PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 39, Fasc. 2 (2019), pp. 237–258
doi:10.19195/0208-4147.39.2.1

CAUCHY–STIELTJES FAMILIES WITH POLYNOMIAL VARIANCE FUNCTIONS AND GENERALIZED ORTHOGONALITY

BY

WŁODZIMIERZ BRYC* (CINCINNATI), RAOUF FAKHFAKH (SFAX),
AND WOJCIECH MŁOTKOWSKI** (WROCŁAW)

Abstract. This paper studies variance functions of Cauchy–Stieltjes Kernel (CSK) families generated by compactly supported centered probability measures. We describe several operations that allow us to construct additional variance functions from known ones. We construct a class of examples which exhausts all cubic variance functions, and provide examples of polynomial variance functions of arbitrary degree. We also relate CSK families with polynomial variance functions to generalized orthogonality.

Our main results are stated solely in terms of classical probability; some proofs rely on analytic machinery of free probability.

2010 AMS Mathematics Subject Classification: Primary: 60E10, 46L54, 62E10; Secondary: 05A15.

Key words and phrases: Kernel families, generalized orthogonality, R-transform, S-transform, Fuss–Catalan numbers, variance functions, free additive convolution, free multiplicative convolution.

1. INTRODUCTION AND MAIN RESULTS

The Cauchy–Stieltjes Kernel (CSK) families of probability measures were introduced in [11] and extended to non-compact setting in [13]. The constructive approach adopted in these papers is based on an idea of kernel family from an unpublished manuscript [46]. The construction emphasizes analogies with exponential families, using the Cauchy–Stieltjes kernel $1/(1-\theta x)$ instead of the exponential kernel $\exp(\theta x)$, and establishing parametrization by the mean. Kernels of the form $h(x\theta)$, including $1/(1-\theta x)^a$, appear also in [28] and the references cited therein.

After reparametrization by the mean, CSK families are also a special case $q=0$ of the q-exponential families from [14]. The non-constructive definition from

* Research partially supported by the Taft Research Center.
** Supported by NCN grant 2016/21/B/ST1/00628.

[14], Section 4, is most convenient for our purposes, as it emphasizes the role of the pseudo-variance function, which appears directly in the definition.

DEFINITION 1.1. The *CSK family with a pseudo-variance function* $\mathbb{V}$ generated by a compactly supported non-degenerate probability measure ν is a family of probability measures

$$\{Q_m(dx) := f(x,m)\nu(dx) : \ m \in (m_-, m_+)\},$$

where

$$f(x,m) := \begin{cases} \frac{\mathbb{V}(m)}{\mathbb{V}(m)+m(m-x)}, & m \neq 0, \\ 1, & m = 0, \mathbb{V}(0) \neq 0, \\ \frac{\mathbb{V}'(0)}{\mathbb{V}'(0)-x}, & m = 0, \mathbb{V}(0) = 0. \end{cases} \tag{1.1}$$

The interval (m_-, m_+) is sometimes called the domain of means, but it will not play a major role here. We will only assume that $0 \in (m_-, m_+)$ and $\mathbb{V}(0) \neq 0$. Then (1.1) is the solution of the difference equation

$$\frac{f(x,m) - f(x,0)}{m} = \frac{x-m}{\mathbb{V}(m)} f(x,m), \quad f(x,0) = 1, \tag{1.2}$$

which is a discrete analog of the differential equation for exponential families noted in [45], Theorem 2 (see also [17], Section 5, and [14]).

It is known that measure ν, if it exists, is uniquely determined (up to the mean) by $\mathbb{V}$, see [11]. It is also known that any non-degenerate compactly supported probability measure ν gives rise to a unique (real analytic) function $\mathbb{V}$, which will sometimes be denoted by $\mathbb{V}_\nu$. On the other hand, not every function $\mathbb{V}$ can appear as a pseudo-variance function. The question of determining whether a given class of functions $\mathbb{V}$ corresponds to some measures ν generated a sizeable literature both for the exponential and more recently for the CSK families. In the theory of exponential families, all quadratic variance functions were determined in [24] and in [37]. All cubic variance functions up to affine transformations are described in [30]. In [20], cubic variance functions are characterized by generalized orthogonality. Numerous non-polynomial variance functions have also been studied, see [29]; see also [14], Section 2.

The literature about the variance functions of the CSK families is less comprehensive. CSK families with quadratic variance functions were determined in [11], [14], see also [18]. Cubic (pseudo) variance functions with $\mathbb{V}(0) = 0$ have been studied in [13] and they correspond to measures without first moment. In contrast to exponential families, CSK families are not invariant under translation, so cubic variance functions with $\mathbb{V}(0) \neq 0$ cannot be reduced to the case studied in [13] and require separate investigation. This paper is devoted solely to the case $\mathbb{V}(0) \neq 0$.

We now recall some formulas and assumptions that we will rely upon. It is known (see [13], Proposition 3.1, or [14], (3.4)) that for $m \neq 0$

$$\int x Q_m(dx) = m, \tag{1.3}$$

so the family $\{Q_m : m \in (m_-, m_+)\}$ is indeed parameterized by the mean. One can show that if ν has all moments, $0 \in (m_-, m_+)$ and $\mathbb{V}(0) \neq 0$, then (1.3) extends by continuity to

$$\int x\nu(dx) = 0. \tag{1.4}$$

We will simply assume (1.4). It is then known, and easy to check, that the pseudo-variance function that appears in (1.1) is indeed the variance function,

$$\mathbb{V}(m) = \int (x-m)^2 Q_m(dx), \tag{1.5}$$

see [13], Proposition 3.2, and [14], (3.4), where a more general case was considered.

Denote by $\mathcal{V}$ the class of variance functions corresponding to probability measures ν such that ν is compactly supported, centered: $\int x\nu(dx) = 0$, with variance $\int x^2\nu(dx) = 1$, so that $\mathbb{V}_\nu(0) = 1$. Denote by $\mathcal{V}_\infty$ the class of those $\mathbb{V} \in \mathcal{V}$ that the function $m \mapsto \mathbb{V}(cm)$ is in $\mathcal{V}$ for every real c.

We begin with some algebraic operations that allow us to build new variance functions from known ones. (Here we write $\mathbb{V}(m)$ for a function, not its value.)

THEOREM 1.1. *Assume that* $\mathbb{V}(m) \in \mathcal{V}$, $\mathbb{V}_1(m), \mathbb{V}_2(m) \in \mathcal{V}_\infty$ *and* $c \geqslant 1$. *Then*:

(i) $\mathbb{V}(m/c) \in \mathcal{V}$;

(ii) $\mathbb{V}(m) + am \in \mathcal{V}$ *and* $\mathbb{V}_1(m) + am \in \mathcal{V}_\infty$ *for any* $a \in \mathbb{R}$;

(iii) $\mathbb{V}_1(m) + \mathbb{V}_2(m) - 1 \in \mathcal{V}_\infty$ *and* $c\mathbb{V}_1(m) - c + 1 \in \mathcal{V}_\infty$;

(iv) $\mathbb{V}_1(m) - m^2 \in \mathcal{V}$;

(v) $\mathbb{V}(m) + m^2 \in \mathcal{V}_\infty$.

The proof of this theorem appears in Section 2.3.

COROLLARY 1.1. *The map* $\mathbb{V}(m) \mapsto \mathbb{V}(m) - m^2$ *is a bijection of* $\mathcal{V}_\infty$ *onto* $\mathcal{V}$.

Next, we describe the class of cubic variance functions.

THEOREM 1.2. *Fix* $a, b, c \in \mathbb{R}$. *A cubic function* $\mathbb{V}(m) = 1 + am + bm^2 + cm^3$ *is in* $\mathcal{V}$ *if and only if* $(b+1)^3 \geqslant 27c^2$. *Furthermore*, $\mathbb{V}$ *is in* $\mathcal{V}_\infty$ *if and only if* $b^3 \geqslant 27c^2$.

The proof of this theorem appears in Section 2.4.

Our final result relates polynomial variance functions for a CSK family to generalized orthogonality. Suppose $\{P_n(x) : n = 0, 1, 2, \dots\}$ is a family of real

Probability and Mathematical Statistics 39, z. 2, 2019

polynomials, indexed by their degree n with $P_0(x) = 1$; it is sometimes convenient to set $P_k(x) = 0$ for $k < 0$.

There is a substantial literature on generalized orthogonality and finite-step recursions for polynomials. We introduce the following generalized orthogonality condition.

DEFINITION 1.2. Fix $d \in \mathbb{N}$ and a probability measure ν with moments of all orders. We say that polynomials $\{P_n\}$ are $(\nu; d)$-*orthogonal* if $\int P_n(x)\nu(dx) = 0$ for all $n \geqslant 1$, and

$$\int P_n(x)P_k(x)\nu(dx) = 0 \qquad \text{for all } n \geqslant 2 + (k-1)d, \quad k = 1, 2, \ldots$$

It is clear that for measures with infinite support, $(\nu; 1)$-orthogonality is just the standard orthogonality. For $d = 2$, we recover [20], Definition 3.1. The concept of d-orthogonality introduced in [42] is different as even for $d = 1$ it has no positivity requirements for the functional/measure. When $d > 2$, the condition of pseudo-orthogonality in [25], [26] is also different. It is somewhat interesting to note that various concepts of generalized orthogonality are related to $(d + 2)$-step recursions for the polynomials, so the distinctions sometimes rely on minute technicalities, see the paragraph above Corollary 3.1.

The following result is a generalization of Theorem 3.2 in [18] to $d > 1$, and a CSK-version of Theorem 3.1 in [20] when $d = 2$.

THEOREM 1.3. *Suppose that $\mathbb{V}$ is a variance function of a CSK family generated by a non-degenerate compactly supported probability measure ν with mean zero and variance one. Consider the family of polynomials $\{P_n(x)\}$ with generating function*

$$f(x, m) = \sum_{n=0}^{\infty} P_n(x)m^n, \tag{1.6}$$

where $f(x, m)$ is given by (1.1). *Then the following statements are equivalent*:

(i) $\mathbb{V}(m) = 1 + \sum_{k=1}^{d+1} a_k m^k$ *is a polynomial of degree at most* $d + 1$.

(ii) *There exist constants $\{b_k : k = 1, \ldots, d+1\}$ such that polynomials $\{P_n\}$ satisfy the recursion*

$$xP_n(x) = P_{n+1}(x) + \sum_{k=1}^{(d+1)\wedge n} b_k P_{n+1-k}(x), \quad n \geqslant 1, \tag{1.7}$$

with initial conditions $P_0(x) = 1$, $P_1(x) = x$.

(iii) *Polynomials $\{P_n(x)\}$ are $(\nu; d)$-orthogonal.*

Note that the upper limit of the sum on the right-hand side of (1.7) is $d + 1$ under the convention that $P_k(x) = 0$ for $k < 0$, and that Proposition 2.3 below provides examples of polynomial variance functions of arbitrarily high degree. The proof of Theorem 1.3 appears in Section 3.1.

The paper is organized as follows. In Section 2 we introduce free probability notation and use it to prove the first two theorems. We also include some additional examples of variance functions. Section 3 is independent of Section 2 and discusses results on polynomials that imply Theorem 1.3. In Section 4 we provide a combinatorial example involving sequences A001764, A098746 and A106228 from OEIS [41]. We also discuss generating functions and sharpness of some results.

2. VARIANCE FUNCTIONS AND FREE PROBABILITY

Recall that a *dilation* $D_t(\nu)$ of a probability measure ν by a non-zero real number t is a measure $\mu(U) = \nu(U/t)$, and $D_{-1}(\nu)$ is called the *reflection* of ν. In the language of probability theory, dilation changes the law of random variable X to the law of tX.

2.1. Notation from free probability. For a probability measure μ on $\mathbb{R}$ we put

$$M_\mu(z) := \int \frac{\mu(dx)}{1-xz}, \quad G_\mu(z) := \int \frac{\mu(dx)}{z-x}, \quad F_\mu(z) := 1/G_\mu(z),$$

where M_μ is called the *moment generating function*, and $G_\mu(z)$ is the *Cauchy–Stieltjes transform*. The *free R-transform* can be defined by the equation

$$R_\mu\big(zM_\mu(z)\big) + 1 = M_\mu(z). \tag{2.1}$$

The coefficients $\kappa_n(\mu)$ in the Taylor expansion $R_\mu(z) = \sum_{n=1}^\infty \kappa_n(\mu)z^n$ are called *free cumulants*. We will also use

$$r_\mu(z) := R_\mu(z)/z. \tag{2.2}$$

Equation (2.1) can also be written as

$$zM_\mu(z)r_\mu\big(zM_\mu(z)\big) = M_\mu(z) - 1. \tag{2.3}$$

Note that for the dilated measure we have

$$M_{D_t(\mu)}(z) = M_\mu(tz), \quad R_{D_t(\mu)}(z) = R_\mu(tz), \quad r_{D_t(\mu)}(z) = tr_\mu(tz). \tag{2.4}$$

If $|z| \neq 0$ is small enough, then

$$G_\mu\left(r_\mu(z) + \frac{1}{z}\right) = z, \quad F_\mu\left(r_\mu(z) + \frac{1}{z}\right) = \frac{1}{z}. \tag{2.5}$$

The sum of two R-transforms is an R-transform and defines the *free additive convolution* of measures $\mu \boxplus \nu$ by $R_{\mu\boxplus\nu}(z) = R_\mu(z) + R_\nu(z)$. For any real $t \geqslant 1$, it is known that $tR_\mu(z)$ is an R-transform and defines an *additive free convolution power* $\mu^{\boxplus t}$ (see [38]).

Probability measure μ is called $\boxplus$-*infinitely divisible* if its free convolution power $\mu^{\boxplus t}$ is well-defined for all real $t > 0$. If $\lambda \neq 0$, then μ is $\boxplus$-infinitely divisible if and only if $D_\lambda(\mu)$ is $\boxplus$-infinitely divisible.

It is known, see [6], [21], that a compactly supported μ with the first moment $m_0 = \int x\mu(dx)$ is $\boxplus$-infinitely divisible if and only if there exists a compactly supported finite measure ω on $\mathbb{R}$ such that $\omega(\mathbb{R}) = \int (x - m_0)^2 \mu(dx)$ and

$$r_\mu(z) = m_0 + z \int \frac{\omega(dx)}{1 - zx}.$$

In particular, if ν is a generating measure of a CSK family, then under our moment assumptions, ν is free-infinitely divisible if and only if there is a compactly supported probability measure ω such that

$$r_\nu(z) = zM_\omega(z). \tag{2.6}$$

For a probability measure $\mu \neq \delta_0$ with support in $[0, \infty)$, the S-*transform* is defined by

$$R_\mu\big(zS_\mu(z)\big) = z \quad \text{or} \quad M_\mu\left(\frac{z}{1+z} S_\mu(z)\right) = 1 + z, \tag{2.7}$$

see e.g. [22], (5). Note that, in particular, $\int x\mu(dx) = 1/S_\mu(0)$.

The product of S-transforms is an S-transform and defines the *multiplicative free convolution* $\mu_1 \boxtimes \mu_2$ by $S_{\mu_1 \boxtimes \mu_2}(z) = S_{\mu_1}(z) S_{\mu_2}(z)$. *Multiplicative free convolution powers* $\mu^{\boxtimes p}$ are defined at least for all $p \geqslant 1$ (see [5], Theorem 2.17) by $S_{\mu^{\boxtimes p}}(z) = S_\mu(z)^p$.

The *Marchenko–Pastur measure* with parameter $\lambda > 0$,

$$\pi_\lambda(dx) = (1 - \lambda)^+ \delta_0 + \frac{\sqrt{4\lambda - (x - 1 - \lambda)^2}}{2\pi x} 1_{x \in [(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2]} dx,$$

plays in free probability the role of the Poisson distribution, see [39]. Since $S_{\pi_\lambda}(z) = 1/(\lambda + z)$, we have

$$S_{D_b(\pi_{1/b})}(z) = \frac{1}{1 + bz}.$$

It is known (see [22], Section 2; [3], Theorem 1.2, and [4], [34]) that if $p > 0$, $b > 0$, then $\mu = \big(D_b(\pi_{1/b})\big)^{\boxtimes p}$ exists if and only if $\max\{p, 1/b\} \geqslant 1$. This measure μ has compact support in $[0, \infty)$ and its S-transform equals

$$S_\mu(z) = \frac{1}{(1 + bz)^p}. \tag{2.8}$$

For additional details and background on free probability we refer to [39] and [44].

2.2. Formulas for variance functions. A variance function $\mathbb{V}$ of a CSK family generated by a compactly supported centered probability measure $\nu \neq \delta_0$ is real analytic at $m = 0$, so it extends to the analytic mapping $z \mapsto \mathbb{V}(z)$ on an open disk near $z = 0$. Our assumptions on the first two moments of ν imply that $r_\nu(z) = z + \kappa_3(\nu)z^2 + \ldots$ is invertible near $z = 0$ and its composition inverse is $z/\mathbb{V}_\nu(z)$ ([11], Theorem 3.3), so that

$$r_\nu(z) = z\mathbb{V}_\nu\big(r_\nu(z)\big). \tag{2.9}$$

Replacing z by $z/\mathbb{V}(z)$, from equation (2.5) we get

$$F_\nu\left(z + \frac{\mathbb{V}_\nu(z)}{z}\right) = \frac{\mathbb{V}_\nu(z)}{z}. \tag{2.10}$$

(This was first noted in [14], (4.4), and exploited in [11]–[13].)

The following result is known but we prove it for completeness.

LEMMA 2.1. *If $z \mapsto \mathbb{V}(z)$ is a variance function, then so is $z \mapsto \mathbb{V}(-z)$.*

Proof. Put $\nu_- := D_{-1}(\nu)$. Then, by (2.4), $r_{\nu_-}(z) = -r_\nu(-z)$ and from (2.9) we have $\mathbb{V}_{\nu_-}(z) = \mathbb{V}_\nu(-z)$. ■

The following relates the class $\mathcal{V}_\infty$ of variance functions to free probability.

PROPOSITION 2.1. *If $\mathbb{V} = \mathbb{V}_\nu$, then $\nu^{\boxplus\lambda^2}$ exists if and only if $\mathbb{V}(z/\lambda) \in \mathcal{V}$. In particular, $\mathcal{V}_\infty$ is the class of those $\mathbb{V}_\nu \in \mathcal{V}$ that ν is $\boxplus$-infinitely divisible.*

Proof. Suppose $\nu^{\boxplus\lambda^2}$ exists and define $\nu_\lambda := D_{1/\lambda}(\nu^{\boxplus\lambda^2})$. Then, by (2.4), we have $r_{\nu_\lambda}(z) = \lambda r_\nu(z/\lambda)$ and

$$\frac{r_{\nu_\lambda}(z)}{\mathbb{V}\big(r_{\nu_\lambda}(z)/\lambda\big)} = \frac{\lambda r_\nu(z/\lambda)}{\mathbb{V}\big(r_\nu(z/\lambda)\big)} = \lambda\frac{z}{\lambda} = z,$$

which proves that $\mathbb{V}_{\nu_\lambda}(z) = \mathbb{V}(z/\lambda)$. Conversely, by the first equality, if $\mathbb{V}(z/\lambda)$ is a variance function of some ν_λ, then $r_{\nu_\lambda}(z)/\lambda = r_\nu(z/\lambda)$, so $\nu^{\boxplus\lambda^2} = D_\lambda(\nu_\lambda)$ exists.

In particular, from Lemma 2.1 we see that $\mathbb{V} \in \mathcal{V}_\infty$ if and only if ν is $\boxplus$-infinitely divisible. ■

From (2.9), (2.6) and (2.1) we get the following.

PROPOSITION 2.2. *A function $\mathbb{V}(z)$ belongs to $\mathcal{V}_\infty$ if and only if there is a compactly supported probability measure ω on $\mathbb{R}$ such that $\mathbb{V}(z) = 1 + R_\omega(z)$.*

We remark that the perturbation theorem in [8] generates a large number of implicit examples of variance functions in $\mathcal{V}_\infty$. In particular, for every $d \geqslant 3$ there is a $\delta > 0$ such that $\mathbb{V}(z) = 1 + z^2 + \sum_{k=3}^{d} c_k z^k$ is in $\mathcal{V}_\infty$ when $\max_k |c_k| < \delta$. Corollary 2.5 in [15] yields explicit characterization of such variance functions for $d = 4$.

2.3. Proof of Theorem 1.1. We need the following lemma that will be used with $\alpha = 1$, $\beta = 0$.

LEMMA 2.2. *If ω is a probability distribution on $\mathbb{R}$, $\alpha > 0$, $\beta \in \mathbb{R}$, then there exists a non-degenerate probability distribution μ such that*

$$M_\mu(z) = \frac{1}{1 - \beta z - \alpha z^2 M_\omega(z)}. \tag{2.11}$$

Conversely, if μ is a probability measure with moments

$$\int x\mu(dx) = \beta, \quad \int (x-\beta)^2 \mu(dx) = \alpha > 0,$$

then there exists a probability measure ω such that (2.11) *holds.*

Proof. For the F-transform of μ we have $M_\mu(z) = 1/\big(zF_\mu(1/z)\big)$, so relation (2.11) becomes

$$F_\mu(z) = z - \beta - \frac{\alpha}{F_\omega(z)}.$$

Now, it suffices to apply Proposition 5.2 from [7] (see also [21], Section 3.3).

To prove the converse, we apply Proposition 5.2 from [7] to the analytic function

$$F(z) := \frac{\alpha}{z - \beta - F_\mu(z)},$$

which becomes the F-transform of a probability measure. To verify the assumptions in [7], we note that since μ is non-degenerate, we have $\Im F_\mu(z) > \Im z$ (see comments below and [31], Proposition 2.1). So F maps $\mathbb{C}_+$ into itself. Series expansion at infinity gives $z - \beta - F_\mu(z) = \alpha/z + o(1/z)$ as $|z| \to \infty$. ∎

Proof of Theorem 1.1. Statement (i) follows from Proposition 2.1, as the free convolution power $\nu^{\boxplus c^2}$ exists for $c \geqslant 1$.

(ii) Let $G(z) = M_\nu(1/z)/z$ be the Cauchy–Stieltjes transform of ν and $F(z) = 1/G(z)$. The continued fraction expansion for G gives

$$F(z) = z - b_0 - \cfrac{c_0}{z - b_1 - \cfrac{c_1}{z - b_2 - \cfrac{c_2}{\ddots}}},$$

where b_n, c_n are the Jacobi coefficients in the three-step recursion for the monic orthogonal polynomials with respect to measure ν,

$$xp_n(x) = p_{n+1}(x) + b_n p_n(x) + c_{n-1}p_{n-1}(x), \quad n \geqslant 0.$$

(This can be read out from [23], Section 2.6. The recursion and the continued fraction terminate at $c_N = 0$ if ν is purely atomic with $N+1$ atoms.)

Define $F_a(z) = F(z-a) + a$. Then $F_a(z)$ has the same continued fraction expansion with the same coefficients c_n, the same coefficient b_0, and for $k \geqslant 1$ the coefficient b_k is replaced by $b_k + a$. Therefore, by Favard's theorem (the usual version, or a finite version when $c_N = 0$; the latter can be read out from the first page of Section 2.5 in [23]), $F_a(z)$ is the inverse of a Cauchy–Stieltjes transform of a probability measure ν_a. The first two moments of ν_a are not affected by the change of $b_1, b_2, \ldots$, so ν_a has mean zero and variance one.

Since F is well-defined outside of the support of ν, we have $F(x) > 0$ for $x > K$ and $F(x) < 0$ for $x < -K$. So F_a also extends to the real axis far away from zero, and therefore ν_a has compact support. (This fact is sometimes called Krein's theorem [27], see e.g. [15], Theorem 3.9.)

Since F satisfies (2.10), the function $\mathbb{V}_a(z) = \mathbb{V}(z) + az$ satisfies the same identity with F_a in place of F, identifying the variance function.

Suppose now that $\mathbb{V} \in \mathcal{V}_\infty$. Then $\mathbb{V}(cz)$ is a variance function for any real c, so by the previous reasoning with a replaced by ac, we see that $\mathbb{V}(cz) + acz = \mathbb{V}_a(cz)$ is in $\mathcal{V}$, i.e., $\mathbb{V}_a \in \mathcal{V}_\infty$.

(iii) We use Proposition 2.2. If $\mathbb{V}_1(z) = \mathbb{V}_{\nu_1}(z) = 1 + R_{\omega_1}(z)$, $\mathbb{V}_2(z) = \mathbb{V}_{\nu_2}(z) = 1 + R_{\omega_2}(z)$, then

$$\mathbb{V}_1(z) + \mathbb{V}_2(z) - 1 = 1 + R_{\omega_1}(z) + R_{\omega_2}(z) = 1 + R_{\omega_1 \boxplus \omega_2}(z),$$

and similarly

$$c\mathbb{V}_1(z) - c + 1 = 1 + cR_{\omega_1}(z) = 1 + R_{\omega_1^{\boxplus c}}(z).$$

(iv) Let $\mathbb{V}_1 = \mathbb{V}_{\nu_1}$ and define $r_1 := r_{\nu_1}$. Then $r_1(z) = zM_\omega(z)$ for some probability measure ω. Using Lemma 2.2, let ν be a probability measure such that $M_\nu(z) = 1/\big(1 - zr_1(z)\big)$. It is clear that ν has mean zero and variance one. Let $M_\nu = M$, $r_\nu := r$ and put $\widetilde{z} := zM(z)$. Then, by (2.3),

$$z = \frac{\widetilde{z}}{M(z)} = \frac{\widetilde{z}}{\widetilde{z}r(\widetilde{z}) + 1}$$

and

$$r_1(z) = \frac{M(z) - 1}{zM(z)} = r(\widetilde{z}).$$

Applying these identities to the equality

$$\frac{r_1(z)}{\mathbb{V}_1\big(r_1(z)\big)} = z$$

yields

$$\frac{r(\widetilde{z})}{\mathbb{V}_1\big(r(\widetilde{z})\big)} = \frac{\widetilde{z}}{\widetilde{z}r(\widetilde{z})+1}$$

or, equivalently,

$$\frac{r(\widetilde{z})}{\mathbb{V}_1\big(r(\widetilde{z})\big) - r(\widetilde{z})^2} = \widetilde{z},$$

which proves that $\mathbb{V}_\nu(z) = \mathbb{V}_1(z) - z^2$.

(v) Let ν be the measure corresponding to $\mathbb{V}$. By the converse part of Lemma 2.2, there exists a compactly supported probability measure ω such that $M_\nu(z) = 1/\big(1 - z^2 M_\omega(z)\big)$. Then (2.6) defines a measure ν_1 with $r_{\nu_1}(z) = zM_\omega(z)$ and the relation $\mathbb{V}(z) = \mathbb{V}_{\nu_1}(z) - z^2$ holds by the proof of part (iv). ■

2.4. Proof of Theorem 1.2. The proof of Proposition 2.3 uses S-transforms from (2.8).

LEMMA 2.3. *Suppose that* $\mu \neq \delta_0$ *is a probability measure with compact support in* $[0,\infty)$ *and that* $S_\mu(0) = 1$. *Then*

$$\mathbb{V}(z) = \frac{1+z}{S_\mu(z)} \tag{2.12}$$

is in $\mathcal{V}_\infty$.

Proof. Define $\omega(dx) = x\mu(dx)$ and note that this is a probability measure since $\omega(\mathbb{R}) = \int x\mu(dx) = 1/S_\mu(0) = 1$. Let ν be the $\boxplus$-infinitely divisible probability measure defined by (2.6). Then $M_\mu(z) = 1 + r_\nu(z)$, so (2.7) gives

$$r_\nu\left(\frac{z}{1+z}S_\mu(z)\right) = z.$$

Recalling that the composition inverse of $z \mapsto r(z)$ is $z/\mathbb{V}_\nu(z)$, in a neighborhood of $z = 0$ we get (2.12). ■

Lemma 2.3 yields a class of variance functions in $\mathcal{V}_\infty$ of the following form.

LEMMA 2.4. *Let* $\beta_1, \ldots, \beta_d > 0$, $p_1, \ldots, p_2 > 0$, *with* $\max\{p_j, 1/\beta_j\} \geqslant 1$. *Then the function*

$$\mathbb{V}(z) = (1+z)\prod_{j=1}^{d}(1+\beta_j z)^{p_j} \tag{2.13}$$

is in $\mathcal{V}_\infty$.

P r o o f. For $j \geqslant 1$, choose μ_j with $S_{\mu_j}(z) = (1+\beta_j z)^{-p_j}$, see [3], [22]. Define $\mu = \mu_1 \boxtimes \mu_2 \boxtimes \ldots \boxtimes \mu_d$ so that

$$S_\mu(z) = \prod_{j=1}^{d} (1+\beta_j z)^{-p_j}.$$

Lemma 2.3 completes the proof. ■

We will deduce sufficiency in Theorem 1.2 from the following general result.

PROPOSITION 2.3. *Let us assume that* $d \in \mathbb{N}$, $a, b \in \mathbb{R}$, $c > 0$, $b_1, \ldots, b_d > 0$, $p_1, \ldots, p_d > 0$ *and that* $\max\{p_j, c/b_j\} \geqslant 1$ *for* $1 \leqslant j \leqslant d$. *Put*

$$\mathbb{V}(z) = az + bz^2 + (1+cz) \prod_{j=1}^{d} (1+b_j z)^{p_j}.$$

If $b \geqslant -1$, *then* $\mathbb{V} \in \mathcal{V}$. *If* $b \geqslant 0$, *then* $\mathbb{V} \in \mathcal{V}_\infty$.

In the present paper we are mainly interested in polynomial variance functions, however here we would like to emphasize that the exponents p_j do not have to be integers; for example, $(1+z)^{\sqrt{2}}$ or $(1+z)^{3/2}(1+2z)^{3/2}$ are variance functions in $\mathcal{V}_\infty$.

P r o o f. Put $\mathbb{V}_1(z) := 1 + az + bz^2$, $\mathbb{V}_2(z) := (1+cz)\prod_{j=1}^{d}(1+b_j z)^{p_j}$. Then $\mathbb{V}_2 \in \mathcal{V}_\infty$ in view of Lemma 2.4 with $\beta_j = b_j/c$. If $b \geqslant 0$, then $\mathbb{V}_1 \in \mathcal{V}_\infty$ (see [11], Theorem 3.2 and the comments therein), and consequently $\mathbb{V}(z) = \mathbb{V}_1(z) + \mathbb{V}_2(z) - 1 \in \mathcal{V}_\infty$ by Theorem 1.1(iii), which proves the second part. When $b \geqslant -1$, we apply Theorem 1.1(iv) to $z^2 + \mathbb{V}_1(z) + \mathbb{V}_2(z) - 1 \in \mathcal{V}_\infty$. ■

We are now ready to prove Theorem 1.2.

P r o o f o f T h e o r e m 1.2. The case $c = 0$ is well understood: $\mathbb{V} \in \mathcal{V}$ if and only if $b + 1 \geqslant 0$ and $\mathbb{V} \in \mathcal{V}_\infty$ if and only if $b \geqslant 0$, see [11]. In view of Lemma 2.1 we can assume that $c > 0$.

Applying Proposition 2.3 with $d = 1$, $b_1 = c$ and $p_1 = 2$, we see that

$$az + bz^2 + (1+cz)^3 = 1 + (a+3c)z + (b+3c^2)z^2 + c^3 z^3$$

is in $\mathcal{V}$ for any $b \geqslant -1$, and in $\mathcal{V}_\infty$ for any $b \geqslant 0$, with any real a, c. Replacing $a + 3c, b + 3c^2, c^3$ by a, b, c, respectively, we get the sufficient conditions for $\mathbb{V} \in \mathcal{V}$ and for $\mathbb{V} \in \mathcal{V}_\infty$, as stated (recall that $c > 0$).

It remains to show that if $b^3 < 27c^2$, then $\mathbb{V}(z) = 1 + az + bz^2 + cz^3$ is not in $\mathcal{V}_\infty$. By Theorem 1.1(ii), without loss of generality we may assume $a = 0$.

We proceed by contradiction. Suppose $\mathbb{V} \in \mathcal{V}_\infty$. If $b > 0$, then by scaling we would get $1 + z^2 + cz^3 \in \mathcal{V}_\infty$ for some (different) $c^2 > 1/27$. By Proposition 2.2

this would mean that there exists a compactly supported probability measure ω with $r_\omega(z) = z + cz^2$, contradicting Corollary 2.5 in [15], which says this to be possible if and only if $c^2 \leqslant 1/27$.

Suppose now that $b \leqslant 0$. Then by Theorem 1.1(iii) with $\mathbb{V}_2(z) = 1 + |b|z^2$ we would get $1 + cz^3 \in \mathcal{V}_\infty$. Since $c > 0$, we would be able to rescale and get, say, $1 + 2z^3 \in \mathcal{V}_\infty$. Using Theorem 1.1(iii) again, we would get $1 + z^2 + 2z^3 \in \mathcal{V}_\infty$, contradicting Corollary 2.5 in [15] again. (In fact, as explained in Remark 4.2 below, $1 + 2z^3$ is not in $\mathcal{V}$.) ■

3. VARIANCE FUNCTIONS AND POLYNOMIALS

In general, if $\mathbb{V}$ is (real) analytic at zero and $\mathbb{V}(0) \neq 0$, it is easy to see that expansion (1.6) holds, and its coefficients are polynomials $\{P_n(x)\}$ which solve the recursion

$$xP_n(x) = P_{n-1}(x) + \sum_{k=0}^{n} \frac{\mathbb{V}^{(k)}(0)}{k!} P_{n+1-k}(x), \quad n \geqslant 0, \tag{3.1}$$

with initial polynomials $P_{-1}(x) = 0$ and $P_0(x) = 1$. (In particular, polynomials $\{P_n\}$ are monic when $\mathbb{V}(0) = 1$.) To derive (3.1), multiply (1.2) by $m\mathbb{V}(m)$, expand $\mathbb{V}$ into the power series at $m = 0$, expand $f(x, m)$ into power series (recall that $\mathbb{V}(0) \neq 0$), and compare the coefficients at the powers of m.

We therefore consider a slightly more general recursion than (1.7). Suppose that polynomials $\{P_n\}$ satisfy the recursion

$$xP_n(x) = P_{n-1}(x) + \sum_{k=0}^{n} a_k P_{n+1-k}(x), \quad n \geqslant 0, \tag{3.2}$$

with $a_0 \neq 0$ and initial polynomials $P_{-1}(x) = 0$ and $P_0(x) = 1$.

PROPOSITION 3.1. *Suppose that there are $A, R > 0$ such that $|a_k| \leqslant AR^k$ for all $k = 0, 1, \ldots$ Define $\mathbb{V}(z) = \sum_{k=0}^{\infty} a_k z^k$ for $|z| < 1/R$.*

(i) *If polynomials $\{P_n\}$ satisfy recursion* (3.2), *then*

$$\sum_{n=0}^{\infty} P_n(x) z^n = \frac{\mathbb{V}(z)}{\mathbb{V}(z) + z(z - x)} \tag{3.3}$$

and the series converges uniformly over $x \in K$ for any compact set $K \subset \mathbb{R}$. That is, there is $r > 0$ that does not depend on $x \in K$ such that the series converges uniformly over $x \in K$ for all $|z| < r$.

(ii) *If polynomials $\{P_n\}$ satisfy recursion* (3.2) *and there is a non-degenerate compactly supported centered probability measure ν such that $\int P_n(x)\nu(dx) = 0$ for all $n \geqslant 1$, then $a_0 = \mathbb{V}(0) > 0$ and $\mathbb{V}(\cdot)$ is the variance function of a CSK family generated by ν.*

(iii) *If* $\mathbb{V}(\cdot)$ *is a variance function of a CSK family generated by a non-degenerate centered compactly supported probability measure* ν *and* $\{P_n\}$ *are polynomials from* (1.6), *then* $\int P_n(x)\nu(dx) = 0$ *for* $n \geqslant 1$.

Proof. (i) Since $P_0(x) = 1$, without loss of generality, we may assume that $A = 1$. Let $M = \sup_{x\in K} |x|$. Choose $C > R$ such that

$$\frac{M}{C} + \frac{1}{C^2} + \frac{R}{C-R} \leqslant |a_0|. \tag{3.4}$$

We now check by induction that with this choice of C we have

$$\sup_{x\in K} |P_n(x)| \leqslant C^n \quad \text{for all } n \geqslant 0. \tag{3.5}$$

Clearly, $|P_0(x)| \leqslant 1 \leqslant C^0$ and $\sup_{x\in K} |P_1(x)| = \sup_{x\in K} |x/a_0| \leqslant M/|a_0| \leqslant C$. Suppose that $N \geqslant 1$ is such that (3.5) holds for all P_n with $n \leqslant N$. From (3.2) we see that

$$\begin{aligned}
&|a_0| \sup_{x\in K} |P_{N+1}(x)| \\
&\leqslant \sup_{x\in K} |xP_N(x)| + \sup_{x\in K} |P_{N-1}(x)| + \sum_{k=1}^{N} R^k \sup_{x\in K} |P_{N+1-k}(x)| \\
&\leqslant MC^N + C^{N-1} + C^{N+1} \sum_{k=1}^{N} \Big(\tfrac{R}{C}\Big)^k \leqslant C^{N+1}\Big(\tfrac{M}{C} + \tfrac{1}{C^2} + \tfrac{R}{C-R}\Big) \leqslant |a_0| C^{N+1}
\end{aligned}$$

by (3.4). This proves (3.5) by induction.

From (3.5) it is clear that with $r = 1/C$ the series (1.6) converges uniformly over $x \in K$ for all (complex) $|m| < r$.

To identify the limit, denote the sum of the series by $\varphi(x,z)$. Multiplying (3.2) by $z^n \neq 0$ and summing over n, we get

$$x\varphi(x,z) = z\varphi(x,z) + \frac{1}{z}\sum_{n=0}^{\infty}\sum_{k=0}^{n} z^k a_k z^{n+1-k} P_{n+1-k}(x). \tag{3.6}$$

Changing the order of summation, we obtain

$$\begin{aligned}
\sum_{n=0}^{\infty}\sum_{k=0}^{n} z^k a_k z^{n+1-k} P_{n+1-k}(x) &= \sum_{k=0}^{\infty} z^k a_k \sum_{n=k}^{\infty} z^{n+1-k} P_{n+1-k}(x) \\
&= \sum_{k=0}^{\infty} z^k a_k \big(\varphi(x,z) - 1\big) = \mathbb{V}(z)\big(\varphi(x,z) - 1\big).
\end{aligned}$$

Inserting this into (3.6), we see that

$$x\varphi(x,z) = z\varphi(x,z) + \frac{\mathbb{V}(z)}{z}\big(\varphi(x,z) - 1\big).$$

The solution of this equation is $\varphi(x,z) = \frac{\mathbb{V}(z)}{\mathbb{V}(z)+z(z-x)}$, as claimed.

(ii) Since the polynomial $P_2(x) = x^2/a_0^2 - a_1 x/a_0^2 - 1/a_0$ integrates to zero, and $\int x\nu(dx) = 0$ by assumption, we see that $a_0 > 0$. So $\mathbb{V}(m) \geqslant 0$ in some neighborhood of zero and on the support of ν the generating function $f(x,m) \geqslant 0$ for m small enough.

Since $\int P_n(x)\nu(dx) = 0$ for $n \geqslant 1$, and by part (i) the series (3.3) converges uniformly on the support of ν, integrating term-by-term we get $\int f(x,m)\nu(dx) = 1$, i.e., $\mathbb{V}$ is the variance function of the CSK family generated by ν.

(iii) Suppose that $\{P_n\}$ are polynomials from (1.6) and $\nu(dx)$ has compact support. Then (3.1) implies (3.2) with $a_k = \mathbb{V}^{(k)}(0)/k!$. Since $\mathbb{V}(0) \neq 0$ and $\mathbb{V}$ is real analytic, one can find $R > 1$ such that $|\mathbb{V}^{(k)}(0)| \leqslant k!\mathbb{V}(0)R^k$, so the assumption on the growth of $|a_k|$ is satisfied. By uniform convergence for all small enough m we can integrate series (1.6) term-by-term. We get

$$1 = \int f(x;m)\nu(dx) = 1 + \sum_{n=1}^{\infty} m^n \int P_n(x)\nu(dx).$$

Thus $\int P_n(x)\nu(dx) = 0$ for all $n \geqslant 1$. ∎

Next, we relate polynomial variance functions to $(\nu; d)$-orthogonality.

PROPOSITION 3.2. *Suppose that $\mathbb{V}$ is a variance function of a CSK family generated by a non-degenerate compactly supported probability measure ν with mean zero and variance one. Consider the family of polynomials $\{P_n(x)\}$ with generating function* (1.6), *where $f(x,m)$ is given by* (1.1). *Then*:

(i) $\int P_n(x)\nu(dx) = 0$ *for* $n \geqslant 1$.

(ii) *Assume that the polynomial $P_2(x)$ is orthogonal in $L_2(\nu)$ to all polynomials $\{P_n(x) : n \geqslant 2 + d\}$. Then the family $\{P_n(x)\}$ is $(\nu; d)$-orthogonal, satisfies recursion* (1.7) *and $\mathbb{V}$ is a polynomial of degree at most $d + 1$.*

(iii) *Conversely, if the variance function $\mathbb{V}$ of a CSK family generated by measure ν is a polynomial of degree at most $d + 1$, then the polynomials from expansion* (1.6) *are $(\nu; d)$-orthogonal.*

Proof. (i) This is included in Proposition 3.1(iii).

(ii) Since $P_2(x) = x^2 - \mathbb{V}'(0)x - 1$ and $\mathbb{V}(m)$ is given by (1.5), we see that

$$\int P_2(x) f(x,m)\nu(dx) = \mathbb{V}(m) + m^2 - \mathbb{V}'(0)m - 1. \tag{3.7}$$

On the other hand, due to uniform convergence (Lemma 3.1), for all small enough m we can integrate series (1.6) term-by-term. Since, by assumption, it follows that $\int P_2(x)P_k(x)\nu(dx) = 0$ for $k \geqslant 2 + d$, we get

$$\begin{aligned} \int P_2(x) f(x,m)\nu(dx) &= \int P_2(x) \sum_{k=0}^{\infty} P_k(x) m^k \nu(dx) \\ &= \sum_{k=0}^{\infty} m^k \int P_2(x)P_k(x)\nu(dx) = \sum_{k=0}^{d+1} m^k \int P_2(x)P_k(x)\nu(dx). \end{aligned} \tag{3.8}$$

Thus, comparing the right-hand sides of (3.7) and (3.8), we see that

$$\mathbb{V}(m) = 1 - m^2 + \mathbb{V}'(0)m + \sum_{k=0}^{d+1} c_k m^k$$

is a polynomial of degree at most $d+1$, where $c_k = \int P_2(x)P_k(x)\nu(dx)$.

(iii) We now prove the converse claim. If $\mathbb{V}$ is a polynomial of degree $d+1$, then recursion (3.1) becomes (1.7). Proposition 3.1(iii) gives $\int P_n(x)\nu(dx) = 0$ for $n \geqslant 1$. Noting that $\{P_j(x) : j \leqslant k\}$ span the same subspace as monomials, to prove $(\nu; d)$-orthogonality it remains to verify that

$$\int x^k P_n(x)\nu(dx) = 0 \quad \text{for all } n \geqslant 2 + (k-1)d \tag{3.9}$$

for all $k \in \mathbb{N}$.

The proof proceeds by induction on k. Consider first the case $k = 1$. From (1.7) we see that xP_n is a linear combination of $P_{n+1}, P_n, \ldots, P_{n-d}$. Thus, $\int xP_n(x)\nu(dx) = 0$ if $n \geqslant d+1$. If $n = 2, \ldots, d$, then (1.7) shows that xP_n is a linear combination of $P_{n+1}, \ldots, P_1$, thus $\int xP_n(x)\nu(dx) = 0$, too.

Suppose now that (3.9) holds for some $k \geqslant 1$. Take $n \geqslant 2 + kd$. Then $n > d+1$, so from (1.7) we see that the polynomial $x^{k+1}P_n(x)$ is a linear combination of polynomials $\{x^k P_j(x) : j = n-d, n-d+1, \ldots, n+1\}$. Since $j \geqslant n - d \geqslant 2 + kd - d = 2 + (k-1)d$, each of the polynomials $x^k P_j(x)$ in the linear combination satisfies the inductive assumption, $\int x^k P_j(x)\nu(dx) = 0$. Thus $\int x^{k+1}P_n(d)\nu(dx) = 0$, proving that (3.9) holds for all $k \in \mathbb{N}$. ■

Combining the above results with Theorem 1.2, we have the following; compare [32], Théorème 2.1, and [42], Theorem 3.1, where the authors study polynomials given by finite recursions under regularity conditions which fail in the case we are interested in. (The paper [16] gives a nice introduction to their theory.)

COROLLARY 3.1. *Let us consider polynomials $\{P_n(x)\}$ given by the four-step recursion*:

$$\begin{aligned} xP_1(x) &= P_2(x) + aP_1(x) + P_0(x),\\ xP_2(x) &= P_3(x) + aP_2(x) + bP_1(x),\\ xP_n(x) &= P_{n+1}(x) + aP_n(x) + bP_{n-1}(x) + cP_{n-2}(x), \qquad n \geqslant 3, \end{aligned}$$

with $P_0(x) = 1, P_1(x) = x$. Then the following conditions are equivalent:

(i) $b^3 \geqslant 27c^2$.

(ii) *Polynomials $\{P_n\}$ are $(\nu; 2)$-orthogonal for some probability measure ν (which then necessarily has mean zero, variance one and compact support).*

Proof. If $b^3 \geqslant 27c^2$, then by Theorem 1.2, $\mathbb{V}(m) = 1 + am + (b-1)m^2 + cm^3$ is a variance function, and (1.6) holds. So Proposition 3.2(iii) implies (ii). Conversely, if (ii) holds, then, by Proposition 3.1(ii), we have $\mathbb{V}(m) = 1 + am + (b-1)m^2 + cm^3 \in \mathcal{V}$, so Theorem 1.2 implies (i). ■

Probability and Mathematical Statistics 39, z. 2, 2019

3.1. Proof of Theorem 1.3. By Proposition 3.1, for a family of monic polynomials $\{P_n(x)\}$, recursion (1.7) holds if and only if the generating function (1.6) is given by (1.1) with

$$\mathbb{V}(m) = 1 + b_1 m + (b_2 - 1)m^2 + \sum_{k=3}^{d+1} b_k m^k. \tag{3.10}$$

Thus, statements (i) and (ii) are equivalent.

Proposition 3.2(ii) gives the implication (iii)⇒(i), as it says that already a special case of $(\nu; d)$-orthogonality implies (i); the implication (i)⇒(iii) is Proposition 3.2(iii).

4. ADDITIONAL RESULTS AND COMMENTS

4.1. A combinatorial example. Consider the probability distribution on $[0, \infty)$, which in [34] was denoted by $\mu(3,1)$. Its moments are $\frac{1}{3n+1}\binom{3n+1}{n}$ (Fuss numbers of order three, A001764 in OEIS) and the moment generating function, denoted by $\mathcal{B}_3(z)$, is

$$\mathcal{B}_3(z) = \frac{3}{3 - 4\sin^2\alpha} = \frac{2\sin\alpha}{\sqrt{3z}},$$

where $\alpha = \frac{1}{3}\arcsin\sqrt{27z/4}$. The first expression was obtained in [35], the second can be obtained by elementary manipulations. The density function was described in [40], [36]. We are going to study a probability distribution which is a transformation of $\mu(3,1)$.

PROPOSITION 4.1. *If μ is a probability measure on $[0,\infty)$, with the moment generating function $M_\mu(z)$, then there exists a probability measure μ_1 on $[0,\infty)$ such that $M_{\mu_1}(z) = \frac{1}{1-zM_\mu(z)}$.*

Proof. This is a consequence of Proposition 6.1 in [7] with

$$\psi(z) = \frac{zM_\mu(z)}{1 - zM_\mu(z)}.$$

Namely, since $M_\mu(z)$ is $\mathbb{C}^+ \to \mathbb{C}^+$, the function

$$\frac{z}{1 - zM_\mu(z)} = \frac{z - |z|^2\overline{M_\mu(z)}}{|1 - zM_\mu(z)|^2}$$

is also $\mathbb{C}^+ \to \mathbb{C}^+$. ■

Let μ denote the probability measure which satisfies

$$M_\mu(z) = \frac{1}{1 - z\mathcal{B}_3(z)} = \frac{3}{3 - 2\sqrt{3z}\sin\alpha},$$

Probability and Mathematical Statistics 39, z. 2, 2019

$\alpha = \frac{1}{3}\arcsin\sqrt{27z/4}$. This identity implies that moments $s(n)$ of μ satisfy the following recurrence relation: $s(0) = 1$ and for $n \geqslant 1$

$$s(n) = \sum_{i=0}^{n-1} \frac{1}{3i+1}\binom{3i+1}{i} s(n-1-i).$$

This sequence appears in OEIS as A098746:

$$1, 1, 2, 6, 23, 102, 495, 2549, 13682, 75714, 428882, \ldots,$$

and counts permutations which avoid patterns 4231 and 42513, see [1], [33]. For $n \geqslant 1$ we have also

$$s(n) = \sum_{i=0}^{n} \frac{n-i}{n+2i}\binom{n+2i}{i}.$$

From the equation $\mathcal{B}_3(z) = 1 + z\mathcal{B}_3(z)^3$ (see [19]) we obtain the identity

$$zM_\mu(z)^2\big(M_\mu(z) - 1\big) = z^2 M_\mu(z)^3 + \big(M_\mu(z) - 1\big)^3, \tag{4.1}$$

which yields the free S-transform

$$S_\mu(z) = \frac{1 + z + \sqrt{(1+z)(1-3z)}}{2(1+z)}.$$

Substituting $zM_\mu(z) \mapsto z$ in (4.1) and applying (2.1), we get

$$z\big(R_\mu(z) + 1\big)R_\mu(z) = z^2\big(R_\mu(z) + 1\big) + R_\mu(z)^3.$$

Putting $R_\mu(z) = zr_\mu(z)$ yields

$$r_\mu(z) - 1 = zr_\mu(z)\big(1 - r_\mu(z) + r_\mu(z)^2\big). \tag{4.2}$$

This implies that $r_\mu(z)$ is the generating function for the sequence A106228:

$$1, 1, 2, 6, 21, 80, 322, 1347, 5798, 25512, 114236, 518848, \ldots,$$

which counts Motzkin paths of a special kind. These are free cumulants of μ, namely $\kappa_n(\mu) = A106228(n-1)$ for $n \geqslant 1$. Note that the shifted sequence

$$1, 2, 6, 21, 80, 322, 1347, 5798, 25512, 114236, 518848, \ldots$$

is not positive definite, for example $\det\big(\kappa_{i+j+2}(\mu)\big)_{i,j=0}^{5} = -3374$, so μ is not $\boxplus$-infinitely divisible, see [39].

From (4.2) one can read out that the centered measure ν with $r_\nu(z) = r_\mu(z) - 1$ (so that ν is the translation of μ by -1) has

$$\mathbb{V}_\nu(z) = 1 + 2z + 2z^2 + z^3$$

and the comment above (or Theorem 1.2) shows that $\mathbb{V}_\nu \notin \mathcal{V}_\infty$.

4.2. More on generating functions. Several authors considered families of polynomials $\{T_n\}$ with the generating function of the form

$$\sum_{n=0}^{\infty} T_n(x)z^n = \frac{M(z)}{N(z)-zx}, \tag{4.3}$$

where $z \mapsto M(z)$ and $z \mapsto N(z)$ are analytic functions in the neighborhood of $0 \in \mathbb{C}$ with $M(0) = N(0) \neq 0$. See [2], Lemma 2, with $u(z) = z/M(z)$ and $f(z) = N(z)/z$ or the generating function in [18], (3.10). (See also [9], [28] and the discussion in [10].)

At first sight (4.3) looks more general than (1.6), but in fact the difference is superficial. The following result was inspired by results in Section 3.2 of [18].

PROPOSITION 4.2. *Let ν be a non-degenerate compactly supported probability measure with mean zero. Suppose the sequence of polynomials $\{T_n\}$ has generating function* (4.3), $\int T_n(x)\nu(dx) = 0$ *for* $n \geqslant 1$, *and* $\int T_n(x)T_1(x)\nu(dx) = 0$ *for* $n \geqslant 2$. *Let $\mathbb{V}$ be the variance function of the CSK family generated by ν.*

Then, with $t = \mathbb{V}(0)/M(0)$ we have

$$M(z) = \mathbb{V}(tz)/t \quad \textit{and} \quad N(z) = \mathbb{V}(zt)/t + tz^2.$$

In particular, $T_n(x) = t^n P_n(x)$ for all $n = 0, 1, 2, \ldots$, where the sequence $\{P_n\}$ is given by expansion (1.6) *for the density of the CSK family generated by ν.*

(Polynomials $\{P_n\}$ are monic if the variance of ν is one.)

We remark that if in addition, $\int T_2(x)T_n(x)\nu(dx) = 0$ for $n \geqslant d+2$, then by Proposition 3.2 the variance function of the CSK family generated by ν is a polynomial of degree at most $d+1$. When $d = 1$, this recovers Corollary 3.6 in [18]. For related results with exponential rather than Cauchy generating functions see [25], [43].

In order to be able to integrate the series term-by-term, we first confirm that the series converges uniformly over x from any compact set. (Compare Proposition 3.1(i).)

LEMMA 4.1. *Fix $M > 0$. Then there is $r > 0$ such that the series* (4.3) *converges for all $|x| < M$ and all $|m| < r$.*

Proof. The x-dependent radius $r(x)$ of convergence of the series is the minimum modulus root of the equation $N(z) - zx = 0$. Since $N(0) \neq 0$, it is clear that for every $M > 0$ there is $r > 0$ such that $|N(z)| > |zx|$ for all $|z| < r$ and all $|x| < M$. So there are no roots in the disk $|z| < r$ and the radius of convergence is at least r. ■

Proof of Proposition 4.2. Choose $r > 0$ such that the series (4.3) converges for all x from the support of ν. Integrating term-by-term with respect to ν, we get

$$\int \Big(\sum_{n=0}^{\infty} T_n(x) z^n \Big) \nu(dx) = \sum_{n=0}^{\infty} \int T_n(x) z^n \nu(dx) = \int T_0(x) \nu(dx)$$
$$= \int \frac{M(0)}{N(0)} \nu(dx) = 1.$$

We therefore get

$$\int \frac{M(z)}{N(z) - zx} \nu(dx) = 1 \tag{4.4}$$

for all real z close enough to zero.

Using this and (4.3), we compute $T_1(x) = \big(x + M'(0) - N'(0)\big)/M(0)$. Since $\int T_1(x)\nu(dx) = 0$, we see that $M'(0) = H'(0)$ and $T_1(x) = \alpha x$ with $\alpha = 1/M(0)$ $\neq 0$.

Since $T_1(x)$ is bounded on the support of ν and the series converges uniformly, integrating term-by-term, we get

$$\int \Big(\sum_{n=0}^{\infty} T_n(x) T_1(x) z^n \Big) \nu(dx) = \sum_{n=0}^{\infty} \int T_n(x) T_1(x) z^n \nu(dx) \tag{4.5}$$
$$= z \int T_1^2(x) \nu(dx) = \alpha^2 \mathbb{V}(0) z,$$

where $\mathbb{V}(0) > 0$ is the variance of ν (recall that ν is non-degenerate). On the other hand, using partial fractions, we get

$$\int \frac{M(z)}{N(z) - zx} T_1(x) \nu(dx) = \alpha \int M(z) \bigg(\frac{N(z)}{z\big(N(z) - xz\big)} - \frac{1}{z} \bigg) \nu(dx) \tag{4.6}$$
$$= \frac{\alpha N(z)}{z} \int \frac{M(z)}{N(z) - xz} \nu(dx) - \frac{\alpha M(z)}{z} \int 1 \nu(dx) = \alpha \frac{N(z) - M(z)}{z}.$$

(Here, we used (4.4) and the fact that ν is a probability measure.) Therefore, with $t = \alpha \mathbb{V}(0) = \mathbb{V}(0)/M(0) \neq 0$, since (4.5) and (4.6) are equal, we get $N(z) = M(z) + tz^2$, and (4.4) takes the form

$$\int \frac{M(z)}{M(z) + tz^2 - zx} \nu(dx) = 1.$$

Substituting $z = m/t$ and setting $\mathbb{V}(m) = tM(m/t)$, we see that

$$\int \frac{\mathbb{V}(m)}{\mathbb{V}(m) + m(m-x)} \nu(dx) = 1.$$

This shows that $\mathbb{V}(m) = tM(m/t)$ is the variance function of the CSK family generated by ν, and it defines the corresponding polynomials $\{P_n\}$ via (1.6).

To relate polynomials T_n and polynomials P_n, we use the above identities to rewrite (4.3) as follows:

$$\sum_{n=0}^{\infty} \frac{T_n(x)}{t^n} m^n = \frac{M(m/t)}{N(m/t) - mx/t} = \frac{\mathbb{V}(m)}{\mathbb{V}(m) + m(m-x)} = \sum_{n=0}^{\infty} P_n(x) m^n. \quad \blacksquare$$

4.3. Sharpness of some results.

REMARK 4.1. *Corollary* 2.5 *in* [15] *implies sharp results about general quartic polynomials. For example, one can deduce that* $1 + az^4 \in \mathcal{V}$ *if and only if* $-1 \leqslant 12a \leqslant 3$.

REMARK 4.2. *Theorem* 1.1(iii) *does not extend to* $\mathbb{V}_1, \mathbb{V}_2 \in \mathcal{V}$. *To see this, consider* $\mathbb{V}_1(z) = \mathbb{V}_2(z) = 1 + z^3/6$, *which is in* $\mathcal{V}$ *by Theorem* 1.2. *Applying the operation* $\mathbb{V}_1 + \mathbb{V}_2 - 1$ *twelve times, we would get* $1 + 2z^3 \in \mathcal{V}$. *The latter is not possible. Using recursion* (1.7) *and Proposition* 3.2(i), *one can compute low order moments of the measure corresponding to the variance function* $1 + cz^3$. *The first six moments are* $(m_1, \ldots, m_6) = (0, 1, 0, 2, c, 5)$. *The* 4×4 *Hankel determinant of these moments is* $1 - c^2$, *so* $1 + 2z^3$ *is not a variance function.*

REMARK 4.3. *Theorem* 1.1(iv) *does not extend to* $\mathbb{V}_1 \in \mathcal{V}$. *To see this, consider* $\mathbb{V}_1(z) = 1 + 4z^2 + 2z^3$, *which is in* $\mathcal{V}$ *by Theorem* 1.2, *and apply the operation four times to get* $1 + 2z^3$, *which is not in* $\mathcal{V}$, *as was already noted in Remark* 4.2.

Acknowledgments. The authors thank Takahiro Hasebe and Kamil Szpojankowski for helpful discussions. Włodzimierz Bryc's research was supported in part by the Charles Phelps Taft Research Center at the University of Cincinnati. Wojciech Młotkowski is supported by NCN grant 2016/21/B/ST1/00628.

REFERENCES

[1] M. H. Albert, R. E. L. Aldred, M. D. Atkinson, H. P. van Ditmarsch, C. C. Handley, and D. A. Holton, *Restricted permutations and queue jumping*, Discrete Math. 287 (1–3) (2004), pp. 129–133.

[2] M. Anshelevich, *Free martingale polynomials*, J. Funct. Anal. 201 (1) (2003), pp. 228–261.

[3] O. Arizmendi and T. Hasebe, *Classical scale mixtures of Boolean stable laws*, Trans. Amer. Math. Soc. 368 (7) (2016), pp. 4873–4905.

[4] T. Banica, S. T. Belinschi, M. Capitaine, and B. Collins, *Free Bessel laws*, Canad. J. Math. 63 (1) (2011), pp. 3–37.

[5] S. T. Belinschi, *Complex Analysis Methods in Noncommutative Probability*, ProQuest LLC, Ann Arbor, MI, 2005.

[6] H. Bercovici and V. Pata, *A free analogue of Hincin's characterization of infinite divisibility*, Proc. Amer. Math. Soc. 128 (4) (2000), pp. 1011–1015.

[7] H. Bercovici and D. Voiculescu, *Free convolution of measures with unbounded support*, Indiana Univ. Math. J. 42 (3) (1993), pp. 733–773.

[8] H. Bercovici and D. Voiculescu, *Superconvergence to the central limit and failure of the Cramér theorem for free random variables*, Probab. Theory Related Fields 103 (2) (1995), pp. 215–222.

[9] M. Bożejko and N. Demni, *Generating functions of Cauchy–Stieltjes type for orthogonal polynomials*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 12 (1) (2009), pp. 91–98.

[10] M. Bożejko and N. Demni, *Topics on Meixner families*, in: *Noncommutative Harmonic Analysis with Applications to Probability II*, M. Bożejko (Ed.), Banach Center Publ., Warsaw 2010, pp. 61–74.

[11] W. Bryc, *Free exponential families as kernel families*, Demonstr. Math. 42 (3) (2009), pp. 657–672.

[12] W. Bryc, R. Fakhfakh, and A. Hassairi, *On Cauchy–Stieltjes kernel families*, J. Multivariate Anal. 124 (2014), pp. 295–312.

[13] W. Bryc and A. Hassairi, *One-sided Cauchy–Stieltjes kernel families*, J. Theoret. Probab. 24 (2) (2011), pp. 577–594.

[14] W. Bryc and M. Ismail, *Approximation operators, exponential, and q-exponential families*, arxiv.org/abs/math.ST/0512224 (2005).

[15] G. P. Chistyakov and F. Götze, *Characterization problems for linear forms with free summands*, arXiv:1110.1527 (2011).

[16] Z. da Rocha, *Shohat–Favard and Chebyshev's methods in d-orthogonality*, Numer. Algorithms 20 (2–3) (1999), pp. 139–164.

[17] A. Di Bucchianico and D. E. Loeb, *Natural exponential families and umbral calculus*, in: *Mathematical Essays in Honor of Gian-Carlo Rota (Cambridge, MA, 1996)*, B. E. Sagan and R. P. Stanley (Eds.), Progr. Math., Vol. 161, Birkhäuser, Boston, MA, 1998, pp. 195–211.

[18] R. Fakhfakh, *Characterization of quadratic Cauchy–Stieltjes kernel families based on the orthogonality of polynomials*, J. Math. Anal. Appl. 459 (1) (2018), pp. 577–589.

[19] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, second edition, Addison-Wesley, Reading, MA, 1994.

[20] A. Hassairi and M. Zarai, *Characterization of the cubic exponential families by orthogonality of polynomials*, Ann. Probab. 32 (3B) (2004), pp. 2463–2476.

[21] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Mathematical Society, Providence, RI, 2000.

[22] M. Hinz and W. Młotkowski, *Free powers of the free Poisson measure*, Colloq. Math. 123 (2) (2011), pp. 285–290.

[23] M. E. H. Ismail, *Classical and Quantum Orthogonal Polynomials in One Variable*, Cambridge University Press, Cambridge 2005.

[24] M. E. H. Ismail and C. P. May, *On a family of approximation operators*, J. Math. Anal. Appl. 63 (2) (1978), pp. 446–462.

[25] C. C. Kokonendji, *Characterizations of some polynomial variance functions by d-pseudo-orthogonality*, J. Appl. Math. Comput. 19 (1–2) (2005), pp. 427–438.

[26] C. C. Kokonendji, *On d-orthogonality of the Sheffer systems associated to a convolution semigroup*, J. Comput. Appl. Math. 181 (1) (2005), pp. 83–91.

[27] M. G. Kreĭn and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*, Transl. Math. Monogr., Vol. 50, American Mathematical Society, Providence, RI, 1977.

[28] I. Kubo, H. Kuo, and S. Namli, *The characterization of a class of probability measures by multiplicative renormalization*, Commun. Stoch. Anal. 1 (3) (2007), pp. 455–472.

[29] G. Letac, *Lectures on Natural Exponential Families and Their Variance Functions*, Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro 1992.

[30] G. Letac and M. Mora, *Natural real exponential families with cubic variance functions*, Ann. Statist. 18 (1) (1990), pp. 1–37.

[31] H. Maassen, *Addition of freely independent random variables*, J. Funct. Anal. 106 (2) (1992), pp. 409–438.

[32] P. Maroni, *L'orthogonalité et les récurrences de polynômes d'ordre supérieur à deux*, Ann. Fac. Sci. Toulouse Math. (5) 10 (1) (1989), pp. 105–139.

[33] M. Martinez and C. Savage, *Patterns in inversion sequences II: Inversion sequences avoiding triples of relations*, J. Integer Seq. 21 (2) (2018), Art. 18.2.2.

[34] W. Młotkowski, *Fuss–Catalan numbers in noncommutative probability*, Doc. Math. 15 (2010), pp. 939–955.

[35] W. Młotkowski and K. A. Penson, *Probability distributions with binomial moments*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 17 (2) (2014), 1450014.

[36] W. Młotkowski, K. A. Penson, and K. Życzkowski, *Densities of the Raney distributions*, Doc. Math. 18 (2013), pp. 1573–1596.

[37] C. N. Morris, *Natural exponential families with quadratic variance functions*, Ann. Statist. 10 (1) (1982), pp. 65–80.

[38] A. Nica and R. Speicher, *On the multiplication of free N-tuples of noncommutative random variables*, Amer. J. Math. 118 (4) (1996), pp. 799–837.
[39] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, Cambridge 2006.
[40] K. A. Penson and A. I. Solomon, *Coherent states from combinatorial sequences*, in: *Quantum Theory and Symmetries (Kraków, 2001)*, World Sci. Publ., River Edge, NJ, 2002, pp. 527–530.
[41] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, Notices Amer. Math. Soc. 50 (8) (2003), pp. 912–915.
[42] J. Van Iseghem, *Approximants de Padé vectoriels*, PhD thesis, Université des sciences et techniques de Lille-Flandres-Artois, 1987.
[43] S. Varma, *A characterization theorem and its applications for d-orthogonality of Sheffer polynomial sets*, arXiv:1603.07261 (2016).
[44] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups*, American Mathematical Society, Providence, RI, 1992.
[45] R. W. M. Wedderburn, *Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method*, Biometrika 61 (1974), pp. 439–447.
[46] J. Wesołowski, *Kernel families*, unpublished manuscript, 1999.

Włodzimierz Bryc
Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH 45221-0025, USA
E-mail: Wlodzimierz.Bryc@uc.edu

Raouf Fakhfakh
Mathematics Department
College of Science and Arts in Gurayat
Jouf University
Gurayat, Saudi Arabia, and
Laboratory of Probability and Statistics
Sfax University, Sfax
Tunisia
E-mail: fakhfakh.raouf@gmail.com

Wojciech Młotkowski
Institute of Mathematics
University of Wrocław
pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland
E-mail: mlotkow@math.uni.wroc.pl

Received on 16.6.2017;
revised version on 30.1.2018