What Are Cumulants ?

Dedicated to Professor Dietrich Morgenstern,

on the occasion of his seventy-fifth birthday

Lutz Mattner

Received: July 5, 1999

Communicated by Friedrich Götze

Abstract. Let $\mathcal{P}$ be the set of all probability measures on $\mathbb{R}$ possessing moments of every order. Consider $\mathcal{P}$ as a semigroup with respect to convolution. After topologizing $\mathcal{P}$ in a natural way, we determine all continuous homomorphisms of $\mathcal{P}$ into the unit circle and, as a corollary, those into the real line. The latter are precisely the finite linear combinations of cumulants, and from these all the former are obtained via multiplication by i and exponentiation.

We obtain as corollaries similar results for the probability measures with some or no moments finite, and characterizations of constant multiples of cumulants as affinely equivariant and convolution-additive functionals. The "no moments"-case yields a theorem of Halász. Otherwise our results appear to be new even when specialized to yield characterizations of the expectation or the variance.

Our basic tool is a refinement of the convolution quotient representation theorem for signed measures of Ruzsa & Székely.

1991 Mathematics Subject Classification: 60E05, 60E10, 60-03.
Keywords and Phrases: Additive functional, characteristic function, character, convolution, equivariance, expectation, Halász, historical note, homomorphism, mean, moment, multiplicative functional, Ruzsa, semi-invariant, semiinvariant, semigroup, Székely, variance.

1 Introduction, results, and easy proofs

1.1 Aim. Cumulants are certain functionals of probability measures. This paper attempts to explain more precisely what they are by characterizing them through their most useful properties. For simplicity, only the one-dimensional case of probability measures on $\mathbb{R}$ is treated. There the most familiar examples of cumulants are the expectation and the variance. Our results yield, in particular, new descriptions of the roles played by these latter two functionals in probability theory.

1.2 Guide. The definition of cumulants is recalled in Subsection 1.4 below, as formula (4). The useful properties of cumulants, referred to above, are the homomorphism property (5) and their transformation behaviour under affine mappings, (14). The relation between cumulants and moments is recalled in Subsection 1.5.

Subsection 1.6 introduces topologies on the domains of definition of the cumulants, with the aim of formulating regularity assumptions in our theorems and corollaries. That some regularity assumptions are actually necessary, at least in the results 1.8 – 1.12, is demonstrated in 1.20.

Theorem 1.8, characterizing the continuous characters of the semigroup $\mathrm{Prob}_\infty(\mathbb{R})$, is the main result of the present paper. Its natural forerunner from the literature, namely the theorem of Halász, is recalled in 1.10 below as a special case of Corollary 1.9.

Another corollary of Theorem 1.8, and perhaps the most interesting result of this paper, is the characterization of the finite linear combinations of cumulants as the continuous, $\mathbb{R}$-valued, and convolution-additive functionals of probability measures, stated in Theorem 1.11 and Corollary 1.12. Such results were conjectured by Kemperman (1972). By restricting the functionals to be $[0,\infty[$-valued, we arrive at a characterization of the variance in 1.14. [A related result of Martin Diaz (1977) is discussed in 1.22.]

Our next results, 1.17 and 1.18, are spezializations of 1.8 and 1.11 to scale equivariant functionals, the definition of which being recalled in 1.16.

As a further corollary, we obtain in 1.19 a characterization of the expectation as the only nontrivial continuous functional homomorphic with respect to additive and multiplicative convolutions.

Historical and etymological remarks on cumulants are given in Subsection 1.21.

Subsection 1.22 discusses some further references related to the present work.

Easy proofs are given immediately after the statement of a result in Section 1. The only difficult proof of this paper, needed for the "only if" part of our main result 1.8, is the content of Section 2. Its basic technical tool, refining the convolution quotient representation theorem for signed measures of Ruzsa & Székely (1983, 1985, 1988), is supplied in Subsection 2.5.

1.3 Some notation and conventions. The positive integers are denoted by $\mathbb{N}$, the nonnegative ones by $\mathbb{N}_0$.

If $\mathcal{X}$ is a set equipped with a σ-algebra $\mathcal{A}$, we let $\mathrm{Prob}(\mathcal{X})$ denote the set of all probability measures defined on $\mathcal{A}$. The real line $\mathbb{R}$ is understood to be equipped with its Borel σ-algebra. The convolution of $P, Q \in \mathrm{Prob}(\mathbb{R})$ is denoted by $P * Q$. We write δ_a for the Dirac measure concentrated at $a \in \mathbb{R}$, and $\delta := \delta_0$ for the one concentrated at zero. For the image measure of a probability measure P under a measurable function f, we use the notation $f \,\square\, P$. We write $\operatorname{supp} P$ for the support [= minimal closed set of probability one] of a $P \in \mathrm{Prob}(\mathbb{R})$.

$\mathrm{Prob}(\mathbb{R})$ will mainly be considered as a semigroup with respect to convolution. Homomorphisms of a semigroup [below always a sub-semigroup of $\mathrm{Prob}(\mathbb{R})$] into the multiplicative group $\mathbb{T}$ of complex numbers of absolute value one will be called *characters*, homomorphisms into the additive group $\mathbb{R}$ will be called *additive* functions.

1.4 Cumulants. We present below the usual introduction of cumulants and their most important properties. For $P \in \mathrm{Prob}(\mathbb{R})$, let $\widehat{P}$ denote its Fourier transform or characteristic function, defined by

$$\widehat{P}(t) \quad := \quad \int e^{itx}\, dP(x) \qquad (t \in \mathbb{R}). \tag{1}$$

The most important reason for considering Fourier transforms of probability measures is multiplicativity with respect to convolution:

$$(P * Q)\widehat{\ }(t) \quad = \quad \widehat{P}(t) \cdot \widehat{Q}(t) \qquad (P, Q \in \mathrm{Prob}(\mathbb{R}),\ t \in \mathbb{R}). \tag{2}$$

Let log denote the usual logarithm defined on, say, $\{z \in \mathbb{C} : |z - 1| < 1\}$. Let $P \in \mathrm{Prob}(\mathbb{R})$. Then $\widehat{P}$ is continuous with $\widehat{P}(0) = 1$, so that $\log \circ \widehat{P}$ is defined in some P-dependent neighbourhood of zero. Now put

$$\mathrm{Prob}_r(\mathbb{R}) \quad := \quad \left\{ P \in \mathrm{Prob}(\mathbb{R}) \,:\, \int |x|^r \, dP(x) < \infty \right\} \qquad (r \in \mathbb{N}_0), \tag{3}$$

and assume that $r \in \mathbb{N}$ and $P \in \mathrm{Prob}_r(\mathbb{R})$. Then $\widehat{P}$ and thus $\log \circ \widehat{P}$ is r times continuously differentiable in the neighbourhood of zero introduced above, and the number

$$\boxed{\kappa_r(P) \,:=\, i^{-r} \left(D^r \log \circ \widehat{P} \right)(0)} \tag{4}$$

is called the rth *cumulant* of P. [Readers wondering about this strange name are referred to Subsection 1.21.] It is easy to show that the cumulants are real-valued functionals. Their most important property, which obviously follows from (2) and (4), is additivity with respect to convolution:

$$\boxed{\kappa_r(P * Q) \,=\, \kappa_r(P) + \kappa_r(Q) \qquad (r \in \mathbb{N},\ P, Q \in \mathrm{Prob}_r(\mathbb{R}))}. \tag{5}$$

In other words: For each $r \in \mathbb{N}$, $(\mathrm{Prob}_r(\mathbb{R}), *)$ is a semigroup on which κ_r is an additive function.

1.5 EXAMPLES, EXPRESSION IN TERMS OF MOMENTS. The two most familiar examples of cumulants are the mean μ and the variance σ^2, since

$$\begin{aligned} \kappa_1(P) &= \mu(P) := \int x\,dP(x) \qquad (P \in \mathrm{Prob}_1(\mathbb{R})), \\ \kappa_2(P) &= \sigma^2(P) := \int (x-\mu(P))^2\,dP(x) \qquad (P \in \mathrm{Prob}_2(\mathbb{R})). \end{aligned}$$

These formulas are special cases of the relation between cumulants and the *moments*

$$\mu_r(P) := \textstyle\int x^r\,dP(x) = i^{-r}(D^n\widehat{P})(0) \qquad (r \in \mathbb{N}_0,\, P \in \mathrm{Prob}_r(\mathbb{R})).$$

One possibility of expressing this relation is to use the recursion

$$\mu_{r+1} = \sum_{l=0}^{r} \binom{r}{l} \mu_{r-l}\kappa_{l+1} \qquad (r \in \mathbb{N}_0), \tag{6}$$

which is easily proved using the Leibniz rule for the differentiation of a product and the representation of the moments as derivatives: For $P \in \mathrm{Prob}_{r+1}(\mathbb{R})$ put $\varphi := \widehat{P}$ and $\psi := \log\varphi$, in a neighbourhood of zero, and compute $D^{r+1}\varphi = D^r(\varphi\cdot D\psi) = \sum_{l=0}^{r}\binom{r}{l}(D^{r-l}\varphi)\cdot(D^{l+1}\psi)$, evaluate the extreme left and right hand sides at zero, and divide by i^{r+1}, to arrive at (6). Since the coefficients of μ_{r+1} and κ_{r+1} in (6) are both one, it follows by induction that

$$\kappa_r = \mu_r + \text{polynomial without constant term in } \mu_1, \ldots, \mu_{r-1} \ (r \in \mathbb{N}), \tag{7}$$

and that corresponding relations hold when μ and κ are interchanged. Various explicit fomulas derived from these relations and some examples of actual computations of cumulants may be found in Chapter 3 of Kendall, Stuart & Ord (1987). We merely note here two further examples, for convenience rewritten in terms of centered moments,

$$\begin{aligned} \kappa_3(P) &= \int (x-\mu(P))^3\,dP(x) \qquad (P \in \mathrm{Prob}_3(\mathbb{R})), \\ \kappa_4(P) &= \int (x-\mu(P))^4\,dP(x) - 3\left(\sigma^2(P)\right)^2 \qquad (P \in \mathrm{Prob}_4(\mathbb{R})). \end{aligned}$$

As one might suspect on seeing these formulas, the variance κ_2 is the only nonnegative cumulant. [This fact follows easily from 1.13 below, as can be seen from the proof of 1.14.]

1.6 Topologies on some subsets of $\mathrm{Prob}(\mathbb{R})$. One of our aims is to show that every "reasonable" homomorphism from $(\mathrm{Prob}_r(\mathbb{R}), *)$ into $(\mathbb{R}, +)$ is a linear combination of cumulants of order at most r. This is the content of Corollary 1.12, where "reasonable" is specified to mean "continuous". To this end we introduce here on each $\mathrm{Prob}_r(\mathbb{R})$ a topology. In order to make the continuity assumption in Corollary 1.12 weak, we have to choose a strong topology on $\mathrm{Prob}_r(\mathbb{R})$. We take the one induced by the weighted total variation metric d_r defined by

$$d_r(P,Q) \quad := \quad \int (1+|x|^r)\, d|P-Q|(x) \qquad (P,Q \in \mathrm{Prob}_r(\mathbb{R})). \tag{8}$$

We further consider

$$\mathrm{Prob}_\infty(\mathbb{R}) \quad := \quad \bigcap_{r\in\mathbb{N}_0} \mathrm{Prob}_r(\mathbb{R}),$$

which is the largest set of probability measures on which every cumulant is defined. We topologize $\mathrm{Prob}_\infty(\mathbb{R})$ by the family of metrics $(d_r : r \in \mathbb{N}_0)$.

1.7 Lemma. a) *Each $\kappa_r|\mathrm{Prob}_r(\mathbb{R})$ is continuous with respect to d_r.*

b) *Let $r \in \mathbb{N}$ and $c \in\ ]0,\infty[$. Then there exists a sequence (P_n) in $\mathrm{Prob}_\infty(\mathbb{R})$ with*

$$\lim_{n\to\infty} d_{r-1}(P_n,\delta) \quad = \quad 0, \tag{9}$$

$$\lim_{n\to\infty} \kappa_l(P_n) \quad = \quad 0 \qquad (l = 1,\dots,r-1), \tag{10}$$

$$\lim_{n\to\infty} \kappa_r(P_n) \quad = \quad c. \tag{11}$$

Proof. a) The functionals $(\mu_l : l = 1,\dots,r)$ are obviously continuous with respect to d_r, and (7) shows in particular that κ_r is a polynomial function of them.

b) We may restrict attention to those $n \in \mathbb{N}$ with $cn^{-r} \le 1$ and put $P_n := (1-cn^{-r})\delta + cn^{-r}\delta_n$. Then $P_n \in \mathrm{Prob}_\infty(\mathbb{R})$, and $d_{r-1}(P_n,\delta) = \frac{c}{n}$ yields (9). By part a), (9) implies (10). Finally, (11) follows from $\mu_l(P_n) = cn^{l-r}$ $(l = 1,\dots,r)$ and (7). □

1.8 Theorem (continuous characters of $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ is a continuous character iff*

$$\boxed{\chi(P) = \exp(i \sum_{l\in\mathbb{N}} c_l \kappa_l(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))} \tag{12}$$

holds for some finitely supported sequence of real numbers $(c_l : l \in \mathbb{N})$. The latter, if existent, is uniquely determined by χ.

PROOF. The proof of the "only if" part is the content of Section 2. The "if" part follows trivially from 1.7 a) and (5).

Finally suppose that we have (12) and an analogous representation of χ involving another finitely supported sequence $(\tilde{c}_l : l \in \mathbb{N})$. Then the sequence $(d_l) := (c_l - \tilde{c}_l)$ yields an analogous representation of the constant character 1. Suppose that not all d_l vanish. Put $r := \min\{l : d_l \neq 0\}$ and apply 1.7 b) with $c := \pi/|d_r|$. Then $1 = \exp(i\sum_{l=1}^{r} d_l \kappa_l(P_n)) \to \exp(\pm i\pi) = -1$ for $n \to \infty$. This contradiction shows that we must have $d_l = 0$ for every $l \in \mathbb{N}$, as was to be proved. □

1.9 COROLLARY. *Let $r \in \mathbb{N}_0$. A function $\chi|\mathrm{Prob}_r(\mathbb{R})$ is a continuous character iff* (12) *holds with $c_l = 0$ for $l > r$, and with* $\mathrm{Prob}_r(\mathbb{R})$ *in place of* $\mathrm{Prob}_\infty(\mathbb{R})$.

PROOF. Again, the "if" part follows from from 1.7 a) and (5). To prove "only if": Let $\chi|\mathrm{Prob}_r(\mathbb{R})$ be a continuous character. Then, by 1.8, its restriction $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ fulfils (12) for some finitely supported sequence (c_l). Assume that $c_l \neq 0$ for some $l > r$. Put $\tilde{r} := \min\{l \in \mathbb{N} : c_l \neq 0\}$. Choose (P_n) according to 1.7 b) with $\tilde{r}$ in place of r and with $c := \pi/|c_{\tilde{r}}|$. Then, since $r < \tilde{r}$, we have $P_n \to \delta$ with respect to d_r. On the other hand, we have $\chi(P_n) \to -1 \neq 1 = \chi(\delta)$. This contradiction to the continuity of χ shows that we must have $c_l = 0$ for $l > r$. It follows that the right hand side of (12) is defined and continuous on $\mathrm{Prob}_r(\mathbb{R})$. Since $\mathrm{Prob}_\infty(\mathbb{R})$ is obviously dense in $\mathrm{Prob}_r(\mathbb{R})$, this implies that (12) also holds with $\mathrm{Prob}_r(\mathbb{R})$ in place of $\mathrm{Prob}_\infty(\mathbb{R})$. □

1.10 THEOREM OF HALÁSZ. The last corollary yields in particular a theorem of Halász, presented on page 132 of Ruzsa & Székely (1988), which reads:

> 1 *is the only character of* $\mathrm{Prob}(\mathbb{R})$ *continuous with respect to weak convergence.*

In fact, the special case $r = 0$ of our Corollary 1.9 is slightly stronger, since our continuity assumption refers to a stronger topology on $\mathrm{Prob}(\mathbb{R})$.

1.11 THEOREM (ADDITIVE FUNCTIONS ON $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ is continuous and additive iff*

$$\boxed{\kappa(P) = \sum_{l \in \mathbb{N}} c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))} \tag{13}$$

holds for some finitely supported family of real numbers $(c_l : l \in \mathbb{N})$. The latter, if existent, is uniquely determined by κ.

Proof. The "if" part and the uniqueness of (c_l) follows via multiplication by i and subsequent exponentiation from the corresponding statements in 1.8.

To prove the "only if" part, let $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ be continuous and additive. Put

$$\chi(P) \quad := \quad \exp(i\kappa(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})).$$

Then χ satisfies the hypothesis of Theorem 1.8, and hence can be represented as in (12). This implies

$$\kappa(P) \quad = \quad \eta(P) + \sum_l c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})),$$

where $\eta|\mathrm{Prob}_\infty(\mathbb{R}) \to 2\pi\mathbb{Z}$. Since η must be additive, $\eta(\delta) = 0$. Since η must be continuous and $\mathrm{Prob}_\infty(\mathbb{R})$ is convex, $\eta(\mathrm{Prob}_\infty(\mathbb{R}))$ must be connected. [Here we have used the obvious fact that for $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ the function $[0,1] \ni t \mapsto tP + (1-t)Q \in \mathrm{Prob}_\infty(\mathbb{R})$ is continuous.] Thus $\eta = 0$. □

1.12 Corollary. *Let $r \in \mathbb{N}_0$. A function $\kappa|\mathrm{Prob}_r(\mathbb{R}) \to \mathbb{R}$ is continuous and additive iff* (13) *holds with $c_l = 0$ for $l > r$ and with* $\mathrm{Prob}_r(\mathbb{R})$ *in place of* $\mathrm{Prob}_\infty(\mathbb{R})$.

Proof. Deduce 1.12 from 1.9, by arguing as in the proof of 1.11. Alternatively, deduce 1.12 from 1.11 by arguing as in the proof of 1.9. □

1.13 Lemma (cumulants of Bernoulli distributions). *For $r \in \mathbb{N}$, let $f_r|[0,1] \to \mathbb{R}$ be defined by*

$$f_r(p) \quad := \quad \kappa_r((1-p)\delta_0 + p\delta_1) \qquad (p \in [0,1]).$$

Then, for each r, f_r is a polynomial function of degree r with r simple zeros in $[0,1]$.

Proof. It is known [for example, from Kendall, Stuart & Ord (1987), exercise 5.1] that

$$f_{r+1}(p) \quad = \quad p \cdot (1-p) \cdot f_r'(p) \qquad (r \in \mathbb{N},\, p \in [0,1]),$$

where the prime denotes differentiation with respect to p. Since $f_1(p) = p$ for $p \in [0,1]$, the claim follows by an induction argument, using Rolle's theorem and the fact that f_r' has at most $r-1$ zeros, counting multiplicity. □

1.14 A characterization of the variance. *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to [0,\infty[$ is continuous and additive iff $\kappa = c\kappa_2$ for some $c \in [0,\infty[$.*

Proof. The "if" claim is trivial. To prove "only if", we may by Theorem 1.11 start from the representation (13). Inserting there $P = \delta_a$ with $a \in \mathbb{R}$, we see that the assumption $\kappa \geq 0$ forces $c_1 = 0$. Thus, except for the trivial case $\kappa = 0$, we have

$$\kappa(P) = \sum_{l=2}^{r} c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))$$

for some $r \geq 2$ with $c_r \neq 0$. Suppose now that $r \geq 3$. Then we may, by the lemma 1.13, choose a Bernoulli distribution $P_0 = (1-p)\delta_0 + p\delta_1$ with $c_r\kappa_r(P_0) < 0$. It follows that $\kappa(P) < 0$ for $P := (x \mapsto ax)\Box P_0$ with $a > 0$ sufficiently large, using (14) below. This contradiction proves our claim. $\square$

1.15 Affine equivariance of cumulants. The second most important property of the cumulants is their behaviour under affine transformations: For $r \in \mathbb{N}$, $P \in \mathrm{Prob}_r(\mathbb{R})$ and $a, b \in \mathbb{R}$, we have

$$\kappa_r((x \mapsto ax+b)\Box P) \quad = \quad \begin{cases} a\kappa_1(P) + b & (r = 1), \\ a^r \kappa_r(P) & (r \geq 2). \end{cases} \tag{14}$$

In particular, each cumulant is affinely equivariant in the sense of the following definition and, by a trivial specialization, also scale equivariant.

1.16 Definition (equivariance). a) Let $\mathcal{X}$ be a set and $\mathcal{T}$ be a set of functions from $\mathcal{X}$ into $\mathcal{X}$. A function $\varphi|\mathcal{X}$ is called *equivariant*, with respect to $\mathcal{T}$, if we have the implication

$$x, y \in \mathcal{X},\, \varphi(x) = \varphi(y),\, T \in \mathcal{T} \quad \Longrightarrow \quad \varphi(T(x)) = \varphi(T(y)). \tag{15}$$

b) For $a, b \in \mathbb{R}$ define $T_{a,b}|\mathrm{Prob}(\mathbb{R}) \to \mathrm{Prob}(\mathbb{R})$ by

$$T_{a,b}(P) \quad := \quad (x \mapsto ax+b)\Box P \qquad (P \in \mathrm{Prob}(\mathbb{R}))$$

and put $\mathcal{T} := \{T_{a,b} : a, b \in \mathbb{R}\}$. Let $\mathcal{P} \subset \mathrm{Prob}(\mathbb{R})$ satisfy the implication $P \in \mathcal{P}, T \in \mathcal{T} \implies T(P) \in \mathcal{P}$. Then a function $\varphi|\mathcal{P}$ is called *affinely equivariant* if it is equivariant with respect to $\mathcal{T}$, in the sense of part a).

c) We define a function $\varphi|\mathcal{P}$ to be *scale equivariant* if it satisfies the definition given in b) above, but with $b = 0$ and $a > 0$ in the definition of $\mathcal{T}$.

1.17 Theorem (equivariant continuous characters of $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ is a scale equivariant continuous character iff*

$$\chi(P) \quad = \quad \exp(ic\kappa_r(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})) \tag{16}$$

for some $r \in \mathbb{N}$ and some $c \in \mathbb{R}$.

Proof. The "if" part is trivial. To prove "only if": Define $S_a(P) := (x \mapsto ax)\Box P$ for $P \in \mathrm{Prob}(\mathbb{R})$ and $a \in \,]0, \infty[$. For $\lambda \in \,]0, \infty[$, let P_λ denote the Poisson distribution with expectation λ. Then

$$\kappa_l(S_a(P_\lambda)) \quad = \quad a^l \lambda \qquad (l \in \mathbb{N},\, a, \lambda \in \,]0, \infty[). \tag{17}$$

Now let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a scale equivariant continuous character. Applying 1.8, we get (12) for some finitely supported sequence $(c_l : l \in \mathbb{N})$, and we have to show that there is at most one $l \in \mathbb{N}$ with $c_l \neq 0$. Using (17), (12) yields in particular

$$\chi(S_a(P_\lambda)) \quad = \quad \exp(i\lambda p(a)) \qquad (a, \lambda \in \,]0, \infty[) \tag{18}$$

where p is the polynomial function defined by

$$p(a) \quad := \quad \sum_{l \in \mathbb{N}} c_l a^l \qquad (a \in \mathbb{C}).$$

Now assume, to get a contradiction, that there are at least two $l \in \mathbb{N}$ with $c_l \neq 0$. Then for arbitrary $a_1, a_2 \in \,]0, \infty[$ with $a_1 \neq a_2$ and arbitrary $\lambda_1, \lambda_2 \neq 0$, there exists a number $b \in \,]0, \infty[$ with

$$\lambda_1 p(ba_1) - \lambda_2 p(ba_2) \quad \notin \quad 2\pi\mathbb{Z}. \tag{19}$$

[Proof: Assume without loss of generality that $a_1 < a_2$. If our claim is false, then the rational function $\mathbb{C} \ni z \mapsto R(z) := p(a_1 z)/p(a_2 z)$ is constant. But by our assumption on p, $\varrho := \sup\{|z| : z \in \mathbb{C},\ p(z) = 0\} > 0$. In view of $0 < a_1 < a_2$, it is obvious that R has a zero, namely on the circle $\{|z| = \varrho/a_1\}$. Hence $R \equiv 0$ and thus $p \equiv 0$, a contradiction.]

Now choose specifically $a_1, a_2 \in \,]0, \infty[$ with $a_1 \neq a_2$ in such a way that $p(a_1) \cdot p(a_2) > 0$. Choose $\lambda_1, \lambda_2 \in \,]0, \infty[$ with

$$\lambda_1 p(a_1) \quad = \quad \lambda_2 p(a_2), \tag{20}$$

choose b as in (19), and put $Q_k := S_{a_k}(P_{\lambda_k})$ for $k = 1, 2$. Then (18) and (20) yield $\chi(Q_1) = \chi(Q_2)$, whereas (18) also yields $\chi(S_b(Q_k)) = \chi(S_{ba_k}(P_{\lambda_k})) = \exp(i\lambda_k p(ba_k))$ for $k = 1, 2$, so that (19) implies $\chi(S_b(Q_1)) \neq \chi(S_b(Q_2))$, in contradiction to the scale equivariance of χ. □

1.18 Theorem (scale equivariant additive functions on $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ is continuous, additive, and scale equivariant, iff there exist $r \in \mathbb{N}$ and $c \in \mathbb{R}$ such that $\kappa = c\kappa_r$.*

Proof. Proceed as in the proof of the "only if" part of Theorem 1.11, but use equivariance of χ and 1.17 in place of 1.8. □

1.19 A characterization of the expectation. Notation: In this subsection only, we write $P \boxplus Q$ for the usual convolution $P * Q$ of $P, Q \in \mathrm{Prob}(\mathbb{R})$, and $P \boxdot Q$ for the multiplicative convolution of $P, Q \in \mathrm{Prob}(\mathbb{R})$, that is, the distribution of $X \cdot Y$ with X, Y being independent random variables with distributions P, Q.

THEOREM. *Let* $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ *be continuous. Then we have both*

$$\kappa(P \boxplus Q) \quad = \quad \kappa(P) + \kappa(Q), \tag{21}$$

$$\kappa(P \boxdot Q) \quad = \quad \kappa(P) \cdot \kappa(Q) \tag{22}$$

for $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$*, iff either* $\kappa = \kappa_1$ *or* $\kappa = 0$.

PROOF. The "if" part is obvious. So assume that κ is continuous and satisfies (21) and (22). By applying (22) to $Q = \delta_a$, for every $a \in]0, \infty[$, we see that κ is scale equivariant. Hence (21) and Corollary 1.18 yield $\kappa = c\kappa_r$ for some $c \in \mathbb{R}$ and some $r \in \mathbb{N}$. Choose $P \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\kappa_r(P) \neq 0$, for example P = Poisson distribution with parameter 1. Insert this P and $Q = \delta_1$ into (22), use $\kappa = c\kappa_r$, and divide by $\kappa_r(P)$. The result is $c = c^2\kappa_r(\delta_1)$. If $r \geq 2$, then $\kappa_r(\delta_1) = 0$, hence $c = 0$ and thus $\kappa = 0$. If $r = 1$, then $\kappa_r(\delta_1) = 1$, hence either again $c = 0$ and $\kappa = 0$, or $c = 1$ and thus $\kappa = \kappa_1$. □

1.20 "COUNTEREXAMPLES". Examples a) and b) below show that the continuity assumptions in 1.8 – 1.12 can not be omitted without substitute. Both a) and b) should be regarded as pathological. On the other hand, the examples in c) show that not only 1.8 – 1.12, but also 1.14 and, using (23), also 1.17 and 1.18 receive non-pathological counterexamples if the continuity assumption is dropped and if the domain of definition of the functionals is taken to be to small. Concerning 1.8 – 1.12, we may also refer to example d), suggested to me by I.Z. Ruzsa, where the domain of definition of κ could be thought of as being not much smaller than $\mathrm{Prob}_\infty(\mathbb{R})$.

a) By the axiom of choice, there exists a discontinuous additive function $f|\mathbb{R} \to \mathbb{R}$. Now $\kappa(P) := f(\mu(P))$ defines a discontinuous additive function $\kappa|\mathrm{Prob}_1(\mathbb{R}) \to \mathbb{R}$.

b) [Ruzsa & Székely (1988), pp. 122-123, 2.3 and 2.4] construct, using the axiom of choice, a homomorphism κ from $(\mathrm{Prob}(\mathbb{R}), *)$ into $(\mathbb{R}, +)$ which extends the expectation κ_1 defined on the subsemigroup $\mathrm{Prob}_1(\mathbb{R})$. They also show that each such κ assumes negative values for some P with support in $[0, \infty[$. It follows that the κ constructed is a discontinuous additive function from $\mathrm{Prob}(\mathbb{R})$ into $\mathbb{R}$.

c) On the semigroup

$$\mathrm{Prob}_c(\mathbb{R}) \quad := \quad \{P \in \mathrm{Prob}(\mathbb{R}) : \mathrm{supp}\, P \text{ compact}\} \quad \subset \quad \mathrm{Prob}_\infty(\mathbb{R})$$

we obtain an additive and nonnegative functional, normalized here as to satisfy additionally condition *ii)* from 1.22 below, by each of the following definitions:

$$\kappa(P) \quad := \quad \frac{1}{2} \cdot (\max \mathrm{supp}\, P - \min \mathrm{supp}\, P) \qquad (P \in \mathrm{Prob}_c(\mathbb{R})), \tag{23}$$

$$\kappa(P) \quad := \quad \frac{\log \widehat{P}(i) + \log \widehat{P}(-i)}{2 \log \cos i} \qquad (P \in \mathrm{Prob}_c(\mathbb{R})). \tag{24}$$

[In (24), we use of course the definition (1) with $\mathbb{C}$ in place of $\mathbb{R}$.]

d) Consider the semigroup

$$\mathcal{P} \quad := \quad \Big\{ P \in \mathrm{Prob}_\infty(\mathbb{R}) \;:\; \widehat{P} \text{ holomorphic near zero} \Big\} \quad \subset \quad \mathrm{Prob}_\infty(\mathbb{R}).$$

Let $(a_l : l \in \mathbb{N})$ be any sequence of real numbers satisfying $a_l = O(\varepsilon^l)$, for every $\varepsilon > 0$. Then

$$\kappa(P) \quad := \quad \sum_{l=1}^{\infty} \frac{a_l}{l!} \kappa_l(P) \qquad (P \in \mathcal{P}) \tag{25}$$

defines an additive function on $\mathcal{P}$. [To see that the series in (25) always converges, observe that $\log \circ \widehat{P}$ is now holomorphic in some P-dependent neighbourhood of zero, so that its Taylor series $\sum_{l=1}^{\infty} \kappa_l(P) \cdot (iz)^l / l!$ converges for $|z|$ sufficiently small.]

1.21 Some early history and etymology. Cumulants were apparently first introduced by T.N. Thiele [1838-1910] under the name of "half-invariants". Hald (1981) describes, on pages 7-10, Thiele's contributions and their insufficient acknowledgement by K. Pearson and R.A. Fisher. According to Hald, cumulants are first defined in the book Thiele (1889). [This I did not check. Hald's formula (4.1), claimed to be Thiele's definition, is, up to an obvious misprint, the now well-known recursion (6), determining κ_{r+1} as a polynomial in the moments μ_l.] In a later and more accessible version of his book, Thiele (1903) essentially gives definition (4). Hald (1998) contains a much more comprehensive early history of cumulants.

Later authors, such as Craig (1931) and Wishart (1929), refer to the cumulants as "semi-invariants of Thiele", while Fisher (1929-30), on page 200 of his paper, simply calls them "semi-invariants", without bothering to name Thiele. But Wishart and Fisher, who obviously new about each others work before publication, prefer to use the new term "cumulative moment functions" instead. The reason for adopting this term is hinted at in Fisher's paper: On page 199, he gives an interesting although perhaps not quite precise definition of rather general "moment functions" of populations, roughly speaking by polynomial estimability, which seems at any rate to be intended to include polynomial functions of finitely many ordinary moments, and hence in particular cumulants. On page 202, Fisher then refers to a "cumulative property" of the logarithm of the Laplace transform which, expressed in terms of the cumulants, is just condition (5). Thus the the adjective "cumulative" refers, in this context, to a homomorphism condition. In particular, it is not used to distinguish a concept related to probability measures from a corresponding concept related to probability densities, as would often be the case in the older statistical literature.

Finally, "cumulative moment function" was abbreviated to "cumulant" by Fisher & Wishart (1931-32) and Fisher (1932), with Hotelling (1933) claiming to have suggested this name, which quickly became the standard one in the english language literature. The first publication having the word "cumulant" in its title seems to be the paper by Cornish & Fisher (1937), who repeat the definition, but already Haldane (1937), page 136, uses "cumulants" without definition or reference.

Readers generally interested in the history of probabilistic or statistical terms are referred to David (1995, 1998) as a useful starting point.

1.22 Related work not discussed above. The following papers have some relation with the present one.

Craig (1931) states on page 160 a forerunner of our Corollary 1.18. Where we assume mere continuity of κ, Craig assumes in particular that κ is a polynomial function of some finite number of moments μ_l. His treatment is not quite rigorous: For example, no domain of definition of κ is specified, his conclusion is $\kappa = \kappa_r$ for some r [instead of the correct conclusion $\kappa = c\kappa_r$ for some r and c], and a proof is offered only for the case where κ is a polynomial function of $\mu_1, \ldots, \mu_4$.

Savage (1971) characterizes moments and more general expectations of exponential polynomials as functionals κ satisfying, on the one hand, conditions like $\kappa(P * Q) = T(\kappa(P), \kappa(Q))$ with T unspecified and, on the other hand, having a representation $\kappa(P) = \int f\, dP$ with f unspecified. His first assumption is more liberal than our homomorphism assumptions, but his second assumption is rather restrictive, excluding for example every cumulant κ_r with $r \geq 2$. Thus the work of Savage is incomparable to the present one.

Martin Diaz (1977), Teorema 4, states a characterization of the variance which may be formulated as follows. We temporarily put $\mathcal{P} := \{P \in \mathrm{Prob}(\mathbb{R}) \,:\, \mathrm{supp}\, P \text{ finite}\}$.

Theorem (Martin Diaz) *Let $\kappa|\mathcal{P} \to [0, \infty[$ and assume:*

i) For every $n \in \mathbb{N}$, the map

$$\mathbb{R}^n \times \left\{ p \in \,]0,1]^n \,:\, \sum_{i=1}^n p_i = 1 \right\} \quad \ni \quad (x,p) \mapsto \kappa(\sum_{i=1}^n p_i \delta_{x_i})$$

is partially continuous in the two variables x and p.

ii) $\kappa(\delta_1) = 0$, $\kappa(\frac{1}{2}(\delta_{-1} + \delta_1)) = 1$.

iii) If we put $\kappa(X) := \kappa(P)$ for every random variable X with distribution $P \in \mathcal{P}$, then

$$\kappa(\sum_{i=1}^n X_i) \quad = \quad \sum_{i=1}^n \kappa(X_i)$$

whenever the X_i *are* pairwise *independent random variables, on a common probability space, with distributions belonging to* $\mathcal{P}$.

Then $\kappa = \kappa_2$.

We observe that the word "pairwise" renders the third assumption rather confining. But without this word, a counterexample would be obtained by restricting to $\mathcal{P}$ either κ from (23) or (24). These examples may be regarded as negative solutions to the problem stated in Martin Diaz (1977) on page 96, while our result 1.14 may be regarded as a kind of positive solution.

Good (1979) speculates about the existence of a useful notion of "fractional cumulants", perhaps to be defined via fractional differentiation of $\log \circ \widehat{P}$ in analogy to (4). Such a definition, if possible, should lead to an additive function on $\mathrm{Prob}_\infty(\mathbb{R})$, and Theorem 1.11 could be taken as an indication that it will not lead to anything new and useful.

Heyer (1981) reviews, among other topics, axiomatic approaches to expectation and variances for probability measures on compact groups, referring to earlier publications of himself and of Maksimov, in particular Maksimov (1980). Although somewhat similar in spirit to the present paper, there is no overlap in the results obtained.

Characterizations of the variance not referring to the semigroup structure of $\mathrm{Prob}(\mathbb{R})$ have been provided by Bomsdorf (1974), by Gil Alvarez (1983), and by Kagan & Shepp (1998). The former two are somewhat similar to the characterization of the Shannon entropy by Fadeev's axioms, as presented in Rényi (1970), page 548.

2 The main proof

2.1 Further notation and conventions. The proof of the "only if" part of Theorem 1.8, given in 2.8 below, is prepared by the introduction of an auxiliary topological vector space $\mathcal{H}$ in 2.2 and the identification of its dual $\mathcal{H}'$ in 2.3. We will use some tools from functional and Fourier analysis as explained in Rudin (1991). In particular, we assume as known the spaces $\mathcal{C}^\infty(\mathbb{R})$, $\mathcal{D}(\mathbb{R})$, $\mathcal{D}'(\mathbb{R})$ with their usual topologies. We depart from the conventions of Rudin (1991) in that here a topological vector space is not necessarily assumed to be Hausdorff.

We let $\mathcal{U}$ denote the set of all open symmetric neighbourhoods of $0 \in \mathbb{R}$. For $U \in \mathcal{U}$, a function $h|U \to \mathbb{C}$ is called *hermitean* if

$$h(t) \quad = \quad \overline{h(-t)} \qquad (t \in U).$$

2.2 The space $\mathcal{H}$ of germs of hermitean $\mathcal{C}^\infty$ functions vanishing at zero. We consider

$$X \quad := \quad \{h \in \mathcal{C}^\infty(\mathbb{R}) \,:\, h \text{ hermitean}, h(0) = 0\}$$

as a topological vector space over $\mathbb{R}$, with the topology inherited from the usual topology of $\mathcal{C}^\infty(\mathbb{R})$. We further consider the vector subspace

$$N \quad := \quad \{h \in X \,:\, \exists U \in \mathcal{U} \text{ with } h|U = 0\}$$

of X, and form the quotient topological vector space

$$\mathcal{H} \quad := \quad X/N.$$

For $h \in X$, we write $[h]$ for the equivalence class $H \in \mathcal{H}$ with $h \in H$. It easy to see, though for our purposes unnecessary to check, that N is not closed, so that $\mathcal{H}$ is not Hausdorff. Since $\mathcal{C}^\infty(\mathbb{R})$ is metrizable, $\mathcal{H}$ is pseudo-metrizable, and a sequence $(H_j \,:\, j \in \mathbb{N})$ converges to $0 \in \mathcal{H}$ iff there exist $h_j \in H_j$ with $h_j \to 0 \in X$. [Proof: The discussion in Sections 1.40, 1.41 of Rudin (1991) applies with obvious changes, necessitated by the nonclosedness of our N. In particular, if d is some tranlation-invariant metric for X, the formula $\varrho([h_1],[h_2]) := \inf\{d(h_1 - h_2, g) : g \in N\}$ defines a translation-invariant pseudo-metric ϱ for $\mathcal{H}$. And if $([h_j]) : j \in \mathbb{N})$ is a sequence in $\mathcal{H}$ with $\lim \varrho([h_j],[0]) = 0$, we may choose $g_j \in N$ with $d(h_j, g_j) \leq 2\varrho([h_j],[0]) + j^{-1}$, yielding $\tilde{h}_j := h_j - g_j \in [h_j]$ with $\tilde{h}_j \to 0$.]

The value at zero of the derivatives $D^l H(0)$ of a $H \in \mathcal{H}$, occuring below, is defined in the obvious way.

2.3 THE DUAL $\mathcal{H}'$ OF $\mathcal{H}$. *A function $\Lambda|\mathcal{H}$ is an $\mathbb{R}$-valued, continuous, and $\mathbb{R}$-linear functional iff there exists an $n \in \mathbb{N}_0$ and a finite sequence of real numbers $(c_l : 1 \leq l \leq n)$ such that*

$$\Lambda(H) \quad = \quad \sum_{l=1}^{n} c_l \cdot i^{-l}(D^l H)(0) \qquad (H \in \mathcal{H}). \tag{26}$$

PROOF. The "if" claim is obviously true. To prove "only if": Let $\Lambda|\mathcal{H} \to \mathbb{R}$ be continuous and $\mathbb{R}$-linear. Define $S|\mathcal{D}(\mathbb{R}) \to \mathbb{R}$ by

$$S(\varphi) \quad := \quad \Lambda([\frac{1}{2}\left(\varphi - \varphi(0) + \overline{\check{\varphi - \varphi(0)}}\right)]) \qquad (\varphi \in \mathcal{D}(\mathbb{R})),$$

where $\check{\psi}(t) := \psi(-t)$. It is obvious that S is well-defined and $\mathbb{R}$-valued, as well as continuous and $\mathbb{R}$-linear. It follows that the functional $T|\mathcal{D}(\mathbb{R}) \to \mathbb{C}$ defined by

$$T(\varphi) \quad := \quad S(\varphi) - iS(i\varphi) \qquad (\varphi \in \mathcal{D}(\mathbb{R}))$$

is continuous and $\mathbb{C}$-linear, that is, a distribution $\in \mathcal{D}'(\mathbb{R})$. It is easily checked that T has support contained in $\{0\}$. Hence, by Rudin (1991), Theorem 6.24 d) and Theorem 6.25, there is an $n \in \mathbb{N}_0$ and a sequence of complex numbers $(b_l \,:\, 0 \leq l \leq n)$ such that

$$T(\varphi) \quad = \quad \sum_{l=0}^{n} b_l \cdot \left(D^l\varphi\right)(0) \qquad (\varphi \in \mathcal{D}(\mathbb{R})).$$

Since $S = \mathrm{Re}\, T$, we get for $H = [h] \in \mathcal{H}$, using the hermitean property of h and $h(0) = 0$,

$$\begin{aligned}
\Lambda(H) &= S(h) \\
&= \mathrm{Re}\, T(h) \\
&= \sum_{l=1}^{n} \mathrm{Re}\, \left(b_l \cdot \left(D^l h\right)(0)\right) \\
&= \sum_{l=1}^{n} \mathrm{Re}\, \left(b_l i^l\right) \cdot i^{-l} \left(D^l h\right)(0),
\end{aligned}$$

and thus (26) with $c_l = \mathrm{Re}\, (b_l i^l)$. □

2.4 Convergence in $\mathrm{Prob}_\infty(\mathbb{R})$. *Let P be an element of and (P_j) be a net in $\mathrm{Prob}_\infty(\mathbb{R})$. Then $\lim P_j = P$, in the topology of $\mathrm{Prob}_\infty(\mathbb{R})$, iff $\lim P_j = P$ with respect to total variation distance and*

$$\lim_j \int x^l \, dP_j(x) \quad = \quad \int x^l \, dP(x) \qquad (l \in \mathbb{N}). \tag{27}$$

Proof. Let first w be any nonnegative measurable function on a measurable space $\mathcal{X}$. Let $P, Q \in \mathrm{Prob}(\mathcal{X})$ with $\int w \, d(P+Q) < \infty$, and fix $a > 0$. Then

$$\begin{aligned}
\int w \, d|P-Q| &\leq \int w \cdot (w \leq a) \, d|P-Q| + \int w \cdot (w > a) \, d(P+Q) \\
&= \int w \cdot (w \leq a) \, d|P-Q| + 2 \int w \cdot (w > a) \, dP \\
&\quad + \int w \, d(Q-P) - \int w \cdot (w \leq a) \, d(Q-P) \\
&\leq 2 \int w \cdot (w \leq a) \, d|P-Q| + 2 \int w \cdot (w > a) \, dP \\
&\quad + \int w \ \, dQ - \int w \ \, dP.
\end{aligned}$$

Now let (P_j) be a net in $\mathrm{Prob}(\mathcal{X})$ with $\int w \, dP_j < \infty$ for every j. The preceding inequality shows that $\lim \int w \, d|P - P_j| = 0$ if both $\lim \int 1 \, d|P - P_j| = 0$ and $\overline{\lim} \int w \, dP_j \leq \int w \, dP$. Applied to $\mathcal{X} = \mathbb{R}$ and $w(x) = 1 + x^{2n}$, for each $n \in \mathbb{N}$, the "if" part follows. The "only if" part is trivial. □

2.5 Quotients of characteristic functions. *Let*

$$\varphi \in \Phi \quad := \quad \{\varphi \in \mathcal{D}(\mathbb{R}) \, : \, \varphi(0) = 1, \varphi \text{ hermitean}\}.$$

a) *There exist $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ with*

$$\varphi \widehat{Q} \quad = \quad \widehat{P}. \tag{28}$$

b) *Let* (φ_j) *be a net in* Φ *with* $\lim \varphi_j = \varphi$ *in the* $\mathcal{D}(\mathbb{R})$*-topology. Then we may choose* $P_j, Q_j \in \mathrm{Prob}_\infty(\mathbb{R})$ *with* $\varphi_j \widehat{Q}_j = \widehat{P}_j$ *and*

$$\lim P_j = P, \qquad \lim Q_j = Q \quad \text{in } \mathrm{Prob}_\infty(\mathbb{R}). \tag{29}$$

REMARK. As said before in 1.2, this basic tool of the present paper is a refinement of a theorem of Ruzsa & Székely. In particular, most of the following proof of part a) is as in Ruzsa & Székely (1988), pages 126-127.

PROOF. We will calculate in

$$M^1(\mathbb{R}) \quad := \quad \text{set of all bounded complex measures on } \mathbb{R},$$

which is well known to be a Banach algebra, with convolution $*$ as multiplication and norm $\|\cdot\|$ defined by

$$\begin{aligned} \|\mu\| \quad &:= \quad \int 1\, d|\mu| \qquad (\mu \in M^1(\mathbb{R})), \qquad (30)\\ |\mu| \quad &:= \quad \text{total variation measure of } \mu. \end{aligned}$$

For a $\mu \in M^1(\mathbb{R})$, its Fourier transform is the continuous function $\widehat{\mu}$ defined by

$$\widehat{\mu}(t) \quad := \quad \int e^{itx}\, d\mu(x) \qquad (t \in \mathbb{R}).$$

We assume as known properties of the Fourier transform as explained, for example, in Chapter 7 of Rudin (1991). All elements of $M^1(\mathbb{R})$ actually occuring below will in fact belong to

$$M^1_\infty(\mathbb{R}) \quad := \quad \left\{ \mu \in M^1(\mathbb{R}) \,:\, \int |x|^l\, d|\mu|(x) < \infty \quad (l \in \mathbb{N}_0) \right\}.$$

For $\mu \in M^1_\infty(\mathbb{R})$, we have $\widehat{\mu} \in \mathcal{C}^\infty(\mathbb{R})$.

a) We have $\varphi = \widehat{\mu}$ with $\mu \in M^1_\infty(\mathbb{R})$, μ real, $\mu(\mathbb{R}) = 1$. [Apply Theorem 7.7 of Rudin (1991).]

Choose $\alpha, \beta \in [0, \infty[$ and $R \in \mathrm{Prob}_\infty(\mathbb{R})$ with

$$\|(\mu - \delta) * R\| \;=\; \alpha \;<\; \beta \tag{31}$$

and

$$R^{*2} \quad \geq \quad \beta R. \tag{32}$$

[For example, if R is any centered normal distribution, then (32) is true with $\beta = 2^{-1/2}$, and for R sufficiently flat (31) is true as well. Alternatively, we may take $\beta = 2^{-1}$ and for R a sufficiently flat uniform distribution on an interval $[-a, a]$.]

Put

$$\begin{aligned} S &:= \beta^{-1}|(\mu-\delta)*R|, &(33)\\ Q &:= (1-\frac{\alpha}{\beta})R^{*2}*\sum_{k=0}^{\infty}S^{*k}, &(34)\\ P &:= \mu * Q. &(35)\end{aligned}$$

Since S is a sub-probability measure with $\|S\| = S(\mathbb{R}) = \alpha/\beta < 1$, the series in (34) is convergent in $M^1(\mathbb{R})$, and $Q \in \mathrm{Prob}(\mathbb{R})$. Also $P(\mathbb{R}) = 1$ and, easily verified,

$$\begin{aligned}(1-\frac{\alpha}{\beta})^{-1}P &= \mu * R^{*2} * \sum_{k=0}^{\infty} S^{*k}\\ &= R^{*2} + R^{*2} * (\mu - \delta + S) * \sum_{k=0}^{\infty} S^{*k},\end{aligned}$$

where, using (32) and (33),

$$\begin{aligned} R^{*2} * (\mu-\delta+S) &\geq R*(R*(\mu-\delta)+\beta S)\\ &\geq 0.\end{aligned}$$

Hence $P \geq 0$ and thus $P \in \mathrm{Prob}(\mathbb{R})$.

By $0 \leq S \leq \beta^{-1}(|\mu|+\delta)*R$, $S \in M^1_\infty(\mathbb{R})$. Hence $\widehat{S} \in \mathcal{C}^\infty(\mathbb{R})$. Since (34) shows that

$$\widehat{Q}(t) = (1-\frac{\alpha}{\beta})\cdot(\widehat{R}(t))^2\cdot(1-\widehat{S}(t))^{-1} \qquad (t \in \mathbb{R}), \tag{36}$$

and since also $\widehat{R} \in \mathcal{C}^\infty(\mathbb{R})$, it follows that $\widehat{Q} \in \mathcal{C}^\infty(\mathbb{R})$. Since (35) implies (28), $\widehat{P}$ is $\mathcal{C}^\infty$ as well, at least in some neighbourhood of zero. Since P, Q are probability measures, it follows that $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$. [Compare, for example, Feller (1971), page 528, problem 15.]

b) We continue to use the notation of the above proof of part a). Let, additionally, μ_j denote the element of $M^1_\infty(\mathbb{R})$ with $\varphi_j = \widehat{\mu}_j$, and

$$\alpha_j := \|(\mu_j - \delta) * R\|.$$

By Theorem 7.7 of Rudin (1991), we have $\lim \mu_j = \mu$ in the Schwartz space $\mathcal{S}(\mathbb{R})$. It follows that

$$\lim \mu_j = \mu \quad \text{with respect to the norms} \quad \|\cdot\|_k \qquad (k \in \mathbb{N}_0), \tag{37}$$

where

$$\|\nu\|_k := \int (1+|x|^k)\, d|\nu|(x) \qquad (k \in \mathbb{N}_0,\, \nu \in M^1_\infty(\mathbb{R})).$$

The particular case $k = 0$ implies $\lim \mu_j = \mu$ with respect to the norm $\|\cdot\|$ from (30), hence $\lim \alpha_j = \alpha$. We may and do assume that $\alpha_j < \beta$ in what follows. Put $S_j := \beta^{-1}|(\mu_j - \delta) * R|$, $Q_j := (1 - (\alpha_j/\beta))R^{*2} * \sum_{k=0}^{\infty} S_j^{*k}$, and $P_j = \mu_j * Q_j$. Then $Q_j, P_j \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\varphi_j \widehat{Q}_j = \widehat{P}_j$, and what remains to be shown is (29).

By (37),

$$\lim S_j = S \quad \text{with respect to the norms} \quad \|\cdot\|_k \qquad (k \in \mathbb{N}_0). \tag{38}$$

Using (38) and the definition of Q_j, P_j, we get $\lim Q_j = Q$ and $\lim P_j = P$ with respect to $\|\cdot\|$. From (38) we also get $\lim \widehat{S}_j = \widehat{S}$ in $\mathcal{C}^\infty(\mathbb{R})$. Since we have (36) with α_j replacing α, $\widehat{Q}_j$ replacing $\widehat{Q}$, and $\widehat{S}_j$ replacing $\widehat{S}$, we may conclude that $\lim \widehat{Q}_j = \widehat{Q}$ in $\mathcal{C}^\infty(\mathbb{R})$. By $\varphi_j \widehat{P}_j = \widehat{Q}_j$, we deduce $\lim \widehat{P}_j|U = \widehat{P}|U$ in $\mathcal{C}^\infty(U)$, for some neighbourhood U of zero. Hence we have in particular (27) and the corresponding statement for (Q_j), so that we reach (29) via 2.4. □

2.6 LEMMA. *Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a character, not necessarily continuous. If $P_1, P_2, Q_1, Q_2 \in \mathrm{Prob}_\infty(\mathbb{R})$, and if there exists an $U \in \mathcal{U}$ with*

$$\widehat{P}_1(t)\widehat{Q}_2(t) \quad = \quad \widehat{P}_2(t)\widehat{Q}_1(t) \qquad (t \in U),$$

then

$$\frac{\chi(P_1)}{\chi(Q_1)} \quad = \quad \frac{\chi(P_2)}{\chi(Q_2)}. \tag{39}$$

PROOF. There exists an $R \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\mathrm{supp}\, \widehat{R} \subset U$. Thus $\widehat{P}_1 \widehat{Q}_2 \widehat{R} = \widehat{P}_2 \widehat{Q}_1 \widehat{R}$ everywhere, so that we successively get

$$\begin{aligned} P_1 * Q_2 * R &= P_2 * Q_1 * R, \\ \chi(P_1)\chi(Q_2)\chi(R) &= \chi(P_2)\chi(Q_1)\chi(R), \end{aligned}$$

and hence (39). □

2.7 FROM χ TO A LINEAR FUNCTIONAL Λ. *Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a continuous character. Then there exists a $\Lambda \in \mathcal{H}'$ with*

$$\chi(P) \quad = \quad \exp(i\Lambda(\log \circ [\widehat{P}])) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})). \tag{40}$$

Here $\log \circ [\widehat{P}]$ of course denotes the element of $\mathcal{H}$ containing the functions $h \in X$ satisfying

$$h(t) \quad = \quad \log \widehat{P}(t) \qquad (t \in U)$$

for some $U \in \mathcal{U}$ with $U \subset \Big\{ t \in \mathbb{R} : |\widehat{P}(t) - 1| < 1 \Big\}$.

PROOF. Follows from Steps 1-5 below. □

STEP 1: CONSTRUCTION OF A FUNCTION $X|\mathcal{H}$. *Let $H \in \mathcal{H}$. Then we may define $X(H) \in \mathbb{T}$ by the construction leading to (42) below, and this definition is independent of the choices of h, U, ω, P, Q made along the way.*

PROOF. Choose $h \in H$. Define $\psi \in \mathcal{C}^\infty(\mathbb{R})$ by

$$\psi(t) \quad := \quad \exp(h(t)) \qquad (t \in \mathbb{R}). \tag{41}$$

Choose $U \in \mathcal{U}$ with compact closure and choose $\omega \in \mathcal{D}(\mathbb{R})$ real and symmetric with $\omega|U = 1$. Define $\varphi \in \mathcal{D}(\mathbb{R})$ by

$$\varphi(t) \quad := \quad \omega \cdot \psi.$$

Then φ is hermitean with $\varphi(0) = 1$, and hence satisfies the assumptions of 2.5. So we may choose $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ satisfying (28), and put

$$X(H) \quad := \quad \frac{\chi(P)}{\chi(Q)}. \tag{42}$$

To show that this definition is independent of the choices made along the way, consider two choices $(h_i, U_i, \omega_i, P_i, Q_i)$, for $i \in \{1, 2\}$, yielding two values $X_i(H)$. There exists a $V \in \mathcal{U}$ with $\varphi_1|V = \varphi_2|V$. Hence (28) applied to φ_i, P_i, Q_i implies $\widehat{P}_1/\widehat{Q}_1 = \widehat{P}_2/\widehat{Q}_2$ on $U := V \cap \{t : \varphi_1(t) \neq 0\}$, so that Lemma 2.6 yields $X_1(H) = X_2(H)$. □

STEP 2: THE RELATION BETWEEN X AND χ. *For $P \in \mathrm{Prob}_\infty(\mathbb{R})$,*

$$\chi(P) \quad = \quad X(\log \circ [\widehat{P}]).$$

PROOF. Changing notation, let $P_1 \in \mathrm{Prob}_\infty(\mathbb{R})$. Put $H := \log \circ [\widehat{P}_1]$. Referring to Step 1 and its notation, let us denote one choice for the computation of $X(H)$ by (h, U, ω, P_2, Q_2), with (ψ, φ) accordingly defined. Then $\varphi = \widehat{P}_1$ in some $\tilde{U} \in \mathcal{U}$. With $Q_1 := \delta$ it follows that $\widehat{P}_1\widehat{Q}_2 = \widehat{P}_2\widehat{Q}_1$ in $\tilde{U}$. Hence (42), Lemma 2.6, and $\chi(\delta) = 1$, successively yield

$$X(H) = \frac{\chi(P_2)}{\chi(Q_2)} = \frac{\chi(P_1)}{\chi(Q_1)} = \chi(P_1).$$

□

STEP 3: *The function $X|\mathcal{H} \to \mathbb{T}$ defined in Step 1 is a character, with respect to addition in $\mathcal{H}$.*

PROOF. We have to prove that

$$X(H_1 + H_2) \quad = \quad X(H_1) \cdot X(H_2) \qquad (H_1, H_2 \in \mathcal{H}).$$

So let $H_1, H_2 \in \mathcal{H}$. Choose $(U_i, h_i, V_i, \omega_i, P_i, Q_i)$ and define (ψ_i, φ_i) as in Step 1 to calculate $X(H_i)$ for $i \in \{1, 2\}$. Then we may use the choice

$$(h_1 + h_2,\, U_1 \cap U_2,\, \omega_1 \cdot \omega_2,\, P_1 * P_2,\, Q_1 * Q_2),$$

leading to $\psi = \psi_1 \cdot \psi_2$ and $\varphi = \varphi_1 \cdot \varphi_2$, to compute $X(H_1 + H_2)$. The result is

$$\begin{aligned} X(H_1 + H_2) &= \chi(P_1 * P_2) \cdot (\chi(Q_1 * Q_2))^{-1} \\ &= \chi(P_1) \cdot \chi(P_2) \cdot (\chi(Q_1) \cdot \chi(Q_2))^{-1} \\ &= X(H_1) \cdot X(H_2). \end{aligned}$$

□

STEP 4: CONTINUITY. *X is continuous.*

PROOF. Since $\mathcal{H}$ is pseudometrizable, it suffices to consider any given convergent sequence $(H_j : j \in \mathbb{N})$, with $\lim H_j = H$. There exist $h \in H$, $h_j \in H_j$, such that

$$\lim h_j \quad = \quad h \qquad \text{in } \mathcal{C}^\infty(\mathbb{R}).$$

Starting from the present h, choose and define, respectively, ψ, U, and ω as in Step 1 around equation (41). Analogously, define ψ_j and then φ_j, using the same U and ω as for ψ, φ. Then $\lim \varphi_j = \varphi$ in $\mathcal{D}(\mathbb{R})$. Now apply part b) of 2.5 to choose P, Q, P_j, Q_j with the properties stated there. Then, using Step 1 and the continuity of χ,

$$X(H_j) \;=\; \tfrac{\chi(P_j)}{\chi(Q_j)} \;\to\; \tfrac{\chi(P)}{\chi(Q)} \;=\; X(H).$$

□

STEP 5: *There exists a $\Lambda \in \mathcal{H}'$ with $X = \exp \circ (i\Lambda)$.*

PROOF. This is always true whenever $\mathcal{H}$ is a topological $\mathbb{R}$-vectorspace with dual $\mathcal{H}'$, and $X|\mathcal{H}$ a continuous character, with respect to the additive group of $\mathcal{H}$. See section (23.32.a) on page 370 of Hewitt & Ross (1979) for a proof assuming, and using, that $\mathcal{H}$ is additionally Hausdorff. For the general case, needed here, apply the special case to the Hausdorff quotient space of $\mathcal{H}$, obtained by identifying points $h_1, h_2 \in \mathcal{H}$ iff $h_2 - h_1$ belongs to the closure of $\{0\}$.
□

2.8 PROOF OF THE "ONLY IF" PART OF THEOREM 1.8. Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a continuous character. Then there exists a linear functional Λ as in 2.7. By 2.3, Λ has a representation as in (26). Inserting this representation into (40) and applying the definition (4) yields (12). □

ACKNOWLEDGEMENT. The present work was supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft. I thank H.A. David, A. Hald, J.H.B. Kemperman, B. Roos, and I.Z. Ruzsa for helpful remarks.

References

An asterisk indicates work I have found discussed in other sources, but have not seen in the original.

Bomsdorf, E. (1974). Zur Charakterisierung von Lokations- und Dispersionsmaßen. *Metrika* 21, 223-229.

Cornish, E.A. & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* 4, 1-14.

Craig, C.C. (1931). On a property of the semi-invariants of Thiele. *Ann. Math. Statist.* 2, 154-164.

David, H.A. (1995). First (?) occurence of common terms in mathematical statistics. *The American Statistician* 49, 121-133.

David, H.A. (1998). First (?) occurence of common terms in probability and statistics-a second list, with corrections. *The American Statistician* 52, 36-40.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II, 2nd. Ed.* Wiley, N.Y.

Fisher, R.A. (1929-30). Moments and product moments of sampling distributions. *Proc. London Math. Soc.* 30, 199-238.

*Fisher, R.A. (1932). *Statistical Methods for Research Workers. 4th. Ed.* Oliver and Boyd, Edinburgh.

Fisher, R.A. & Wishart, J. (1931-32). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. London Math. Soc.* (2) 33, 195-208.

Gil Alvarez, María Angeles (1983). Caracterización axiomática para la varianza. *Trab. Estad. Invest. Oper.* 34, 40-51.

Good, I.J. (1979). Fractional moments and cumulants: some unsolved problems. *J. Statist. Comp. Simulation* 9, 314-315.

Hald, A. (1981). T. N. Thiele's contributions to statistics. *Int. Statist. Review* 49, 1-20.

Hald, A. (1998). The early history of the cumulants and the Gram-Charlier series. Preprint, Department of Theoretical Statistics, University of Copenhagen. (25 pages)

Haldane, J.B.S. (1937). The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small. *Biometrika* 29, 133-143. Correction note in *Biometrika* 31 (1939), 220.

Hewitt, E. & Ross, K.A. (1979). *Abstract Harmonic Analysis I*, 2nd ed. Springer, Berlin.

Heyer, H. (1981). Moments of probability measures on a group. *Int. J. Math. Sci.* 4, 231-249.

HOTELLING, H. (1933). Review of Fisher (1932). *J. Amer. Statist. Ass.* 28, 374-375.

KAGAN, A. & SHEPP, L.A. (1998). Why the variance? *Statist. Probab. Letters* 38, 329-333.

KEMPERMAN, J.H.B. (1972). Problem P 92. *Aeq. Math.* 8, 172.

KENDALL, M., STUART, A. & ORD, J.K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1, Distribution Theory.* Griffin, London.

MAKSIMOV, V.M. (1980). Mathematical expectations for probability distributions on compact groups. *Math. Z.* 174, 49-60. [MR 82g:60020]

MARTIN DIAZ, MIGUEL (1977). Caracterización de la varianza. *Trab. Estad. Invest. Oper.* 28, Nos. 2 and 3, 85-97. [MR 58, #31513]

RÉNYI, A. (1970). *Probability Theory.* North-Holland, Amsterdam, and Akadémiai Kiadó, Budapest.

RUDIN, W. (1991). *Functional Analysis, 2nd. Ed.* McGraw-Hill, N.Y.

RUZSA, I.Z. & SZÉKELY, G.J. (1983). Convolution quotients of nonnegative functions. *Mh. Math.* 95, 235-239.

RUZSA, I.Z. & SZÉKELY, G.J. (1985). No distribution is prime. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 70, 263-269.

RUZSA, I.Z. & SZÉKELY, G.J. (1988). *Algebraic Probability Theory.* Wiley, Chichester.

SAVAGE, L.J. (1971). The characteristic function characterized and the momentousness of moments. In: *Studi di probabilità, statistica e ricerca operativa in onore di Giuseppe Pompilj*, Tipografia Oderisi Editrice, Gubbio, pp. 131-141. Reprinted in: *The Writings of Leonard Jimmie Savage*, The American Statistical Association and The Institute of Mathematical Statistics, pp. 615-625 (1981).

*THIELE, T.N. (1889). *Forelæsninger over Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode.* Reitzel, København.

THIELE, T.N. (1903). *The theory of observations.* C. & E. Layton, London. Reprinted 1931 in: *Ann. Math. Statist.* 2, 165-308.

WISHART, J. (1929). A problem in combinatorial analysis giving the distribution of certain moment statistics. *Proc. London Math. Soc.* 29, 309-321.

Lutz Mattner
Universität Hamburg
Institut für Mathematische Stochastik
Bundesstr. 55
20146 Hamburg
Germany
mattner@math.uni-hamburg.de