n(z)) = \frac{\omega_n(z)}{\sigma^2}(zG_X(z) - 1). \tag{5.23}$$

From Proposition 2.12, ω_n is uniquely determined by the equation

$$F_{X_{n,n}}(\omega_n(z)) = F_{X_{n,n}+X_n^{(n)}}(z) = F_X(z).$$

From Proposition 2.7 applied in sequence to $F_{X_{n,n}}$ and F_X, for z in an appropriate truncated cone Γ this is equivalent to $\omega_n(z) = F_{X_{n,n}}^{-1} \circ F_X(z)$, and the range of this function contains another truncated cone Γ'. (These truncated cones a priori vary with n; since $X_{n,n} \rightharpoonup 0$ by Lemma 5.9, it follows from Proposition 2.8 that uniform truncated cones exist here.)

Let $w = F_{X_{n,n}}^{-1} \circ F_X(z)$; then for $z \in \Gamma'$ we equivalently have $z = F_X^{-1} \circ F_{X_{n,n}}(w)$. Thus, for $w \in \Gamma'$, (5.23) says

$$\begin{aligned} G_{X_{n,n}^{\square}}(w) &= \frac{w}{\sigma^2}\left(F_X^{-1} \circ F_{X_{n,n}}(w) \cdot G_X(F_X^{-1} \circ F_{X_{n,n}}(w)) - 1\right) \\ &= \frac{w}{\sigma^2}\left(\frac{F_X^{-1} \circ F_{X_{n,n}}(w)}{F_{X_{n,n}}(w)} - 1\right) \end{aligned}$$

where the second equality comes from the fact that $G_X \circ F_X^{-1}(\zeta) = 1/\zeta$. As $X_{n,n} \rightharpoonup 0$, $F_{X_{n,n}}(w)$ converges to $F_{\delta_0}(w) = w$ uniformly on compact subsets of $\mathbb{C}_+$ by Proposition 2.4(1). Hence

$$G_{X_{n,n}^{\square}}(w) \to \frac{w}{\sigma^2}\left(\frac{F_X^{-1}(w)}{w} - 1\right) = \frac{1}{\sigma^2}\left(F_X^{-1}(w) - w\right) = \frac{1}{\sigma^2}\varphi_X(w) \tag{5.24}$$

uniformly on compact subsets of Γ'.

Taking stock: we have now used the "replace one" property, and the definitions of the square bias and free zero bias transforms, to show that the Cauchy transform of $X_{n,n}^{\square}$ converges to the analytic function φ_X/σ^2 on some domain $\Gamma' \subseteq \mathbb{C}_+$. We would like to use Proposition 2.4(2) to complete the proof of the theorem that $X_{n,n}^{\square}$ converges in distribution to a probability measure. This requires two more steps: first we must show that the convergence in (5.24) occurs on all of $\mathbb{C}_+$, and then we must show that no mass is lost at infinity, that is, the resulting limit is a Cauchy transform with total mass 1.

For each n, $G_{X_{n,n}^{\square}}$ is analytic on all of $\mathbb{C}_+$; similarly, since X is freely infinitely divisible, φ_X is also analytic on all of $\mathbb{C}_+$ (cf. [8, Theorem 5.10]). Since Γ' is an open subset of $\mathbb{C}_+$, it now follows from Montel's theorem that $G_{X_{n,n}^{\square}}$ converges to φ_X/σ^2 on all of $\mathbb{C}_+$ (in fact uniformly on compact subsets). We therefore conclude by Proposition 2.4(2) that φ_X is the Cauchy transform of a finite measure ν with $\nu(\mathbb{R}) \leq 1$, and that the law of $X_{n,n}^{\square}$ converges vaguely to ν. It remains only to show that $\nu(\mathbb{R}) = 1$ (since a vaguely convergent sequence of probability measures whose limit is a probability measure converges in distribution). In other words, we merely need to show that $wG_\nu(w) = \frac{w}{\sigma^2}\varphi_X(w) \to 1$ as $w \to \infty$ along some ray in the upper half plane.

To motivate this last step, consider what happens in the special case that X is compactly supported. In this case, Speicher's theory of free cumulants [40] gives a Taylor series expansion $R_X(z) = \sum_{n\geq 0} \kappa_{n+1}[X]z^n$ where $\kappa_1[X] = E[X] = 0$ and $\kappa_2[X] = \mathrm{Var}[X] = \sigma^2$. Hence $\varphi_X(w) = \sigma^2/w + O(1/w^2)$, and thence it follows immediately that $\frac{w}{\sigma^2}\varphi_X(w) \to 0$ as $w \to \infty$. To extend this argument rigorously to the more generally setting where X need not be compactly supported, but only assumed to have two finite moments, we make use of some results due to Maassen [37].

F_X is the reciprocal Cauchy transform of a probability measure on $\mathbb{R}$ with finite variance σ^2 and mean 0. Therefore, by [37, Proposition 2.2(b)] it follows that $F_X(z) = z - \sigma^2 G_\rho(z)$ for some probability measure ρ on $\mathbb{R}$. Thus, for $w \in \Gamma$,

$$w = F_X(F_X^{-1}(w)) = F_X^{-1}(w) - \sigma^2 G_\rho(F_X^{-1}(w))$$

and hence

$$\varphi_X(w) = F_X^{-1}(w) - w = \sigma^2 G_\rho(F_X^{-1}(w)) = \sigma^2 G_\rho(w + \varphi_X(w)). \tag{5.25}$$

Multiplying both sides by $(w + \varphi_X(w))/\sigma^2$ yields

$$\frac{w}{\sigma^2}\varphi_X(w) + \frac{\varphi_X^2(w)}{\sigma^2} = (w + \varphi_X(w))G_\rho(w + \varphi_X(w)). \tag{5.26}$$

Now, again using only the fact that X has mean 0 and variance σ^2, [37, Lemma 2.4] yields that $|\varphi_X(w)| \leq 2\sigma^2/|\mathrm{Im}\, w|$. Thus, along any path in Γ with imaginary part tending to ∞, the left-hand-side of (5.26) has the same limit as $\frac{w}{\sigma^2}\varphi_X(w)$, while the right-hand-side tends to 1 (since G_ρ is the Cauchy transform of a probability measure). This concludes the proof that ν is a probability measure.

Let V_0 be a random variable whose law is ν. Hence, we have shown that $G_{X_{n,n}^{\boxplus}} \to G_{V_0}$ on $\mathbb{C}_+$. As established in (5.24), we also have $G_{X_{n,n}^{\boxplus}} \to \varphi_X/\sigma^2$ on $\mathbb{C}_+$, and hence, we have shown that $G_{V_0} = \varphi_X/\sigma^2$ on $\mathbb{C}_+$, establishing (5.3) (in the mean 0 case that is the context of this theorem). This concludes the proof. □

The calculations following (5.25) above give a useful way to compute the mean of any L^2 random variable from its Voiculescu transform.

Corollary 5.16. *If $X \in L^2$, then $\lim_{y\to\infty} \varphi_X(iy) = E[X]$.*

Proof. Since $\varphi_{X-E[X]}(z) = \varphi_X(z) - E[X]$, it suffices to prove the lemma in the case $E[X] = 0$. As shown just below (5.26), $|\varphi_X(iy)| \leq 2\mathrm{Var}[X]/y$ for $y > 0$, and hence for any probability measure ρ, $G_\rho(iy + \varphi_X(iy)) \to 0$ as $y \to \infty$. The result now follows from (5.25). □

We may now conclude the proof of our main Theorem 1.6.

Proof of Theorem 1.6: Theorem 5.15 shows that a) implies b) (by applying to the centered random variable $X - m$ and noting that $\varphi_{X-m}(z) = \varphi_X(z) - m$), Theorem 5.1 shows that b) implies c), and the equivalence of (5.1) and (5.2) provided by Theorem 5.1 and Proposition 5.12 show that c) implies a). □

We now present the 'size bias' version of Theorem 5.15.

Theorem 5.17. *Let X be a non-negative, positively freely infinitely divisible random variable with finite mean $m > 0$. For $n \geq 1$, let $\mathcal{L}(X_{n,n})$ be the distribution of the n freely independent, identically distributed summands satisfying $\mathcal{L}(X_{n,n})^{\boxplus n} = \mathcal{L}(X)$. Then the weak limit V_+ of $X_{n,n}^s$ exists and uniquely satisfies (5.7).*

Proof. The proof follows that of Theorem 5.15 with minor changes. To begin, let $\omega_n = \omega_{X_{n,n},X_n^{(n)}}$ be the subordinator used before, which is characterized by $F_{X_{n,n}}(\omega_n(z)) = F_X(z)$. The "replace one" property for size bias, Theorem (3.21), in the present case yields, by (3.26),

$$G_{X^s}(z) = G_{X^s_{n,n}}(\omega_n(z)).$$

Lemma 3.9(d) for the Cauchy transform of a size bias thus yields

$$G_{X^s_{n,n}}(\omega_n(z)) = \frac{zG_X(z) - 1}{m}. \tag{5.27}$$

But $G_X(z) = 1/F_X(z) = 1/F_{X_{n,n}}(\omega_n(z))$. Using the same (uniform over n) truncated cones Γ and Γ' from the proof of Theorem 5.15, we then express $\omega_n(z) = F^{-1}_{X_{n,n}} \circ F_X(z)$ for $z \in \Gamma$, and then express (5.27) as

$$G_{X^s_{n,n}}(w) = \frac{1}{m}\left(\frac{F_X^{-1} \circ F_{X_{n,n}}(w)}{F_{X_{n,n}}(w)} - 1\right)$$

for $w = \omega_n(z) \in \Gamma'$. As before, since $X_{n,n} \rightharpoonup 0$, $F_{X_{n,n}}(w) \to w$ uniformly on compact sets, and it follows just as before that

$$G_{X^s_{n,n}}(w) \to \frac{1}{m}\left(\frac{F_X^{-1}(w)}{w} - 1\right) = \frac{1}{mw}(F_X^{-1}(w) - w) = \frac{1}{mw}\varphi_X(w) \tag{5.28}$$

uniformly on compact subsets of Γ' as $n \to \infty$. The remainder of the proof after (5.28) follows mutatis mutandis from the proof of Theorem 5.15 following (5.24). □

Proof of Theorem 1.8: Theorem 5.17 shows that a) implies b), Theorem 5.3 shows that b) implies c), and the equivalence of (5.5) and (5.6) given by Theorem 5.3, and Proposition 5.14, show that c) implies a). □

Proof of Theorem 1.10: For the forward direction, suppose (1) holds, and let $X' = X + \alpha$ be a translation of X which is positively freely infinitely divisible, and further large enough that $E[X'] > 0$. We restrict to the case $\mathrm{Var}[X'] > 0$ (otherwise X and X' are constant and the result is straightforward). For such X', both Theorems 1.6 and 1.8 apply. Note that $V_0 = V_0(X) = V_0(X')$ by Remark 1.11. Let $V_+ = V_+(X')$. Since both (1.17) and (1.20) hold for X', we have

$$E[X'] + \mathrm{Var}[X']G_{V_0}(z) = E[X']zG_{V_+}(z)$$

and thence

$$z\left(zG_{V_+}(z) - 1\right) = \frac{\mathrm{Var}[X']}{E[X']}zG_{V_0}(z). \tag{5.29}$$

Applying Lemma 2.3 to the limit as $z \to \infty$ in (5.29), we find that $V_+ \in L^1$ with mean $E[V_+] = \mathrm{Var}[X']/E[X']$. We may now rewrite (5.29) to obtain

$$G_{V_0}(z) = \frac{zG_{V_+}(z) - 1}{E[V_+]},$$

and via Lemma 3.9(d), recognize that $V_0 = V_+^s$. In particular, $V_0 \geq 0$ and by Lemma 3.9(g) it possesses a finite inverse first moment as claimed, thus verifying condition (2) and the remaining conclusions of the theorem.

For the reverse direction, suppose (2) holds. Since V_0 is non-negative and has an inverse first moment, the distribution $V := V_0^{-s}$ exists and satisfies $V^s \stackrel{\mathrm{d}}{=} (V_0^{-s})^s \stackrel{\mathrm{d}}{=} V_0$ by Lemma 3.9(f). Noting that $V = V_0^{-s}$ cannot be trivial and hence $E[V]$

must be positive, by Lemma 3.9(d) the size bias relation between V and V_0 can be expressed in terms of Cauchy transforms as

$$\frac{zG_V(z)-1}{E[V]} = G_{V_0}(z).$$

Applying Theorem 1.6(b) to X now yields

$$\begin{aligned}\varphi_X(z) = E[X] + \mathrm{Var}[X]G_{V_0}(z) &= E[X] + \mathrm{Var}[X]\left(\frac{zG_V(z)-1}{E[V]}\right)\\ &= E[X] - \frac{\mathrm{Var}[X]}{E[V]} + \frac{\mathrm{Var}[X]}{E[V]}zG_V(z).\end{aligned}$$

Now, set $X' = X - E[X] + \mathrm{Var}[X]/E[V] + \alpha$ for some $\alpha > 0$ large enough that $X' \geq 0$ (which must exists since X is assumed to be bounded below). Then

$$\varphi_{X'}(z) = \alpha + \beta z G_V(z)$$

where $\beta = \mathrm{Var}[X]/E[V] > 0$.

Noting that the point mass δ_0 has Cauchy transform $G_{\delta_0}(z) = 1/z$, we reformulate this as

$$\varphi_{X'}(z) = \alpha z G_{\delta_0} + \beta z G_V(z) = (\alpha+\beta)z\left[\frac{\alpha}{\alpha+\beta}G_{\delta_0}(z) + \frac{\beta}{\alpha+\beta}G_V(z)\right]$$

and this convex combination inside the brackets is the Cauchy transform $G_{V_+}(z)$ of a random variable V_+ whose law is a mixture of δ_0 and $\mathcal{L}(V)$:

$$\mathcal{L}(V_+) = \frac{\alpha}{\alpha+\beta}\delta_0 + \frac{\beta}{\alpha+\beta}\mathcal{L}(V). \tag{5.30}$$

Hence, we conclude that

$$\varphi_{X'}(z) = (\alpha+\beta)zG_{V_+}(z) \tag{5.31}$$

which, by Theorem 1.8 together with the positivity of X', proves that X' is positively freely infinitely divisible. Since X' is a translate of X, this proves part (1) holds true.

Since V_+ is a mixture of V with a point mass at 0, by Lemma 3.9(b), $(V_+)^s \overset{\mathrm{d}}{=} V^s \overset{\mathrm{d}}{=} V_0$ as claimed. Finally, taking $z = iy$ and sending $y \to \infty$ in (5.31) shows, by Corollary 5.16, that $E[X'] = \alpha + \beta$. Hence (5.30) shows that $E[V_+] = \frac{\beta}{\alpha+\beta}E[V] = \beta E[V]/E[X']$. Since $\beta = \mathrm{Var}[X]/E[V]$, we have

$$E[V_+] = \frac{\beta}{\alpha+\beta}E[V] = \frac{\mathrm{Var}[X]}{\alpha+\beta} = \frac{\mathrm{Var}[X]}{E[X']} = \frac{\mathrm{Var}[X']}{E[X']}$$

yielding the final statement of the theorem, concluding the proof. □

Remark 5.18. *When X is supported on a compact set, we have the following probabilistic way of obtaining the correspondence of Theorem 1.10, that is,*

$$V_0 = \lim_{n\to\infty}(X_{n,n} - m/n)^{\square} = \lim_{n\to\infty} X_{n,n}^{\square} = \lim_{n\to\infty}(X_{n,n}^s)^s = V_+^s,$$

where the first inequality holds due to Theorem 5.15, the second by of Lemma 3.6(c), the third by the definition of $\square$, and the last by Theorem 5.17 and Lemma 3.9(c); compactness is only used to ensure the convergence in expectation when invoking Lemmas 3.6 and 3.9 in the second and final equality.

In addition, as $X_{n,n} \rightharpoonup 0$ *and* $E[X_{n,n}] \to 0$*, we have*

$$E[V_+] = \lim_{n\to\infty} E[X^s_{n,n}] = \lim_{n\to\infty} E[X^s_{n,n} - X_{n,n}]$$
$$= \lim_{n\to\infty} \left(\frac{E[X^2_{n,n}]}{E[X_{n,n}]} - E[X_{n,n}] \right) = \lim_{n\to\infty} \frac{\mathrm{Var}(X_{n,n})}{E[X_{n,n}]} = \frac{\mathrm{Var}[X]/n}{E[X]/n},$$

where we have used the size bias relation (1.2) *for the third equality.*

5.4. **The Compactly-Supported Case.** The proofs of Theorems 5.15 and 5.17 use [8, Theorem 5.10]: the fact that the Voiculescu transform φ_X of a freely-infinitely divisible random variable X has an analytic continuation to the whole upper half plane; the continuation is given by the Lévy–Khintchine formula itself, as in the statement of Theorem 1.6(b). We can avoid requiring a priori knowledge of this important fact, if we restrict to the case that X is compactly-supported. To do so, we need the following ancillary result, which is of some independent interest, and may well be known in the literature but not in literature known to the authors.

Lemma 5.19. *Let* μ *be a compactly-supported probability measure on* $\mathbb{R}$*. There is a compact interval* J *that contains the supports of all* $\boxplus$*-convolution roots of* μ*. That is: for any* $n \in \mathbb{N}$*, if* ν *is a probability measure satisfying* $\nu^{\boxplus n} = \mu$*, then* ν *is compactly supported with* $\operatorname{supp} \nu \subseteq J$*.*

Proof. To begin, note that we may safely assume μ is centered (which implies that ν is centered); otherwise we can center it by translating by the mean, which distributes to translating ν by its mean ($1/n$ times the mean of μ), then apply the result in the centered case and translate the intervals back.

So, let μ be a centered probability distribution, with support contained in a compact interval $[-R, R]$, and let σ^2 denote its variance. For $n \geq 1$, since $\mu = \nu \boxplus \nu^{\boxplus(n-1)}$, we have the subordination relation $F_\mu(z) = F_\nu(\omega_n(z))$ where ω_n is the subordinator from ν to $\nu^{\boxplus(n-1)}$. In [6, Theorem 2.5(3)], the authors show that this subordinator has an elegant form:

$$\omega_n(z) = \frac{z}{n} + \left(1 - \frac{1}{n}\right) F_\mu(z). \tag{5.32}$$

(See also [47, Section 2.3] for a concise presentation.) For $n \geq 2$, $F_\mu = F_{\nu^{\boxplus n}}$ has a continuous extension to $\mathbb{C}_+ \cup \mathbb{R}$ [6, Theorem 2.5(4)], so we may discuss the behavior of F_μ and F_ν on $\mathbb{R}$. From the Stieltjes inversion formula, F_μ is real-valued on $\mathbb{R} \setminus [-R, R]$.

By [37, Proposition 2.2(b)], since F_μ is the reciprocal Cauchy transform of a centered measure with variance $\sigma^2 < \infty$, there is a probability measure ρ on $\mathbb{R}$ such that

$$F_\mu(z) = z - \sigma^2 G_\rho(z), \qquad z \in \mathbb{C}_+$$

and the continuity of F_μ on $\mathbb{C}_+ \cup \mathbb{R}$ extends this relation similarly. Note that F_μ is real-valued on $\mathbb{R}$ prescisely where G_ρ is, and thus ρ is also supported in $[-R, R]$. Equation 5.32 therefore becomes

$$\omega_n(z) = z - \left(1 - \frac{1}{n}\right) \sigma^2 G_\rho(z), \qquad z \in \mathbb{C}_+ \cup \mathbb{R}.$$

Outside $[-R, R]$, the function G_ρ is analytic, and for $x > R$,

$$|G'_\rho(x)| = \int_{-R}^{R} \frac{\rho(dt)}{(x-t)^2} \leq \frac{1}{(x-R)^2}.$$

In particular $G'_\rho(x) \to 0$ as $x \to \infty$, and so for $x > R$,

$$|\omega_n(x) - x| = \left(1 - \frac{1}{n}\right)\sigma^2 \left| \int_x^\infty G'_\rho(y)\, dy \right| \leq \sigma^2 \int_x^\infty \frac{1}{(y-R)^2}\, dy = \frac{\sigma^2}{x-R}.$$

A similar analysis for $x < -R$ shows that $|\omega_n(x) - x| \leq \sigma^2/|x+R|$ there.

In particular, if $x > R + \sigma^2$, $|\omega_n(x) - x| < 1$. Since ω_n is continuous, it follows that its image on $[R+\sigma^2, \infty)$ contains $[R+\sigma^2+1, \infty)$ (and derivative calculations show that it is one-to-one there as well).

Hence every point $y \in [R + \sigma^2 + 1, \infty)$ has the form $y = \omega_n(x)$ for some $x \in [R+\sigma^2, \infty)$ with $|y - x| < 1$, and so $F_\mu(x) = F_\nu(\omega_n(x)) = F_\nu(y)$ shows that $F_\nu(y)$ is real since F_μ is real valued off $[-R, R]$. A similar argument shows that F_ν is real-valued on $[-\infty, -R - \sigma^2 - 1)$. This shows that the support of ν is contained in $[-R - \sigma^2 - 1, R + \sigma^2 + 1]$, as desired. □

With this lemma in hand, we can provide a much simpler proof of Theorems 5.15 and 5.17 in the case of compact support. We focus on Theorem 5.15; the alternate treatment Theorem 5.17 is entirely analogous.

Proof of Theorem 5.15 *in the case that* X *has compact support*: Since X has compact support, by Lemma 5.19, the support of the summand distributions $X_{n,n}, n \geq 1$ in (5.15), with X replacing X_n, are contained in a compact interval J. As the distribution of $X^{\square}_{n,n}$ is absolutely continuous with respect to that of $X_{n,n}$ by (1.3), the supports of $X^{\square}_{n,n}, n \geq 1$ are also contained in J. Any sequence of probability measures supported in the fixed compact interval J is tight; hence every subsequence of $\mathcal{L}(X^{\square}_{n,n})$ has a limit $\mathcal{L}(V_0)$ (which a priori depends on the subsequence). As X_n in (5.15) equals X for all $n \geq 1$, the remaining hypotheses of Lemma 5.10 hold trivially, and invoking the lemma shows that (5.3) holds for the pair X, V_0, and Theorem 5.1 shows V_0 is unique. Therefore, each subsequence has a convergent subsequence and all subsequential limits are the same V_0; it follows the full sequence must also converge to V_0. □

6. New Examples of freely infinitely divisible Distributions

In this final section, we use Theorems 1.6 and 1.8 to (computationally effectively) produce previously unknown examples of freely infinitely divisible distributions. We begin with our new formulation — Theorem 1.6(c) — of Bercovici–Voiculescu's free Lévy–Khintchine formula, which categorizes the laws of all L^2 freely infinitely divisible random variables in terms of their free Lévy–Khintchine measure V_0 and their free zero bias X°.

6.1. **Using the Free Zero Bias.** Proposition 1.9 shows that for every law $\mathcal{L}(V_0)$, every $m \in \mathbb{R}$ and every $\sigma^2 \in (0, \infty)$ there exists a freely infinitely divisible X with whose Voiculescu transform is given by (1.17) (i.e. the free Lévy–Khintchine formula). Our main Theorem 1.6 gives an equivalent characterization in terms of the free zero bias, (1.18). Indeed, from the definition (1.7) of the free zero bias,

noting that the contribution of the mean is simply a translation of the measure, (1.18) yields

$$G_{(X-m)^\circ}(z+m) = G_{(X-m)^\circ+m}(z) = G_{X^\circ}(z) \\ = G_{V_0^\flat}(1/G_X(z)) = G_{V_0^\flat}(1/G_{X-m}(z+m)).$$

Squaring both sides and using the definition (3.4) of the El Gordo transform then gives the equality

$$\frac{1}{\sigma^2}\left((z+m)G_{X-m}(z+m)-1\right) = G_{X-m}(z+m)G_{V_0}(1/G_{X-m}(z+m)).$$

and now noting that $G_{X-m}(z+m) = G_X(z)$, we have for each V_0 and σ^2 a one-parameter family of solutions for $G_X(z)$ that are translates by m:

$$(z+m)G_X(z) - 1 = \sigma^2\, G_X(z)G_{V_0}(1/G_X(z)). \tag{6.1}$$

We will use (6.1) in the examples below, and restrict to the $m = 0$ case, all other solutions being translations.

Remark 6.1. *The free Lévy–Khintchine formula (1.17) shows that the measure $\mathcal{L}(V_0)$ is symmetric iff $\mathcal{L}(X-m)$ is symmetric. Indeed, $\mathcal{L}(X)$ is symmetric if and only if, for $z \in \mathbb{C}\setminus\mathbb{R}$*

$$G_X(-z) = E\left[\frac{1}{-z-X}\right] = -E\left[\frac{1}{z+X}\right] = -E\left[\frac{1}{z-X}\right] = -G_X(z)$$

where the penultimate equality follows from the assumption $\mathcal{L}(X) = \mathcal{L}(-X)$; the reverse implication follows from the Stieltjes inversion formula, cf. Proposition 2.4. Reciprocating, the same holds for F_X, and using (2.3), it follows that $\mathcal{L}(X)$ is symmetric iff $\varphi_X(-z) = -\varphi_X(z)$ on its domain. Since $\varphi_X(z) - m = \varphi_{X-m}(z)$, the equivalence of symmetry for $X-m$ and V_0 follows from (1.17). This equivalence will be useful in several examples below.

Example 6.2. *Let V_0 have the centered semicircle distribution with variance t, $G_{V_0}(z) = \frac{1}{2t}(z - \sqrt{z^2-4t})$. From (6.1), we find that the associated freely infinitely divisible random variable X with mean 0 and variance σ^2 satisfies*

$$zG_X(z) - 1 = \sigma^2 G_X(z)\frac{1}{2t}\left(\frac{1}{G_X(z)} - \sqrt{\frac{1}{G_X^2(z)} - 4t}\right) \\ = \frac{\sigma^2}{2t}\left(1 - G_X(z)\sqrt{\frac{1}{G_X^2(z)} - 4t}\right).$$

Isolating the square root on the right hand side and then squaring both sides, we obtain the quadratic equation

$$\left(\frac{2t}{\sigma^2}(zG_X(z)-1) - 1\right)^2 = 1 - 4tG_X^2(z),$$

which simplifies to

$$(tz^2+\sigma^4)G_X^2(z) - (2t+\sigma^2)zG_X(z) + \sigma^2 + t = 0. \tag{6.2}$$

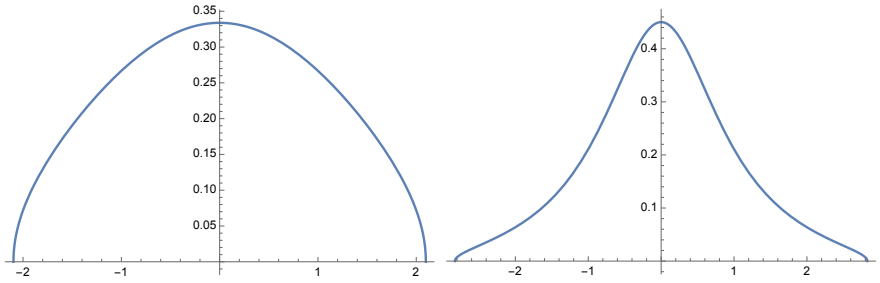

FIGURE 4. The density (6.3) of the freely infinitely divisible law with mean 0, variance 1, and free Lévy–Khintchine measure $\mathcal{L}(V_0)$ a semicircle of variance $t = 0.1$ on the left and $t = 1$ on the right.

Solving, and choosing the negative square root to produce the required cancellation (to yield $\lim_{z\to\infty} zG_X(z) = 1$*), we obtain*

$$\begin{aligned} G_X(z) &= \frac{(2t+\sigma^2)z - \sqrt{(2t+\sigma^2)^2 z^2 - 4(\sigma^2+t)(tz^2+\sigma^4)}}{2(tz^2+\sigma^4)} \\ &= \frac{(2t+\sigma^2)z - \sigma^2\sqrt{z^2 - 4(\sigma^2+t)}}{2(tz^2+\sigma^4)}. \end{aligned}$$

As te denominator is strictly positive for all $z = x \in \mathbb{R}$*, this function has a continuous extension to* $\mathbb{C}_+ \cup \mathbb{R}$*; taking* $z = x \in \mathbb{R}$ *we see that* $G_X(x)$ *is purely real if* $x^2 \geq 4(\sigma^2 + t)$*, and hence the density is given by*

$$\varrho_X(x) = -\frac{1}{\pi}\lim_{y\downarrow 0} G_X(x+iy) = \frac{\sigma^2\sqrt{4(\sigma^2+t) - x^2}}{2\pi(tx^2+\sigma^4)}\mathbb{1}_{|x|\leq 2\sqrt{\sigma^2+t}}. \tag{6.3}$$

Note that we recover the semicircle density as the law of X *when* $t = 0$*, i.e. when the free Lévy–Khintchine measure* $\mathcal{L}(V_0) = \delta_0$*. Indeed, in this case Equation (6.2) reduces to* $\sigma^2 G_X^2(z) - G_X(z) + 1 = 0$ *which is the standard quadratic equation for the Cauchy transform of the semicircle law. This aligns with Example 5.2.*

Example 6.3. *Let* $V_0 \overset{\mathrm{d}}{=} \frac{1}{2}(\delta_1 + \delta_{-1})$*, a Rademacher distribution. Here*

$$G_{V_0}(w) = \frac{1}{2}\left(\frac{1}{w-1} + \frac{1}{w+1}\right) = \frac{w}{w^2-1}.$$

In this case (6.1) becomes

$$zG_X(z) - 1 = \sigma^2 G_X(z) \cdot \frac{1/G_X(z)}{(1/G_X(z))^2 - 1}$$

which simplifies to the cubic equation

$$zG_X^3(z) + (\sigma^2 - 1)G_X^2(z) - zG_X(z) + 1 = 0. \tag{6.4}$$

While analytic expressions for the general solution to this cubic are tractable, they become much simpler in the case $\sigma^2 = 1$*, where (6.4) becomes the depressed cubic equation*

$$G_X(z)^3 - G_X(z) + \frac{1}{z} = 0. \tag{6.5}$$

For the general depressed cubic equation $w^3+pw+q=0$ with $p,q\in\mathbb{C}$, the three solutions $w_0, w_{\pm 1}$ can be expresses as follows. Let $\xi = e^{2\pi i/3} = \frac{-1+\sqrt{3}i}{2}$ be the cube root of unity in $\mathbb{C}_+$; then

$$w_k = \xi^k r - \frac{p}{3\xi^k r}, \quad \text{where} \quad r = \sqrt[3]{-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}. \tag{6.6}$$

Here the square and cube roots have branch cuts along the positive real axis (i.e. $(re^{i\theta})^{1/n} = r^{1/n}e^{i\theta/n}$ where $\theta\in[0,2\pi)$). For our case of interest (6.5), $p=-1$ and $q=1/z$, so

$$r = r(z) = \sqrt[3]{-\frac{1}{2z}+\sqrt{\frac{1}{4z^2}-\frac{1}{27}}} =: \sqrt[3]{-a(z)+b(z)}. \tag{6.7}$$

To determine which of the three roots is the Cauchy transform we seek, we look at asymptotics: using Newton's Binomial theorem (twice), we compute that

$$r(z) = -\sqrt{3}i + \frac{3}{2}\cdot\frac{1}{z} + O\left(\frac{1}{z^2}\right) \quad \text{and} \quad \frac{1}{r(z)} = \frac{i}{\sqrt{3}} + \frac{1}{2}\cdot\frac{1}{z} + O\left(\frac{1}{z^2}\right).$$

Hence, (6.6) yields

$$w_k = w_k(z) = (\xi^k - \xi^{-k})\frac{i}{\sqrt{3}} + (\xi^k+\xi^{-k})\frac{1}{2z} + O\left(\frac{1}{z^2}\right).$$

Note that if $k=\pm 1$ then $\xi^k \neq \xi^{-k}$, and so $\lim_{z\to\infty} w_k(z) = (\xi^k-\xi^{-k}) \neq 0$, hence w_k is not a Cauchy transform. On the other hand, we see that $w_0(z) = \frac{1}{z}+O(1/z^2)$, exhibiting the correct asymptotic behavior for a Cauchy transform of a probability measure. Since we know that one of the roots is $G_X(z)$*, we conclude that it is $w_0(z)$, i.e.*

$$G_X(z) = r(z) + \frac{1}{3r(z)} \tag{6.8}$$

with $r(z)$ as given in (6.7).

Note that the quantity inside the cube root in $r(z)$ is never 0; as such, the function r has a continuous extension to $\mathbb{C}_+\cup(\mathbb{R}\setminus\{0\})$. Thus $\mathcal{L}(X)$ is a combination of a density and potentially a point mass at 0.

Referring to (6.7),

$$r(x) = \sqrt[3]{-a(x)+b(x)} = \sqrt[3]{-(a(x)-b(x))}.$$

Take $0<x\leq\frac{3}{2}\sqrt{3}$, in which case $1/4x^2-1/27\geq 0$ and $a(x)-b(x)>0$. Given the choice of branch cut for the cube root, for a positive number δ, $\sqrt[3]{-\delta} = e^{i\pi/3}\sqrt[3]{\delta} = \frac{1+\sqrt{3}i}{2}\sqrt[3]{\delta}$ and hence

$$-\frac{1}{\pi}\operatorname{Im} r(x) = -\frac{\sqrt{3}}{2\pi}\sqrt[3]{a(x)-b(x)}. \tag{6.9}$$

An analogous argument applied to $1/r(x)$, noting that $e^{-i\pi/3} = \frac{1-\sqrt{3}i}{2}$, yields

$$-\frac{1}{\pi}\operatorname{Im}\frac{1}{3r(x)} = \frac{\sqrt{3}}{2\pi}\cdot\frac{1}{3}\sqrt[-3]{a(x)-b(x)}, \qquad 0<x\leq\frac{3}{2}\sqrt{3}. \tag{6.10}$$

Notice that, in this interval,

$$\sqrt[3]{a(x)-b(x)}\sqrt[3]{a(x)+b(x)} = \sqrt{a^2(x)-b^2(x)} = \frac{1}{3}. \tag{6.11}$$

Hence

$$\frac{1}{3}\,{}^{-3}\sqrt{a(x)+b(x)} = \sqrt[3]{a(x)-b(x)}.$$

In light of (6.8)*, we can thence combine* (6.9) *and* (6.10)*, and apply the symmetry principle of Remark 6.1, to yield the density*

$$\varrho_X(x) = \frac{\sqrt{3}}{2\pi}\left(\sqrt[3]{\frac{1}{2|x|}+\sqrt{\frac{1}{4x^2}-\frac{1}{27}}} - \sqrt[3]{\frac{1}{2|x|}-\sqrt{\frac{1}{4x^2}-\frac{1}{27}}}\right) \tag{6.12}$$

on the punctured interval $[-\frac{3}{2}\sqrt{3},\frac{3}{2}\sqrt{3}]\setminus\{0\}$.

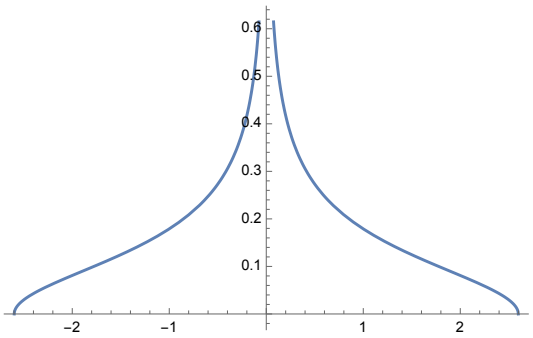

FIGURE 5. The density in (6.12) of the ⊞-infinitely divisible 'Azadi Tower' distribution with Rademacher *Free*− Lévy measure.

For $|x| > \frac{3}{2}\sqrt{3}$, $1/4x^2 - 1/27 < 0$ *and so*

$$r(x) = \sqrt[3]{-\frac{1}{2x}+i\sqrt{\frac{1}{27}-\frac{1}{4x^2}}}.$$

An analogous calculation to (6.11)*, relying on the fact that* $\sqrt[3]{\alpha\beta} = \sqrt[3]{\alpha}\sqrt[3]{\beta}$ *and* $\sqrt[3]{\overline{\alpha}} = \overline{\sqrt[3]{\alpha}}$ *for* $\alpha,\beta\in\mathbb{C}_+$*, shows that*

$$\frac{1}{3r(x)} = \sqrt[3]{-\frac{1}{2x}-i\sqrt{\frac{1}{27}-\frac{1}{4x^2}}} = \overline{r(x)}.$$

Hence, for $|x| > \frac{3}{2}\sqrt{3}$, $G_X(x) = r(x) + 1/3r(x) = r(x) + \overline{r(x)}$ *is real-valued, and therefore* $\mathcal{L}(X)$ *has no mass there.*

Finally: A direct calculation (with the aid of Mathematica) shows that the density (6.12) *integrates to* 1 *on the interval* $[-\frac{3}{2}\sqrt{3},\frac{3}{2}\sqrt{3}]$*, which shows that there is no mass at* 0 *and hence we conclude that* (6.12) *is the density of* X*, supported on the full interval* $[-\frac{3}{2}\sqrt{3},\frac{3}{2}\sqrt{3}]$.

Before proceeding to the next example, it will be useful to record the formula for the complex square root, in the next (long-known) lemma.

Lemma 6.4. *Let $z \mapsto \sqrt{z}$ denote the square root with branch cut along the positive real line:*

$$\sqrt{re^{i\theta}} = r^{1/2}e^{i\theta/2}, \qquad r \geq 0, \;\; \theta \in [0, 2\pi).$$

For any complex number $z = c + di$ with $c, d \in \mathbb{R}$,

$$\sqrt{c+di} = \operatorname{sign}(d)\sqrt{\frac{\sqrt{c^2+d^2}+c}{2}} + i\sqrt{\frac{\sqrt{c^2+d^2}-c}{2}} \tag{6.13}$$

where $\operatorname{sign}(d) = d/|d|$ when $d \neq 0$, and $\operatorname{sign}(0) = 1$.

Note: (6.13) may disagree with formulas found in textbooks (and Wikipedia), where the $\operatorname{sign}(d)$ term appears with the imaginary part instead of the real part. That alternate formula is correct for a different branch cut, along the *negative* real line, i.e. where the polar variable is taken with $\theta \in [-\pi, \pi)$. Different sources call either that square root or this one the "principal branch" confusingly; we stick to this one, cut along the *positive* real line, throughout this paper.

Example 6.5. *Let V_0 have the Cauchy distribution, with density $\varrho_{V_0}(x) = \frac{1}{\pi(1+x^2)}$, cf. Example 2.5. Thus $G_{V_0}(z) = \frac{1}{z+i}$. The corresponding centered, variance σ^2 freely infinitely divisible distribution $\mathcal{L}(X)$ then satisfies*

$$zG_X(z) - 1 = \sigma^2 G_X(z)\frac{1}{1/G_X(z) + i}$$

cf. (6.1). The solution is

$$G_X(z) = \frac{z - i - \sqrt{(z+i)^2 - 4\sigma^2}}{2(\sigma^2 - iz)} \tag{6.14}$$

where the square root is the principal branch, and the negative sign is the correct root to yield $O(1/z)$ asymptotic behavior at infinity.

To compute the density, we realize the denominator,

$$\begin{aligned} G_X(z) &= \frac{z - i - \sqrt{(z+i)^2 - 4\sigma^2}}{2(\sigma^2 - iz)} \frac{\sigma^2 + i\bar{z}}{\sigma^2 + i\bar{z}} \\ &= \frac{(z - i - \sqrt{(z+i)^2 - 4\sigma^2})(\sigma^2 + i\bar{z})}{2(\sigma^4 + |z|^2)}. \end{aligned} \tag{6.15}$$

Taking the imaginary part, expanding the numerator, expressing $z = x + iy$ and letting $a(z) = \operatorname{Re}\sqrt{(z+i)^2 - 4\sigma^2}$ and $b(z) = \operatorname{Im}\sqrt{(z+i)^2 - 4\sigma^2}$ we obtain

$$\begin{aligned} &\operatorname{Im}\left(i|z|^2 - \bar{z} + \sigma^2 z - i\sigma^2 - (\sigma^2 + i\bar{z})\sqrt{(z+i)^2 - 4\sigma^2}\right) \\ &\quad = x^2 + y^2 + (1+\sigma^2)y - \sigma^2 - \operatorname{Im}\left[(\sigma^2 + y + ix)(a(z) + ib(z))\right] \\ &\quad = x^2 + y^2 + (1+\sigma^2)y - \sigma^2 - xa(z) - (\sigma^2 + y)b(z) \end{aligned}$$

Substituting into (6.15) gives

$$\operatorname{Im}(G_X(z)) = \frac{x^2 + y^2 + (1+\sigma^2)y - \sigma^2 - xa(z) - (\sigma^2 + y)b(z)}{2(\sigma^4 + x^2 + y^2)}. \tag{6.16}$$

To compute the functions $a(z)$ and $b(z)$, we use the formula (6.13) for the real and imaginary part of the complex square root; in this case $c + di = (z+i)^2 - 4\sigma^2 =$

$x^2-(y+1)^2-4\sigma^2+2x(y+1)\,i$. *As* $y>0$, $\operatorname{sign}(d)=\operatorname{sign}(x)$. *Hence*

$$\sqrt{2}\,a(z)=\operatorname{sign}(x)\sqrt{\sqrt{(x^2-(y+1)^2-4\sigma^2)^2+4x^2(y+1)^2}+x^2-(y+1)^2-4\sigma^2}$$
$$\sqrt{2}\,b(z)=\sqrt{\sqrt{(x^2-(y+1)^2-4\sigma^2)^2+4x^2(y+1)^2}-x^2+(y+1)^2+4\sigma^2}$$

These expressions, and the other terms in (6.16)*, all have pointwise limits as* $y\downarrow 0$*, and so combining and using the symmetry principle of Remark 6.1 to restrict to* $x>0$ *and then replace* x *by* $|x|$*, we see that the density of* X *is*

$$\varrho_X(x)=\frac{1}{2\pi}\,\frac{|x|a_0(x)+\sigma^2(b_0(x)+1)-x^2}{\sigma^4+x^2} \tag{6.17}$$

where

$$\sqrt{2}\,a_0(x)=\sqrt{\sqrt{(x^2-1-4\sigma^2)^2+4x^2}+x^2-1-4\sigma^2}$$
$$\sqrt{2}\,b_0(x)=\sqrt{\sqrt{(x^2-1-4\sigma^2)^2+4x^2}-x^2+1+4\sigma^2}$$

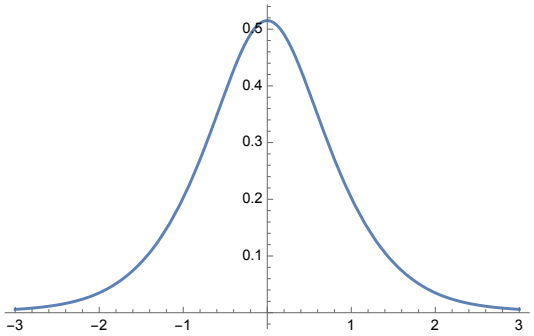

FIGURE 6. The density in (6.17) of the mean 0, variance $\sigma^2=1$, freely infinitely divisible distribution with Cauchy *Free*− Lévy measure.

Note that, like the Cauchy distribution itself, this freely infinitely divisible law has a strictly positive density on all of $\mathbb{R}$*. The law of* V_0 *(the Cauchy distribution) has no finite integer moments; nevertheless, it is the free Lévy–Khintchine measure of the above freely infinitely divisible distribution, which has (exactly) two finite integer moments:* ϱ_X *is continuous and bounded, and* $\lim_{x\to\infty} x^4\cdot\varrho_X(x)=\sigma^2/\pi$.

6.2. **Using the Size Bias.** In this final subsection we produce examples of *positively* freely infinitely divisible distributions using Theorem 1.8(c). Namely, by Proposition 5.14, any non-negative random variable $V_+\geq 0$ is associated to a one-parameter family of positively freely infinitely divisible distributions $\mathcal{L}(X)$ as in (1.21), parametrized by the mean $m>0$ of X. Unpacking (1.21) using Lemma 3.9(d) for the action of size bias on the Cauchy transform, we have

$$zG_X(z)-1=m\,G_{V_+}(1/G_X(z)). \tag{6.18}$$

Hence, specifying V_+ we have an implicit equation for the Cauchy transform of X which, in many cases, can be solved explicitly to determine the distribution of X.

Example 6.6. *In the simplest possible case, take* $\mathcal{L}(V_+) = \delta_\alpha$ *to be a point-mass at some point* $\alpha \geq 0$*. Then* $G_{V_+}(z) = \frac{1}{z-\alpha}$*, and* (6.18) *becomes*

$$zG_X(z) - 1 = m\frac{1}{1/G_X(z) - \alpha} = \frac{mG_X(z)}{1 - \alpha G_X(z)}.$$

Dividing through by $G_X(z)$ *and rearranging, this yields*

$$z = \frac{m}{1 - \alpha G_X(z)} + \frac{1}{G_X(z)}. \tag{6.19}$$

This yields a quadratic equation for G_X *except in the case* $\alpha = 0$ *where it reduces to* $G_X(z) = \frac{1}{z-m}$ *meaning* $G_X(z) = \frac{1}{z-m}$*: i.e. if* $V_+ = 0$ *then* X *is constant (equal to its mean). Otherwise, comparing* (6.19) *to* (2.8)*, we see that* $X \overset{\mathrm{d}}{=} \mathrm{FP}(m/\alpha, \alpha)$ *has a free Poisson distribution. Indeed, this is consistent with Lemma 5.6, taking the jump distribution* $U \overset{\mathrm{d}}{=} \delta_\alpha$*, and noting that for a single positive point mass* $U^s = U$*.*

For our final example, we consider "doubling down" on the free Poisson that appears as the law of X when $V_+(X)$ is a constant, and instead consider the distribution X when the law of $V_+(X)$ is a free Poisson $\mathrm{FP}(\lambda, \alpha)$. This yields a new two-parameter family of positively freely infinitely divisible distributions with interesting behavior as the parameters vary.

Example 6.7. *Let* $\lambda > 0$ *and* $\alpha > 0$*, and take* $V_+ \overset{\mathrm{d}}{=} \mathrm{FP}(\lambda, \alpha)$ *have the free Poisson distribution, whose Cauchy transform is given in* (2.9)*:*

$$G_{V_+}(z) = \frac{z + (1-\lambda)\alpha - \sqrt{(z - (1+\lambda)\alpha)^2 - 4\alpha^2\lambda}}{2\alpha z}.$$

In this case, (6.18) *yields*

$$\frac{zG_X(z) - 1}{m} = \frac{1/G_X(z) + (1-\lambda)\alpha - \sqrt{(1/G_X(z) - (1+\lambda)\alpha)^2 - 4\alpha^2\lambda}}{2\alpha/G_X(z)}$$

resulting in the quadratic equation

$$\alpha z\left[z - (1-\lambda)m\right] G_X^2(z) + \left[-(2\alpha + m)z + m((1-\lambda)\alpha + m)\right] G_X(z) + (\alpha + m) = 0.$$

Applying the quadratic formula yields

$$G_X(z) = \frac{(2\alpha + m)z - m(\alpha(1-\lambda) + m) - m\sqrt{z^2 - 2(\alpha(1+\lambda) + m)z + (\alpha(1-\lambda) + m)^2}}{2\alpha z\left(z - (1-\lambda)m\right)}, \tag{6.20}$$

where we have taken the negative root, based on the approximation to first order: for $z \to \infty$ *in* $\mathbb{C}_+$*,*

$$zG_X(z) = \frac{(2\alpha + m)z \pm mz}{2\alpha z} + o(z).$$

To recover the law of X *from* (6.20)*, we employ the Stieltjes inversion formula. To that end (for readability) it is convenient to introduce the following notation for the constants:*

$$\delta^2 = 4\alpha\lambda(\alpha + m), \qquad c = \alpha(1+\lambda) + m, \qquad r = (1-\lambda)m.$$

Completing the square in the quadratic function inside the square root in (6.20), we have

$$G_X(z) = \tilde{G}_X(z) - \frac{m}{2\alpha}\frac{\sqrt{(z-c)^2-\delta^2}}{z(z-r)} \tag{6.21}$$

where

$$\tilde{G}_X(z) = \frac{(2\alpha+m)z - m(\alpha(1-\lambda)+m)}{2\alpha z(z-r)}$$

is a rational expression that is continuous on $\mathbb{C}_+ \sqcup (\mathbb{R}\setminus\{0,r\})$, and takes real values on $\mathbb{R}\setminus\{0,r\}$. Ergo $\lim_{\epsilon\downarrow 0}\mathrm{Im}\tilde{G}_X(x+i\epsilon) = 0$ at all real $x \neq 0, r$, and so $\tilde{G}_X$ does not contribute to the law of X at any points other than possible $\{0,r\}$. Similarly, the second summand in (6.21) is continuous for $z \in \mathbb{C}_+ \setminus \{0,r\}$, and so

$$\lim_{\epsilon\downarrow 0} -\frac{1}{\pi}\mathrm{Im}G_X(x+i\epsilon) = \frac{m}{2\pi\alpha}\frac{\sqrt{\delta^2-(x-c)^2}}{x(x-r)}\mathbf{1}\{|x-c|\leq\delta\} =: \rho_X(x). \tag{6.22}$$

The function ρ_X is integrable and ≥ 0. Indeed: elementaty computations show that

$$c^2-\delta^2 = (\alpha(1-\lambda)+m)^2, \qquad (c-r)^2-\delta^2 = (\alpha(1-\lambda)-\lambda m)^2 \tag{6.23}$$

from which it follows in particular that $c-\delta \geq \max\{0,r\}$. Since the denominator of ρ_X is positive except on the interval between 0 and r, and since the numerator is non-negative everywhere, it follows that $\rho_X \geq 0$. For integrability, we note that the function is uniformly bounded on the support interval $[c-\delta, c+\delta]$ provided that both 0 and r are strictly $< c-\delta$. *A simple calculation shows that this holds true except in two cases:*

$$\text{if } \lambda = 1+\frac{m}{\alpha}, \quad c-\delta = 0; \qquad \text{if } \frac{1}{\lambda} = 1+\frac{m}{\alpha}, \quad c-\delta = r.$$

Aside from these two values for λ, ρ_X is bounded and compactly supported, hence integrable. At these values, ρ_X has $(\,\cdot\,)^{-1/2}$ behavior at the pole that has collided with the edge of its support, and thus still remains integrable.

A fairly involved (and entertaining) calculus exercise shows that

$$\int_{\mathbb{R}} \rho_X(x)\,dx = \begin{cases} \frac{m}{\alpha(1/\lambda-1)} & 0<\lambda<\frac{1}{1+m/\alpha} \\ 1 & \frac{1}{1+m/\alpha}\leq\lambda\leq 1+\frac{m}{\alpha} \\ \frac{m}{\alpha(\lambda-1)} & \lambda > 1+\frac{m}{\alpha}. \end{cases} \tag{6.24}$$

In other words: the total mass of the density ρ_X is 1 if λ is between $1+\frac{m}{\alpha}$ and its reciprocal, and is otherwise equal to $\frac{m}{\alpha(\max\{\lambda,1/\lambda\}-1)}$:

$$\int_{\mathbb{R}} \rho_X(x)\,dx = \min\left\{\frac{m}{\alpha(\max(\lambda,1/\lambda)-1)}, 1\right\}. \tag{6.25}$$

It remains to understand what happens at the poles 0 and r. To that end, we note that for any $x \in \mathbb{R}$,

$$\lim_{z\downarrow x}\sqrt{(z-c)^2-\delta^2} = \mathrm{sign}(x-c)\sqrt{(x-c)^2-\delta^2}.$$

(This follows from (6.13).) In particular, since $x = 0 < c$, using (6.23) we have

$$\lim_{z\downarrow 0}\sqrt{(z-c)^2-\delta^2} = -\sqrt{c^2-\delta^2} = -|\alpha(1-\lambda)+m|$$

and hence, at least when $\lambda \neq 1$,

$$\begin{aligned}\lim_{z\downarrow 0} zG_X(z) &= \frac{(2\alpha+m)\cdot 0 - m(\alpha(1-\lambda)+m) + m|\alpha(1-\lambda)+m|}{2\alpha(0-(1-\lambda)m)} \\ &= \frac{\alpha(1-\lambda)+m-|\alpha(1-\lambda)+m|}{2\alpha(1-\lambda)} \\ &= \begin{cases} 1-\frac{m}{\alpha(\lambda-1)} & \lambda > 1+\frac{m}{\alpha} \\ 0 & \lambda \leq 1+\frac{m}{\alpha}. \end{cases}\end{aligned}$$

This shows that X *has a point mass at* 0 *of weight* $1-\frac{m}{\alpha(\lambda-1)}$ *when* $\lambda > 1+\frac{m}{\alpha}$, *but no mass at* 0 *otherwise.*

Similarly, since $r < c$ *as well, using* (6.23) *we find*

$$\lim_{z\downarrow r} \sqrt{(z-c)^2-\delta^2} = \sqrt{(r-c)^2-\delta^2} = -|\alpha(1-\lambda)-\lambda m|$$

and hence, at least when $\lambda \neq 1$,

$$\begin{aligned}\lim_{z\downarrow r}(z-r)G_X(z) &= \frac{(2\alpha+m)r - m(\alpha(1-\lambda)+m) + m|\alpha(1-\lambda)-\lambda m|}{2\alpha r} \\ &= \frac{(2\alpha+m)(1-\lambda)m - m(\alpha(1-\lambda)+m) + m|\alpha(1-\lambda)-\lambda m|}{2\alpha m(1-\lambda)} \\ &= \frac{m(\alpha(1-\lambda)-\lambda m) + m|\alpha(1-\lambda)-\lambda m|}{2\alpha m(1-\lambda)} \\ &= \begin{cases} 1-\frac{m}{\alpha(1/\lambda-1)} & \lambda < \frac{1}{1+m/\alpha} \\ 0 & \lambda \geq \frac{1}{1+m/\alpha}. \end{cases}\end{aligned}$$

This shows that X *has a point mass at* r *of weight* $1-\frac{m}{\alpha(1/\lambda-1)}$ *when* $\lambda < \frac{1}{1+m/\alpha}$, *bot no mass at* r *otherwise.*

Combining with (6.24), *and the formula* (6.22) *for the density* ρ_X, *we thus have the complete description of the law of* X*:*

$$\begin{aligned}\rho_X(x) &= \frac{m}{2\pi\alpha}\frac{\sqrt{(4\alpha\lambda(\alpha+m)-(x-m-\alpha(1+\lambda))^2)_+}}{x(x-(1-\lambda)m)} \\ \mathcal{L}(X) &= \rho_X(x)\,dx + \left(1-\tfrac{m}{\alpha(\lambda-1)}\right)_+ \delta_0 + \left(1-\tfrac{m}{\alpha(1/\lambda-1)}\right)_+ \delta_{(1-\lambda)m}\end{aligned} \tag{6.26}$$

where, as usual, $f_+ = f\mathbf{1}_{\{f\geq 0\}}$.

Notice that the two values of λ *where the nature of* $\mathcal{L}(X)$ *changes are precisely the values for which the density* ρ_X *is unbounded. These are also the precise values of* λ *where a point mass collides with the support of* ρ_X.

Acknowledgments

The authors would like to thank Serban Belinschi and Ching Wei Ho for several very useful conversations clarifying core ideas of this work.

A few comments on the name 'El Gordo' (Lemma 3.4), which is Spanish for 'The Fat One'. As shown in some examples, e.g. Example 3.3(b), the El Gordo transform tends to spread the support out. As we discussed at the beginning of Section 3.3, $X \mapsto X^\flat$ is a free probability analogue of $X \mapsto UX$ where U is a Unif$[0,1]$ random variable independent from X, and this operation also has the effect of spreading out X. We could give other reasons why this name is sensible, but honestly we

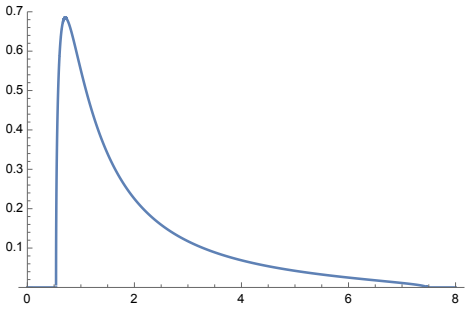

FIGURE 7. The law in (6.26) with $\alpha = \lambda = 1, m = 2$ (in which case there are no point masses, only the density.

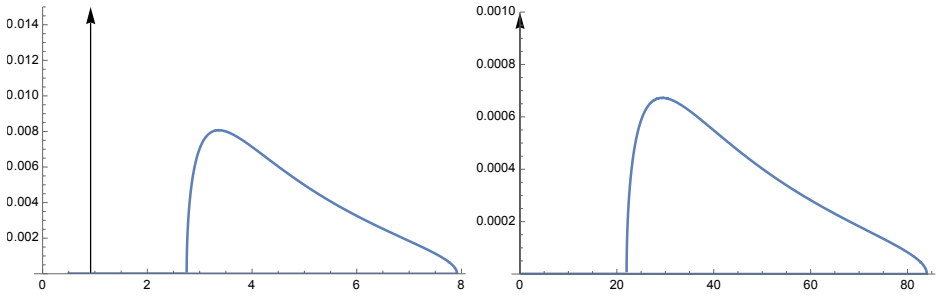

FIGURE 8. The law in (6.26) with $\alpha = 4$, $m = 1$, and $\lambda = 1/12$ on the left and $\lambda = 12$ on the right. In these complementary cases, most of the mass is concentrated at an atom with weight $\frac{43}{44}$, located at the point $\frac{11}{12}$ on the left, and at 0 on the right.

chose this name because the insight to introduce this and related transforms came to the authors over many excellent lunches at the incomparable *Tacos El Gordo* in Chula Vista, California.

REFERENCES

1. Octavio Arizmendi, Takahiro Hasebe, and Noriyoshi Sakuma, *On the law of free subordinators*, ALEA Lat. Am. J. Probab. Math. Stat. (2013). (Cited on pages 8 and 22.)
2. Benjamin Arras and Christian Houdré, *On Stein's method for infinitely divisible laws with finite first moment*, Springer, 2019. (Cited on page 5.)
3. Richard Arratia, Larry Goldstein, and Louis Gordon, *Poisson approximation and the Chen-Stein method*, Statistical Science (1990), 403–424. (Cited on page 2.)
4. Richard Arratia, Larry Goldstein, and Fred Kochman, *Size bias for one and all*, Probability Surveys (2019), 16: 1–61. (Cited on pages 2, 5, and 34.)
5. Andrew D Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, Oxford University Press, 1992. (Cited on pages 2 and 35.)
6. S. T. Belinschi and H. Bercovici, *Atoms and regularity for measures in a partially defined free convolution semigroup*, Math. Z. **248** (2004), no. 4, 665–674. MR 2103535 (Cited on page 59.)
7. Hari Bercovici and Vittorino Pata, *Stable laws and domains of attraction in free probability theory*, Ann. of Math. (2) **149** (1999), no. 3, 1023–1060, With an appendix by Philippe Biane. MR 1709310 (Cited on pages 7, 8, 21, and 51.)

8. Hari Bercovici and Dan Voiculescu, *Free convolution of measures with unbounded support*, Indiana University Mathematics Journal **42** (1993), no. 3, 733–773. (Cited on pages 6, 7, 8, 16, 17, 21, 44, 55, and 59.)
9. ———, *Regularity questions for free convolution*, Nonselfadjoint operator algebras, operator theory, and related topics, Oper. Theory Adv. Appl., vol. 104, Birkhäuser, Basel, 1998, pp. 37–47. MR 1639647 (Cited on page 19.)
10. Hari Bercovici and Dan-Virgil Voiculescu, *Lévy-Hinčin type theorems for multiplicative and additive free convolution*, Pacific journal of mathematics **153** (1992), no. 2, 217–248. (Cited on pages 7, 8, and 44.)
11. P Biane, *Processes with free increments*, Mathematische Zeitschrift **227** (1998), 143–174. (Cited on pages 19 and 20.)
12. T Cacoullos and V Papathanasiou, *Lower variance bounds and a new proof of the central limit theorem*, Journal of multivariate analysis **43** (1992), no. 2, 173–184. (Cited on page 10.)
13. Guillaume Cébron, Max Fathi, and Tobias Mai, *A note on existence of free Stein kernels*, Proceedings of the American Mathematical Society **148** (2020), no. 4, 1583–1594. (Cited on pages 10 and 11.)
14. Sourav Chatterjee, *Stein's method for concentration inequalities*, Probab. Theory Related Fields **138** (2007), no. 8, 305–321. (Cited on page 3.)
15. ———, *Fluctuations of eigenvalues and second order poincaré inequalities*, Probability Theory and Related Fields **143** (2009), no. 1-2, 1–40. (Cited on page 10.)
16. Sourav Chatterjee and Partha S Dey, *Applications of Stein's method for concentration inequalities*, The Annals of Probability **38** (2010), no. 6, 2443–2485. (Cited on page 3.)
17. Louis HY Chen, *Poisson approximation for dependent trials*, The Annals of Probability **3** (1975), no. 3, 534–545. (Cited on pages 2, 3, and 35.)
18. Louis HY Chen, Larry Goldstein, and Qi-Man Shao, *Normal approximation by Stein's method*, Springer Science & Business Media, 2010. (Cited on pages 3, 38, and 39.)
19. Murat A Erdogdu, Lee H Dicker, and Mohsen Bayati, *Scaled least squares estimator for GLMs in large-scale problems*, Advances in Neural Information Processing Systems **29** (2016). (Cited on page 3.)
20. Max Fathi, Larry Goldstein, Gesine Reinert, and Adrien Saumard, *Relaxing the Gaussian assumption in shrinkage and SURE in high dimension*, arXiv preprint arXiv:2004.01378 (2020). (Cited on page 3.)
21. Max Fathi and Brent Nelson, *Free Stein kernels and an improvement of the free logarithmic sobolev inequality*, Advances in Mathematics **317** (2017), 193–223. (Cited on pages 10 and 11.)
22. Subhankar Ghosh and Larry Goldstein, *Concentration of measures via size-biased couplings*, Probability Theory and Related fields **149** (2011), no. 1, 271–278. (Cited on page 3.)
23. B. Gnedenko and A. Kolmogorov, *Poisson approximation*, Addison-Wesley Publ. Co., 1954. (Cited on page 21.)
24. Larry Goldstein and Gesine Reinert, *Stein's method and the zero bias transformation with application to simple random sampling*, The Annals of Applied Probability **7** (1997), no. 4, 935–952. (Cited on pages 3, 5, 33, 38, 39, and 40.)
25. ———, *Distributional transformations, orthogonal polynomials, and Stein characterizations*, Journal of Theoretical Probability **18** (2005), 237–260. (Cited on pages 3 and 29.)
26. ———, *Total variation distance for poisson subset numbers*, Annals of Combinatorics **10** (2006), no. 3, 333–341. (Cited on page 35.)
27. Larry Goldstein and Yosef Rinott, *Multivariate normal approximations by Stein's method and size bias couplings*, Journal of Applied Probability **33** (1996), no. 1, 1–17. (Cited on page 34.)
28. Larry Goldstein and Uwe Schmock, *Probabilistic interpretation of a theorem of Kolmogorov using the zero bias transformation*, preprint (2023+). (Cited on pages 5, 6, and 52.)
29. Larry Goldstein and Xiaohan Wei, *Non-Gaussian observations in nonlinear compressed sensing via Stein discrepancies*, Information and Inference: A Journal of the IMA **8** (2019), no. 1, 125–159. (Cited on page 3.)
30. David A. Jekel, Todd A. Kemp, and Evangelos A. Nikitopoulos, *A martingale approach to noncommutative stochastic calculus*, Preprint (2023). (Cited on page 3.)
31. Tobias Johnson and Erol Peköz, *Concentration inequalities from monotone couplings for graphs, walks, trees and branching processes*, arXiv preprint arXiv:2108.02101 (2021). (Cited on page 3.)

32. Tosio Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1980 edition. MR 1335452 (Cited on page 3.)
33. Todd Kemp, *Introduction to random matrix theory*, unpublished, 2013. (Cited on page 15.)
34. Qiang Liu, Jason Lee, and Michael Jordan, *A kernelized Stein discrepancy for goodness-of-fit tests*, International conference on machine learning, PMLR, 2016, pp. 276–284. (Cited on page 3.)
35. Qiang Liu and Dilin Wang, *Stein variational gradient descent: A general purpose Bayesian inference algorithm*, Advances in Neural Information Processing Systems **29** (2016). (Cited on page 3.)
36. Qingwei Liu and Aihua Xia, *Geometric sums, size biasing and zero biasing*, Electronic Communications in Probability **27** (2022), 1–13. (Cited on page 5.)
37. Hans Maassen, *Addition of freely independent random variables*, J. Funct. Anal. **106** (1992), no. 2, 409–438. MR 1165862 (Cited on pages 16, 56, and 59.)
38. James A. Mingo and Roland Speicher, *Free probability and random matrices*, Fields Institute Monographs, vol. 35, Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017. MR 3585560 (Cited on page 11.)
39. Alexandru Nica and Roland Speicher, *On the multiplication of free N-tuples of noncommutative random variables*, Amer. J. Math. **118** (1996), no. 4, 799–837. MR 1400060 (Cited on pages 18 and 48.)
40. ———, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series, vol. 335, Cambridge University Press, Cambridge, 2006. MR 2266879 (Cited on pages 11 and 56.)
41. Evangelos A. Nikitopoulos, *Noncommutative C^k functions and Fréchet derivatives of operator functions*, Expo. Math. **41** (2023), no. 1, 115–163. MR 4557276 (Cited on page 3.)
42. Victor Pérez-Abreu and Noriyoshi Sakuma, *Free infinite divisibility of free multiplicative mixtures of the wigner distribution*, Journal of Theoretical Probability **25** (2012), 100–121. (Cited on page 22.)
43. Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory, University of California Press, 1972, pp. 583–602. (Cited on page 3.)
44. FW Steutel, *Some recent results in infinite divisibility*, Stochastic processes and their applications **1** (1973), no. 2, 125–143. (Cited on pages 5 and 6.)
45. Dan Voiculescu, *Addition of certain noncommuting random variables*, J. Funct. Anal. **66** (1986), no. 3, 323–346. MR 839105 (Cited on page 16.)
46. ———, *The analogues of entropy and of Fisher's information measure in free probability theory. I*, Comm. Math. Phys. **155** (1993), no. 1, 71–92. MR 1228526 (Cited on page 19.)
47. Jiun-Chau Wang, *Local limit theorems in free probability theory*, Ann. Probab. **38** (2010), no. 4, 1492–1506. MR 2663634 (Cited on page 59.)

[1]Department of Mathematics, University of Southern California, Los Angeles
Email address: larry@usc.edu

[2]Department of Mathematics, University of California, San Diego
Email address: tkemp@ucsd.edu