Winkler:

Moment Sets of Bell-Shaped Distributions: Extreme Points, Extremal Decomposition and Chebysheff Inequalities

Sonderforschungsbereich 386, Paper 121 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Moment Sets of Bell-Shaped Distributions: Extreme Points, Extremal Decomposition and Chebysheff Inequalities

Gerhard Winkler

Institute of Biomathematics and Biometry
GSF - National Research Center for Environment and Health
D-85764 Neuherberg, Germany
e-mail: gwinkler@gsf.de

Abstract

The paper deals with sets of distributions which are given by moment conditions for the distributions and convex constraints on derivatives of their c.d.fs. A general albeit simple method for the study of their extremal structure, extremal decomposition and topological or measure theoretical properties is developed. Its power is demonstrated by the application to bell-shaped distributions. Extreme points of their moment sets are characterized completely (thus filling a gap in the previous theory) and inequalities of Tchebysheff type are derived by means of general integral representation theorems.

Some key words: Moment sets, Tschebysheff inequalities, extremal bell-shaped distributions

1 Introduction

This paper is devoted to the study of sets of distributions on the real line defined by both, moment constraints and convex constraints on derivatives (in the distributional sense). Of particular interest are their topological and measure theoretical properties and the characterization of extremal elements. Integral representations $\mu(B) = \int \nu(B)\, dp(\nu)$, where p is a probability measure on the set of (known) extreme points, are of interest, as well, since combined with the characterization of extreme points they immediately give sharp inequalities of Tchebysheff type.

Important examples are moment sets of distributions which are bell-shaped to some order with fixed turning points, for example unimodal with fixed mode. In the context of finite mixture distributions ([14]), there is increasing interest in distributions with more than one interval of modality, as well. Further examples are distributions which are arbitrary on some interval $(-\infty, x_1]$ with n^{th} concave or convex derivatives on the rest for even or odd n, respectively, or of distributions where the total variation of the n^{th} derivative is bounded by some prescribed constant ([10], [6], [15], [8], [5]).

In the first Part of this paper, the linear map from finite signed measures to their n^{th} derivatives is examined since in the present context the analysis of derivatives is much simpler than that of the original distributions. The derivatives are characterized as finite signed measures fulfilling natural moment conditions, the inverse map is computed and moment conditions on the distributions are transformed into additional moment conditions on the derivatives. Plainly, everything is based on integration by parts, but one has to proceed cautiously, since Fubini's Theorem is applied to the unbounded Lebesgue measure and signed measures.

In a large class of problems, the derivatives can be transformed into probability measures by a flip of the negative part followed by a suitable normalization. This amounts to the construction of a density s such that $s\mu$ is a probability measure for all derivatives μ in question. If such a construction works then moment sets of derivatives are affinely isomorphic to moment sets in the space of *all* distributions which are much easier to handle than the original ones. Such transformations are studied in the second Part.

The sketched paradigm is illustrated in Part 3 by way of the example of bell-shaped distributions with fixed turning points. The extremal ones are explicitely characterized in the cases of presence or absence of moment conditions. It is shown that the respective moment sets are even (weakly) homeomorphic and hence well-known integral representation results for general moment sets can be carried over to the case of bell-shaped distributions. These in turn yield Tchebysheff type inequalities.

Results of this type seem to have appeared first in the pioneering paper [10] by MULHOLLAND and ROGERS (1956) although some of the ideas can be found in earlier work, for example [4] from 1951. An integral representation theorem for moment sets of bell-shaped measures is proved in [10] by *ad hoc* methods and a characterization of extremal elements is formulated but not verified (cf. Remark 4.17; this seems to have been overlooked by various authors quoting this paper, cf. [5], [9], [6], [1]). The program is carried out for the much simpler set of distributions which have convex or concave derivatives on the right of some point x_1 and which are arbitrary elsewhere in [6]. This set can be transferred to the set of (nearly) *all* distributions on the real line, the extreme points of which are the point measures. Therefore the inverse of the corresponding transformation is even induced by a Markov kernel, which allows an elegant and simple treatment. KEMPERMAN announces an application of his method to bell-shaped distributions in [6] but this seems never to have appeared. The case of bell-shaped distributions is much more intricate as will be seen below, basically since there may be several exceptional points and not only one.

We plan to work out some more of the examples mentioned above in future work, in the first place distributions with derivatives of uniformly bounded variation.

2 Derivatives and their Moments

In this section, derivatives of distributions are defined and - if they exist - characterized by moment conditions. It is more convenient to work with functions than with measures in this context and we shall do so.

2.1 Derivatives of Distributions

Let f be a real function on the real line R. The *variation* of f is denoted by $\int |df|$ and f is of *finite variation* if $\int |df| < \infty$. Since only the measure defined by a function will be of interest, the following definition is reasonable: A real function F on R is *differentiable (in the wide sense)* with *(generalized) derivative* $f = F'$ if there is a right-continuous real function $f \in L^1(dx)$ of finite variation such that

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \text{ for every } x \in \mathsf{R}.$$

Note that F automatically has finite variation $\int |dF| = \int |f(y)|\, dy$ (cf. [3], Thm. III.2.20, p. 114) and is continuous if F' exists. Set

$$f(x+) = \lim_{y \to x, y > x} f(y), f(x-) = \lim_{y \to x, y < x} f(y)$$

and write $f(\pm\infty)$ for the limits at $\pm\infty$ (if the respective limits exist). Let further $\mathcal{V}_b$ denote the space of right-continuous real functions on R of finite variation such that $f(-\infty) = 0$ and $\mathcal{M}_f$ the space of finite signed measures on the Borel-σ-field $\mathcal{B}$ of the real line. An element of $\mathcal{M}_f$ will briefly be called a *measure* in the sequel.

Remark 2.1 *The spaces $\mathcal{V}_b$ and $\mathcal{M}_f$ are linearly isomorphic, the isomorphism and its inverse induced by*

$$\nu((-\infty, x]) = f(x), x \in \mathsf{R}.$$

Plainly, each *function f of finite variation induces some $\nu \in \mathcal{M}_f$ by $\nu((x, y]) = f(y+) - f(x+)$. Moreover, it has an at most countable set of jumps and hence the right-continuous regularization $f(\cdot+)$ coincides with f Lebesgue almost everywhere. Hence we may work with the regularization from the beginning and require f to be right-continuous. This requirement (and that f vanishes at $-\infty$) forces the map $f \mapsto \nu$ to be one-to-one.*

Assume now that $f = F'$ exists and denote the signed measures corresponding to F and f by μ and ν, respectively. Let further $\varphi \in \mathcal{C}_c^\infty(\mathsf{R})$ be a test function. Then integration by parts yields

$$-\int \varphi'\, d\mu = -\int \varphi'\, dF = -\int f(x)\varphi'(x)\, dx = -\int f\, d\varphi = \int \varphi\, df = \int \varphi\, d\nu.$$

Hence the distributional derivative of a measure is again a measure in our setting.

For most parts of this section, right-continuity of derivatives is not essential except in Section 2.4. Therefore the formulae will be given in a form which is correct for functions of finite variation not necessarily right-continuous, as well. This requires some extra '+'-signs but allows notational symmetry in right- and left-hand limits.

2.2 Integration by Parts

A technical condition is formulated first.

Definition 2.2 *A real function f fulfills condition (R) if there is $R > 0$ such that f is monotone on the intervals $(-\infty, -R]$ and $[R, \infty)$, respectively. Let $f \in \mathcal{R}$ if it is right-continuous, has finite variation, fulfills (R) and $f(x) \to 0$ as $|x| \to \infty$.*

The following integration by parts formula is convenient in the present context.

Theorem 2.3 *Let $f \in \mathcal{R}$. Then for every $k \geq 1$,*

$$\int_0^\infty x^k\, df(x) = -k \int_0^\infty x^{k-1} f(x)\, dx, \quad \int_{-\infty}^0 x^k\, df(x) = -k \int_{-\infty}^0 x^{k-1} f(x)\, dx.$$

where in each equation both sides may attain the values $\pm\infty$. In particular,

$$\int x^k\, df(x) = -k \int x^{k-1} f(x)\, dx \tag{1}$$

with the usual convention that that either both integrals exist – possibly with values $\pm\infty$ – or both sides do not exist.

The following simple observation is useful.

Lemma 2.4 *If $f \to 0$ for $|x| \to \infty$ then $f(x+), f(x-) \to 0$ if $|x| \to \infty$.*

Proof (of Theorem 2.3)**.** The measure df either is positive or it is negative on $[R, \infty)$. Since Lebesgue measure dx is positive and σ-finite on $[0, \infty)$ the product measure $dydf(x)$ is defined on the rectangle $[0, R)^2$ as a finite signed measure, and on $[0, \infty) \times [R, \infty)$ where it is positive or negative σ-finite. Letting I denote one of these intervals Fubini's theorem gives

$$\int_I \int_0^\infty \mathbf{1}_{(0,x)}(y) y^{k-1}\, dy\, df(x) = \int_0^\infty \int_I \mathbf{1}_{(0,x)}(y)\, df(x)\, y^{k-1}\, dy.$$

For $I = [0, R)$ both sides are finite; if $I = [R, \infty)$ then both sides are finite or equal to $\pm\infty$ depending on the sign of the product measure. Hence in view of $f(\infty) = 0$ (Lemma 2.4),

$$\begin{aligned} \int_0^\infty x^k\, df(x) &= k \int_0^\infty \int_0^x y^{k-1}\, dy\, df(x) = k \int_0^\infty \int_0^\infty \mathbf{1}_{(0,x)}(y) y^{k-1}\, dy\, df(x) \\ &= k \int_0^\infty \int_0^\infty \mathbf{1}_{(y,\infty)}(x)\, df(x)\, y^{k-1}\, dy = -k \int_0^\infty y^{k-1}\, f(y)\, dy. \end{aligned}$$

This proves the first equation. The second one is verified by the same computation. If one of the both sides of the third equation exists, then the expressions in the first and second equation are finite and hence the third identity holds. This completes the proof.

Remark 2.5 *The only task of requirement (R) is to ensure existence of a product measure and applicability of Fubini's theorem. The condition $f(x) \to 0$, $|x| \to \infty$, removes additive constants in the integration by parts formula.*

Lemma 2.6 *Let $f \in \mathcal{R}$ and suppose $y^k \in L^1(df(y))$. Then*

$$x^k f(x+), x^k f(x-) \longrightarrow 0 \ \text{ as } |x| \to \infty.$$

Proof. Since $y^k \in L^1(df(y))$ one can apply familiar integration by parts on compact intervals and let the interval boundaries tend to infinity. In particular,

$$\int_0^x y^k \, df(y) = x^k f(x+) - \int_0^x f(y) \, dy^k = x^k f(x+) - k \int_0^x y^{k-1} f(y) \, dy.$$

By Theorem 2.3 the integrals coincide in the limit $x \to \infty$. Hence $\lim_{x\to\infty} x^k f(x+)$ exists and even vanishes. Integration over $[x, 0)$ gives $x^k f(x-) \to 0$ for $x \to \infty$. This implies the result.

2.3 Moments of Derivatives

Two simple observations will be useful. The terms 'in-' or 'decreasing' will be used in the sense 'nonde-' and 'nonincreasing', respectively.

Lemma 2.7 *If $f = F'$ then*

$$f(x) \longrightarrow 0, \ |x| \to \infty.$$

Proof. Every real function f on R of finite variation is bounded and there are real numbers c_l and c_r such that

$$f(x) \longrightarrow c_l, \ x \to -\infty \ \text{ and } \ f(x) \longrightarrow c_r, \ x \to \infty.$$

In fact, there are bounded increasing functions f^+ and f^- such that $f = f^+ - f^-$ (choose for instance the minimal functions

$$f^+ = \frac{1}{2}\left(\int_{-\infty}^x |df| + f(x)\right), \ f^- = \frac{1}{2}\left(\int_{-\infty}^x |df| - f(x)\right),$$

called the *upper* and *lower* variation ([3], Lemma III.6.21, p. 154). Since $\int f(x)\,dx$ exists, the left and right integrals $\int_{-\infty}^0 f(x)\,dx$ and $\int_0^\infty f(x)\,dx$ are finite. This can hold only if $c_l = 0 = c_r$.

In the sequel, we are concerned with higher derivatives. Define $F^{(1)} = F'$ and, recursively, $F^{(n)} = \left(F^{(n-1)}\right)'$ for $n > 1$ provided the derivatives exist. If $F^{(n)}$ exists it is called the n^{th} *derivative* of F and F is called n *times differentiable*. For convenience of notation set $F^{(0)} = F$.

Lemma 2.8 *The following holds:*
(a) If $f = F'$ fulfills (R) then F fulfills (R) as well.
(b) If $F^{(n)}$ fulfills (R) then $F^{(k)} \in \mathcal{R}$ for every $k = 1, \ldots, n$.

Proof. If f, for instance, decreases on, say, $[R, \infty)$ then Lemma 2.7 implies $f \geq 0$ on $[R, \infty)$. Hence for $R \leq x < y < \infty$,

$$F(y) - F(x) = \int_x^y f(z) \, dz \geq 0$$

and F increases. The other three cases are treated similarly. The rest follows from this and Lemma 2.7

The first main result reads:

Theorem 2.9 *Let $n \geq 1$. If F is n times differentiable and $F^{(n)}$ fulfills condition (R) then $F^{(n)} \in \mathcal{R}$, $x^k \in L^1(dF^{(n)}(x))$ for $k = 0, \ldots, n$ and*

$$\int x^k \, dF^{(n)}(x) = 0 \text{ for } k = 0, \ldots, n-1, \quad \frac{(-1)^n}{n!} \int x^n \, dF^{(n)}(x) = \lim_{x \to \infty} F(x). \qquad (2)$$

Remark 2.10 *Since $\lim_{x \to -\infty} F(x) = 0$ the total mass $\int F'(x) \, dx$ of R is given by the second equality in (2).*

Proof. Let $n = 1$. In view of Lemma 2.8, Theorem 2.3

$$\int dF(x) = \int F^{(1)}(x) \, dx = -\int x \, dF^{(1)}(x)$$

and hence $x \in L^1(dF^{(1)}(x))$. By induction, $x^n \in L^1(dF^{(n)}(x))$ for every $n > 1$. This in turn implies $x^k \in L^1(dF^{(n)}(x))$ for all $n \geq 1$ and $k = 0, \ldots, n$.

For $n = 1$ and $k = 0$ the first integral in (2) boiles down to $\int dF^{(1)}(x)$. This integral is finite since $F^{(1)}$ has finite variation. Hence it may be written as $\lim_{x \to \infty} F^{(1)}(x+) - F^{(1)}(x-)$ which vanishes by Lemma 2.8 (b). Again induction using Theorem 2.3 yields $\int x^{n-1} \, dF^{(n)}(x) = 0$ for every $n \geq 1$. The first equation in (2) follows from the just proved result and

$$\int x^k \, dF^{(n)}(x) = \int x^k \, d\left(F^{(n-k-1)}\right)^{(k+1)}.$$

Similarly, starting with $n = 1$ and $k = 1$,

$$\lim_{x \to \infty} F(x) - \lim_{x \to -\infty} F(x) = \int dF(x) = -\int x \, dF^{(1)}(x),$$

the second identity is verified by induction (note that $\lim_{x \to -\infty} F(x) = 0$).

In the sequel, the symbol $g(x)|_a^b$ will denote $g(b+) - g(b-)$ if (partial) integration is carried out over a compact interval $[a, b]$, or $g(b-) - g(a+)$ for an open interval (a, b) and so on. Let us agree that $\int_a^b$ denotes integration over $[a, b]$, $\int_{b-}^{a+}$ over (a, b) etc. The following result establishes the inverse of the map $F \mapsto F^{(n)}$.

Theorem 2.11 *Suppose that F is n times differentiable for $n \geq 1$ and $F^{(n)}$ fulfills (R). Then*

$$F(x) = \frac{1}{k\,!} \int_{-\infty}^{x} (x - y)^k \, dF^{(k)}(y), \quad k = 1, \ldots n.$$

Proof. By Theorem 2.9 the integrals $\int (y - x)^k \, dF^{(k)}(y)$, $0 \leq k \leq n$, exist. For $k = 1$ the desired identity holds since $F^{(1)} \in \mathcal{R}$ by Lemma 2.8 and hence Lemma 2.6 implies

$$-\int_{-\infty}^{x} (y - x) \, dF^{(1)}(y) = -\left.(y - x) F^{(1)}(y)\right|_{-\infty}^{x} + \int_{-\infty}^{x} F^{(1)}(y) \, d(y - x) = F(x).$$

Let now $2 \leq k \leq n$ and assume that the identity holds for $k - 1$. Then by the same arguments

$$\begin{aligned} \int_{-\infty}^{x} (y - x)^k \, dF^{(k)}(y) &= \left.(y - x)^k F^{(k)}(y)\right|_{-\infty}^{x} - \int_{-\infty}^{x} F^{(k)}(y) \, d(y - x)^k \\ &= -k \int_{-\infty}^{x} (y - x)^{k-1} \, dF^{(k-1)}(y) = F(x). \end{aligned}$$

This proves the result.

A final result allows to transform moments of functions to moments of their derivatives. The *essential range* $\mathrm{essr}(G)$ of a function G is the set of those points $x \in \mathsf{R}$ such that there is (i) $y \leq x$ such that $G(y) \neq G(x+)$ and there is (ii) $y \geq x$ such that $G(y) \neq G(x-)$.

Remark 2.12 *(a) Plainly, the essential range is an interval where an endpoint belongs to the range if and only if G has a jump there.*
(b) If $G = F^{(n)}$ and the interiour of $\mathrm{essr}(G)$ *is (a,b) then* $\mathrm{essr}(F) = (a,b)$. *This follows from the observations that* $\mathrm{essr}(F)$ *is open since F is continuous, F is constant on each open interval on which $F^{(n)} \equiv 0$ and $F^{(n)}$ vanishes on any open interval on which F is constant.*

Theorem 2.13 *Suppose that F is n times differentiable and $F^{(n)}$ fulfills (R). For each $g \in L^1(dF(x))$ the function*

$$g^*(x) = \frac{(-1)^n}{(n-1)!} \int_0^x g(y)(x-y)^{n-1}\, dy \tag{3}$$

is defined for every x in the essential range of $F^{(n)}$ and

$$\int g(x)\, dF(x) = \int g^*(x)\, dF^{(n)}(x). \tag{4}$$

If g is locally integrable w.r.t. Lebesgue measure then g^* is defined everywhere.

Proof. By the same justification as for Theorem 2.3 Fubini's theorem applies. For $x \geq 0$ one computes

$$\begin{aligned}
& \frac{(-1)^n}{(n-1)!} \int_0^\infty \int_0^x g(y)(x-y)^{n-1}\, dy\, dF^{(n)}(x) \\
= & \frac{(-1)^n}{(n-1)!} \int_0^\infty g(y) \int_y^\infty (x-y)^{n-1}\, dF^{(n)}(x)\, dy \\
= & \frac{(-1)^{(n-1)}}{(n-1)!} \int_0^\infty g(y) \int_{-\infty}^y (x-y)^{n-1}\, dF^{(n)}(x)\, dy \\
= & \int_0^\infty g(y) \frac{1}{(n-1)!} \int_{-\infty}^y (x-y)^{n-1}\, dF^{(n)}(x)\, dy \\
= & \int_0^\infty g(y) F^{(1)}(y)\, dy = \int_0^\infty g(x)\, dF(x)
\end{aligned}$$

where the integrand $g^*(x)$ in the first line exists for $dF^{(n)}(x)$ almost every x. The third identity holds since the integrals w.r.t. $dF^{(n)}(x)$ of polynomials of degree less than n vanish. Addition of the corresponding identity on $(-\infty, 0)$ gives the desired formula for $F^{(n)}$ almost all x.

If $g^*(z)$ exists for one particular value z then it exists for each x inbetween 0 and z. Hence the set $D(g^*)$ where $g^*(x)$ exists is an interval containing 0. If now the essential range of $F^{(n)}$ is an interval with interiour (a,b), where $b \geq 0$ then $(c,b]$ is not a $dF^{(n)}(x)$-nullset for each $c < b$ and there is some $x \in (c,b]$ such that $g^*(x)$ is defined. This shows that $g^*(x)$ is defined on $[0,b)$ and even on $(0,b]$ if $F^{(n)}$ jumps at b. Now either $a \geq 0$ – and then we are done – or $a < 0$ and then the same arguments prove the result. Therefore the identity holds on $\mathrm{essr}(F^{(n)})$. The rest is clear.

Example 2.14 *Of particular interest are the powers* $g(x) = x^r$*, for natural numbers* $r \geq 0$*. For them one computes*

$$g^*(x) = (-1)^n \frac{r!}{(r+n)!} x^{r+n} \tag{5}$$

(which is defined everywhere). The proof is straightforward. Denote g^* *by* $g_r^{(n)}$ *to indicate the order of differentiability. Then* $g_r^{(1)}(x) = -x^{r+1}/(r+1)$ *and (usual) integration by parts gives the recursion* $g_r^{(n)} = -g_{r+1}^{(n-1)}$ *which proves the identity. In particular, even if* g *is bounded then* g^* *in general is not bounded since* $\mathbf{1}^* = (-1)^n x^n/n!$.

2.4 The Inverse

Next higher derivatives of differentiable functions are characterized. This amounts to the inverse of Theorem 2.9.

Theorem 2.15 *Let* $G \in \mathcal{R}$ *and suppose that for* $n \geq 1$, $x^n \in L^1(dG(x))$ *and*

$$\int x^k \, dG(x) = 0, \ \ k = 0, \ldots, n-1.$$

Then G *is the* n^{th} *derivative of some* $F \in \mathcal{V}_b$.

Remark 2.16 *Let* G *be of finite variation only. The moment condition for* $k = 0$ *amounts to* $\lim_{x \to -\infty} G(x-) = \lim_{x \to \infty} G(x+)$ *(and the existence of these limits). The requirement* $G \in \mathcal{R}$ *implies the normalization* $\lim_{x \to \infty} G(x+) = 0$.

Proof. Let $1 \leq k \leq n$. By assumption and Theorem 2.3, integration by parts applies and gives

$$\int_a^x (y-x)^k \, dG(y) = -(a-x)^k G(a-) - k \int_a^x (y-x)^{k-1} G(y) \, dy$$

for every $a \in \mathsf{R}$. One may let a tend to $-\infty$ which in view of Lemma 2.6 results in

$$\int_{-\infty}^x (y-x)^k \, dG(y) = -k \int_{-\infty}^x (y-x)^{k-1} G(y) \, dy.$$

Hence the integrals

$$F_{n-1}(x) = -\int_{-\infty}^x (y-x) \, dG(x) = \int_{-\infty}^x G(y) \, dy$$

exist. Moreover

$$\int x^{k-1} \, dF_{n-1}(x) = \int x^{k-1} G(x) \, dx = \frac{-1}{k} \int x^k \, dG(x)$$

for $k = 0, \ldots, n$ which implies

$$\frac{(-1)^{n-1}}{(n-1)!} \int x^{n-1} \, dF_{n-1}(x) = \frac{(-1)^n}{n!} \int x^n \, dG(x), \ \int x^k \, dF_{n-1}(x) = 0, \, k = 0, \ldots, n-2.$$

Finally,

$$\begin{aligned} F_{n-1}(\infty) &= \int G(x)\,dx = -\int x\,dG(x) = 0 \text{ if } n \geq 2, \\ F_{n-1}(\infty) &= \int G(x)\,dx = -\int x\,dG(x) \text{ if } n = 1. \end{aligned}$$

In summary, $F_{n-1} \in \mathcal{R}$ with the correct moment conditions and derivative G if $n \geq 2$ and $G = F'$ if $n = 1$ where $F = F_0$. Hence (backward) induction works and the theorem is proved.

Here is a summary of the previously proved results. Let $\mathcal{D}_n$ denote the set of those functions $F \in \mathcal{V}_b$ which are n-times differentiable such that $F^{(n)}$ fulfills (R). Let further $\mathcal{D}^n$ denote the set of those $G \in \mathcal{R}$ which fulfill the moment conditions

$$x^n \in L^1(dG(x)), \quad \int x^k\,dG(x) = 0, \ \ k = 0, \ldots, n-1. \tag{6}$$

Theorem 2.17 *The map*

$$\Phi : \mathcal{D}_n \longrightarrow \mathcal{V}_b, \ \ F \longmapsto F^{(n)}.$$

is a linear isomorphism from $\mathcal{D}_n$ onto $\mathcal{D}^n$. Its inverse is given by

$$\Phi^{-1} : \mathcal{D}^n \longrightarrow \mathcal{D}_n, \ G \longmapsto F = \Phi^{-1}(G), \ \ F(x) = \frac{1}{n\,!}\int_{-\infty}^{x} (x-y)^n\,dG(y)$$

and, moreover,

$$F(\infty) = \frac{(-1)^n}{n!}\int x^n\,dF^{(n)}.$$

Proof. This is a summary of Theorems 2.9, 2.15, 2.11 and formula (2).

3 Switching Measures

The characterization of extreme points of a convex set frequently can be reduced to the construction of an affine bijection onto another convex set with known extreme points. The simple method to be introduced below allows to transform certain functions of finite variation to probability measures on some Borel sets S of R. This is convenient since the extremal probability measures on S just are the Dirac measures $\varepsilon_x(A) = \mathbf{1}_A(x)$, $x \in S$, where A runs through the Borel sets $\mathcal{B}(S)$ of S.

Notation gets simpler if one switches from functions f of finite variation to the associated signed measures μ and we shall do so.

3.1 Convex Sets and Extreme Points

The main instrument for the characterization of extreme points will be the purely geometric Theorem 3.1 proved in [19]. Some definitions will be needed. A subset K of a linear space L is convex if it contains with any two points x and y the (compact) *line-segment* $[x, y] = \{z \in L : z = \alpha x + (1-\alpha)y, 0 \leq \alpha \leq 1\}$ (other types of intervals are also defined like on the real line). Suppose now that K is convex. An element

$x \in K$ is an *extreme point* of K if it is not in the interior (x, y) of a line-segment in K; the set of extreme points is denoted by ex$K$. K is said to be *linearly compact* if its intersection with a straight line in L either is empty or a line-segment compact in the (order-) topology of the line. Finally, K is a *(Choquet-) simplex* if the cone $C = \{\alpha \cdot (x, 1) \in L \times \mathsf{R} : \alpha \geq 0, x \in K\}$ is a lattice cone in its own order (i.e. the linear space $C - C$ is a vector lattice with nonnegative cone taken to be C, cf. [2], §28).

Theorem 3.1 *Suppose that K is a convex and linearly compact subset of a real linear space. Suppose further that $A : K \to \mathsf{R}^n$ is an affine map and W is a convex subset of $A[K]$. Set $H = A^{-1}(W)$. Then*
(a) H is a convex subset of K and

$$\text{ex}H \subset \quad \Big\{x \in H : x = \sum_{i=1}^{m} \alpha_i e_i,\ e_i \in \text{ex}K,\ \alpha_i > 0,\ \sum_{i=1}^{m} \alpha_i = 1,$$
$$\{A(e_1), \ldots, A(e_m)\} \text{ affinely independent, } 1 \leq m \leq n+1\Big\}.$$

(b) If in addition K is a simplex and W is a singleton then equality of sets holds in (a).

Linear compactness is inherited by special subsets: a convex subset F of K is a *face* if $(x, y) \cap F \neq \emptyset$ for $x, y \in K$ implies $x, y \in F$.

Lemma 3.2 *Let F be a face of the convex set K . Then*
(a) ex$F = F \cap$ ex$K$.
(b) If K is linearly compact then F is linearly compact.
(c) If K is a simplex then F is a simplex.

Proof. If $x \in F$ is contained in an open line-segment I of K then $I \subset F$ and x is not extremal in F. Hence each extreme point of F is extremal in K. The rest of (a) is obvious. To verify (b) denote the intersections of F and K with a fixed line by I and J, respectively. If I is empty or a singleton there is nothing to prove. Otherwise there is an inner point in $I \subset J$. Since F is a face, $I = J$ and I is compact since K is linearly compact. For the last part, choose $x \in F$ and assume that $(x, 1)$ dominates $\alpha(y, 1)$, $y \in K$, $\alpha \geq 0$, in the own order of C (defined above). This means that there are $z \in K$, $\beta \geq 0$, such that $x = \alpha y + \beta z$. Let h denote the linear functional on the linear span of C taking the constant value 1 on $K \times \{1\}$. Then $\alpha + \beta = h((x, 1)) = 1$ and x is a convex combination of y and z. Since F is a face, $y, z \in F$ and, in particular, $\alpha(y, 1)$ is an element of the cone generated by $F \times \{1\}$. Hence this cone inherits the lattice property from C (cf. [2]) and F is a simplex.

Lemma 3.3 *The image of a linearly compact set under an affine isomorphism is linearly compact.*

Proof. Let K be linearly compact. An affine isomorphism on K induces an affine isomorphism Υ between the affine spaces aff$K$ and aff$\Upsilon(K)$ generated by K and $\Upsilon(K)$, respectively, which restricted to a line is a homeomorphism onto the image of the line (in the line-topologies). Hence the intersection of a line L in $\mathrm{aff}\Upsilon(K)$ with $\Upsilon(K)$ is compact if and only if the intersection of $\Upsilon^{-1}(L)$ with K is compact. The latter holds by assumption and the assertion is proved.

Let $\mathcal{P}(\Omega)$ denote the set of probability measures on a measurable space $(\Omega, \mathcal{F})$ and $\mathcal{D}(\Omega)$ the set of Dirac measures; for $\mathcal{K} \subset \mathcal{P}(\Omega)$ and a family $\mathcal{G}$ of $\mathcal{F} - \mathcal{B}$–measurable functions set

$$\mathcal{K}^{\mathcal{G}} = \{\rho \in \mathcal{K} : \mathcal{G} \subset L^1(\rho)\}.$$

Lemma 3.4 *$\mathcal{P}^{\mathcal{G}}(\Omega)$ is a face in $\mathcal{P}(\Omega)$. In particular, it is linearly compact and a simplex. If, moreover, $\mathrm{ex}\mathcal{P}(\Omega) = \mathcal{D}(\Omega)$ then $\mathrm{ex}\mathcal{P}^{\mathcal{G}}(\Omega) = \mathcal{D}(\Omega)$.*

Proof. $\mathcal{P}^{\mathcal{G}}(\Omega)$ is a face since $\rho \in (\rho_1, \rho_2)$ implies $\rho_i \leq \alpha\rho$ for some $\alpha > 0$ and hence $\int |f|\, d\rho_i \leq \alpha \int |f|\, d\rho < \infty$. Linear compactness follows from Lemma 3.2 since $\mathsf{R}_+ \cdot \mathcal{P}(\Omega)$ is a lattice-cone in its own order and hence $\mathcal{P}(\Omega)$ is linearly compact ([7], condition 2^0 and p. 369). By the same lemma, $\mathcal{P}^{\mathcal{G}}(\Omega)$ inherits the simplex property from $\mathcal{P}(\Omega)$. Plainly, $\varepsilon_x \in \mathcal{P}^{\mathcal{G}}(\Omega)$ for every $x \in \Omega$ and hence $\mathcal{D}(\Omega) \subset \mathrm{ex}\mathcal{P}^{\mathcal{G}}(\Omega)$. The reverse inclusion follows from Lemma 3.2

Remark 3.5 *On arbitrary measurable spaces the extreme probability measures are those taking values 0 and 1 only. Fortunately, in most practical cases they are the Dirac-measures (cf. [19], Examples 2.1). This holds in particular for subsets Ω of Euclidean spaces endowed with the Borel-sigma field $\mathcal{F} = \mathcal{B}(\Omega)$.*

3.2 Switch Functions

A Borel function s will be called a *switch function* for $\mu \in \mathcal{M}_f$ with support $S \in \mathcal{B}$ if $s \in L^1(\mu)$, $\mathsf{R}\backslash S$ is a μ-nullset, $s \neq 0$ on S and $s\mu \in \mathcal{P}(S)$ where $s\mu$ is given by

$$s\mu(A) = \int_A s\, d\mu, \;\; A \in \mathcal{B}(S).$$

Given s and S let $\mathcal{K}_S$ be the collection of all $\mu \in \mathcal{M}_f$ for which s is a switch function. Finally, set $t = 1/s$ on S. We shall identify measures on S and their canonical extensions to R if convenient.

Lemma 3.6 *Let $\mathcal{G}$ be a family of Borel functions.*

(a) $\mathcal{K}_S$ and $\mathcal{K}_S^{\mathcal{G}}$ are convex. The map

$$\Upsilon : \mathcal{K}_S \longrightarrow \mathcal{P}^{\{t\}}(S), \; \mu \longmapsto s\mu$$

is an affine bijection with inverse

$$\Upsilon^{-1} : \mathcal{P}^{\{t\}}(S) \longrightarrow \mathcal{K}_S, \; \rho \longmapsto t\rho.$$

Moreover, $\Upsilon(\mathcal{K}_S^{\mathcal{G}}) = \mathcal{P}^{\{t\}\cup\mathcal{G}}(S)$.

(b) The extreme points of both, $\mathcal{K}_S$ and $\mathcal{K}_S^{\mathcal{G}}$, are the point measures $t(x)\varepsilon_x$, $x \in S$, and both sets are linearly compact and simplices.

Proof. Convexity of sets and affinity of Υ are clear. Since s does not vanish on S and $\mathsf{R}\backslash S$ is a null-set for all $\mu \in \mathcal{K}_S$, the trivial identities

$$\mu = \int t \cdot s \, d\mu, \;\; \rho = \int s \cdot t \, d\rho$$

for $\mu \in \mathcal{K}_S$ and $\rho \in \mathcal{P}^{\{t\}}(S)$ show that Υ is bijective. A measurable function h is μ-integrable if and only if the restriction of $t \cdot h$ to S is $s\mu$-integrable. This proves the equality of sets in (a). The sets $\mathcal{P}^{\{t\}}(S)$ and $\mathcal{P}^{\{t\}\cup\mathcal{G}}(S)$ are linearly compact and simplices by Lemma 3.4 and hence $\mathcal{K}_S$ and $\mathcal{K}_S^{\mathcal{G}}$ as well by Lemma 3.3. Finally, the extreme points of $\mathcal{P}^{\{t\}}(S)$ and $\mathcal{P}^{\{t\}\cup\mathcal{G}}(S)$ are the Dirac measures ε_x on S by Lemma 3.4 and Remark 3.5. These are transformed by Υ^{-1} to $t(x)\varepsilon_x$, $x \in S$.

Remark 3.7 *(a) In general, the construction does not yield an affine isomorphism onto* $\mathcal{P}(S)$ *since there may exist* $\rho \in \mathcal{P}(S)$ *for which* t *is not integrable.*
(b) Without the condition that S *is a nullset, the map* Υ *may fail to be one-to-one; cf. Example 4.11.*

4 Bell-Shaped Distributions: Extreme Points

Important special unimodal distributions are the bell-shaped ones. They will be introduced below and extreme points of their moment sets will be characterized. The concept in [10] is embedded into the present setting.

4.1 Definition and Basic Properties

Let us introduce a notion of bell-shaped distributions.

Definition 4.1 *A function* $F \in \mathcal{D}_n$ *is* bell-shaped to the n^{th} order *with turning-points* $x_1 < \cdots < x_n$ *if* $F^{(n)}$ *is continuous at* $x_1, \ldots, x_n$ *and* $(-1)^r F^{(n)}$ *increases on* (x_r, x_{r+1}), $r = 0, 1, \ldots, n$, *where* $x_0 = -\infty$ *and* $x_{n+1} = \infty$.

Note that bell-shaped functions are bounded since they are of finite variation. Let us give some simple examples.

Example 4.2 *We are particularly interested in cumulative distribution functions (c.d.f.)* F. *C.d.f. of uniform distributions on intervals are bell-shaped to the first order but not to the second one; those with a triangular density are bell-shaped to the second order but not to the third one; etc. Let* d *be the well-known Cantor function on* $[0, 1]$ *('devil's staircase') which is continuous, increasing from* $d(0) = 0$ *to* $d(1) = 1$ *and differentiable outside Cantor's discontinuum with vanishing derivative (cf. [12], p.145). Let* $f = 1 - d$ *on* $(0, 1]$, $f(x) = d(x + 1)$ *on* $[-1, 0]$ *and* $f(x) = 0$ *off* $[-1, 1]$. *This function* f *is the probability density of a c.d.f.* F *which is bell-shaped to the first but not to the second order.*

Normal c.d.f. F *are bell-shaped to any order. This holds since* $F^{\{n\}} = h \cdot H_{n-1}$ *where* h *does not vanish anywhere and* H_n *is the* n^{th} *Hermite polynomial (which has precisely* n *real roots). All normal distributions with mean* x_1 *are bell-shaped to the first order with turning point* x_1, *there is precisely one of the second order with* *turning points x_1, x_2 (having mean $m = (x_1 + x_2)/2$ and standard deviation $x_2 - m$) and at most one of order $n \geq 3$ with prescribed turning points $x_1, \ldots, x_n$.*

Let us agree that a (possibly higher) derivative f of a bell-shaped function *changes direction precisely r times* if there are real numbers $z_1 < \cdots < z_r$ such that $(-1)^k f$ increases and is not constant in (x_k, x_{k+1}), where $z_0 = -\infty$, $z_{r+1} = \infty$ and $0 \leq k \leq r$.

Lemma 4.3 *Suppose that F is bell-shaped to the n^{th} order and is not identically constant. Then:*
(a) $F^{(n)}$ changes direction precisely n times.
(b) F is bell-shaped to each order $1 \leq k \leq n$.
(c) $F(\infty)$ is finite and strictly positive; the measure μ associated to F is positive with $\mu(\mathsf{R}) = F(\infty)$.

All arguments in the following proof appear in a similar form in the proofs of Lemma 4 and its Corollary in [10].

Proof. Let $n = 1$. Since $f = F^{(1)}$ vanishes at $\pm\infty$ and does not vanish identically bell-shapedness implies that f is strictly positive somewhere. Moreover, continuity at x_1 implies $f(x_1) > 0$ and $f(x) > 0$ for some point $x < x_1$ and some point $x > x_1$. Hence f increases on $(-\infty, x_1)$ and decreases on (x, ∞) and is not constant on any of these intervals. This shows that f changes direction precisely once. Let now $n \geq 2$. Suppose that $f = F^{(n)}$ changes direction precisely r times. Plainly, $r \leq n$ and f changes sign at most $r - 1$ times. Hence $F^{(n-1)}$ changes direction at most $r - 1$ times. Similarly, $F^{(1)}$ changes direction at most $r - n + 1$ times and at least once. Hence $r \leq n$ and $r - n + 1 \geq 1$ and thus $r = n$. This proves (a). Since derivatives of order $k < n$ are continuous part (b) holds as well. Finally, $F^{(1)}$ is positive, nonconstant and fulfills $F(\infty) = \int f(x)\, dx$ which implies (c).

Remark 4.4 *Since $(-1)^r F^{(n)}$ increases on $[x_r, x_{r+1})$ and is right-continuous, the function $(-1)^r F^{(n-1)}$ is convex on this interval. Since F is bell-shaped to the order one F is convex on some interval $(-\infty, \xi)$ and concave on (ξ, ∞). Hence it is unimodal with mode ξ (for example in the sense of [4]).*

4.2 Bell-Shaped Distributions Extremal under Constraints

We are going to characterize the extreme points of sets $\mathcal{H} = \mathcal{H}(x_1, \ldots, x_n)$ of c.d.f. which are bell-shaped to the n^{th} order with prescribed turning points $x_1 < \cdots < x_n$ and of subsets defined by restrictions on their moments. For $F \in \mathcal{H}$ let $\mu = \Psi(F^{(n)})$ be the measure associated to $F^{(n)}$, $\mathcal{J} = \Psi \circ \Phi(\mathcal{H})$ and

$$s(x) = \frac{1}{n!} \prod_{i=1}^{n} (x_i - x), \;\; t(x) = \frac{1}{s(x)}, \;\; S = \mathsf{R} \backslash \{x_1, \ldots, x_n\}. \tag{7}$$

Lemma 4.5 *The function s is a common switch function for all $\mu \in \mathcal{J}$. More precisely,*

$$\mathcal{K}_S \supset \mathcal{J} = \{\mu \in \mathcal{K}_S^{\mathcal{G}} : \int f\, d\mu = 0, f \in \mathcal{G}\}, \;\; \textit{where } \mathcal{G} = \{x^k : 0 \leq k \leq n-1\}.$$

Proof. Choose $F \in \mathcal{H}$ and let $\mu = \Psi(F^{(n)})$. Since $F^{(n)}$ is continuous at each x_i the turning points form a μ-nullset. By Theorem 2.17, $F^{(n)} \in \mathcal{D}^n$ with n^{th} moment equal to $(-1)^n n!$. Hence $s \in L^1(\mu)$ and $\int s\,d\mu = 1$. By definition, the measures $(-1)^r\mu$ are positive on (x_r, x_{r+1}) and s has sign $(-1)^r$ on this interval. Since s does not vanish on S we conclude that s is a switch function for μ. This proves $\mathcal{J} \subset \mathcal{K}_S$ and the validity of the moment equalities.

To verify the remaining inclusion, choose $\mu \in \mathcal{K}_S$ satisfying the moment conditions and let $G = \Psi^{-1}(\mu)$. Plainly $x^n \in L^1(\mu)$ and hence $G = F^{(n)}$ for some $F \in \mathcal{V}_b$ by Theorem 2.17. G is continuous at all x_i since $s(x_i) = 0$. By the form of s and since $s\mu$ is a positive measure $(-1)^r G$ increases on (x_r, x_{r+1}). In summary, F is bell-shaped to the n^{th} order. It is even a c.d.f. since by Lemma 4.3 the moment conditions and $\int s\,d\mu = 1$ imply $F(\infty) = ((-1)^n/n!)\int x^n\,d\mu = 1$.

Things can be put together now to characterize extreme points of moment sets in $\mathcal{H}$. Given $H \subset \mathcal{M}_f$, a collection $\mathcal{G}$ of Borel-measurable functions and a subset D of $\mathsf{R}^{\mathcal{G}}$ a *generalized moment set* is defined by

$$H(\mathcal{G}, D) = \left\{\mu \in \mathcal{H} : \mathcal{G} \subset L^1(\mu), \left(\int g\,d\mu\right)_{g\in\mathcal{G}} \in D\right\}.$$

If $\mathcal{G} = \emptyset$ then $H(\mathcal{G}, D) = H$. In most applications $\mathcal{G}$ is finite and D a singleton representing the prescribed values of moments or – more generally - a product $\prod_{g\in\mathcal{G}}(-\infty, a_g]$. If $\mathcal{G} = \{g_1, \ldots, g_p\}$ and $D = \{(d_1, \ldots d_p)\}$ we shall write $H(\mathcal{G}, d)$ for $H(\mathcal{G}, D)$. The case $\mathcal{G} = \emptyset$ is included setting $p = 0$ and $H(\emptyset, d) = H$.

Theorem 4.6 *Let $\mathcal{G} = \{g_1, \ldots g_p\}$ and $d = (d_1, \ldots, d_p)$. Precisely those functions F in $\mathcal{H}(\mathcal{G}, d)$ are extremal which are of the form*

$$F(x) = \sum_{\xi_k \le x} \gamma_k (x - \xi_k)^n, \;\; \xi_0 < \cdots < \xi_m, \; \gamma_k \neq 0, \;\; 0 \le k \le m,$$
$$\left\{\left(1, \xi_k, \ldots \xi_k^n, g_1^*(\xi_k), \ldots, g_p^*(\xi_k)\right) : 0 \le k \le m\right\} \text{ linearly independent,}$$
$$n \le m \le n + p.$$

Each interval (x_i, x_{i+1}) contains at least one of the points ξ_k. If some of the moment conditions are given by inequalities $\int g_i\,d\mu \le d_i$ then each extremal element has the stated form.

This is one (and the more heavy) half of Theorem 6 in [10] (cf. Remark 4.17).

Remark 4.7 *(a) Plainly, the mass of the measure associated to an extremal c.d.f. F is concentrated on $[\xi_0, \xi_m]$.*
(b) The extremal c.d.fs.

$$F(x) = \sum_{k=0}^{n} \gamma_k \mathbf{1}_{(\xi_k,\infty)}(x)(x - \xi_k)^n$$

are polynomial splines of n^{th} order with knots ξ_k given in terms of the highest order elements $\mathbf{1}_{[\xi_k,\infty)}(x)(x - \xi_k)^n$ of the natural base (cf. [13], p. 111).

For later use, part of the proof is formulated separately.

Lemma 4.8 *The map $\Psi \circ \Phi$ is an affine isomorphism from $\mathcal{H}(\mathcal{G}, D)$ onto*

$$\Psi \circ \Phi(\mathcal{H}(\mathcal{G}, D)) = \mathcal{J}(\mathcal{G}^*, D^*)$$

where

$$\mathcal{G}^* = \left\{\mathbf{1}, x, \ldots, x^{n-1}\right\} \cup \{g^* : g \in \mathcal{G}\}, \quad D^* = \{0\} \times D \subset \mathsf{R}^{n+p}.$$

Proof. By Theorem 2.13

$$\int g \, d\rho = \int g^* \, d\mu$$

for each $g \in L^1(d\rho)$ where $\rho \in \mathcal{H}(\mathcal{G}, d)$, $\mu = \Psi \circ \Phi(\rho)$, g^* is defined over $\mathrm{essr}(\Phi(\rho))$ and given there by (3). Thus g^* is defined even on

$$\bigcup \left\{\mathrm{essr}(\Phi(\rho)) : \rho \in \mathcal{H}^{\{\mathcal{G}\}}\right\}.$$

This implies the assertion.

Proof of the Theorem. The moment set is lifted to the level of n^{th} derivatives or rather the associated measures. Because of Lemma 4.8 the proof is a straightforward application of Theorem 3.1(b) to

$$\begin{aligned} K &= \mathcal{K}_S^{\mathcal{G}^*}, \\ A &: K \longrightarrow \mathsf{R}^{n+p}, \ \mu \longmapsto \left(\left(\int x^k \, d\mu, \ 0 \le k \le n-1\right), \left(\int g_i^* \, d\mu : 1 \le i \le p\right)\right), \\ W &= \{d^*\} \subset \mathsf{R}^{n+p}, \ H = A^{-1}(W) = \mathcal{J}(\mathcal{G}^*, d^*). \end{aligned}$$

The assumptions of Theorem 3.1 are met by Lemmata 3.6 and 4.5. Since by Lemma 3.6(b) the extreme points of K are the point measures $t(x)\varepsilon_x$ the extremal elements ν of $\mathcal{J}(\mathcal{G}^*, d^*)$ are characterized by the conditions

$$\nu = \sum_{k=0}^{m} \alpha_k t(\xi_k) \varepsilon_{\xi_k}, \ \alpha_k > 0, \ \sum_{k=0}^{m} \alpha_k = 1, \ \xi_0 < \cdots < \xi_m \in S, \tag{8}$$

$$\left\{\left(1, t(\xi_k), t(\xi_k)\xi_k, \ldots, t(\xi_k)\xi_k^{n-1}, g_1^*(\xi_k)t(\xi_k), \ldots, g_p^*(\xi_k)t(\xi_k)\right) : 0 \le k \le m\right\} \tag{9}$$

linearly independent, $0 \le m \le n + p$.

Multiplication by $s(\xi_k)$ and taking into account the form of s shows that the vectors in (9) can be replaced by

$$\left\{\left(1, \xi_k, \ldots, \xi_k^n, g_1^*(\xi_k), \ldots, g_p^*(\xi_k)\right) : 0 \le k \le m\right\}.$$

By Lemma 4.3 (a) and bell-shapedness, each of the measures in (8) must charge each interval (x_i, x_{i+1}) and hence $n \le m \le n + p$.

The representation of ν in (8) can be simplified as well. Each ν is of the form $\sum_{k=0}^{m} \beta_k \varepsilon_{\xi_k}$ with real numbers $\beta_k \neq 0$. On the other hand, such an element of $\mathcal{J}$ can be written in the form (8) setting $\alpha_k = s(\xi_k)\beta_k$ since s is a switch function and hence $s(\xi_k)\beta_k > 0$ and $\sum s(\xi_k)\beta_k = 1$.

The corresponding statement for inequality constraints follows in the same way and application of the inversion formula completes the proof.

Extremal bell-shaped functions (without restrictions) can be characterized explicitely.

Theorem 4.9 *A c.d.f. F is an extreme point of $\mathcal{H}$ if and only if it is of the form*

$$F(x) = \sum_{\xi_k \le x} \gamma_k (x-\xi_k)^n, \;\; x_k < \xi_k < x_{k+1}, \;\; \gamma_k = \frac{1}{\prod_{j\neq k}(\xi_j - \xi_k)}, \;\; 0 \le k \le n.$$

Proof. We shall continue with the proof of Theorem 4.6 setting $\mathcal{G} = \emptyset$ and $p = 0$. Plainly, we have $m = n$. The extreme points of $\mathcal{J}$ are precisely those elements in $\mathcal{J}$ which have the form

$$\nu = \sum_{k=0}^{n} \beta_k \varepsilon_{\xi_k}, \; \beta_k \neq 0, \; \xi_0 < \cdots < \xi_m \in S, \;\; \{(1, \xi_k, \ldots, \xi_k^n)\} \text{ linearly independent,} \tag{10}$$

Since each interval is charged by ν, $x_k < \xi_k < x_{k+1}$. The vectors in (10) are automatically linearly independent, since the values ξ_i are mutually distinct and thus the Vandermonde determinant V (of the Vandermonde matrix $\left(\xi_k^j\right)$) does not vanish.

The condition $\nu \in \mathcal{J}$ can be removed as well. For a measure $\nu = \sum_{k=0}^n \beta_k \varepsilon_{\xi_k}$ the moment conditions boil down to the system of linear equations

$$\sum_{k=0}^{n} \beta_k \xi_k^j = 0, \;\; 0 \le j \le n-1; \;\; \sum_{k=0}^{n} \beta_k \xi_k^n = (-1)^n n\,!.$$

Given $x_k < \xi_k < x_{k+1}$, Cramer's rule gives the unique solution

$$\beta_k = V^{-1}(-1)^n n\,!(-1)^{n+1+k+1} \prod_{i<j;i,j\neq k} (\xi_j - \xi_i) = \frac{n\,!}{\prod_{j\neq k}(\xi_j - \xi_k)}. \tag{11}$$

In summary, a discrete measure $\sum_{k=0}^n \beta_k \varepsilon_{\xi_k}$ is in $\mathcal{J}$ and even extremal there if and only if its coefficents are given by (11). Application of the inversion formula completes the proof.

Remark 4.10 *For any extreme point μ of $\mathcal{J}$ and any Borel function h the integral*

$$\int h \, d\mu = \sum_{k=0}^{n} \frac{h(\xi_k)}{\prod_{j\neq k}(\xi_j - \xi_k)} \tag{12}$$

is the n^{th} divided difference of h w.r.t. the nodes ξ_k (cf. [13]). This is intimately connected with the moment conditions (6) since they mean that the functional defined by (12) is 'exact of degree $n-1$'.

Finally, Remark 3.7 is completed.

Example 4.11 *The map $\mu \mapsto s\mu$ is not one-to-one on $\{\mu \in \mathcal{M}_f : s\mu \in \mathcal{P}^{\{t\}}(S)\}$ since mass in turning points is annihilated. For instance, in the case $n = 1$, let $\xi_0 < x_1$, $\mu = (x_1 - \xi_0)^{-1}\varepsilon_{\xi_0}$ and $\nu = (x_1 - \xi_0)^{-1}\varepsilon_{\xi_0} + (\xi_0 - x_1)^{-1}\varepsilon_{x_1}$. Then $s\mu = \varepsilon_{\xi_0} = s\nu$. On the other hand, $t(s\mu) = \nu = t(s\nu)$.*

4.3 Related Notions

The following definition was given in [10] and adopted by various authors (for instance [8] or [5]). Classical n^{th} (left- and right-hand) derivatives will be denoted by $F_-^{\{n\}}$, $F_+^{\{n\}}$ and $F^{\{n\}}$, respectively. For simplicity of notation let $F^{\{0\}} = F$.

Definition 4.12 (i) *A function F is* smooth to the n^{th}-order, *$n \geq 1$, $F^{\{n\}}$ exists and is continuous, except perhaps in a finite number of points, and $F_-^{\{n\}}$ and $F_+^{\{n\}}$ exist everywhere and are left- and right-continuous, respectively.*
(ii) *F is* bell-shaped to the n^{th} order in the narrow sense *(of [10]) with turning-points $x_1 < \cdots < x_n$ if it is smooth to the n^{th} order, $F^{\{n\}}$ is continuous at $x_1, \ldots, x_n$ and $(-1)^r F_-^{\{n\}}$ and $(-1)^r F_+^{\{n\}}$ increase on (x_r, x_{r+1}), $r = 0, 1, \ldots, n$, where $x_0 = -\infty$ and $x_{n+1} = \infty$.*

In contrast to the previous definition the last one allows unbounded functions. Under the additional requirement of boundedness Definition 4.12 is more restrictive than Definition 4.1.

Proposition 4.13 *Let F be bell-shaped to the n^{th} order in the narrow sense and bounded. Then F is n times differentiable with $F^{(k)} = F_+^{\{k\}}$ for all $k = 1, \ldots, n$ and $F^{(n)}$ fulfills (R). In particular, F is bell-shaped.*

For the proof of Proposition 4.13, Lemma 4 from [10] will be borrowed. Note that it is a (simple) isolated result derived directly from the definitions. Note further that the requirement on F to be a c.d.f. is not needed since boundedness is sufficient (c.f. p. 210 in the same reference).

Lemma 4.14 *Let F be bell-shaped to the n^{th} order in the narrow sense and bounded. Then:*

(a) $F_+^{\{n\}}(x) \to 0$ as $|x| \to \infty$.
(b) F is bell-shaped to the k^{th} order for all $k = 1, \ldots, n$.

Proof (of Proposition 4.13)**.** $F_+^{\{n\}}$ is piecewise monotone by definition. It is bounded by the smoothness assumption and by Lemma 4.14 (a). Hence it is of finite variation. By part (b) of the lemma, this property is inherited by each $F_+^{\{k\}}$. For $k < n$ these functions are continuous and their derivatives exist in all except perhaps a finite number of points. Hence they are absolutely continuous and the $F_+^{\{k+1\}}$ are generalized derivatives. Since all functions in question are right-continuous we conclude that $F_+^{\{k\}} = F^{(k)}$ for all $k = 1, \ldots, n$. Finally, $F^{(n)}$ fulfills (R) by the monotonicity requirement.

Let now $\tilde{\mathcal{H}}$ denote the set of c.d.f. bell-shaped to the n^{th} order in the sense of Definition 4.12 and $\mathcal{H}$ as defined on page 13. If a distribution μ has a bell-shaped c.d.f. then μ will be addressed as a *bell-shaped distribution*

Example 4.15 *There are bell-shaped distributions which are not bell-shaped in the narrow sense, i.e. $\tilde{\mathcal{H}} \subset \mathcal{H}$, $\tilde{\mathcal{H}} \neq \mathcal{H}$. The density constructed from devil's staircase in Example 4.2, for example, violates the differentiability conditions.*

On the other hand, $\tilde{H}$ is a pleasant subset of $\mathcal{H}$.

Proposition 4.16 *$\tilde{\mathcal{H}}$ is a face in $\mathcal{H}$ and* ex$\tilde{\mathcal{H}}$ = ex$\mathcal{H}$.

Proof. By Proposition 4.13, $\tilde{\mathcal{H}} \subset \mathcal{H}$. The inclusion ex$\mathcal{H} \subset \tilde{\mathcal{H}}$ is obvious. Therefore and by Lemma 3.2 it is sufficient to verify that $\tilde{\mathcal{H}}$ is a face in $\mathcal{H}$. To this end, choose $F \in \tilde{\mathcal{H}}$ and assume that $F = \alpha G + (1-\alpha)H$ for $G, H \in \mathcal{H}$ and $\alpha \in (0,1)$. Since the derivatives $G^{(k)}$ and $H^{(k)}$ are continuous for $k = 0, \ldots, n-1$, they are derivatives in the classical sense. The n^{th} derivative $F^{(n)} = F_+^{\{n\}}$ exists and is continuous except on the finite set J of finite jumps. At a point x the functions $G^{(n)}$ and $H^{(n)}$ either are continuous or have a finite jump since they are of finite variation. If $F^{(n)}$ is continuous at x then $G^{(n)}$ and $H^{(n)}$ are continuous at x too since by bell-shapedness possible jumps of these two functions have the same direction. Hence $G^{(n)}$ or $H^{(n)}$ may jump at points $x \in J$ only. At points $x \notin J$ the function $G^{(n-1)}$ thus is continuously differentiable and $G^{\{n\}}(x) = G^{(n)}(x)$. Suppose now that $g = G^{(n)}$ has a jump at ξ. We claim that $g(\xi) = G_+^{\{n\}}(\xi)$. There is $\varepsilon > 0$ such that g is continuous on $[\xi, \xi + 2\varepsilon]$. Choose a sequence $\xi_n \to \xi$ in $(\xi, \xi+\varepsilon)$ and set $h_n(x) = G^{(n-1)}(x+\xi_n)$ and $h(x) = G^{(n-1)}(x+\xi)$. Then $h_n \to h$ on $[0,\varepsilon]$ pointwise by continuity of $G^{(n-1)}$. Each h_n is differentiable on $[0,\varepsilon]$, and $h_n^{\{1\}}(x) = g(x+\xi_n) \to g(x+\xi)$ uniformly on $[0,\varepsilon]$ since g is continuous on the compact interval $[\xi, \xi+\varepsilon]$. This proves that g is the derivative of $G^{(n-1)}$ on $[\xi, \xi+\varepsilon]$ and, in particular, that the right-hand derivative at ξ exists and is right-continuous there. The corresponding property of left-hand derivatives is verified similarly. We conclude that G and also H are elements of $\tilde{\mathcal{H}}$ and hence this set is a face of $\mathcal{H}$. This completes the proof.

Remark 4.17 Mulholland *and* Rogers *(1958) ([10])* claim*:*
A c.d.f. $F \in \tilde{\mathcal{H}}$ is an extreme point of $\tilde{\mathcal{H}}$ if and only if it is of the form

$$F(x) = \sum_{\xi_i \le x} \gamma_i (x-\xi_i)^n,\ \xi_i, \gamma_i \in \mathsf{R},\ 0 \le i \le m,$$
$$\{(1, \xi_i, \ldots, \xi_i^n),\ 0 \le i \le m\}\ \textit{linearly independent},\ \ 0 \le m \le n,$$

cf. their Theorem 6. They prove *that every c.d.f. bell-shaped in the narrow sense is a mixture of c.d.f. having the just specified form. They further claim that all extreme points of $\tilde{\mathcal{H}}$ are such functions and argue that this can be proved along the same lines as for ordinary moment sets of c.d.f. Inspection of their respective proof reveals that they crucially appeal to the fact, that if an extreme point is the barycenter of a probability measure on the extremal set then this measure is the Dirac measure in the extreme point. A common sufficient condition for this is measure convexity of the set in question (cf. Remark 5.5). On the other hand, $\tilde{\mathcal{H}}$ definitely is* not *measure convex as will be shown in Remark 5.5. Hence additional arguments – for instance those given above – are needed to identify the extreme points.*

5 Bell-Shaped Distributions: Extremal Decomposition and Bounds for Moments

Suppose that for each element μ of some set H of measures there is a decomposition

$$\mu(A) = \int_M \nu(A)\, dp_\mu(\nu), \ \ A \in \mathcal{A}, \tag{13}$$

where M is a subset of H, $\mathcal{A}$ is the σ-algebra on which the measures in H live and p_μ is a probability measure on the evaluation σ-algebra $\Sigma(M)$ on M generated by the functions $\nu \mapsto \nu(B)$, $B \in \mathcal{B}$. Suppose further that for some cost function C on H

$$C(\mu) \leq \int_M C(\nu)\, dp_\mu(\nu).$$

Then

$$\sup_{\mu \in H} C(\mu) \leq \sup_{p \in \mathcal{P}(M)} \int_M C(\nu)\, dp(\nu) \leq \sup\{C(\nu) : \nu \in M\} = U(C)$$

and $U(C)$ is a best upper bound of C on H. Frequently $C(\mu)$ is the mass $\mu(A)$ of a set, a moment $\int x^k\, d\mu(x)$ or a generalized moment $\int g\, d\mu$, for instance with $g(x) = \mathbf{1}_{[a,\infty)}(x-a), a > 0$ in the case of stop-loss premiums in insurance (cf. [17]). By this paradigm best upper and lower bounds for functionals on sets of measures can be determined.

In the present setting, H is defined by moment and by differentiability conditions. Integral representations as in (13) will be obtained by a transformation to a pure moment set i.e. the differentiability conditions are transformed into moment conditions by the previously developed methods. For such moment sets integral representation results with $M = \mathrm{ex}H$ exist and can be applied. Combined with the identification of extremal elements in the previous sections they provide the desired bounds for functionals.

5.1 Measure Theoretic Preparations

We shall show now that the map $F \mapsto F^{(n)} \mapsto s\mu$ where μ is the signed measure induced by $F^{(n)}$, i.e. in the previously introduced notation

$$\Theta : F \longmapsto s\mu = s(\Upsilon \circ \Psi \circ \Phi(F)),$$

has the best properties we can hope for.

Some notation is needed. If Ω is a Borel set of the real line with Borel-σ-algebra $\mathcal{B}(\Omega)$ then the functionals $\nu \mapsto \int \phi\, d\nu$ with bounded continuous functions ϕ induce the *weak topology* on $\mathcal{P}(\Omega)$. For any subset H of $\mathcal{P}(\Omega)$ let $\upsilon(H)$ be trace of the weak topology on H and $\mathcal{B}(H)$ the corresponding (Borel-)σ-algebra. Further, recall the definition of the switch function s and the set S in (7) and note that measures will be identified with their c.d.f. where this makes sense and is convenient.

Theorem 5.1 *The map Θ is an affine weak homeomorphism from $\mathcal{H}(x_1, \ldots, x_n)$ onto*

$$\Theta(\mathcal{H}) = \left\{ \mu \in \mathcal{P}(S) : \int \frac{x^k}{\prod_{i=1}^n (x_i - x)}\, d\mu = 0;\ k = 0, \ldots, n-1 \right\}.$$

.

Proof. By Theorem 2.17 and Lemma 3.6, the image $\Theta(\mathcal{H})$ has the asserted form and Θ is one-to-one and onto. To verify continuity of Θ^{-1} choose a sequence μ_n converging weakly to μ in $\Theta(\mathcal{H})$ (all spaces in question are metrizable). The sequence $\Theta^{-1}(\mu_n)$ converges to $\Theta^{-1}(\mu)$ if and only if

$$\int \phi\, d\Theta^{-1}(\mu_n) \longrightarrow \int \phi\, d\Theta^{-1}(\mu),\ \ n \to \infty,$$

for each function ϕ on S with compact support $\mathrm{supp}(\phi)$ contained in one of the open intervals (x_i, x_{i+1}). By (4),

$$\int \phi \, d\Theta^{-1}(\nu) = \int \frac{n}{\prod_{i=1}^{n}(x - x_i)} \int_0^x \phi(y)(x-y)^{n-1} \, dy \, d\nu(x)$$

where ν either denotes μ or μ_n. By the special form of ϕ the integrand $\phi^* \cdot t$ is defined everywhere, continuous and has compact support in (x_i, x_{i+1}) equal to $\mathrm{supp}(\phi)$. Hence

$$\int \phi \, d\Theta^{-1}(\mu_n) = \int \phi^* \cdot t \, d\mu_n \longrightarrow \int \phi^* \cdot t \, d\mu = \int \phi \, d\Theta^{-1}(\mu)$$

and thus Θ^{-1} is continuous.

Now choose μ_n weakly convergent to μ in $\mathcal{H}$ and a test function ϕ of the above type. We may and shall assume that ϕ even is infinitely often differentiable. Then

$$\int \phi \, d\Theta(\nu) = (-1)^n \int \phi^{\{n\}} \cdot s \, d\nu,$$

again with ν equal to μ or μ_n. The integrand

$$\tilde{\phi}(x) = \frac{1}{n!} \prod_{1=1}^{n} (x - x_i) \cdot \phi^{\{n\}}(x)$$

on the right-hand side is bounded and continuous and hence

$$\int \phi \, d\Theta(\mu_n) = \int \tilde{\phi} \, d\mu_n \longrightarrow \int \tilde{\phi} \, d\mu = \int \phi \, d\Theta(\mu)$$

which shows that Θ is continuous.

Remark 5.2 *In [6], (5.37), there is a map T which in a similar but simpler setting corresponds to our map Θ^{-1}. To verify that T is a homeomorhism erroneously weak compactness of $\mathcal{P}(\mathsf{R})$ is assumed (and the set $T(\mathcal{P}(\mathsf{R}))$ corresponding to our set $\mathcal{H}$ is* not *compact). On the other hand, the proof can be rectified by the arguments in Theorem 5.1. Unfortunately, the mentioned (minor) error has propagated through the literature, see for example [1], p. 62.*

Plainly, the set $\mathcal{H}$ is not weakly closed in $\mathcal{P}(\mathsf{R})$ but it inherits all pleasant topological and measure theoretical properties from the pure moment set $\Theta(\mathcal{H})$.

Proposition 5.3 *On $\Theta(\mathcal{H})$ and $\mathcal{H}$ the evaluation σ-algebra and the Borel σ-algebra coincide:*

$$\Theta(\mathcal{H}) \in \Sigma(\mathcal{P}(S)) = \mathcal{B}(\mathcal{P}(S)), \;\; \Sigma(\Theta(\mathcal{H})) = \mathcal{B}(\Theta(\mathcal{H})), \;\; \Sigma(\mathcal{H}) = \mathcal{B}(\mathcal{H}).$$

Furthermore, ex$\mathcal{H}$ *is a G_δ set in $\mathcal{H}$ and Θ is affine and bimeasurable w.r.t. $\Sigma(\Theta(\mathcal{H}))$ and $\Sigma(\mathcal{H})$.*

Proof. The first relations were proved in Theorem 3 of [16]. By the same reference, the extreme points form a G_δ-set in $\Theta(\mathcal{H})$ and by Theorem 5.1 the extremal set in $\mathcal{H}$ shares this property. Again by this result and the just mentioned equalities of σ-algebras Θ is bimeasurable.

5.2 Decomposition and Tchebyscheff Inequalities

Extremal decompositions of bell-shaped distributions will be derived now. If H is a convex set of nonnegative measures and N a Borel set then H_N denotes the subset of those $\mu \in H$ with $\mu(N) = 0$. Since H_N is a face in H one has $\mathrm{ex}H_N = H \cap \mathrm{ex}H$. By Proposition 4.16, the following result generalizes Theorem 6 of [10]. In particular, the proof there is completed (cf. Remark 4.17).

Theorem 5.4 *Suppose that $\mathcal{G}$ is a countable set of Borel functions and D a closed and convex subset of $\mathsf{R}^{\mathcal{G}}$. Choose $\mu \in \mathcal{H}(\mathcal{G}, D)$, let N denote the complement of* $\mathrm{essr}F^{(n)}$ *- with the c.d.f. F of μ - augmented by the turning points x_i and*

$$M = \{\nu \in \mathrm{ex}\mathcal{H}(\mathcal{G}, D) : \nu(N) = 0\}.$$

Then there is a probability measure P on $\Sigma(M)$ such that

$$\mu(B) = \int_M \nu(B)\, dP(\nu), \;\; B \in \mathcal{B}. \tag{14}$$

A straightforward monotone class argument shows that the *barycentrical formula* (14) is equivalent to

$$\mu(h) = \int_M \nu(h)\, dP(\nu), \;\; g \text{ measurable and bounded,}$$

where $\nu(h) = \int h \, d\nu$, and to

$$F(x) = \int_M G(x)\, dP(G), \;\; x \in \mathsf{R}, \tag{15}$$

with the respective c.d.fs. F and G of μ and ν. The distribution μ is called the *barycenter* of P and P is said to *represent* μ (on M).

Remark 5.5 *(a) In general, $\mathcal{H}$ is no (Choquet) simplex (cf. page 9), or, which is equivalent, the representing probability measure P of μ is not unique ([16], Theorem 2). In fact by Proposition 5.3, $\mathcal{H}$ shares this property if and only if $\Theta(\mathcal{H})$ does. A most simple example can be constructed for $n = 1$. Let $x_1 = 0$ and $\nu_1 = \varepsilon_{-1}/3 + 2 \cdot \varepsilon_2/3$, $\nu_2 = 2 \cdot \varepsilon_{-2}/3 + \varepsilon_1/3$. Then*

$$\mu = \frac{1}{2} \cdot (\nu_1 + \nu_2) = \frac{2}{6} \cdot \varepsilon_{-2} + \frac{1}{6} \cdot \varepsilon_{-1} + \frac{1}{6} \cdot \varepsilon_1 + \frac{2}{6} \cdot \varepsilon_2.$$

But $\mu = \rho_1/3 + 2 \cdot \rho_2/3$ for $\rho_1 = (\varepsilon_{-1} + \varepsilon_1)/2$ and $\rho_2 = (\varepsilon_{-2} + \varepsilon_2)/2$. Since $\mu \in \mathcal{H}$ and $\nu_i, \rho_i \in \mathrm{ex}\mathcal{H}$, we have a member of $\mathcal{H}$ with two different extremal decompositions.
(b) Whereas each P on $\Sigma(\mathcal{H})$ has a barycenter in the weak closure of $\mathcal{H}$ (since $\mathcal{M}_f$ is a complete and locally convex linear space in the weak topology, cf. [18], Corollary 1.2.3.), the set $\mathcal{H}$ is not measure convex *i.e. this barycenter is not necessarily in $\mathcal{H}$. Assuming $n = 1$ and $x_1 = 0$, for example, the measure $P = \sum_{k=1}^{\infty} 2^{-k} \varepsilon_k$ where ε_k is concentrated in a distribution with rectangular density equal to 2^{k+1} on $[-2^{-k-2}, 2^{-k-2})$ does not have a barycenter in $\mathcal{H}$.*
(c) The theorem shows that if there is any bell-shaped distribution fulfilling the moment conditions then there is also an extremal one doing so.

The following partial result is of independent interest.

Lemma 5.6 *Under the hypothesis of Theorem 5.4,*

$$\Theta(\mathcal{H}(\mathcal{G},D)) = \left\{\rho \in \mathcal{P}(S) : \tilde{\mathcal{G}} \subset L^1(\rho), \int \tilde{g}\, d\rho \in D^*, \tilde{g} \in \tilde{\mathcal{G}}\right\},$$

where $\tilde{\mathcal{G}} = \{tg^* : g^* \in \mathcal{G}^*\}$, and Θ is an affine isomorphism from $\mathcal{H}(\mathcal{G},D)$ onto this set bimeasurable w.r.t. the respective evaluation-σ-algebras.

Proof. Combine 2.13 and Proposition 5.3.

Now the last Theorem can be proved.

Proof (of Theorem 5.4)**.** Let H be the moment set in Lemma 5.6. By Lemma 5.6 and [16], Corollary 2, for each $\rho \in H$ there is a probability measure Q on $\Sigma(\mathrm{ex}H)$ such that

$$\rho(B) = \int_{\mathrm{ex}H} \kappa(B)\, dQ(\kappa), \;\; B \in \mathcal{B}(S).$$

Let now $\rho = \Theta(\mu)$ and P denote the image measure $Q \circ \Theta$ of Q under Θ^{-1} defined on $\Sigma(\mathrm{ex}\mathcal{H}(\mathcal{G},D))$ by $P(A) = Q(\Theta(A))$. For any bounded Borel function h the associated function $\tilde{h} = t \cdot h^*$ with h^* from Theorem 2.13 is defined everywhere and integrable for each $\kappa \in H$. Hence the barycentrical formula holds for $\tilde{h}$ by [19], Proposition 3.1, i.e.

$$\rho(\tilde{h}) = \int_{\mathrm{ex}H} \kappa(\tilde{h})\, dQ(\kappa).$$

Since

$$\Theta^{-1}(\mathrm{ex}H) = \{\nu \in \mathrm{ex}\mathcal{H}(\mathcal{G},D) : \nu(S) = 1\} =: M'$$

this implies

$$\mu(h) = \rho(\tilde{h}) = \int_{\mathrm{ex}H} \kappa(\tilde{h})\, dQ(\rho) = \int_{\mathrm{ex}H} \Theta^{-1}(\kappa)(h)\, dQ(\kappa) = \int_{M'} \nu(h)\, dP. \tag{16}$$

Since $\mu(N) = 0$ the identity (16) implies $P\{\nu : \nu(N) = 0\} = 1$ and P can be restricted to

$$\mathrm{ex}\mathcal{H}(\mathcal{G},D)_N = \mathcal{H}(\mathcal{G},D)_N \cap \mathrm{ex}\mathcal{H}(\mathcal{G},D) = M.$$

This completes the proof.

A functional C on $\mathcal{H}(\mathcal{G},D)$ is called *measure affine* if it is integrable for each probability measure on $\Sigma(M)$ with barycenter μ in $\mathcal{H}(\mathcal{G},D)$ and fulfills the barycentrical formula

$$C(\mu) = \int_M C(\nu)\, dP(\nu).$$

In view of the introductory remarks one has

Corollary 5.7 *Under the hypothesis of Theorem 5.4 each measure affine functional fulfills*

$$\begin{aligned} \inf\{C(\mu) : \mu \in \mathcal{H}(\mathcal{G},D)\} &= \inf\{C(\mu) : \mu \in M\}, \\ \sup\{C(\mu) : \mu \in \mathcal{H}(\mathcal{G},D)\} &= \sup\{C(\mu) : \mu \in M\}. \end{aligned} \tag{17}$$

This holds in particular if $C(\mu) = \int \phi\, d\mu$, where ϕ is a Borel function integrable for each $\mu \in \mathcal{H}(\mathcal{G},D)$.

Proof. The general part is clear and $\mu \mapsto \mu(\phi)$ is measure affine by standard monotone class arguments ([18], Proposition 1.1.2).

In particular, for $F \in \mathcal{H}(\mathcal{G}, D)$,

$$L(x) = \inf\{G(x) : G \in M\} \leq F(x) \leq \sup\{G(x) : G \in M\} = U(x),$$

and L and U are best possible lower and upper envelops for $\mathcal{H}(\mathcal{G}, D)$ (cf. [8]). Note that L and U increase and can be used to estimate fractiles.

Remark 5.8 *Theorem 5.4 and Corollary 5.7 hold for bell-shaped distributions with unspecified turning points as well. In this case the set M has to be replaced by the set*

$$M' = \bigcup\{\nu \in \mathcal{H}(x_1, \ldots, x_n; \mathcal{G}, D) : x_1 < \cdots x_n\}$$

(in self-explaining notation).

If $\mathcal{G}$ is finite and the moment conditions are given by equalities, i.e. D is singleton $\{d\}$ then the concrete description of M can be plugged into (17) to obtain more explicit bounds.

Example 5.9 *By Lemma 5.6, $\mathcal{H}(\mathcal{G}, D) \neq \emptyset$ if and only if*

$$\left\{\rho \in \mathcal{P}(S) : \tilde{\mathcal{G}} \subset L^1(\rho), \int \tilde{g}\, d\rho \in D^*, \tilde{g} \in \tilde{\mathcal{G}}\right\} \neq \emptyset.$$

To be more specific, consider the standard moment conditions

$$\int x^r\, d\mu(x) = d_r,\ \ 1 \leq r \leq p. \tag{18}$$

Letting $g_r(x) = x^r$, the identity (5) gives

$$\tilde{g}_{n+r}(x) = t(x)(x^r)^* = (-1)^n t(x) \frac{r!}{(r+n)!} x^{r+n} = \binom{r+n}{n} \frac{x^{r+n}}{\prod_{i=1}^n (x - x_i)}.$$

Hence there is some μ, bell-shaped to the n^{th} order with turning points $x_1, \ldots, x_n$ and satisfying (18), if and only if there is some (general) distribution ρ satisfying

$$\begin{aligned} \int \frac{x^i}{\prod_{i=1}^n (x - x_i)}\, d\rho(x) &= 0,\ \ 0 \leq i \leq n-1, \\ \int \frac{x^{r+n}}{\prod_{i=1}^n (x - x_i)}\, d\rho(x) &= \binom{r+n}{n} d_r,\ \ 1 \leq r \leq p. \end{aligned}$$

This generalizes Theorem 1 in [4], where this result is proved for $n = 1$ and $x_1 = 0$. In this case the above identities boil down to

$$\int x^{-1}\, d\rho(x) = 0,\ \ \int x^r\, d\rho(x) = (r+1) d_r,\ \ 1 \leq r \leq p.$$

By these observations, the program in [4] could be carried out for arbitrary n and arbitrary turning points, presumably at the expense of heavy calculations (In this paper, the condition $\int x^{-1}\, d\rho = 0$ does not appear, since there a c.d.f. is unimodal if it is convex on the left and concave on the right of a fixed point. Hence it needs not to be differentiable.).

By special methods, Mallows *in [8] obtains precise bounds (17) under moment conditions (18) in the cases $(n, r) = (0, 2q), (1, 2q)$ and $(n, 2)$.*

We thank H.v. Weizsäcker for helpful and encouraging discussions and R. Lasser who provided the economic basis to write this paper.

References

[1] Anastassiou G.A. (1993): *Moments in Probability and Approximation Theory.* Longman Scientific & Technical, Burnt Mill, Harlow Essex

[2] Choquet G. (1969). *Lectures in Analysis, Vol. II.* W.A. Benjamin, Inc. New York, Amsterdam.

[3] Dunford N. and Schwartz J.T. (1957): *Linear Operators Part I: General Theory.* Interscience Publishers, A Division of Wiley & and Sons: New York, London, Sidney

[4] Johnson N.L. and Rogers C.A. (1951). The Moment Problem for Unimodal Distributions. *Ann. Math. Statist.* **22**, 433-439

[5] Karlin S. and Studden W.J. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics.* Interscience Publishers, A Division of Wiley & and Sons: New York, London, Sidney

[6] Kemperman J.H.B. (1971): Moment Problems with Convexity Conditions I. In: *Optimizing Methods in Statistics* (J.S. Rustagi Ed.), Academic Press: New York, 115-178

[7] Kendall D.G. (1962): Simplexes and Vector Lattices. *J. London Math. Soc.* **37**, 365-371

[8] Mallows C.L. (1956): Generalizations of Tchebycheff's inequalities. *J. Roy. Statist. Soc. (B)* **18**, 139-176

[9] Mallows C.L. (1963): A Generalization of the Chebyshev Inequalities. *Proc. London Math. Soc. Ser. 13*, **8**, 385-412

[10] Mulholland H.P. and Rogers C.A (1958): Representation Theorems for Distribution Functions. *Proc. London Math. Soc. Ser. 3*, **8**, 177-223

[11] Rockafellar R.T. (1970): *Convex Analysis.* Princeton University Press. Princeton: New Jersey

[12] Rudin W. (1987): *Real and Complex Analysis.* 3^{rd} edition, McGraw-Hill, Inc.: New York etc.

[13] Schumaker L.L. (1981): *Spline Functions: Basic Theory.* John Wiley & Sons: New York etc.

[14] Titterington D.M, Smith A.F.M and Makov U.E. (1985): *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons

[15] Weizsäcker H.v. (1991): The Extreme Points of a Set of Probability Densities of Uniformly Bounded Variation (Preliminary Version). Preprint, University of Kaiserslautern, Germany

[16] Weizsäcker H.v. and Winkler G. (1979): Integral Representation in the Set of Solutions of a Generalized Moment Problem. *Math. Ann.* **246** 23-32

[17] Winkler G. (1982): Integral Representations and Upper Bounds for Stop-Loss Premiums under Constraints Given by Inequalities. *Scand. Actuarial J.* 15-21

[18] Winkler G. (1985): *Choquet Order and Simplices with Applications in Probabilistic Models.* Lecture Notes in Mathematics 1145, Springer Verlag: Berlin, Heidelberg

[19] Winkler G. (1988): Extreme Points of Moment Sets. *Mathematics of Operations Research* **13**, 581-587