Cite as: Enrique Miranda, Gert de Cooman, and Erik Quaeghebeur. "Finitely additive extensions of distribution functions and moment sequences: The coherent lower prevision approach" *International Journal of Approximate Reasoning*, Vol. 48, No. 1, p. 132–155, April 2008.
DOI: 10.1016/j.ijar.2007.07.007

FINITELY ADDITIVE EXTENSIONS OF DISTRIBUTION FUNCTIONS AND MOMENT SEQUENCES: THE COHERENT LOWER PREVISION APPROACH

ENRIQUE MIRANDA, GERT DE COOMAN, AND ERIK QUAEGHEBEUR

ABSTRACT. We study the information that a distribution function provides about the finitely additive probability measure inducing it. We show that in general there is an infinite number of finitely additive probabilities associated with the same distribution function. Secondly, we investigate the relationship between a distribution function and its given sequence of moments. We provide formulae for the sets of distribution functions, and finitely additive probabilities, associated with some moment sequence, and determine under which conditions the moments determine the distribution function uniquely. We show that all these problems can be addressed efficiently using the theory of coherent lower previsions.

1. INTRODUCTION

This paper consists of two parts, each devoted to one of two specific, but related, problems.

The first problem is: *To what extent does a distribution function determine a probability measure?* This question has a well-known answer when we are talking about probability measures that are σ-additive. We believe the corresponding problem for probability measures that are only finitely additive has received much less attention. This paper tries to remedy that situation somewhat by studying the particular case of finitely additive probability measures on the real unit interval $[0,1]$ (or equivalently, after an appropriate transformation, on any compact real interval). For this study, it will be very convenient to use the mathematical machinery behind Walley's [27] theory of coherent lower previsions, for which we introduce the basics in Section 2.

We shall see that, generally speaking, there is an infinite (closed and convex) set $\mathscr{M}(F)$ of finitely additive probability measures that correspond to a given distribution function F. However, by their very nature, and contrary to the sigma-additive case, finitely additive probabilities on an infinite set that extend 'something' are usually *inconstructibles*, meaning that they cannot actually be constructed, but that their existence may be inferred from the Hahn–Banach Theorem (or even stronger, the Axiom of Choice); see [25, Sections 12.31 and 6.6] for more details. It was one of Walley's achievements to show that we can efficiently and *constructively* deal with them not by looking at the members of $\mathscr{M}(F)$ individually, but by working with their lower envelope $\underline{E}_F$, which in his language is called the *natural extension* of the distribution function F.[1] Not only can this lower envelope always be constructed explicitly, but it is the closest we can get in a constructive manner to the finitely additive probabilities themselves. It turns out to be a coherent lower prevision with very special properties.

Key words and phrases. Coherent lower prevision, lower distribution function, lower Riemann–Stieltjes integral, complete monotonicity, moment sequence.

[1]This natural extension is quite closely related to the Minkowski functional that appears in the more usual formulations of the Hahn–Banach theorem. Not surprisingly, it also makes its appearance, although in a different guise, as the lower bound in de Finetti's Fundamental Theorem of Probability [13, Sections 3.10–12].

The set of finitely additive probabilities with a given distribution function has been considered before by Bruno de Finetti [13, Chapter 6], who seems to suggest[2] that the value $\underline{E}_F(f)$ of this natural extension in a function f is actually given by the lower Riemann–Stieltjes integral $(RS)\underline{\int}_0^1 f(x)\,\mathrm{d}F(x)$ of f with respect to the distribution function F. We study the relationship between $\underline{E}_F$ and lower Riemann–Stieltjes integrals in Section 3, and we shall see in Theorem 1 that de Finetti's suggestion needs some qualification, as it is essentially only correct when F is right-continuous.

In the second part, from Section 4 onwards, we address the second question: *Is a distribution function uniquely determined by the corresponding sequence of moments?* In a companion paper [21], we have studied the set of finitely additive probabilities $\mathscr{M}(m)$ that produce a given moment sequence m. The fundamental step we take there is analogous to the one followed in the present paper for distribution functions. It consists of not considering the finitely additive probabilities in $\mathscr{M}(m)$ themselves, but to study their *lower envelope* $\underline{E}_m$, which is a coherent lower prevision with very special properties too. In answering the second question, we build on those results, but by looking at distribution functions we are also able to extend them.

We shall establish that a distribution function F uniquely determines its moment sequence m. In Section 4 then, we investigate to what extent, conversely, a moment sequence determines the distribution function. For distribution functions coming from σ-additive probability measures, the relation between moment sequences and distribution functions is well-known to be one-to-one, but again, the answer is not so clear when we let go of the assumption of σ-additivity. We shall prove that in general there may be an infinite number of distribution functions with the same moment sequence, and investigate under which conditions the distribution function is unique.

It will perhaps not come as too much of a surprise, at this point, that we can show that the set of finitely additive probability measures $\mathscr{M}(m)$ that corresponds to a moment sequence is the union of the sets $\mathscr{M}(F)$ over all the distribution functions F that are compatible with the moment sequence m. This is also done in Section 4 (see Theorem 6). In Section 5, we further exploit the connection between distribution functions and moment sequences to come up with a number of quite interesting formulae expressing $\underline{E}_m$ (i) as a convex mixture of a lower Riemann–Stieltjes integral and so-called *lower oscillation* functionals associated with point probability masses (see Theorem 14); and (ii) as a σ-additive convex mixture of completely monotone lower previsions associated with neighbourhood filters, which express that all probability mass is concentrated in arbitrarily small neighbourhoods of elements of the unit interval (see Theorem 16). In passing, we give an alternative, constructive proof of the F. Riesz Representation Theorem (see Theorem 8 and Remark 4).

2. COHERENT AND COMPLETELY MONOTONE LOWER PREVISIONS

Let us give a short introduction to those concepts from the theory of coherent lower previsions that we shall use in this paper. We refer to Walley's book [27] for their behavioural interpretation, and for a much more complete introduction and treatment.

Consider a non-empty set Ω. Then a *gamble* on Ω is a bounded real-valued function on Ω. We denote the set of all gambles on Ω by $\mathscr{L}(\Omega)$. It is a linear space, and actually a Banach space when provided with the topology of uniform convergence of gambles.

[2]See [13, Section 6.4.11] where de Finetti states for the bounds on the prevision of a random quantity obtained for a given distribution function from his Fundamental Theorem of Probability, that "we are, of course, dealing with the upper and lower integrals in the Riemann sense". Before, in [13, Section 6.4.4] he refers to these same bounds as "in the Riemann–Stieltjes sense", so the omission of 'Stieltjes' in the first quote appears to be an oversight.

A *lower prevision* $\underline{P}$ is a real-valued map defined on some subset $\mathscr{K}$ of $\mathscr{L}(\Omega)$. If the domain $\mathscr{K}$ of $\underline{P}$ only contains indicators I_A of events A, then $\underline{P}$ is also called a *lower probability*. We also write $\underline{P}(I_A)$ as $\underline{P}(A)$, the lower probability of the event A. The *conjugate upper prevision* $\overline{P}$ of $\underline{P}$ is defined on $-\mathscr{K}$ by $\overline{P}(f) := -\underline{P}(-f)$ for every $-f$ in $\mathscr{K}$. If the domain of $\overline{P}$ contains indicators only, then $\overline{P}$ is also called an *upper probability*.

A lower prevision $\underline{P}$ defined on the set $\mathscr{L}(\Omega)$ of all gambles is called *coherent* if, with f,g in $\mathscr{L}(\Omega)$, it is super-additive: $\underline{P}(f+g) \geq \underline{P}(f)+\underline{P}(g)$, positively homogeneous: $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for all $\lambda \geq 0$, and positive: $\underline{P}(f) \geq \inf f$. A lower prevision $\underline{P}$ on an arbitrary domain $\mathscr{K}$ is then called *coherent* if it can be extended to some coherent lower prevision on all gambles. This is the case if and only if $\sup\left[\sum_{i=1}^{n} f_i - m f_0\right] \geq \sum_{i=1}^{n} \underline{P}(f_i) - m\underline{P}(f_0)$ for any $n,m \geq 0$ and f_0, f_1, ..., f_n in $\mathscr{K}$. For a coherent lower prevision $\underline{P}$, defined on a set $\mathscr{K}$, it holds that $\underline{P}(f) \leq \overline{P}(f)$ for all $f \in \mathscr{K} \cap -\mathscr{K}$. Also, a coherent lower prevision is monotone: $f \leq g \Rightarrow \underline{P}(f) \leq \underline{P}(g)$, and uniformly continuous: if a sequence of gambles $f_n, n \geq 0$ converges uniformly to another gamble f, then $\underline{P}(f_n) \to \underline{P}(f)$.

A *linear prevision* P on $\mathscr{L}(\Omega)$ is a self-conjugate coherent lower prevision: $P(-f) = -P(f)$. In other words, a linear prevision is a positive and normalised ($P(1) = 1$) linear functional (we also use 1 as the constant function with value 1). A functional defined on an arbitrary subset $\mathscr{K}$ of $\mathscr{L}(\Omega)$ is called a *linear prevision* if it can be extended to a linear prevision on $\mathscr{L}(\Omega)$. This is the case if and only if $\sup\left[\sum_{i=1}^{n} f_i - \sum_{j=1}^{m} g_j\right] \geq \sum_{i=1}^{n} P(f_i) - \sum_{j=1}^{m} P(g_j)$ for any $n,m \geq 0$ and f_1, ..., f_n, g_1, ..., g_m in $\mathscr{K}$. We let $\mathbb{P}(\Omega)$ denote the set of all linear previsions on $\mathscr{L}(\Omega)$.

The restriction Q of a linear prevision P on $\mathscr{L}(\Omega)$ to the set of all events is a finitely additive probability (probability charge). Linear previsions are completely determined by the values they assume on events; they are simply expectations with respect to finitely additive probabilities. This can be expressed using a Dunford integral (see, for instance, [3]): for any gamble h in $\mathscr{L}(\Omega)$ we have $P(h) = (D)\int h\,\mathrm{d}Q$.

The *natural extension* $\underline{E}_{\underline{P}}$ to $\mathscr{L}(\Omega)$ of a coherent lower prevision $\underline{P}$ defined on $\mathscr{K}$, is the point-wise smallest coherent lower prevision that extends $\underline{P}$ to all gambles. It is equal to the lower envelope of the set $\mathscr{M}(\underline{P})$ of all linear previsions that point-wise dominate $\underline{P}$ on its domain $\mathscr{K}$: for any gamble f in $\mathscr{L}(\Omega)$

$$\underline{E}_{\underline{P}}(f) = \min_{Q \in \mathscr{M}(\underline{P})} Q(f).$$

Observe that the set $\mathscr{M}(\underline{P})$ is convex, and closed (compact) in the relativisation to $\mathbb{P}(\Omega)$ of the weak* topology on the topological dual $\mathscr{L}(\Omega)^*$ of the Banach space $\mathscr{L}(\Omega)$. Moreover, $\mathscr{M}(\underline{E}_{\underline{P}}) = \mathscr{M}(\underline{P})$. Indeed, if $\underline{P}$ is a coherent lower prevision on $\mathscr{L}(\Omega)$ and $\overline{P}$ is its conjugate upper prevision, then for any gamble f and for any $a \in [\underline{P}(f), \overline{P}(f)]$ there exists a linear prevision $P \in \mathscr{M}(\underline{P})$ such that $P(f) = a$.

The procedure of natural extension is transitive: if we consider $\underline{E}_1$ the point-wise smallest coherent lower prevision on some domain $\mathscr{K}_1 \supseteq \mathscr{K}$ that dominates $\underline{P}$ on $\mathscr{K}$ (i.e., the natural extension of $\underline{P}$ to $\mathscr{K}_1$) and then the natural extension $\underline{E}_2$ of $\underline{E}_1$ to all gambles, then $\underline{E}_2$ is also the natural extension of $\underline{P}$ to $\mathscr{L}(\Omega)$. Moreover, $\mathscr{M}(\underline{E}_2) = \mathscr{M}(\underline{E}_1) = \mathscr{M}(\underline{P})$. In particular, if P is a linear prevision on a negation invariant $\mathscr{K}$ that has a unique extension P_1 to some larger negation invariant domain $\mathscr{K}_1$, then a linear prevision on all gambles will dominate (agree with) P on $\mathscr{K}$ if and only if it dominates (agrees with) P_1 on $\mathscr{K}_1$.

Let us introduce next the notion of n-monotonicity. A thorough study of the properties of n-monotone coherent lower previsions can be found in earlier papers [9; 10]. Here, we only mention those properties that we shall need further on.

A lower prevision defined on a *lattice* $\mathscr{K}$ of gambles (a set of gambles closed under point-wise minima $\wedge$ and maxima $\vee$) is called *n-monotone* if, for all $1 \leq p \leq n$, and all f, $f_1, \dots, f_p$ in $\mathscr{K}$ it holds that

$$\sum_{I \subseteq \{1,\dots,p\}} (-1)^{|I|} \underline{P}\left(f \wedge \bigwedge_{i \in I} f_i\right) \geq 0.$$

A lower prevision is *completely monotone* when it is n-monotone for any $n \geq 1$. This is for instance the case for linear previsions. Another example is given by the so-called *vacuous* previsions. The vacuous prevision $\underline{P}_A$ relative to an event A is given by $\underline{P}_A(f) = \inf_{x \in A} f(x)$ for any gamble f. A convex combination or a Moore–Smith limit of completely monotone and coherent lower previsions is again a completely monotone and coherent lower prevision.

We can easily characterise the natural extension of a completely monotone coherent lower prevision $\underline{P}$. If $\underline{P}$ is defined on a lattice of events $\mathscr{A}$ that includes $\emptyset$ and Ω, its natural extension to all events is again completely monotone, and coincides with its *inner set function* $\underline{P}_*$, where

$$\underline{P}_*(A) = \sup\left\{\underline{P}(B) \colon B \in \mathscr{A}, B \subseteq A\right\}.$$

Moreover, given a completely monotone coherent lower prevision $\underline{P}$ defined on a linear lattice of gambles $\mathscr{K}$ that contains all constant gambles, its natural extension $\underline{E}$ to all gambles coincides with its *inner extension* $\underline{P}_*$, where

$$\underline{P}_*(f) = \sup\left\{\underline{P}(g) \colon g \in \mathscr{K}, g \leq f\right\},$$

and $\underline{E}$ is again completely monotone.

A completely monotone coherent lower prevision $\underline{P}$ on all gambles satisfies a number of interesting properties. First, it is *comonotone additive*: we have $\underline{P}(f+g) = \underline{P}(f) + \underline{P}(g)$ for any two gambles f and g that are *comonotone*, meaning that for all ω and ϖ in Ω if $f(\omega) < f(\varpi)$ then also $g(\omega) \leq g(\varpi)$. Secondly, it is completely determined by the values it assumes on events. Actually, it is equal to the Choquet functional associated with the set function (a completely monotone coherent lower probability) that is the restriction of $\underline{P}$ to events: for all gambles f on Ω

$$\underline{P}(f) = (C)\int f \,\mathrm{d}\underline{P} = \inf f + (R)\int_{\inf f}^{\sup f} \underline{P}(\{h \geq t\})\,\mathrm{d}t,$$

where the first integral is a Choquet and the second a Riemann integral. Thirdly, the class of *$\underline{P}$-integrable gambles*, that is, those gambles h satisfying $\underline{P}(h) = \overline{P}(h)$, is a uniformly closed linear lattice that contains all constant gambles. In particular, the class of $\underline{P}$-integrable events is a field. Interestingly, a gamble h is $\underline{P}$-integrable if and only if its cut sets $\{f \geq t\} := \{x \in [0,1] \colon f(x) \geq t\}$ are $\underline{P}$-integrable for all but a countable number of t.

3. The natural extension of lower and upper distribution functions

Since we shall be dealing with the unit interval and its subintervals throughout, it will be well to establish a number of relevant conventions here. We consider the (Euclidean) topology $\mathscr{T}$ of open sets on $[0,1]$ that is the relativisation to $[0,1]$ of the Euclidean topology on the set of real numbers $\mathbb{R}$. By an *open interval* we shall mean a subinterval of $[0,1]$ that is open (belongs to $\mathscr{T}$), or in other words, that is the intersection of $[0,1]$ with some open interval of $\mathbb{R}$. Thus for x and y in $[0,1]$, (x,y) is an open interval, but so are $[0,1]$, $[0,x)$ and $(y,1]$. For any set $A \subseteq [0,1]$, we denote its topological interior by $\operatorname{int}(A)$ and its topological closure by $\operatorname{cl}(A)$.

We are now ready to tackle the first problem, mentioned in the Introduction: To what extent does a distribution function determine a finitely additive probability?

3.1. **A precise distribution function.** Since we are dealing in this paper with the unit interval, this shall be the domain we consider for the notion of distribution function:

Definition 1. A *distribution function* on $[0,1]$ is a non-decreasing function $F\colon [0,1] \to [0,1]$ that satisfies the normalisation condition $F(1) = 1$.

The interpretation of such a distribution function is as follows: we consider a random variable $X : [0,1] \to [0,1]$ and assume that F provides information about the accumulated probability of X. This means that we can define a functional P_F (the probability induced by the random variable X) such that for any $x \in [0,1]$, the (lower and upper) probability $P_F([0,x])$ of $[0,x]$ is equal to $F(x)$. Consequently, the probability $P_F((x,1])$ of $(x,1]$ is equal to $1-F(x)$. In other words, specifying a distribution function F is tantamount to specifying a set function P_F on the set of events

$$\mathscr{H} := \{[0,x]\colon x \in [0,1]\} \cup \{(x,1]\colon x \in [0,1]\}, \tag{1}$$

and since F satisfies the properties of a distribution function, this P_F can be seen as a linear prevision on (the set of indicator functions of) the elements of $\mathscr{H}$ [26, Lemma 3.58]. This linear prevision can be uniquely extended to a linear prevision on the lattice $\mathscr{Q}$ of subsets of $[0,1]$ generated by $\mathscr{H}$,[3] where all elements of $\mathscr{Q}$ have the form

$$[0,x_1] \cup (x_2,x_3] \cup \cdots \cup (x_{2n-2},x_{2n-1}] \cup (x_{2n},1], \text{ or}$$
$$(x_2,x_3] \cup \cdots \cup (x_{2n-2},x_{2n-1}] \cup (x_{2n},1],$$

where $0 \leq x_1 < x_2 \leq x_3 < \cdots \leq x_{2n-1} < x_{2n} \leq 1$[4]. If we also denote this unique linear prevision on $\mathscr{Q}$ by P_F, then we have that

$$P_F([0,x_1] \cup (x_2,x_3] \cup \cdots \cup (x_{2n-2},x_{2n-1}] \cup (x_{2n},1]) \\ = F(x_1) + \sum_{k=1}^{n-1} [F(x_{2k+1}) - F(x_{2k})] + 1 - F(x_{2n}), \tag{2}$$

and similarly

$$P_F((x_2,x_3] \cup \cdots \cup (x_{2n-2},x_{2n-1}] \cup (x_{2n},1]) = \sum_{k=1}^{n-1} [F(x_{2k+1}) - F(x_{2k})] + 1 - F(x_{2n}). \tag{3}$$

The natural extension $\underline{E}_F$ of P_F is the smallest coherent lower prevision on all gambles that extends P_F. It is the lower envelope of the set $\mathscr{M}(F) := \mathscr{M}(P_F)$ of all linear previsions Q with distribution function F, i.e., for which $Q([0,x]) = F(x)$, $x \in [0,1]$. For any gamble h on $[0,1]$, $[\underline{E}_F(h), \overline{E}_F(h)]$ is the range of the value $Q(h)$ for all such linear previsions Q.

Since the domain $\mathscr{Q}$ of P_F is a lattice of events containing both $\emptyset$ and $[0,1]$, and since any linear prevision on such a lattice of events is in particular completely monotone, we deduce from the discussion in Section 2 that (i) the natural extension $\underline{E}_F$ is a completely

---

[3] To see this, observe that (i) there is a unique extension as a linear (finitely additive) set function, and (ii) there always is an extension to a linear prevision on all gambles by [27, Theorem 3.4.2], and in particular to $\mathscr{Q}$, because P_F is a linear prevision.

[4] As remarked by one of the referees, there is also a unique extension as a linear prevision to the algebra generated by $\mathscr{H}$. See [14, Proposition 2.10] and [5, Theorem 11.2.2]. For the purposes of this paper, it suffices to use the expression of the unique extension to the lattice $\mathscr{Q}$ given in Equations (2) and (3).

monotone and comonotone additive coherent lower prevision; (ii) that the restriction of $\underline{E}_F$ to events is the inner set function $P_{F,*}$ of P_F, given by

$$P_{F,*}(A) = \sup\{P_F(B) \colon B \in \mathcal{Q}, B \subseteq A\} \tag{4}$$

for all $A \subseteq [0,1]$; and (iii) that for all gambles h on $[0,1]$,

$$\underline{E}_F(h) = (C)\int h \,\mathrm{d}\underline{E}_F = \inf h + (R)\int_{\inf h}^{\sup h} P_{F,*}(\{h \geq t\})\,\mathrm{d}t. \tag{5}$$

We can also draw a number of conclusions about the gambles h to which the linear prevision P_F can be extended uniquely:

Definition 2. A gamble h on $[0,1]$ is said to be F-*integrable* when $\underline{E}_F(h) = \overline{E}_F(h)$. The set of F-integrable gambles is denoted by $\mathscr{L}_F$.

Then we also know that (iv) $\mathscr{L}_F$ is a uniformly closed linear lattice containing all constant gambles, and that (v) a gamble h is F-integrable if and only if its cut sets $\{h \geq t\}$, or equivalently its strict cut sets $\{h > t\}$, are F-integrable for all but a countable number of t in $\mathbb{R}$.

Remark 1 (The non-uniqueness of finitely additive probability measures with a given distribution function). If we consider the set $\mathbb{Q} \cap [0,1]$ of all rational numbers between zero and one, then it is clear that $\{0\} = [0,0]$ is the only element of $\mathcal{Q}$ that is included in this set, and therefore $\underline{E}_F(\mathbb{Q} \cap [0,1]) = P_{F,*}(\mathbb{Q} \cap [0,1]) = F(0)$. On the other hand, $\emptyset$ is the only element of $\mathcal{Q}$ that is included in its complement $(\mathbb{Q} \cap [0,1])^c$, which is the set of all irrational numbers between zero and one, so we see that $\overline{E}_F(\mathbb{Q} \cap [0,1]) = 1 - \underline{E}_F((\mathbb{Q} \cap [0,1])^c) = 1$. This shows that the natural extension of any distribution function F is not a linear prevision (precise probability) unless all the probability mass is concentrated in 0. So, unless $F(0) = 1$, there is an uncountable infinity of linear previsions (finitely additive probabilities) Q with distribution function F, and for each $a \in [F(0), 1]$, there is some such Q for which $Q(\mathbb{Q} \cap [0,1]) = a$. To put it differently, a linear prevision on $\mathscr{L}([0,1])$ is not completely determined by its distribution function unless it corresponds to the degenerate distribution on 0. ♦

Example 1. Define the *simple break function* $\beta(\cdot; d, a) \colon [0,1] \to [0,1]$ by

$$\beta(x; d, a) := \begin{cases} 0 & \text{if } x < d \\ a & \text{if } x = d \\ 1 & \text{if } x > d \end{cases}$$

for d and a in $[0,1]$. If $d < 1$ then $\beta(\cdot; d, a)$ is a distribution function on $[0,1]$, which has one 'break' (discontinuity) at d unless $d = 0, a = 1$. For $d = 1$, $\beta(\cdot; d, a)$ is a distribution function if and only if $a = 1$. In the language of de Finetti [13, Section 6.5], the distribution function $\beta(\cdot; d, a)$ has adherent mass 1 at d: any open interval that contains d has probability 1, but we do not know exactly (due to the lack of σ-additivity) the probability of $\{d\}$: it may be 1, but then it may also be 0, and all the mass may then be left- or right-adherent to d. In general, the adherent mass at d will distribute between the left-adherent mass at d, the right-adherent one, and $P(d)$. See also Remark 6.

As we shall show in Section 5, in the case of the distribution function of the probabilities with a sequence of moments m, we know the masses adherent to any of the discontinuity points, but not exactly the mass allocated *at* the discontinuity point. De Finetti argues that we should regard the distribution functions as indeterminate in those discontinuity points.

Let us, as an example, determine the natural extension $\underline{E}_F$ when $F = \beta(\cdot;d,a)$ where $0 < d < 1$. Clearly $F = (1-a)\beta(\cdot;d,0) + a\beta(\cdot;d,1)$, and using Lemma 9 further on, we see that

$$\underline{E}_F = (1-a)\underline{E}_{\beta(\cdot;d,0)} + a\underline{E}_{\beta(\cdot;d,1)},$$

so it suffices to determine $\underline{E}_{\beta(\cdot;d,0)}$ and $\underline{E}_{\beta(\cdot;d,1)}$. Since $\beta(\cdot;d,0)$ and $\beta(\cdot;d,1)$ only assume the values 0 and 1, so do the restrictions of their natural extensions to events; see Equations (2)–(4). For any event A, we have that $\underline{E}_{\beta(\cdot;d,0)}(A) = 1$ if and only if $(d,x) \subseteq A$ for some $d < x \leq 1$, and it then follows from Equation (5) that

$$\underline{E}_{\beta(\cdot;d,0)}(h) = \sup_{d<x\leq 1} \inf_{z\in(d,x)} h(z).$$

Similarly, $\underline{E}_{\beta(\cdot;d,1)}(A) = 1$ if and only if $(x,d] \subseteq A$ for some $0 \leq x < d$, and therefore

$$\underline{E}_{\beta(\cdot;d,1)}(h) = \sup_{0\leq x<d} \inf_{z\in(x,d]} h(z) = \min\left\{h(d), \sup_{0\leq x<d} \inf_{z\in(x,d)} h(z)\right\}.$$

We shall come back to these break functions in Section 4. ♦

As we already stated in the Introduction, de Finetti [13, Section 6.4.4, p. 235] suggests that what we call the lower natural extension $\underline{E}_F$ of a distribution function F coincides with the lower Riemann–Stieltjes integral with respect to that distribution function. We devote some attention to (lower) Riemann–Stieltjes integrals in the next section.

3.2. **Lower and upper Riemann–Stieltjes integrals.** With a distribution function F, we can also associate integrals of the Riemann–Stieltjes type. Let us recall briefly how this is done. We refer to [19] for an excellent and more detailed exposition of this and other types of integrals. Consider a *subdivision* of $[0,1]$, i.e., a finite collection S of adjacent closed intervals $[0,x_1]$, $[x_1,x_2]$, …, $[x_{n-1},x_n]$, $[x_n,1]$ that cover $[0,1]$, where $0 = x_0 < x_1 < x_2 < \dots < x_n < x_{n+1} = 1$. Say that a subdivision S_2 *refines* a subdivision S_1, which we denote as $S_2 \succeq S_1$, if every closed interval in S_2 is a subset of some closed interval in S_1. Then the refinement relation $\succeq$ is reflexive and transitive, and the set $\mathscr{S}$ of all subdivisions is *directed* under the refinement relation, meaning that for any two subdivisions S_1 and S_2 there is a third subdivision S_3 that refines both: $S_3 \succeq S_1$ and $S_3 \succeq S_2$. This implies that we can consider Moore–Smith limits with respect to this directed set; see [22] for more information. Consider, for a gamble h on $[0,1]$, the net $\{I_S(h;F) : S \in \mathscr{S}\}$ which associates the real number

$$I_S(h;F) := \sum_{k=0}^{n} [F(x_{k+1}) - F(x_k)] \inf_{z\in[x_k,x_{k+1}]} h(z), \tag{6}$$

with any subdivision S of $[0,1]$. This net is bounded above by $\sup h$ and increasing: if $S_2 \succeq S_1$ then $I_{S_2}(h;F) \geq I_{S_1}(h;F)$. This implies that it Moore–Smith-converges to some real number (its Moore–Smith limit), and this real number is called the *lower Riemann–Stieltjes integral* of h with respect to F, denoted by:

$$(RS)\underline{\int_0^1} h(x)\,\mathrm{d}F(x) = \lim_{S\in\mathscr{S}} I_S(h;F) = \sup_{S\in\mathscr{S}} I_S(h;F). \tag{7}$$

The real functional that maps any gamble h in $\mathscr{L}([0,1])$ to its lower Riemann–Stieltjes integral $(RS)\underline{\int_0^1} h(x)\,\mathrm{d}F(x)$ can be interpreted as a lower prevision. It is not difficult to show, using Equations (6) and (7), that it is super-additive and positively homogeneous, and that moreover

$$[F(1)-F(0)]\inf h \leq (RS)\underline{\int}_0^1 h(x)\,\mathrm{d}F(x) \leq [F(1)-F(0)]\sup h,$$

so $(RS)\underline{\int}_0^1 \cdot\,\mathrm{d}F(x)$ will be a coherent lower prevision on $\mathscr{L}([0,1])$ if and only if $F(1)-F(0)=1$, or equivalently, $F(0)=0$. We see from Equations (6) and (7) that *in that case* $(RS)\underline{\int}_0^1 \cdot\,\mathrm{d}F(x)$ is a point-wise limit of convex mixtures of vacuous lower previsions. Such vacuous lower previsions are completely monotone, and so is, therefore, the lower Riemann–Stieltjes integral; see [9; 10; 11] for more details.

But when $F(0)=0$, its coherence and complete monotonicity allows us to say much more interesting things about the associated lower Riemann–Stieltjes integral. Indeed, as we have already had occasion to mention before, it ensures that this lower integral is the Choquet integral with respect to its restriction to events. Moreover, let $\mathscr{C}$ be the lattice of events generated by all closed intervals of $[0,1]$, i.e., the set consisting of $\emptyset$ and all finite unions of closed intervals of $[0,1]$, and for any $C=[x_1,x_2]\cup\cdots\cup[x_{2n-1},x_{2n}]$ in $\mathscr{C}$ with $x_1\leq x_2\leq\cdots\leq x_{2n}$ in $[0,1]$, let

$$Q_F(C) := (RS)\underline{\int}_0^1 I_C(x)\,\mathrm{d}F(x) = \sum_{k=1}^{n}[F(x_{2k})-F(x_{2k-1})], \tag{8}$$

and let $Q_F(\emptyset)=0$. Then $\mathscr{C}$ is a lattice of events containing both $\emptyset$ and $[0,1]$, and the set function Q_F is the restriction of the lower Riemann–Stieltjes integral to $\mathscr{C}$, and is therefore a coherent and completely monotone lower probability. Moreover, we infer from Equations (6) and (7) that for any event A

$$(RS)\underline{\int}_0^1 I_A(x)\,\mathrm{d}F(x) = \sup\{Q_F(C)\colon C\in\mathscr{C}, C\subseteq A\} = Q_{F,*}(A),$$

so the lower Riemann–Stieltjes integral coincides on events with the inner set function (the natural extension) $Q_{F,*}$ of Q_F. Finally, since the lower Riemann–Stieltjes integral is a coherent and completely monotone lower prevision, it coincides with the Choquet integral of its restriction to events, whence

$$(RS)\underline{\int}_0^1 h(x)\,\mathrm{d}F(x) = (C)\int h\,\mathrm{d}Q_{F,*} = \inf h + (R)\int_{\inf h}^{\sup h} Q_{F,*}(\{h\geq t\})\,\mathrm{d}t$$

for any gamble h on $[0,1]$.

The *upper Riemann–Stieltjes integral* $(RS)\overline{\int}_0^1 h(x)\,\mathrm{d}F(x)$ is defined similarly, with the infima in Equation (6) replaced by suprema. Alternatively, because

$$(RS)\overline{\int}_0^1 h(x)\,\mathrm{d}F(x) = -(RS)\underline{\int}_0^1 -h(x)\,\mathrm{d}F(x),$$

it can be seen as a conjugate upper prevision. If the lower and the upper Riemann–Stieltjes integrals coincide for some gamble h, we say that h is *Riemann–Stieltjes integrable* with respect to F, and we call the common value the *Riemann–Stieltjes integral* of h with respect to F. It follows from the complete monotonicity and coherence of F (when $F(0)=0$) that the set of all Riemann–Stieltjes integrable gambles constitutes a uniformly closed linear lattice, and that a gamble is Riemann–Stieltjes integrable if and only if (the indicators of) its cut sets $\{h\geq t\}$ are Riemann–Stieltjes integrable for all but a countable number of real numbers t.

We are now able to investigate de Finetti's suggestion that the natural extension $\underline{E}_F$ of a distribution function F can be written as the lower Riemann–Stieltjes integral with respect to F. The following theorem shows that this is not always the case! It was first proven by Troffaes in his doctoral dissertation [26, Theorem 4.59]. We give an alternative proof here that is much shorter than his, because we are able to harness the power of the mathematical machinery behind completely monotone coherent lower previsions.

Theorem 1. *For any distribution function F on $[0,1]$, $\underline{E}_F(h) = (RS)\underline{\int}_0^1 h(x)\,\mathrm{d}F(x)$ for all gambles h on $[0,1]$ if and only if F is right-continuous (i.e., it has no right adherent masses) and $F(0)=0$.*

Proof. We begin with the 'necessity' part. Take any $x \in [0,1]$. It follows from the definition of the lower Riemann–Stieltjes integral that $(RS)\underline{\int}_0^1 I_{(x,1]}(t)\,\mathrm{d}F(t) = 1-F(x+)$ whereas $\underline{E}_F((x,1]) = P_F((x,1]) = 1-F(x)$. This shows that F must be right-continuous on $[0,1]$. Similarly, consider $[0,x]$ for any x in $[0,1]$, then $(RS)\underline{\int}_0^1 I_{[0,x]}(t)\,\mathrm{d}F(t) = F(x)-F(0)$ and $\underline{E}_F([0,x]) = P_F([0,x]) = F(x)$, so we also must have that $F(0)=0$.

We now turn to the 'sufficiency' part. Assume that $F(0)=0$ and that F is right-continuous. First check, using Equations (2), (3) and (8) that in this case $P_F(B) = Q_{F,*}(B)$ for any B in $\mathscr{Q}$ and $Q_F(C) = P_{F,*}(C)$ for any C in $\mathscr{C}$. Then for any $A \subseteq [0,1]$,

$$\begin{aligned}
P_{F,*}(A) &= \sup\{P_F(B) \colon B \in \mathscr{Q}, B \subseteq A\} = \sup\{Q_{F,*}(B) \colon B \in \mathscr{Q}, B \subseteq A\} \\
&= \sup_{B \in \mathscr{Q}, B \subseteq A} \sup\{Q_F(C) \colon C \in \mathscr{C}, C \subseteq B\} \\
&\leq \sup\{Q_F(C) \colon C \in \mathscr{C}, C \subseteq A\} = Q_{F,*}(A),
\end{aligned}$$

and a completely symmetrical argument shows that $Q_{F,*}(A) \leq P_{F,*}(A)$. Hence the coherent lower probabilities $P_{F,*}$ and $Q_{F,*}$ coincide on all events, and so do therefore their natural extensions $\underline{E}_F$ and $(RS)\underline{\int}_0^1 \cdot\,\mathrm{d}F(x)$ on all gambles. $\square$

3.3. **Moments of a distribution function.** Interestingly, *any distribution function F produces precise moments*, i.e., the polynomials p^k defined by $p^k(x) := x^k$, $k>0$ and $p^0(x) := 1$ are always F-integrable. To see this, verify that for $k>0$, $\{p^k > t\}$ is equal to $(t^{\frac{1}{k}},1]$ if $t \geq 0$ and to $[0,1]$ if $t<0$, and that $\{p^0 > t\}$ equals $[0,1]$ if $t<1$ and $\emptyset$ if $t \geq 1$, so all the strict cut sets belong to $\mathscr{H}$, and are therefore F-integrable.

Using Equation (5), we find for the corresponding moments m_k that for $k>0$, after an appropriate change of variables in the Riemann integral, and integration by parts,

$$\begin{aligned}
m_k := \underline{E}_F(p^k) = \overline{E}_F(p^k) &= (R)\int_0^1 [1-F(t^{\frac{1}{k}})]\,\mathrm{d}t \\
&= 1-(R)\int_0^1 kx^{k-1}F(x)\,\mathrm{d}x = (RS)\int_0^1 x^k\,\mathrm{d}F(x),
\end{aligned} \tag{9}$$

since $P_{F,*}((t^{\frac{1}{k}},1]) = P_F((t^{\frac{1}{k}},1]) = 1-F(t^{\frac{1}{k}})$. For $k=0$ on the other hand, we have that $m_0 = 1$ and that

$$(RS)\int_0^1 x^0\,\mathrm{d}F(x) = (RS)\int_0^1 1\,\mathrm{d}F(x) = F(1)-F(0)$$

so we see that $m_0 = (RS)\int_0^1 x^0\,\mathrm{d}F(x)$ if and only if $F(0)=0$.

Let us therefore assume that $F(0)=0$. Then all polynomials p on $[0,1]$ are both F-integrable and Riemann–Stieltjes integrable with respect to F. Since we have seen that both $\underline{E}_F$ and $(RS)\underline{\int}_0^1 \cdot\,\mathrm{d}F(x)$ are coherent and completely monotone lower previsions, it follows that both the F-integrable and the Riemann–Stieltjes integrable gambles constitute a uniformly closed linear lattice. This implies that all continuous gambles are both F-integrable and Riemann–Stieltjes integrable with respect to F. We conclude that for all continuous gambles $h \in \mathscr{C}([0,1])$,

$$\underline{E}_F(h) = \overline{E}_F(h) = (RS)\int_0^1 h(x)\,\mathrm{d}F(x),$$

and we can use the Riemann–Stieltjes integral to calculate the natural extension of F to continuous gambles. We might be tempted to extrapolate this result and surmise that more generally, we can use the lower Riemann–Stieltjes integral to calculate $\underline{E}_F$ for all gambles. Theorem 1 tells us however that we cannot expect this to be the case unless F is right-continuous besides $F(0) = 0$. We shall have occasion to come back to the intriguing connection between (lower) Riemann–Stieltjes integrals and the natural extensions of distribution functions (and moment sequences) in the following sections.

3.4. **Lower and upper distribution functions.** Let us now turn to a more general problem. Suppose we have two maps $\underline{F}, \overline{F}\colon [0,1] \to [0,1]$, which we interpret as a *lower* and an *upper distribution function*, respectively. This means that $\underline{F}$ and $\overline{F}$ determine a lower probability $\underline{P}_{\underline{F},\overline{F}}$ on the set $\mathscr{H}$ given by Equation (1) as follows:

$$\underline{P}_{\underline{F},\overline{F}}([0,x]) = \underline{F}(x) \text{ and } \underline{P}_{\underline{F},\overline{F}}((x,1]) = 1 - \overline{F}(x)$$

for all $x \in [0,1]$. Walley has mentioned [27, Section 4.6.6] and Troffaes [26, Theorem 3.59, p. 93] has shown that $\underline{P}_{\underline{F},\overline{F}}$ is a coherent lower probability if and only if $\underline{F} \leq \overline{F}$ and both $\underline{F}$ and $\overline{F}$ are distribution functions, i.e., non-decreasing and normalised. We shall assume in what follows that these conditions are satisfied. Lower probabilities of this type are sometimes called *probability boxes*, see for instance [16].

The natural extension $\underline{E}_{\underline{F},\overline{F}}$ of the coherent lower probability $\underline{P}_{\underline{F},\overline{F}}$ to all gambles is the smallest coherent lower probability that coincides with $\underline{P}_{\underline{F},\overline{F}}$ on $\mathscr{H}$, or in other words, that has lower and upper distribution functions $\underline{F}$ and $\overline{F}$. It is the lower envelope of the set $\mathscr{M}(\underline{F},\overline{F}) := \mathscr{M}(\underline{P}_{\underline{F},\overline{F}}) = \mathscr{M}(\underline{E}_{\underline{F},\overline{F}})$ of all linear previsions whose distribution function F satisfies $\underline{F} \leq F \leq \overline{F}$. In fact, we have the following result.[5] Denote by

$$\Phi(\underline{F},\overline{F}) = \left\{F\colon F \text{ distribution function and } \underline{F} \leq F \leq \overline{F}\right\} \tag{10}$$

the set of all distribution functions (non-decreasing and normalised) on $[0,1]$ that lie between $\underline{F}$ and $\overline{F}$.

Theorem 2. *$\mathscr{M}(\underline{F},\overline{F}) = \bigcup_{F \in \Phi(\underline{F},\overline{F})} \mathscr{M}(F)$, and so $\underline{E}_{\underline{F},\overline{F}}$ is the lower envelope of all natural extensions $\underline{E}_F$ of the distribution functions F in $\Phi(\underline{F},\overline{F})$: for all gambles h on $[0,1]$,*

$$\underline{E}_{\underline{F},\overline{F}}(h) = \inf\left\{\underline{E}_F(h)\colon F \in \Phi(\underline{F},\overline{F})\right\}.$$

Proof. Recall that for any linear prevision Q on $\mathscr{L}([0,1])$, Q has distribution function F if and only if $Q \in \mathscr{M}(F)$. Now $Q \in \mathscr{M}(\underline{F},\overline{F})$ if and only if the distribution function of Q lies between $\underline{F}$ and $\overline{F}$, so if and only if there is some F in $\Phi(\underline{F},\overline{F})$ such that $Q \in \mathscr{M}(F)$. This means that indeed $\mathscr{M}(\underline{F},\overline{F}) = \bigcup_{F \in \Phi(\underline{F},\overline{F})} \mathscr{M}(F)$. Taking lower envelopes yields the desired expression involving the natural extensions, since $\underline{E}_{\underline{F},\overline{F}}$ is the lower envelope of $\mathscr{M}(\underline{F},\overline{F})$ and $\underline{E}_F$ is the lower envelope of $\mathscr{M}(F)$. □

[5] This result was mentioned by Walley [27, Section 4.6.6]; we give a (straightforward) proof here for the sake of completeness.

4. Distribution functions determined by a moment sequence

We have seen that a distribution function F always has precise moments, that is, if we know the values that the prevision P_F takes in $\mathscr{H}$, there is a unique extension to the class of polynomials, and as a consequence also to the class of continuous gambles, in which the set of polynomial gambles is uniformly dense. In this section, we are going to study the converse problem: to what extent do the values of the moments of a finitely additive probability determine its distribution function? This is related to the so-called *moment problem*, which we now turn our attention to.

4.1. **Basic results for the moment problem.** Let P be a linear prevision on the set

$$\mathscr{V}_p([0,1]) := \left\{p^k\colon k \geq 0\right\}.$$

The value $m_k := P(p^k)$ is called the (raw) moment of order k of the distribution P. Then, using linearity, we can determine the value of P in the set $\mathscr{V}([0,1])$ of all polynomial gambles on $[0,1]$, and since any continuous gamble is the uniform limit of a sequence of polynomials, these determine the values of P on all elements of the set $\mathscr{C}([0,1])$ of continuous gambles on $[0,1]$. Since trivially a linear prevision on $\mathscr{C}([0,1])$ determines the values of all the moments, we see that there is a one-to-one correspondence between linear previsions on $\mathscr{V}_p([0,1])$ and those on $\mathscr{C}([0,1])$.

In a companion paper [21], we have investigated to which extent a sequence of moments determines a finitely additive probability measure. Let us give a short survey of the results we found there, as they will be useful in addressing the problem at hand.

First, we recalled a number of necessary and sufficient conditions for a real sequence $m := (m_k)_{k\geq 0}$ in $[0,1]$ to be the sequence of moments of some finitely additive probability measure on the subsets of $[0,1]$. One such condition is the complete monotonicity of the sequence m:

Definition 3. A sequence m in $[0,1]$ is said to be *completely monotone* when $m_0 = 1$ and $(-1)^n\Delta^n m_k \geq 0$ for all $k,n \geq 0$, the $\Delta^n m_k$ are the n-th order differences defined recursively by $\Delta^n m_k := \Delta^{n-1}m_{k+1} - \Delta^{n-1}m_k$ and $\Delta^0 m_k := m_k$.

We shall also call a completely monotone sequence a *Hausdorff moment sequence*, referring to Hausdorff's [17; 18] original study of the moment problem for σ-additive probabilities. In these works, Hausdorff proved that the complete monotonicity of m is also necessary and sufficient for the existence of a σ-additive probability measure with this sequence of moments, which is moreover unique.

Secondly, we also studied to which extent a Hausdorff moment sequence m determines its inducing probability measure. Observe that such a sequence uniquely determines a linear prevision $\hat{P}_m$ on the set $\mathscr{C}([0,1])$. This implies that the linear previsions on all gambles with the given moment sequence m are precisely those linear previsions that extend $\hat{P}_m$. Let $\mathscr{M}(m)$ denote the set of all these linear previsions. The lower and upper envelopes of $\mathscr{M}(m)$ are given for any gamble h on $[0,1]$ by

$$\begin{aligned}\underline{E}_m(h) &= \sup\left\{\hat{P}_m(g)\colon g \in \mathscr{C}([0,1]), g \leq h\right\}\\ \overline{E}_m(h) &= \inf\left\{\hat{P}_m(g)\colon g \in \mathscr{C}([0,1]), h \leq g\right\}.\end{aligned} \tag{11}$$

Any linear prevision on $\mathscr{L}([0,1])$ induces the moment sequence m if and only if it dominates $\underline{E}_m$. Note that only one of the restrictions of these finitely additive probabilities in $\mathscr{M}(m)$ to the Borel sets is also σ-additive. We shall denote this probability by P_m^σ, and by F_m^σ its (right-continuous) distribution function.

Next, we list some properties of $\underline{E}_m$ and $\overline{E}_m$ (proven in [21]). For this, let us define, for any gamble h on $[0,1]$, the gambles

$$\begin{aligned} h^{\uparrow}(x) &= \sup\{g(x)\colon g \in \mathscr{C}([0,1]), g \leq h\}, \\ h^{\downarrow}(x) &= \inf\{g(x)\colon g \in \mathscr{C}([0,1]), h \leq g\}. \end{aligned} \tag{12}$$

Theorem 3. *[21] Consider a Hausdorff moment sequence m, and let $\underline{E}_m$ be the functional given by Equation* (11)*. The following statements hold.*

1. *For any gamble h on $[0,1]$, $\underline{E}_m(h) = \underline{E}_m(h^{\uparrow})$ and $\overline{E}_m(h) = \overline{E}_m(h^{\downarrow})$. In particular, for any event $A \subseteq [0,1]$, $\underline{E}_m(A) = \underline{E}_m(\mathrm{int}(A))$ and $\overline{E}_m(A) = \overline{E}_m(\mathrm{cl}(A))$.*
2. *For any set A,*
$$\underline{E}_m(A) = \underline{E}_m(\mathrm{int}(A)) = \sum_{I \in \mathscr{I}(A)} \underline{E}_m(I), \tag{13}$$
where $\mathscr{I}(A)$ is a countable family of disjoint open intervals whose union is $\mathrm{int}(A)$.
3. *$\underline{E}_m$ is a completely monotone and comonotone additive coherent lower prevision on $\mathscr{L}([0,1])$, and for all gambles h on $[0,1]$,*
$$\begin{aligned} \underline{E}_m(h) = (C)\int h \,\mathrm{d}\underline{E}_m &:= \inf h + (R)\int_{\inf h}^{\sup h} \underline{E}_m(\{h \geq t\})\,\mathrm{d}t \\ &= \inf h + (R)\int_{\inf h}^{\sup h} \underline{E}_m(\{h > t\})\,\mathrm{d}t, \end{aligned}$$
where the first integral is the Choquet integral associated with the restriction of $\underline{E}_m$ to events, and the second and third integrals are Riemann integrals.

Definition 4. Let m be a Hausdorff moment sequence, and let $\underline{E}_m$ and $\overline{E}_m$ be the lower and upper previsions defined in Equation (11). The associated *lower distribution function* $\underline{F}_m$ and *upper distribution function* $\overline{F}_m$ on $[0,1]$ are given by

$$\underline{F}_m(x) := \underline{E}_m([0,x]) \quad \text{and} \quad \overline{F}_m(x) := \overline{E}_m([0,x])$$

for all $x \in [0,1]$.

As we said before, a linear prevision has moment sequence m if and only if it belongs to $\mathscr{M}(m)$; in that case, its distribution function belongs to the set $\Phi(\underline{F}_m, \overline{F}_m)$ that we can define using Equation (10). We shall see in Theorem 6 later on that the converse also holds: a linear prevision whose distribution function belongs to $\Phi(\underline{F}_m, \overline{F}_m)$ will always produce the moment sequence m.

For any function f on $[0,1]$ and any $x \in [0,1]$ let $f(x-) := \lim_{t \to x, t<x} f(t)$ denote the *left limit* of f in x (if it exists) when $x > 0$, and let $f(0-) := f(0)$. Similarly, let $f(x+) := \lim_{t \to x, t>x} f(t)$ denote the *right limit* of f in x (if it exists) when $x < 1$, and let $f(1+) := f(1)$. Let then $\mathscr{D}_{\underline{F}_m} := \{x \in [0,1]\colon \underline{F}_m(x-) \neq \underline{F}_m(x+)\}$ denote the set of all points of discontinuity of $\underline{F}_m$, and $\mathscr{D}_{\overline{F}_m} = \left\{x \in [0,1]\colon \overline{F}_m(x-) \neq \overline{F}_m(x+)\right\}$ denote the set of points where $\overline{F}_m$ is not continuous. Let $\mathscr{D}_m := \mathscr{D}_{\underline{F}_m} \cup \mathscr{D}_{\overline{F}_m}$ denote their union. It follows from the non-decreasing character of $\underline{F}_m$ and $\overline{F}_m$ that $\mathscr{D}_{\underline{F}_m}$ and $\mathscr{D}_{\overline{F}_m}$ are countable subsets of $[0,1]$, and as a consequence so is their union $\mathscr{D}_m$.

Proposition 4. *[21] Let m be a Hausdorff moment sequence, and let $\underline{F}_m, \overline{F}_m$ be its associated lower and upper distribution functions. The following statements hold:*

(1) For any $x \in [0,1]$, $\underline{F}_m(x+) = \overline{F}_m(x) = \overline{F}_m(x+)$.
(2) For any $x \in (0,1)$, $\underline{F}_m(x-) = \underline{F}_m(x) = \overline{F}_m(x-)$.
(3) $\underline{F}_m(1-) = \overline{F}_m(1-) \leq \underline{F}_m(1) = \overline{F}_m(1) = 1$.
(4) $\underline{F}_m(0-) = \underline{F}_m(0) = 0 \leq \overline{F}_m(0-) = \overline{F}_m(0)$.

We are now ready to find out what are the distribution functions that correspond to a given Hausdorff moment sequence m. We shall see that the coherent lower prevision $\underline{E}_m$ helps us solve this problem.

4.2. **First results.**

Definition 5. A gamble h on $[0,1]$ is called *m-integrable* when $\underline{E}_m(h) = \overline{E}_m(h)$. We shall denote by E_m the restriction of $\underline{E}_m$ (or $\overline{E}_m$) to the class of m-integrable gambles.

The transitivity of the natural extension ensures that $\underline{E}_m$ is the natural extension of E_m. Since it follows from Theorem 3 that $\underline{E}_m$ is determined by its restriction to events and these are in turn determined by the values in open intervals, we only need to be interested in the values that E_m takes in the lattice $\mathscr{O}_m$ generated by the open m-integrable intervals. Denote by $\tilde{P}_m$ the restriction of E_m (and therefore also $\underline{E}_m$) on $\mathscr{O}_m$. It is easy to see that the elements of this lattice take the form

$$O = [0,x_1) \cup (x_2,x_3) \cup \cdots \cup (x_{2n-2},x_{2n-1}) \cup (x_{2n},1] \tag{14}$$

where $0 \leq x_1 \leq x_2 < x_3 \leq \cdots \leq x_{2n-2} < x_{2n-1} \leq x_{2n} \leq 1$, $x_k \notin \mathscr{D}_m$, and that

$$\tilde{P}_m(O) = \underline{F}_m(x_1) + \sum_{k=1}^{n-1} [\underline{F}_m(x_{2k+1}) - \overline{F}_m(x_{2k})] + 1 - \overline{F}_m(x_{2n}). \tag{15}$$

We now proceed to show that $\underline{E}_m$ is actually equal to the natural extension $\underline{E}_{\underline{F}_m,\overline{F}_m}$ of the lower and upper distribution functions $\underline{F}_m$ and $\overline{F}_m$, or in other words, that these two functions already capture, in a very specific way, *all* the information that is present in the moments. We first cite the following lemma, which follows immediately from Proposition 4.

Lemma 5. *Consider a Hausdorff moment sequence m, and let $F \in \Phi(\underline{F}_m,\overline{F}_m)$. Then*
1. *$\underline{F}_m(x) = F(x) = \overline{F}_m(x)$ for all $x \notin \mathscr{D}_m$;*
2. *$F(x-) = \underline{F}_m(x-) = F_m^\sigma(x-)$ for all $x \in (0,1]$;*
3. *$F(x+) = \overline{F}_m(x) = F_m^\sigma(x)$ for all $x \in [0,1]$.*

We see then that the distribution functions of the finitely additive probabilities with a given sequence of moments m may only differ in the countable set $\mathscr{D}_m$ of discontinuity points of $\underline{F}_m,\overline{F}_m$. On such points d, the difference between the distribution functions will come from the distribution of the mass jumps between the left-adherent and right-adherent parts, and $P(d)$. That $\Phi(\underline{F}_m,\overline{F}_m)$ has such structure[6] can be perhaps better understood if we think of the moments produced by a distribution function by means of a Riemann-Stieltjes integral, and the fact that this integral 'flattens out' adherent masses. We shall be more precise about this in Proposition 15 further on.

Remark 2 (The uniqueness of the σ-additive probability measure with a given moment sequence). This lemma allows for a very simple proof of the fact that there is only one σ-additive probability with a given moment sequence m that satisfies the Hausdorff moment condition, or in other words that there is only one σ-additive probability measure that extends a linear prevision on the set of all continuous gambles on $[0,1]$ (which is, essentially, the F. Riesz Representation Theorem in the form mentioned by Feller [15, Section V.1]). On the one hand, by the first statement in Proposition 4, the distribution function $\overline{F}_m$ is right-continuous, and the associated σ-additive probability measure has moment sequence m. On the other hand, let P^σ be any σ-additive probability on the Borel sets of $[0,1]$ with moment sequence m. Then its distribution function F^σ is right-continuous (by σ-additivity), and it

[6]We are grateful to one of the referees for drawing our attention to this.

must belong to $\Phi(\underline{F}_m,\overline{F}_m)$. By the third statement of Lemma 5, we see that $F^\sigma=\overline{F}_m$, so F is uniquely determined, and therefore so is P^σ.[7] ♦

Theorem 6. *Consider a Hausdorff moment sequence m, and let $\Phi(\underline{F}_m,\overline{F}_m)$ be given by Equation* (10)*. Then the following statements hold.*

1. *For all F in $\Phi(\underline{F}_m,\overline{F}_m)$, the restriction of $\underline{E}_F$ to the lattice of events $\mathscr{O}_m$ generated by the m-integrable open intervals is equal to $\tilde{P}_m$.*
2. *For all F in $\Phi(\underline{F}_m,\overline{F}_m)$, $\underline{E}_F$ dominates $\underline{E}_m$ and therefore all m-integrable gambles are also F-integrable: $\mathscr{L}_m\subseteq\mathscr{L}_F$.*
3. $\underline{E}_m=\inf\left\{\underline{E}_F\colon F\in\Phi(\underline{F}_m,\overline{F}_m)\right\}=\underline{E}_{\underline{F}_m,\overline{F}_m}$.

Proof. We begin with the first statement. Consider any distribution function F in the set $\Phi(\underline{F}_m,\overline{F}_m)$ and any finite union $O\in\mathscr{O}_m$ of m-integrable open intervals. Such a union has the form given by Equation (14). If we now apply Equations (2) and (4), we find in particular that for this union, if $x_1>0$,

$$\begin{aligned}\underline{E}_F(O)=P_{F,*}(O)&=F(x_1-)+\sum_{k=1}^{n-1}[F(x_{2k+1}-)-F(x_{2k})]+1-F(x_{2n})\\&=\underline{F}_m(x_1)+\sum_{k=1}^{n-1}[\underline{F}_m(x_{2k+1})-\overline{F}_m(x_{2k})]+1-\overline{F}_m(x_{2n}),\end{aligned}$$

where the last equality follows from Lemma 5 and Proposition 4. Similarly, if $x_1=0$, we get

$$\begin{aligned}\underline{E}_F(O)=P_{F,*}(O)&=\sum_{k=1}^{n-1}[F(x_{2k+1}-)-F(x_{2k})]+1-F(x_{2n})\\&=\sum_{k=1}^{n-1}[\underline{F}_m(x_{2k+1})-\overline{F}_m(x_{2k})]+1-\overline{F}_m(x_{2n}).\end{aligned}$$

If we compare these expressions with Equation (15), we see that $\underline{E}_F(O)=\tilde{P}_m(O)$, so $\tilde{P}_m$ (and therefore $\underline{E}_m$) and $\underline{E}_F$ coincide on $\mathscr{O}_m$. This proves the first statement.

Since $\underline{E}_m$ is the natural extension of $\tilde{P}_m$, and therefore the smallest coherent lower prevision that extends $\tilde{P}_m$, we see that $\underline{E}_F\geq\underline{E}_m$. So for any gamble h on $[0,1]$, $\underline{E}_m(h)\leq\underline{E}_F(h)\leq\overline{E}_F(h)\leq\overline{E}_m(h)$. If h is m-integrable gamble, then $\underline{E}_m(h)=\overline{E}_m(h)$, whence also $\underline{E}_F(h)=\overline{E}_F(h)$, so h is F-integrable as well. This completes the proof of the second statement.

To prove the third statement, use Theorem 2 to deduce from $\underline{E}_F\geq\underline{E}_m$ that $\underline{E}_{\underline{F}_m,\overline{F}_m}\geq\underline{E}_m$. For the converse inequality, recall that since the coherent lower prevision $\underline{E}_m$ has lower distribution function $\underline{F}_m$ and upper distribution function $\overline{F}_m$, it must dominate the smallest coherent lower prevision $\underline{E}_{\underline{F}_m,\overline{F}_m}$ with these lower and upper distribution functions, so $\underline{E}_m\geq\underline{E}_{\underline{F}_m,\overline{F}_m}$. □

We see from this Theorem that, given a distribution function F in $\Phi(\underline{F}_m,\overline{F}_m)$, and a linear prevision P with distribution function F, the linear prevision P belongs to the

[7] An astute reader might worry at this point about the appearance of F_m^σ in Lemma 5, which might lead him to suspect that our reasoning here is circular. But there is no real cause for concern: in no essential part of the development so far have we needed the existence nor uniqueness of a σ-additive probability measure that produces the moment sequence m. We could essentially have dropped every mention of P_m^σ and F_m^σ until now, and used Lemma 5 to prove their existence and uniqueness.

class $\mathscr{M}(\underline{E}_F) \subseteq \mathscr{M}(\underline{E}_m)$, and as a consequence the moment sequence of P is m. Hence, $\Phi(\underline{F}_m, \overline{F}_m)$ is exactly the class of distribution functions whose moment sequence is m.

Next, we establish a number of necessary and sufficient conditions for the equality $\underline{F}_m = \overline{F}_m$, or, equivalently, for the uniqueness of the distribution function with a given sequence of moments. As it could be expected, it amounts to the m-integrability of all the sets in $\mathscr{H}$, which in turn means that the only distribution function inducing these moments is (except for maybe at 1) continuous.

Corollary 7. *Consider a Hausdorff moment sequence m. Then the following statements are equivalent.*

1. $\underline{F}_m = \overline{F}_m$ *whence in particular* $\overline{F}_m(0) = 0$*;*
2. $\underline{F}_m$*,* $\overline{F}_m$ *and* F_m^σ *are continuous on* $[0,1)$*;*
3. $\underline{E}_m = \underline{E}_F$ *for some* $F \in \Phi(\underline{F}_m, \overline{F}_m)$*;*
4. $\underline{E}_m = \underline{E}_F$ *for all* $F \in \Phi(\underline{F}_m, \overline{F}_m)$*;*
5. $\underline{E}_m = (RS)\underline{\int}_0^1 \cdot \, \mathrm{d}F(x)$ *for all* $F \in \Phi(\underline{F}_m, \overline{F}_m)$*.*

Proof. It is clear from Proposition 4 that the first two statements are equivalent.

We now give a circular proof of the equivalence of statements 1, 3 and 4. The fourth statement implies the third. To show that the third statement implies the first, consider any $x \in [0,1]$ and the distribution function F for which $\underline{E}_m = \underline{E}_F$. Then it follows from the assumption that $\underline{F}_m(x) = \underline{E}_m([0,x]) = \underline{E}_F([0,x]) = F(x)$. But it follows by conjugacy that also $\overline{E}_m = \overline{E}_F$, so $\overline{F}_m(x) = \overline{E}_m([0,x]) = \overline{E}_F([0,x]) = F(x)$. This means that $\underline{F}_m = \overline{F}_m$. So we are left to show that the first statement implies the fourth. It follows from $\underline{F}_m = \overline{F}_m$ that $\Phi(\underline{F}_m, \overline{F}_m)$ only contains one distribution function $\underline{F}_m = \overline{F}_m$, and Theorem 6 then tells us that indeed $\underline{E}_m = \underline{E}_{\underline{F}_m} = \underline{E}_{\overline{F}_m}$.

To complete the proof, assume that any (and hence all) of the first four statements hold. Then in particular $\underline{E}_m = \underline{E}_{\overline{F}_m}$ (statement 4). Now, since $\overline{F}_m$ is right-continuous and satisfies $\overline{F}_m(0) = 0$ (statement 1), we know from Theorem 1 that also $\underline{E}_{\overline{F}_m} = (RS)\underline{\int}_0^1 \cdot \, \mathrm{d}\overline{F}_m(x)$. Since $\overline{F}_m$ is the only element of $\Phi(\underline{F}_m, \overline{F}_m)$ (statement 1), we see that the fifth statement holds. Conversely, it follows from the assumption that $\underline{E}_m = (RS)\underline{\int}_0^1 \cdot \, \mathrm{d}\overline{F}_m(x)$. In particular, $1 = \underline{E}_m(1) = (RS)\underline{\int}_0^1 1 \, \mathrm{d}\overline{F}_m(x) = \overline{F}_m(1) - \overline{F}_m(0)$, so $\overline{F}_m(0) = 0 = \underline{F}_m(0)$. Moreover, for all $0 < x \leq 1$, $\underline{F}_m(x) = (RS)\underline{\int}_0^1 I_{[0,x]}(t) \, \mathrm{d}\overline{F}_m(t) = \overline{F}_m(x) - \overline{F}_m(0) = \overline{F}_m(x)$. $\square$

Remark 3 (Distribution functions are more informative than moment sequences). Let us then argue that, for finitely additive probabilities, specifying a distribution function F is generally speaking more informative than specifying a moment sequence (contrary to what we are used to for σ-additive probabilities). Indeed, let F be a distribution function on $[0,1]$. We have seen in Section 3.3 that F produces a precise moment sequence m with $m_0 = 1$ and $m_k = (RS)\int_0^1 x^k \, \mathrm{d}F(x)$, $k > 0$; and it is clear that this moment sequence satisfies the Hausdorff moment condition. By Theorem 6, $\underline{E}_F \geq \underline{E}_m$ and $\mathscr{L}_m \subseteq \mathscr{L}_F$, so $\underline{E}_F$ is indeed generally more informative than $\underline{E}_m$. And Corollary 7 makes us conclude that $\underline{E}_m = \underline{E}_F$ only if $\underline{E}_F$ is equal to the lower Riemann–Stieltjes integral associated with F, which (due to Theorem 1) holds if and only if F is continuous on $[0,1)$ and $F(0) = 0$. ♦

In Section 3.1, we have studied the relationship between the natural extension of a distribution function F and the lower and upper Riemann-Stieltjes integrals. We now consider the more general situation where our information is given by a moment sequence m. We have already argued that in that case the linear previsions with that moment sequence are those that lie between $\underline{E}_m$ and $\overline{E}_m$, and that the corresponding distribution functions are those that lie between $\underline{F}_m$ and $\overline{F}_m$. We may be tempted to think that $\underline{E}_m$ and $\overline{E}_m$ coincide with the lower and upper Riemann–Stieltjes integrals with respect to $\underline{F}_m$ and $\overline{F}_m$, respectively. However, this is generally not the case. The relationship between them is given by the following theorem.

Theorem 8. *Consider a Hausdorff moment sequence m. For any $F \in \Phi(\underline{F}_m, \overline{F}_m)$ such that $F(0) = 0$ and any gamble h on $[0,1]$,*

$$\underline{E}_m(h) \leq (RS)\underline{\int_0^1} h(x)\,\mathrm{d}F(x) \leq (RS)\overline{\int_0^1} h(x)\,\mathrm{d}F(x) \leq \overline{E}_m(h).$$

Moreover, we have that

$$\underline{E}_m(h) = \inf_{F \in \Phi(\underline{F}_m, \overline{F}_m)} (RS)\underline{\int_0^1} h(x)\,\mathrm{d}F(x)$$

for all gambles h on $[0,1]$ if and only if $\overline{F}_m(0) = 0$, or equivalently, $0 \notin \mathscr{D}_m$.

Proof. We begin with the first part. Consider any finite union $O \in \mathscr{O}_m$ of m-integrable open intervals, which always has the form given by Equation (14). Consider any $F \in \Phi(\underline{F}_m, \overline{F}_m)$, then it follows from the definition of the lower Riemann–Stieltjes integral that if $x_1 > 0$, $(RS)\underline{\int_0^1} I_O(x)\,\mathrm{d}F(x)$ is equal to

$$\begin{aligned}
&F(x_1-) - F(0) + \sum_{k=1}^{n-1}[F(x_{2k+1}-) - F(x_{2k}+)] + 1 - F(x_{2n}+)\\
&= \underline{F}_m(x_1) - F(0) + \sum_{k=1}^{n-1}[\underline{F}_m(x_{2k+1}) - \overline{F}_m(x_{2k})] + 1 - \overline{F}_m(x_{2n})\\
&= \tilde{P}_m(O) - F(0),
\end{aligned}$$

where the first equality follows from Lemma 5, and the fact that $x_k \notin \mathscr{D}_m$ for all k, and the second one from Equation (15). A similar reasoning allows us to deduce that $\tilde{P}_m(O) = (RS)\underline{\int_0^1} I_O(x)\,\mathrm{d}F(x)$ if $x_1 = 0$. So if $F(0) = 0$, we see from Section 3.2 that $(RS)\underline{\int_0^1} \cdot\,\mathrm{d}F(x)$ is a coherent lower prevision, and the above developments imply that it coincides with the coherent lower probability $\tilde{P}_m$ on $\mathscr{O}_m$, and therefore dominates the smallest coherent lower prevision $\underline{E}_m$ that coincides with $\tilde{P}_m$ on $\mathscr{O}_m$: $(RS)\underline{\int_0^1} \cdot\,\mathrm{d}F(x) \geq \underline{E}_m$. The other inequalities follow immediately from conjugacy.

We now turn to the equality involving lower Riemann–Stieltjes integrals. Consider $0 < a < 1$. Then, for any $F \in \Phi(\underline{F}_m, \overline{F}_m)$, $(RS)\underline{\int_0^1} I_{[0,a)}(x)\,\mathrm{d}F(x) = F(a-) - F(0) = \underline{F}_m(a) - F(0)$ by Lemma 5 and the second statement of Proposition 4, so

$$\inf_{F \in \Phi(\underline{F}_m, \overline{F}_m)} (RS)\underline{\int_0^1} I_{[0,a)}(x)\,\mathrm{d}F(x) = \underline{F}_m(a) - \overline{F}_m(0)$$

and this is equal to $\underline{E}_m([0,a)) = \underline{F}_m(a)$ [use Proposition 4 and the monotonicity of $\underline{E}_m$] only if $\overline{F}_m(0) = 0$. Hence, the condition is necessary. Let us prove now that it is also sufficient. Assume therefore that $\overline{F}_m(0) = 0$, or equivalently, $0 \notin \mathscr{D}_m$. Consider any F in $\Phi(\underline{F}_m, \overline{F}_m)$. Then by assumption $F(0) = 0$, so we know from Section 3.2 that the lower Riemann–Stieltjes integral with respect to F is a completely monotone lower prevision on all gambles, which is therefore the natural extension of its restriction $Q_{F,*}$ to events. $Q_{F,*}$ is the natural (inner) extension to events of Q_F, which is defined on the lattice $\mathscr{C}$ of events generated by all closed intervals of $[0,1]$ by Equation (8). Now observe that for any event $C=[x_1,x_2]\cup\cdots\cup[x_{2n-1},x_{2n}]$ in $\mathscr{C}$ we have, taking into account Equations (2) and (3), $F(0)=0$ and $F(1)=1$, that

$$Q_F(C)=\sum_{k=1}^{n}[F(x_{2k})-F(x_{2k-1})]=P_F((x_1,x_2]\cup\cdots\cup(x_{2n-1},x_{2n}])\leq P_{F,*}(C).$$

Consequently, we find for any event $A\subseteq[0,1]$ that

$$\begin{aligned}Q_{F,*}(A)&=\sup\{Q_F(C)\colon C\in\mathscr{C},C\subseteq A\}\leq\sup\{P_{F,*}(C)\colon C\in\mathscr{C},C\subseteq A\}\\&=\sup_{C\in\mathscr{C},C\subseteq A}\sup\{P_F(B)\colon B\in\mathscr{Q},B\subseteq C\}\leq\sup\{P_F(B)\colon B\in\mathscr{Q},B\subseteq A\}\\&=P_{F,*}(A),\end{aligned}$$

and therefore also $(RS)\underline{\int}_0^1 h(x)\,\mathrm{d}F(x)\leq\underline{E}_F(h)$ for all gambles h on $[0,1]$. From Theorem 6, we then deduce that

$$\underline{E}_m(h)=\inf_{F\in\Phi(\underline{F}_m,\overline{F}_m)}\underline{E}_F(h)\geq\inf_{F\in\Phi(\underline{F}_m,\overline{F}_m)}(RS)\underline{\int}_0^1 h(x)\,\mathrm{d}F(x)$$

for any gamble h. The converse inequality follows from the first part. □

Remark 4 (On the F. Riesz Representation Theorem). It follows from Theorem 8 that if a gamble h is m-integrable, then it is Riemann–Stieltjes integrable with respect to any $F\in\Phi(\underline{F}_m,\overline{F}_m)$ such that $F(0)=0$, and moreover

$$\underline{E}_m(h)=\overline{E}_m(h)=(RS)\int_0^1 h(x)\,\mathrm{d}F(x)=\underline{E}_F(h)=\overline{E}_F(h). \tag{16}$$

This holds in particular for all continuous gambles on $[0,1]$, which strengthens the conclusions in Section 3.3. We shall be able to further strengthen this statement in Corollary 17 below.

But Equation (16) for continuous gambles is actually a statement of the original form of the F. Riesz Representation Theorem ([23], see also [24, Section 50]). Indeed, we already know that specifying a Hausdorff moment sequence m is equivalent to considering a positive (normalised) linear functional $\hat{P}_m$ on the set $\mathscr{C}([0,1])$ of all continuous gambles on $[0,1]$. And for such a functional, we now see that $\hat{P}_m=(RS)\int_0^1\cdot\,\mathrm{d}F(x)$ for all distribution functions F in $\Phi(\underline{F}_m,\overline{F}_m)$ such that $F(0)=0$. Since it is clear that there are such distribution functions (for instance $\underline{F}_m$), we have indeed proven that any positive linear functional on $\mathscr{C}([0,1])$ can be written as the Riemann–Stieltjes integral with respect to some non-decreasing function.[8] Conversely, it is trivial that the Riemann–Stieltjes integral associated with a distribution function F such that $F(0)=0$ is the restriction of the lower Riemann–Stieltjes integral (a coherent lower prevision) to the uniformly closed linear lattice of all Riemann–Stieltjes integrable gambles. Hence, the Riemann–Stieltjes integral is a linear prevision on this lattice, and therefore a positive linear functional.

Observe that this proof is, as far as we can see, constructive, because it is based on the constructible natural extension $\underline{E}_m$ (and on the constructive version of the Stone–Weierstraß theorem using approximations of continuous gambles by Bernstein polynomials). Contrary to Banach's fairly well-known unconstructive proof [2] it does not rely on the Hahn–Banach Theorem. Observe, by the way, that there is a small and easily correctable mistake in Banach's proof which involves, interestingly and tellingly, the assumption $F(0)=0$.[9] It is also of historical interest to note that in F. Riesz's approach, as reported in [24, Section 50],

[8] Normalisation is not an issue here.

[9] We owe this remark to Eric Schechter.

as well as in Daniell's [6] more general treatment of the extension problem, the proof proceeds by analogy of Dedekind's construction of the reals (see for instance [7, Chapter 2] for an interesting discussion of so-called Dedekind–MacNeille completion of partially ordered sets to complete lattices) showing that the linear functional on the continuous gambles can be extended uniquely to a linear functional satisfying monotone convergence[10] on the convex cone of lower semi-continuous gambles, and then proceeding from there using the, by now familiar, inner and outer extensions. Since we cannot generally assume monotone convergence for our *finitely* additive extensions, we do not have uniqueness of the extending positive linear functionals on the cone of lower semi-continuous functions, and we have to use inner extension earlier in the process to get to $\underline{E}_m$. But interestingly, we do still have that the natural extension $\underline{E}_m$ is completely determined by the values it assumes on lower semi-continuous gambles; see Theorem 3. In fact, we shall see further on in Theorem 16 that $\underline{E}_m$ coincides on the convex cone of lower semi-continuous gambles with the unique extension satisfying monotone convergence constructed in the manner of F. Riesz and Daniell described above. Dual (conjugate) statements can be made for $\overline{E}_m$. ♦

5. INTERESTING EXPRESSIONS FOR $\underline{E}_m$

We are now going to combine all the previous results in order to derive a very elegant expression for $\underline{E}_m$. In order to get there, we only need to take a closer look at distribution functions and their discontinuity points.

Consider any distribution function F in $\Phi(\underline{F}_m,\overline{F}_m)$. Then since F is non-decreasing its set of discontinuities is a countable subset of $[0,1]$. From Lemma 5, it is moreover a subset of $\mathscr{D}_m$. Let us introduce a new distribution function F^- by letting $F^-(x) := F(x-)$ if $x \in (0,1]$ and $F^-(0) := 0$.[11] Then we may infer from Lemma 5 that the sum of the probability masses concentrated in the discontinuity points

$$\sum_{d\in\mathscr{D}_m}[F(d+)-F^-(d)] = \sum_{d\in\mathscr{D}_m}[\overline{F}_m(d)-\underline{F}_m(d-)] =: \mu_m \tag{17}$$

is the same for every F in $\Phi(\underline{F}_m,\overline{F}_m)$, and completely determined by $\underline{F}_m$ and $\overline{F}_m$ (and therefore by the moment sequence m). Since μ_m is the sum of the jumps of F at its discontinuity points, we must have that $0 \leq \mu_m \leq 1$.

Then we can write F as a convex mixture

$$F = \mu_m F_b + (1-\mu_m)F_c \tag{18}$$

of a continuous distribution function F_c and a 'pure break function' F_b, which is a uniformly and absolutely convergent sum (convex mixture) of simple break functions.[12] Explicitly, we

[10] We say that a linear functional L satisfies monotone convergence on some set of gambles $\mathscr{K}$ if for any monotone sequence of gambles f_n in $\mathscr{K}$ that converges *point-wise* to some gamble f in $\mathscr{K}$, it holds that $L(f_n) \to L(f)$.

[11] We introduce this new notation because if d is a discontinuity point of F then the mass jump in d is $F(d+)-F(d-)$ if $d > 0$. But if $d = 0$, then this mass jump is $F(0+)$, and we introduced the convention before that $F(0-) := F(0)$, which may be different from zero whereas $F^-(0)$ is defined to be zero. So the new notation allows us to write $F(d+)-F^-(d)$ for the mass jump in every d.

[12] This idea is explained more extensively in [13, Section 6.2] and [19, Section II.13].

have for all $x \in [0,1]$

$$\begin{aligned}\mu_m F_b(x) &:= \sum_{d\in\mathscr{D}_m}[F(d+)-F^-(d)]\beta\left(x;d,\frac{F(d)-F^-(d)}{F(d+)-F^-(d)}\right)\\ &= \sum_{d\in\mathscr{D}_m}[\overline{F}_m(d)-\underline{F}_m(d-)]\beta(x;d,s_m(F,d))\\ &= F(x)-\underline{F}_m(x-)+\sum_{d\in\mathscr{D}_m,d<x}[\overline{F}_m(d)-\underline{F}_m(d-)] \qquad (19)\end{aligned}$$

using Lemma 5, where we let

$$s_m(F,d) := \frac{F(d)-\underline{F}_m(d-)}{\overline{F}_m(d)-\underline{F}_m(d-)}, \qquad (20)$$

and β is the simple break function defined in Example 1. On the other hand,

$$(1-\mu_m)F_c(x) := F(x)-\mu_m F_b(x) = \underline{F}_m(x-)-\sum_{d\in\mathscr{D}_m,d<x}[\overline{F}_m(d)-\underline{F}_m(d-)].$$

We see that the continuous part F_c is the same for all distributions F in $\Phi(\underline{F}_m,\overline{F}_m)$, and completely determined by the lower and upper distribution functions $\underline{F}_m$ and $\overline{F}_m$. We shall denote it by F_m. Observe that $F_m(0)=0$. The pure break parts are identical in all the continuity points of $\underline{F}_m$ and $\overline{F}_m$, and differ only by the values $F(d)$ they assume in the countably many discontinuity points $d\in\mathscr{D}_m$. Indeed, we can get to all F in $\Phi(\underline{F}_m,\overline{F}_m)$ by for each such break point d assigning to $F(d)$ any value in $[\underline{F}_m(d),\overline{F}_m(d)]$ (independently of all the other break points) if $d<1$, or equivalently, assigning to $s_m(F,d)$ any value in $[0,1]$; if $d=1$ is a break point, then there is of course only one possible choice $F(1)=1$, or $s_m(F,1)=1$.

We then find in particular for $F=\underline{F}_m$ and $F=\overline{F}_m$ that, with obvious notations,

$$\underline{F}_m = \mu_m\underline{F}_{m,b}+(1-\mu_m)F_m \quad\text{and}\quad \overline{F}_m=\mu_m\overline{F}_{m,b}+(1-\mu_m)F_m \qquad (21)$$

where for all $x\in[0,1)$

$$\mu_m\overline{F}_{m,b}(x) = \sum_{d\in\mathscr{D}_m,d\leq x}[\overline{F}_m(d)-\underline{F}_m(d-)] \qquad (22)$$

and

$$\mu_m\underline{F}_{m,b}(x) = \sum_{d\in\mathscr{D}_m,d<x}[\overline{F}_m(d)-\underline{F}_m(d-)], \qquad (23)$$

and where we have also used the second statement in Proposition 4. In particular, for any break point $d<1$, we have that $s_m(\underline{F}_m,d)=0$ and $s_m(\overline{F}_m,d)=1$. It is not hard to see (check also Proposition 13 further on) that $\underline{F}_{m,b}$ and $\overline{F}_{m,b}$ are exactly the lower and upper distribution functions produced by the moment sequence m_b, where

$$(m_b)_k = \sum_{d\in\mathscr{D}_m}\frac{\overline{F}_m(d)-\underline{F}_m(d-)}{\mu_m}d^k = \sum_{d\in\mathscr{D}_m}[\overline{F}_{m,b}(d)-\underline{F}_{m,b}(d-)]d^k$$

which corresponds to a discrete σ-additive probability measure with probability mass $\overline{F}_{m,b}(d)-\underline{F}_{m,b}(d-)$ concentrated in the elements d of $\mathscr{D}_m$.

Indeed, we shall see that we can decompose the moment sequence m into a ‘continuous’ part m_c and a ‘discrete’ part m_b. This is due to the convexity property of the natural extension established in the following two lemmas.

Lemma 9. *Let F_1 and F_2 be two distribution functions on $[0,1]$, and let, for $\alpha \in [0,1]$, the distribution function $F = \alpha F_1 + (1-\alpha)F_2$ be a convex mixture of F_1 and F_2. Then $(RS)\underline{\int}_0^1 \cdot \, \mathrm{d}F(x) = \alpha (RS)\underline{\int}_0^1 \cdot \, \mathrm{d}F_1(x) + (1-\alpha)(RS)\underline{\int}_0^1 \cdot \, \mathrm{d}F_2(x)$ and $\underline{E}_F = \alpha \underline{E}_{F_1} + (1-\alpha)\underline{E}_{F_2}$.*

Proof. For the lower Riemann–Stieltjes integral, observe that the subdivisions of $[0,1]$ form a directed set under the refinement relation, and that consequently, such a lower integral is a Moore–Smith limit. Since the limit of a convex mixture is the convex mixture of the limits, the result follows.

For the natural extensions, the reasoning is similar, if somewhat more involved. First of all, consider the natural extension to all events. Since the set $\mathscr{Q}$ is a lattice of events, it is closed under unions, and therefore constitutes a directed set under the inclusion relation. This ensures that the supremum in Equation (4) is actually a Moore–Smith limit. Since the limit of a convex mixture is the convex mixture of the limits, the result follows for the natural extension to events. For the natural extension to gambles, the proof is now an immediate consequence of Equation (5) and the linearity of the Riemann integral. □

Lemma 10. *Let m' and m'' be two Hausdorff moment sequences, and consider, for any $\alpha \in [0,1]$, the moment sequence $m := \alpha m' + (1-\alpha)m''$ (a convex mixture). Then m satisfies the Hausdorff moment condition as well, and $\underline{E}_m = \alpha \underline{E}_{m'} + (1-\alpha)\underline{E}_{m''}$.*

Proof. First of all, $m_0 = \alpha 1 + (1-\alpha)1 = 1$ and moreover $(-1)^n \Delta^n m_k = \alpha(-1)^n \Delta^n m'_k + (1-\alpha)(-1)^n \Delta^n m''_k \geq 0$ for any $k,n \geq 0$, so m satisfies the Hausdorff moment condition. Observe (i) that $\hat{P}_m = \alpha \hat{P}_{m'} + (1-\alpha)\hat{P}_{m''}$, (ii) that Equation (11) tells us that the natural extension of a moment sequence is a Moore–Smith limit, and (iii) that the limit of a convex mixture is the convex mixture of the limits. □

Applying these results, we deduce that

$$\begin{aligned} m_k &= (1-\mu_m)(m_c)_k + \mu_m (m_b)_k \\ &= (1-\mu_m)(RS)\int_0^1 x^k \, \mathrm{d}F_m(x) + \sum_{d \in \mathscr{D}_m} [\overline{F}_{m,b}(d) - \underline{F}_{m,b}(d-)] d^k, \end{aligned}$$

for all $k \geq 0$ [for the first term, observe that $F_m(0) = 0$ and recall the results of Section 3.3] and that

$$\underline{E}_m = (1-\mu_m)\underline{E}_{m_c} + \mu_m \underline{E}_{m_b} = (1-\mu_m)\underline{E}_{F_m} + \mu_m \underline{E}_{m_b}.$$

Remark 5. It is instructive to derive these results in an alternative manner. We may infer from Theorem 8 that any distribution function F in $\Phi(\underline{F}_m, \overline{F}_m)$ produces the moment sequence m. But then Equation (9) leads to the conclusion that for $k > 0$,

$$m_k = (RS)\int_0^1 x^k \, \mathrm{d}F(x) = (RS)\int_0^1 x^k \, \mathrm{d}\underline{F}_m(x) = (RS)\int_0^1 x^k \, \mathrm{d}\overline{F}_m(x).$$

Now Equations (18)–(23), together with a property of Riemann–Stieltjes integrals, which gives a decomposition for the Riemann–Stieltjes integral as a convex mixture of a continuous and a break part (see for instance [19, Theorem 13.8, p. 60]), allow us to rewrite any of these Riemann–Stieltjes integrals as

$$(1-\mu_m)(RS)\int_0^1 x^k \, \mathrm{d}F_m(x) + \sum_{d \in \mathscr{D}_m} [\overline{F}_{m,b}(d) - \underline{F}_{m,b}(d-)] d^k. \blacklozenge$$

Proposition 13 below provides a 'converse' to these results. Before we can prove it, we need to introduce some additional concepts.

Definition 6. Define, for $d \in [0,1]$, the functionals $\underline{\mathrm{osc}}_d$ and $\overline{\mathrm{osc}}_d$ on $\mathscr{L}([0,1])$ by

$$\underline{\mathrm{osc}}_d(h) := \sup_{d\in B\in\mathscr{T}} \inf_{z\in B} h(z) \quad \text{and} \quad \overline{\mathrm{osc}}_d(h) := \inf_{d\in B\in\mathscr{T}} \sup_{z\in B} h(z)$$

for all gambles h on $[0,1]$, where $\mathscr{T}$ is the topology of the open subsets of $[0,1]$.

The functional $\underline{\mathrm{osc}}_d$ is a completely monotone coherent lower prevision on $\mathscr{L}([0,1])$ and $\overline{\mathrm{osc}}_d$ is its conjugate upper prevision. Indeed, for any open interval $B \subset [0,1]$, the vacuous lower prevision $\underline{P}_B(h) = \inf_{z\in B} h(z)$ is coherent and completely monotone, and $\underline{\mathrm{osc}}_d$ is a Moore–Smith limit of such vacuous lower previsions, which is consequently also completely monotone and coherent. It is easy to prove that for any gamble h on $[0,1]$,

$$\overline{\mathrm{osc}}_d(h) - \underline{\mathrm{osc}}_d(h) = \inf_{d\in B\in\mathscr{T}} \sup_{z,z'\in B} |h(z) - h(z')| := \mathrm{osc}_d(h)$$

is the so-called *oscillation* of h in d (see for instance [25, Section 18.28]), and it is known that h is continuous in d if and only if $\mathrm{osc}_d(h) = 0$, or in other words if $\underline{\mathrm{osc}}_d(h) = \overline{\mathrm{osc}}_d(h)$. Because of this, we shall call $\underline{\mathrm{osc}}_d(h) \leq h(d)$ the *lower oscillation* of h in d, and $\overline{\mathrm{osc}}_d(h) \geq h(d)$ the *upper oscillation*. Also, if h has a left and right limit in d, we get $\underline{\mathrm{osc}}_d(h) = \min\{h(d-), h(d), h(d+)\}$.

The following lemma tells us that the gamble $\underline{\mathrm{osc}}(h)$ which maps any x in $[0,1]$ to $\underline{\mathrm{osc}}_x(h)$ is the point-wise greatest lower semi-continuous gamble that is dominated by h, and similarly that $\overline{\mathrm{osc}}(h)$ is the point-wise smallest upper semi-continuous gamble that dominates h. They coincide moreover with the gambles $h^{\uparrow}, h^{\downarrow}$ defined in Equation (12). Also, the gamble $\mathrm{osc}(h)$ that maps any x in $[0,1]$ to $\mathrm{osc}_x(h)$ is upper semi-continuous as the sum of two upper semi-continuous gambles $\overline{\mathrm{osc}}(h)$ and $-\underline{\mathrm{osc}}(h) = \overline{\mathrm{osc}}(-h)$.

Lemma 11. *Consider any gamble h on $[0,1]$. Then $h^{\uparrow} = \underline{\mathrm{osc}}(h)$ is the point-wise greatest lower semi-continuous gamble on $[0,1]$ that is dominated by h. Similarly, $h^{\downarrow} = \overline{\mathrm{osc}}(h)$ is the point-wise smallest upper semi-continuous gamble on $[0,1]$ that dominates h.*

Proof. It follows from the definition of $\underline{\mathrm{osc}}(h)$ that for any real t, $\{\underline{\mathrm{osc}}(h) > t\} = \mathrm{int}(\{h > t\})$. This implies that $\{\underline{\mathrm{osc}}(h) > t\}$ is open, so $\underline{\mathrm{osc}}(h)$ is lower semi-continuous. Now let g be a lower semi-continuous gamble on $[0,1]$ that is dominated by h. Then for any real t, $\{g > t\} \subseteq \{h > t\}$, whence $\{g > t\} = \mathrm{int}(\{g > t\}) \subseteq \mathrm{int}(\{h > t\}) = \{\underline{\mathrm{osc}}(h) > t\}$. This implies that for any $d \in [0,1]$,

$$g(d) = \sup\{t \colon d \in \{g > t\}\} \leq \sup\{t \colon d \in \{\underline{\mathrm{osc}}(h) > t\}\} = \underline{\mathrm{osc}}_d(h),$$

so $g \leq \underline{\mathrm{osc}}(h)$. This already tells us that $\underline{\mathrm{osc}}(h)$ is the point-wise greatest lower semi-continuous gamble on $[0,1]$ that is dominated by h. We now prove that $h^{\uparrow} = \underline{\mathrm{osc}}(h)$. Observe that $h^{\uparrow}$ is lower semi-continuous, as a point-wise supremum of continuous gambles. Hence $h^{\uparrow} \leq \underline{\mathrm{osc}}(h)$. To prove the converse inequality, consider any d in $(0,1)$, and consider a $B \in \mathscr{T}$ that contains d. Then there is some $B' \subset B$ in $\mathscr{T}$ that also contains d and such that $\inf B' > \inf B$ and $\sup B' < \sup B$. Define the gamble g to be constant and equal to $\inf h$ outside B, constant and equal to $\inf_{z\in B} h(z)$ on B' and linear on the intervals $(\inf B, \inf B']$ and $[\sup B', \sup B)$. Then $g \leq h$ and g is continuous, so it follows from the definition of $h^{\uparrow}$ that $h^{\uparrow}(d) \geq g(d) = \inf_{z\in B} h(z)$. Hence $h^{\uparrow}(d) \geq \sup_{d\in B\in\mathscr{T}} \inf_{z\in B} h(z) = \underline{\mathrm{osc}}_d(h)$. The case where $d \in \{0,1\}$ is similar. $\square$

Lemma 12. *Let h be a gamble on $[0,1]$. Then for any $d \in [0,1]$,*

$$(R)\int_{\inf h}^{\sup h} I_{\mathrm{int}(\{h\geq t\})}(d)\,\mathrm{d}t = \underline{\mathrm{osc}}_d(h) - \inf h.$$

Proof. It follows from the equality $\{\underline{\mathrm{osc}}_d(h) > t\} = \mathrm{int}(\{h > t\})$, valid for all t. $\square$

Proposition 13. *Consider a countable subset $\mathscr{D}$ of $[0,1]$, and (strictly) positive real α_d, $d \in \mathscr{D}$ that sum to one. Let the moment sequence m be given by $m_0 = 1$ and $m_k = \sum_{d\in\mathscr{D}} \alpha_d d^k$, $k > 0$. Then this moment sequence satisfies the Hausdorff moment condition. Moreover, $\mu_m = 1$, $\mathscr{D}_m = \mathscr{D}$ and for any $x \in [0,1]$,*

$$\underline{F}_m(x-) = \underline{F}_{m,b}(x-) = \sum_{d\in\mathscr{D}, d<x} \alpha_d \quad \textit{and} \quad \overline{F}_m(x) = \overline{F}_{m,b}(x) = \sum_{d\in\mathscr{D}, d\leq x} \alpha_d. \tag{24}$$

Finally, for any gamble h on $[0,1]$,

$$\underline{E}_m(h) = \sum_{d\in\mathscr{D}} \alpha_d \, \underline{\mathrm{osc}}_d(h) \tag{25}$$

Proof. Let P^σ be the σ-additive probability measure on the Borel sets of $[0,1]$ with probability mass $P^\sigma(\{d\}) = \alpha_d$ in $d \in \mathscr{D}$. Then this probability measure has moment sequence m, so m must satisfy the Hausdorff moment condition, and P^σ is the only σ-additive probability measure with this moment sequence, i.e., $P^\sigma = P^\sigma_m$. We denote the (right-continuous) distribution function of this P^σ_m by F^σ_m. By σ-additivity, we have for all $x \in [0,1]$ that

$$F^\sigma_m(x) = P^\sigma_m([0,x]) = \sum_{d\in\mathscr{D}, d\leq x} \alpha_d,$$

and Equation (24) now follows from Proposition 4 and Lemma 5. The set $\mathscr{D}_{\underline{F}_m}$ of discontinuity points of $\underline{F}_m$ is therefore given by $\mathscr{D}$ and similarly $\mathscr{D}_{\overline{F}_m} = \mathscr{D} \setminus \{0\}$. Hence $\mathscr{D}_m = \mathscr{D}_{\underline{F}_m} \cup \mathscr{D}_{\overline{F}_m} = \mathscr{D}$. For any $d \in \mathscr{D}$ we also infer from Equation (24) that $\overline{F}_m(d) - \underline{F}_m(d-) = \alpha_d$, so $\mu_m = \sum_{d\in\mathscr{D}} \alpha_d = 1$ by Equation (17). Then, because $\mu_m = 1$, it also holds that $\underline{F}_m = \underline{F}_{m,b}$ and $\overline{F}_m = \overline{F}_{m,b}$.

We now prove Equation (25). Use Theorem 3 and Equation (24) to show that for any $B \in \mathscr{T}$, $\underline{E}_m(B) = \sum_{d\in B\cap\mathscr{D}} \alpha_d$. We can then use Theorem 3 to find that for any $A \subseteq [0,1]$,

$$\underline{E}_m(A) = \sum_{d\in\mathscr{D}\cap\mathrm{int}(A)} \alpha_d = \sum_{d\in\mathscr{D}} \alpha_d I_{\mathrm{int}(A)}(d).$$

(Recall that $\underline{\mathrm{osc}}(I_A) = I_A^{\uparrow} = I_{\mathrm{int}(A)}$.) Now consider any gamble h on $[0,1]$, and label the elements of $\mathscr{D}$ with natural numbers, so $\mathscr{D} = \{d_k \colon k \geq 0\}$. Define the gambles g_n on $\mathbb{R}$ by $g_n(t) := \sum_{k=0}^n \alpha_{d_k} I_{\mathrm{int}(\{h\geq t\})}(d_k)$. Then $0 \leq g_n \leq 1$, so this sequence is uniformly bounded. Moreover, for each $t \in \mathbb{R}$,

$$\lim_{n\to\infty} g_n(t) = \sum_{k=0}^{\infty} \alpha_{d_k} I_{\mathrm{int}(\{h\geq t\})}(d_k) = \sum_{d\in\mathscr{D}} \alpha_d I_{\mathrm{int}(\{h\geq t\})}(d) = \underline{E}_m(\{h \geq t\}).$$

Since we know by Theorem 3 that $\underline{E}_m$ is the Choquet functional associated with its restriction to events, we can invoke a known convergence result for Riemann integrals (Osgood's Theorem, see for instance [19, Theorem 15.6, pp. 71–74]) to pull the limit through the integral and deduce that

$$\begin{aligned}
\underline{E}_m(h) &= \inf h + (R)\int_{\inf h}^{\sup h} \underline{E}_m(\{h \geq t\}) \,\mathrm{d}t \\
&= \inf h + \lim_{n\to\infty} \sum_{k=0}^{n} \alpha_{d_k} (R)\int_{\inf h}^{\sup h} I_{\mathrm{int}(\{h\geq t\})}(d_k) \,\mathrm{d}t \\
&= \sum_{d\in\mathscr{D}} \alpha_d \left[\inf h + (R)\int_{\inf h}^{\sup h} I_{\mathrm{int}(\{h\geq t\})}(d) \,\mathrm{d}t\right] = \sum_{d\in\mathscr{D}} \alpha_d \, \underline{\mathrm{osc}}_d(h),
\end{aligned}$$

where the last equality follows from Lemma 12. $\square$

If we combine all the foregoing results in this section, we get the following central result, which is one of the 'interesting formula for $\underline{E}_m$' we promised to derive.

Theorem 14. *Consider a Hausdorff moment sequence m. Then for any gamble h on* $[0,1]$

$$\underline{E}_m(h) = (1-\mu_m)\underline{E}_{F_m}(h) + \sum_{d\in\mathscr{D}_m} [\overline{F}_m(d) - \underline{F}_m(d-)]\,\underline{\mathrm{osc}}_d(h), \tag{26}$$

where $\underline{\mathrm{osc}}$ *is the lower oscillation given by Definition 6. Moreover, it holds in general that* $\underline{E}_{F_m} = (RS)\underline{\int}_0^1 \cdot\, \mathrm{d}F_m(x)$*, but we have*

$$\underline{E}_m = \inf\left\{(RS)\underline{\int}_0^1 \cdot\, \mathrm{d}F(x) \colon F \in \Phi(\underline{F}_m, \overline{F}_m)\right\} \Leftrightarrow \overline{F}_m(0) = 0.$$

Proof. The first statement follows from Lemmas 9 and 10, and Proposition 13; the second, from Corollary 7. Finally, the equivalence is a consequence of Theorem 8. □

Proposition 15. *Consider a continuous distribution function F on* $[0,1]$ *such that* $F(0)=0$*, and the associated moment sequence m given by* $m_k = (RS)\int_0^1 x^k\,\mathrm{d}F(x)$*,* $k \geq 0$*. Then* $\underline{F}_m = \overline{F}_m = F_m = F$ *and* $\underline{E}_m = \underline{E}_F = (RS)\underline{\int}_0^1 \cdot\,\mathrm{d}F(x)$*. Moreover, the following statements are equivalent for any gamble h on* $[0,1]$*:*

1. *h is m-integrable;*
2. *h is F-integrable;*
3. *h is Riemann–Stieltjes-integrable with respect to F.*

Finally, for all gambles f on $[0,1]$*,*

$$\begin{aligned}\underline{E}_m(f) = \underline{E}_F(f) &= (RS)\underline{\int}_0^1 f(x)\,\mathrm{d}F(x)\\ &= (RS)\underline{\int}_0^1 \underline{\mathrm{osc}}_x(f)\,\mathrm{d}F(x) = (LS)\int_0^1 \underline{\mathrm{osc}}_x(f)\,\mathrm{d}F(x).\end{aligned}$$

Proof. From Lemma 5, we deduce that $\underline{F}_m(x) = \overline{F}_m(x)$ for all $x \in (0,1]$. Moreover, the lemma also implies that $\overline{F}_m(0) = F(0) = 0$. The equivalence between the first three statements follows then from Corollary 7.

Finally, the first three equalities in the last chain follow from Theorem 3 and the first part of this proposition. For the last equality, let us prove that $\underline{E}_m$ and the (LS) integral coincide on lower semi-continuous gambles.

First, consider an open set A. Then, there is a countable family of pair-wise disjoint open intervals I_n such that $A = \bigcup_n I_n$. For any n, we know that $(LS)\int_0^1 I_{I_n}(x)\,\mathrm{d}F(x) = F(\sup I_n) - F(\inf I_n) = \underline{E}_m(I_n)$, where the last equality follows from the first part of this corollary. Moreover,

$$\begin{aligned}(LS)\int_0^1 I_A(x)\,\mathrm{d}F(x) &= (LS)\int_0^1 I_{\bigcup_n I_n}(x)\,\mathrm{d}F(x) = \lim_{n\to\infty}(LS)\int_0^1 I_{\bigcup_{k=1}^n I_k}(x)\,\mathrm{d}F(x)\\ &= \lim_{n\to\infty}\sum_{k=1}^n (LS)\int_0^1 I_{I_k}(x)\,\mathrm{d}F(x) = \lim_{n\to\infty}\sum_{k=1}^n \underline{E}_m(I_k) = \underline{E}_m(A),\end{aligned}$$

where the second equality is a consequence of the monotone convergence of the Lebesgue–Stieltjes functional and the last one follows from Equation (13).

Now, let us consider a gamble f on $[0,1]$, and let $\underline{\mathrm{osc}}(f)$ be its lower oscillation. Then, the strict cut sets $\{\underline{\mathrm{osc}}(f) > t\}$ of $\underline{\mathrm{osc}}(f)$ are open for all real t. Since both $\underline{E}_m$ and the Lebesgue–Stieltjes integral operator are completely monotone and coherent functionals (see [10]), they are equal to the Choquet integrals with respect to their restrictions to events. From this, we deduce that $\underline{E}_m(f) = (LS)\int_0^1 \underline{\mathrm{osc}}_x(f)\,\mathrm{d}F(x)$, also taking into account that $\underline{\mathrm{osc}}(f)$ is Borel-measurable (its strict cut sets are), and therefore Lebesgue–Stieltjes integrable. □

Theorem 16. *Consider a Hausdorff moment sequence m. Then for all gambles h on* $[0,1]$*,*

$$\underline{E}_m(h) = (LS)\int_0^1 \underline{\mathrm{osc}}_x(h)\,\mathrm{d}F_m^\sigma(x). \tag{27}$$

Proof. It follows from Theorem 14 and Proposition 15 that, since F_m is by construction continuous and satisfies $F_m(0) = 0$,

$$\underline{E}_m(h) = (1-\mu_m)(LS)\int_0^1 \underline{\mathrm{osc}}_x(h)\,\mathrm{d}F_m(x) + \sum_{d\in\mathscr{D}_m}[\overline{F}_m(d) - \underline{F}_m(d-)]\,\underline{\mathrm{osc}}_d(h)$$

for any gamble h on $[0,1]$. Now using Equations (21) and (22), and the fact that $\overline{F}_m = F_m^\sigma$ [Lemma 5], we see that for all $x \in [0,1)$

$$F_m^\sigma(x) = (1-\mu_m)F_m(x) + \sum_{d\in\mathscr{D}_m, d\leq x}[\overline{F}_m(d) - \underline{F}_m(d-)].$$

Combining these two equalities leads to the desired result. □

Corollary 17. *Consider a Hausdorff moment sequence m. Then the following statements are equivalent for any gamble h on* $[0,1]$*:*

1. *h is m-integrable;*
2. *h is continuous in all the discontinuity points* $d \in \mathscr{D}_m$*, as well as Riemann–Stieltjes-integrable with respect to* F_m *(or equivalently* F_m*-integrable) if* $\mu_m < 1$*;*
3. $(LS)\int_0^1 \mathrm{osc}_x(h)\,\mathrm{d}F_m^\sigma(x) = 0$*, i.e., h is continuous almost everywhere with respect to the unique* σ*-additive probability measure induced by the moment sequence m.*
4. $\overline{E}_m(\mathrm{osc}(h)) = 0$*.*

Moreover, for any m-integrable gamble h we have

$$\begin{aligned} E_m(h) &= (RS)\int_0^1 h(x)\,\mathrm{d}\underline{F}_m(x) = (RS)\int_0^1 h(x)\,\mathrm{d}\overline{F}_m(x) \\ &= (1-\mu_m)(RS)\int_0^1 h(x)\,\mathrm{d}F_m(x) + \sum_{d\in\mathscr{D}_m}[\overline{F}_m(d) - \underline{F}_m(d-)]h(d). \end{aligned}$$

Proof. We derive from Equations (26) and (27) that

$$\begin{aligned} \overline{E}_m(h) - \underline{E}_m(h) &= (1-\mu_m)[\overline{E}_{F_m}(h) - \underline{E}_{F_m}(h)] + \sum_{d\in\mathscr{D}_m}[\overline{F}_m(d) - \underline{F}_m(d-)]\,\mathrm{osc}_d(h) \\ &= (LS)\int_0^1 \mathrm{osc}_x(h)\,\mathrm{d}F_m^\sigma(x), \end{aligned}$$

also using that $\mathrm{osc}(h)$ is upper semi-continuous and therefore Borel-measurable. This shows that the first three statements are equivalent. We now prove that the first statement implies the fourth. Since $\overline{E}_m$ is coherent and therefore monotone and sub-additive, we get

$$\begin{aligned} 0 \leq \overline{E}_m(\mathrm{osc}(h)) &= \overline{E}_m(\overline{\mathrm{osc}}(h) - \underline{\mathrm{osc}}(h)) \leq \overline{E}_m(\overline{\mathrm{osc}}(h)) + \overline{E}_m(-\underline{\mathrm{osc}}(h)) \\ &= \overline{E}_m(\overline{\mathrm{osc}}(h)) - \underline{E}_m(\underline{\mathrm{osc}}(h)) = \overline{E}_m(h) - \underline{E}_m(h), \end{aligned}$$

where the last equality follows from Theorem 3 and Lemma 11. So if h is m-integrable, then $\underline{E}_m(h) = \overline{E}_m(h)$ and therefore also $\overline{E}_m(\mathrm{osc}(h)) = 0$. Let us also prove that the fourth statement implies the third; the rest of the proof is then obvious. Observe that the positive and normed linear continuous functional on the Borel-measurable gambles on $[0,1]$, given by $(LS)\int_0^1 \cdot\, \mathrm{d}F_m^\sigma(x)$ has moment sequence m, and is therefore dominated by $\overline{E}_m$ on all Borel-measurable gambles. Since $\operatorname{osc}(h) \geq 0$,

$$\overline{E}_m(\operatorname{osc}(h)) \geq (LS)\int_0^1 \operatorname{osc}_x(h)\,\mathrm{d}F_m^\sigma(x) \geq 0,$$

for any gamble h on $[0,1]$. □

Remark 6 (On discrete probability mass). If follows in particular from Proposition 13 that for a given $d \in [0,1]$, the natural extension of the moment sequence $m_k = d^k$, $k \geq 0$ is given by $\underline{E}_m = \underline{\operatorname{osc}}_d$. Similarly, suppose we have a linear prevision P_d on the set of all continuous gambles $\mathscr{C}([0,1])$ given by $P_d(h) = h(d)$. Then the natural extension of this linear prevision to the set of all gambles is again $\underline{\operatorname{osc}}_d$.

Intuitively, the situation above may be described by the phrase "all the mass of the probability distribution is concentrated in d". The discussion in this remark aims at making this interesting case more intelligible to the reader.

Let us consider a non-empty set $\mathscr{F}$ of subsets of $[0,1]$, and define the lower probability $\underline{P}_{\mathscr{F}}$ by

$$\underline{P}_{\mathscr{F}}(A) := \begin{cases} 1 & \text{if } A \in \mathscr{F} \\ 0 & \text{if } A \notin \mathscr{F} \end{cases}$$

Then $\underline{P}_{\mathscr{F}}$ is coherent if and only if $\mathscr{F}$ is a *proper filter*, i.e., a proper subset of the powerset of $[0,1]$ that is increasing and closed under finite intersections.[13] Its natural extension to the set of all gambles on $[0,1]$ will also be denoted by $\underline{P}_{\mathscr{F}}$, and is given by[14]

$$\underline{P}_{\mathscr{F}}(f) = \sup\{t \in \mathbb{R} \colon \{f \geq t\} \in \mathscr{F}\} = \sup_{A \in \mathscr{F}} \inf_{x \in A} f(x). \tag{28}$$

If we consider the *neighbourhood filter* $\mathscr{N}_d$ *of* d, i.e., the filter of all neighbourhoods of d, or in other words, of all subsets of $[0,1]$ that include some open interval containing d, then it follows from Equation (28) that

$$\underline{P}_{\mathscr{N}_d} = \underline{\operatorname{osc}}_d,$$

so $\underline{\operatorname{osc}}_d$ is actually the smallest coherent lower prevision that assumes the value one on any neighbourhood of d (and zero elsewhere). So "all probability mass concentrated in d" should actually be formulated more exactly as "all probability mass located within any neighbourhood of d".

But there is more. A linear prevision Q coincides with P_d, or in other words, satisfies $Q(h) = h(d)$ for all continuous gambles h, if and only if it dominates $\underline{P}_{\mathscr{N}_d} = \underline{\operatorname{osc}}_d$, and it is not so difficult to show that[15]

$$\mathscr{M}(\underline{\operatorname{osc}}_d) = \mathscr{M}(P_d) = \overline{\operatorname{co}}\{P_{\mathscr{U}} : \mathscr{U} \to d\}. \tag{29}$$

where $\overline{\operatorname{co}}$ denotes 'convex closure' in the weak* topology, the $\mathscr{U}$ denote *ultrafilters*, or maximal proper filters, and $\mathscr{U} \to d$ means that $\mathscr{N}_d \subseteq \mathscr{U}$, or in the language of topology, that $\mathscr{U}$ *converges to* d. This means that the linear previsions $P_{\mathscr{U}}$ with $\mathscr{U} \to d$ constitute

[13] See Walley's book [27, Section 2.9.8] for a proof. Walley also shows there that $\underline{P}_{\mathscr{F}}$ is a linear prevision if and only if $\mathscr{F}$ is actually an ultrafilter.

[14] To see this, check that the coherent lower probability $\underline{P}_{\mathscr{F}}$ is actually completely monotone, so its natural extension is the Choquet integral associated with this lower probability, which is again completely monotone. Evaluating this Choquet integral then yields Equation (28), if we also take into account that the lower probability $\underline{P}_{\mathscr{F}}$ only assumes the values zero and one.

[15] To see this, combine Theorems 3.6.2 and 3.6.4 in [27, Section 3.6]. The linear previsions $\underline{P}_{\mathscr{U}}$ with $\mathscr{F} \subseteq \mathscr{U}$ are the extreme points of the convex weak*-compact set $\mathscr{M}(\underline{P}_{\mathscr{F}})$.

the extreme points of the convex weak*-closed set of all linear previsions with moments $m_k = d^k$.

Among the ultrafilters $\mathscr{U}$ converging to d, there is only one for which $P_{\mathscr{U}}$ is σ-additive on events, namely the *fixed ultrafilter* $\mathscr{U} = \{A \subseteq [0,1] \colon d \in A\}$. Any other ultrafilter $\mathscr{U}$ converging to d is *free*, meaning that the intersection of all the sets in $\mathscr{U}$ is the empty set. For those, the corresponding linear previsions $P_{\mathscr{U}}$ are only finitely additive, because σ-additivity of $P_{\mathscr{U}}$ is easily seen to imply that $\{d\} \in \mathscr{U}$.

If for a given ultrafilter $\mathscr{U}$ we define $d := \inf\{x \in [0,1] \colon [0,x) \in \mathscr{U}\}$, then $\mathscr{U} \to d$, so every ultrafilter converges to some element of $[0,1]$. Moreover, for any ultrafilter $\mathscr{U} \to d$ one of the three mutually exclusive possibilities holds: (i) $[0,d) \in \mathscr{U}$; (ii) $(d,1] \in \mathscr{U}$; or (iii) $\{d\} \in \mathscr{U}$. Case (iii) singles out the unique fixed ultrafilter, with $P_{\mathscr{U}}(A) = 1$ if and only if $d \in A$. The distribution function for this linear prevision is given by $\beta(\cdot; d, 1)$, where β is the simple break function defined in Example 1. For case (i), we see that $P_{\mathscr{U}}((d-\varepsilon, d)) = 1$ for all $\varepsilon > 0$ and $P_{\mathscr{U}}([d,1]) = 0$, and the distribution function for this linear prevision is given by $\beta(\cdot; d, 1)$. We say that $\mathscr{U}$ represents probability mass *left-adherent to d*. In the language of non-standard analysis, we can say that all probability mass is concentrated in some non-standard real number infinitesimally close to, and to the left of, d. Similarly, case (ii) describes probability mass that is *right-adherent to d*, with distribution function $\beta(\cdot; d, 0)$. It follows from these considerations and Equation (29) that the distribution function of any linear prevision with moments $m_k = d^k$ is $\beta(\cdot; d, a)$ with $a \in [0,1]$. ♦

Remark 7 (On Choquet–Maaß representation). We can now extend the results mentioned in the previous remark to general moment sequences m, and not just the ones associated with 'discrete probability mass'. Indeed, it is a consequence of results by Choquet [4, Section 45] and Maaß [20, Section 2.4] that any coherent and completely monotone lower prevision can be written as a 'σ-additive convex mixture' of the extreme points of the set of all coherent and completely monotone lower previsions. Now, it follows, again from results by Choquet [4, Section 43.7], that the extreme coherent and completely monotone lower previsions are precisely the lower previsions $\underline{P}_{\mathscr{F}}$ associated with proper filters $\mathscr{F}$. If we rewrite Equation (27) as follows

$$\underline{E}_m(h) = (LS)\int_0^1 \underline{P}_{\mathscr{N}_x}(f)\,\mathrm{d}F_m^{\sigma}(x),$$

we see that for the completely monotone and coherent natural extension $\underline{E}_m$ of the moment sequence m, we can actually identify the 'σ-additive convex mixture' and the extreme points that participate in it: the mixture is precisely the one associated with the unique σ-probability measure induced by the moment sequence m, and the extreme points are the lower previsions associated with the neighbourhood filters $\mathscr{N}_x$, $x \in [0,1]$. As we have seen above, the latter express that all probability mass is located within any neighbourhood of x. Note that the representation in terms of extreme points of the constructible $\underline{E}_m$ is constructible as well.

We want to point out here that this result can be generalised quite easily. If P is a linear prevision on the set of continuous gambles $\mathscr{C}(K)$ on some metrisable compact space K, then the lower envelope $\underline{E}_P$ of all linear previsions that extend P to $\mathscr{L}(K)$ is given by $\underline{E}_P(f) = (L)\int \underline{P}_{\mathscr{N}_{\cdot}}\,\mathrm{d}\mu$, where μ is the unique σ-additive 'extension' of P to all Borel-measurable gambles on K, and $\mathscr{N}_x$ the neighbourhood filter of $x \in K$. See [8] for more details. ♦

6. Conclusion

A σ-additive probability measure is uniquely determined by its distribution function, and also by its sequence of moments. In this paper, together with [21], we have investigated if the same holds when we consider finitely additive probability measures.

Our results show that, in terms of the amount of information they provide, distribution functions are located between probability measures and sequences of moments. On the one hand, for any given distribution function there is an infinite number of finitely additive probability measures inducing it. Only one of these, of course, is σ-additive. On the other hand, a distribution function uniquely determines a sequence of moments, but in general there will be an infinite number of different distribution functions with the same moments. Again, only one of these distribution functions corresponds to a σ-additive probability measure. Interestingly, that is also the greatest distribution function with those moments. This is because of the assumption of right-continuity made in the (classical) definition of a distribution function.

We have also investigated under which conditions the moments uniquely determine the distribution function. We have proven that they do if and only if the distribution function is continuous on $[0,1)$. In that case, we can characterise the (infinite) set of probability measures associated to the distribution by means of a Riemann-Stieltjes integral.

More generally, the complete monotonicity of the linear prevision we can associate with a distribution function allows us to represent the corresponding set of linear previsions by means of a Choquet integral, which in turn can be expressed in terms of a Riemann integral. The complete monotonicity also implies that we can characterise this set by the corresponding restrictions to events. This then provides an alternative equivalent representation of the information given by a distribution function.

Acknowledgements

We acknowledge the support of the projects MTM2004-01269, TSI2004-06801-C04-01 and by the research grant G.0139.01 of the Flemish Fund for Scientific Research (FWO). Erik Quaeghebeur's research is financed by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

We want to express our gratitude to two anonymous referees who came up with many detailed suggestions for improving the readability of this quite technical paper.

References

[1] S. Banach. *Théorie des Opérations Linéaires*. Number 1 in Monografje Matematyczne. Warszava-Lwów, 1932.

[2] S. Banach. *Theory of Linear Operators*. Number 38 in Mathematical Library. North-Holland, Amsterdam, 1987. English translation by F. Jellet of [1].

[3] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.

[4] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.

[5] L. Crisma. *Introduzione alla teoria delle probabilità coerenti*. EUT, Trieste, 2006.

[6] P. J. Daniell. A general form of integral. *The Annals of Mathematics*, 19:279–294, 1918.

[7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.

[8] G. de Cooman and E. Miranda. The F. Riesz Representation Theorem and finite additivity. 2006. Submitted for publication.

[9] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n-Monotone lower previsions and lower integrals. In F. G. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA 2005 – Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, pages 145–154. SIPTA, 2005.

[10] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n-Monotone exact functionals. 2006. Submitted for publication.

[11] G. de Cooman, M. C. M. Troffaes, and E. Miranda. A unifying approach to integration for bounded positive charges. 2006. Submitted for publication.

[12] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[13] B. de Finetti. *Theory of Probability*, volume 1. John Wiley & Sons, Chichester, 1974. English translation of [12].

[14] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.

[15] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley and Sons, New York, 1971.

[16] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.

[17] F. Hausdorff. Summationmethoden und Momentfolgen I. *Mathematische Zeitschrift*, 9:74–109, 1921.

[18] F. Hausdorff. Summationmethoden und Momentfolgen II. *Mathematische Zeitschrift*, 9:280–299, 1921.

[19] T. H. Hildebrandt. *Introduction to the Theory of Integration*, volume 13. Academic Press, New York, 1963.

[20] S. Maaß. *Exact functionals, functionals preserving linear inequalities, Lévy's metric*. PhD thesis, University of Bremen, 2003.

[21] E. Miranda, G. de Cooman, and E. Quaeghebeur. The Hausdorff moment problem under finite additivity. *Journal of Theoretical Probability*, 2007. In press.

[22] E. H. Moore and H. L. Smith. A general theory of limits. *American Journal of Mathematics*, 44:102–121, 1922.

[23] F. Riesz. Sur les opérations fonctionelles linéaires. *Comptes Rendus de l'Académie des Sciences de Paris*, 149:974–977, 1909.

[24] F. Riesz and B. Szökefalvi-Nagy. *Functional Analysis*. Frederick Ungar Publishing Co., New York, 1955. English translation of the second French edition of "Leçons d'analyse fonctionelle". Reprinted by Dover Publications in 1990.

[25] E. Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, CA, 1997.

[26] M. C. M. Troffaes. *Optimality, Uncertainty, and Dynamic Programming with Lower Previsions*. PhD thesis, Ghent University, Ghent, Belgium, March 2005.

[27] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

REY JUAN CARLOS UNIVERSITY, DEPT. OF STATISTICS AND OPERATIONS RESEARCH. C-TULIPÁN, S/N 28933, MÓSTOLES, SPAIN
E-mail address: enrique.miranda@urjc.es

GHENT UNIVERSITY, SYSTEMS RESEARCH GROUP, TECHNOLOGIEPARK – ZWIJNAARDE 914, 9052 ZWIJNAARDE, BELGIUM
E-mail address: gert.decooman@ugent.be

GHENT UNIVERSITY, SYSTEMS RESEARCH GROUP, TECHNOLOGIEPARK – ZWIJNAARDE 914, 9052 ZWIJNAARDE, BELGIUM
E-mail address: erik.quaeghebeur@ugent.be