

THE HAUSDORFF MOMENT PROBLEM REVISITED

ENRIQUE MIRANDA, GERT DE COOMAN, AND ERIK QUAEGHEBEUR

ABSTRACT. We investigate to what extent finitely additive probability measures on the unit interval are determined by their moment sequence. We do this by studying the lower envelope of all finitely additive probability measures with a given moment sequence. Our investigation leads to several elegant expressions for this lower envelope, and it allows us to conclude that the information provided by the moments is equivalent to the one given by the associated lower and upper distribution functions.

1. INTRODUCTION

To what extent does a sequence of moments determine a probability measure? This problem 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).

The question refers to *the moment problem*. Classically, there are three of these: the Hamburger moment problem for probability measures on \mathbb{R} ; the Stieltjes moment problem for probability measures on $[0, +\infty)$; and the Hausdorff moment problem for probability measures on compact real intervals, which is the one we consider here.

Let us first look at the moment problem for σ -additive probability measures. Consider a sequence m of real numbers $m_k, k \geq 0$. It turns out (see, for instance, [12, Section VII.3] for an excellent exposition) that there then is a σ -additive probability measure P_m^σ defined on the Borel sets of $[0, 1]$ such that

$$\int_{[0,1]} x^k dP_m^\sigma = m_k, \quad k \geq 0,$$

if and only if $m_0 = 1$ (normalisation) and the sequence m is *completely monotone*. This means that for all $n, k \geq 0$

$$(-1)^n \Delta^n m_k \geq 0,$$

where the differences $\Delta^n m_k$ are defined recursively through $\Delta^n m_k = \Delta^{n-1} m_{k+1} - \Delta^{n-1} m_k$ for $n \geq 1$ and $\Delta^0 m_k = m_k$. The crucial step towards the proof of this important result was taken by Hausdorff [13, 14]. We shall call this necessary and sufficient condition on the moment sequence the *Hausdorff moment condition*, and any moment sequence that satisfies it a *Hausdorff moment sequence*. The existence of a σ -additive probability measure P_m^σ

Key words and phrases. Hausdorff moment problem, coherent lower prevision, lower distribution function, lower Riemann–Stieltjes integral, complete monotonicity.

with a given sequence of moments m implies its uniqueness, by virtue of the F. Riesz Representation Theorem.¹

When we let go of σ -additivity, uniqueness is no longer guaranteed. In general, there will be an infinite (closed and convex) set $\mathcal{M}(m)$ of finitely additive probabilities that correspond to a given Hausdorff moment sequence m . One of the objectives of this paper is to study this set more closely. We shall see that it is very convenient to do so using the mathematical machinery behind Walley's [23] theory of coherent lower previsions, for which we introduce the basics in Section 2.

In Section 3 we formulate the problem under study using the language of coherent lower previsions. The fundamental step we take, is not to consider the finitely additive probabilities in $\mathcal{M}(m)$ themselves, but to study their *lower envelope* \underline{E}_m , which generally turns out to be a non-linear (but super-additive, normed, positive and positively homogeneous) real functional on the set $\mathcal{L}([0, 1])$ of all bounded functions on $[0, 1]$. A functional with these properties is called a *coherent lower prevision* in Walley's theory. We show that, perhaps surprisingly, this functional is σ -additive on the lattice of open sets (see Propositions 9 and 10).

The decision to study \underline{E}_m , rather than the elements of $\mathcal{M}(m)$, may seem to lead us away from the actual problem at hand, but that is only an illusion. Indeed, 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 [21] 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 $\mathcal{M}(m)$ individually, but by working with their lower envelope, which in his language is called the *natural extension* of the 'something' they extend.² 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. In the present case, we shall see that a finitely additive probability has moment sequence m if and only if it dominates \underline{E}_m on all (indicators of) events. In this precise sense, the natural extension \underline{E}_m of the moment sequence m completely determines the solution to the Hausdorff moment problem for finitely additive probabilities. Much of what we shall do in this paper is to give useful formulae and methods for calculating \underline{E}_m .

Why do we devote so much attention to a problem that may appear perhaps, to some readers, to have only limited technical interest? First of all, the problem is more general than its formulation in terms of moments may seem to suggest: what we do here, is to infer as much as we can in a 'constructive' manner about all the positive linear functionals that extend a given positive linear functional defined on the set $\mathcal{C}([0, 1])$ of all continuous real functions on a compact real interval $[0, 1]$. Obviously, this is because specifying a Hausdorff moment sequence is equivalent to specifying a positive linear functional on $\mathcal{C}([0, 1])$ (also see Theorem 1, normalisation obviously isn't an issue here). By appropriate

¹This is because the set of polynomials on $[0, 1]$ is uniformly dense in the set of continuous functions on $[0, 1]$, so a Hausdorff moment sequence corresponds to a unique positive linear functional on the set of bounded continuous functions on $[0, 1]$. The F. Riesz Representation Theorem in its more modern form (see, for instance, [12, Section V.1]) allows us to extend this functional uniquely (under σ -additivity) to all Borel-measurable functions. More details are also given in Section 3.

²This 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 [10, Sections 3.10–12].

transformations, these results can even be extended from $[0, 1]$ to arbitrary compact real intervals.

Secondly, when extending linear functionals to larger domains, it is current practice in many mathematical fields (and most notably in probability and measure theory) to restrict attention to those domains where the functional still has a unique linear extension, or in other words, to concentrate on linear functionals only. While this attitude may have perfectly respectable historical antecedents, we want to show here, by working out the details in a number of specific examples, that it is unfortunate and perhaps even unproductive, because *it tends to hide very interesting mathematical structure* that quite often becomes apparent only when leaving the linear ambit. Let us mention two examples that are explained in much more detail further on.

In integration theory, it is quite common to define an integral by first looking at lower and upper integrals, and then to conveniently forget about them by zooming in on integrable functions. These are the functions for which the lower and upper integral coincide, and for which an integral can therefore be uniquely defined. Often, it is only this integral that is considered to be of any interest or even to have any meaning, and much effort is then devoted to studying its properties. But, as we have argued in general in earlier papers [6, 7], the lower and upper integrals are of considerable interest in themselves as well. They quite often have mathematical properties that are worthy of consideration *per se*, but which in addition allow us to derive results about integrals and integrability in a straightforward manner. This is often much harder to do when limiting the attention to integrals alone.

Another case in point are the linear extensions of a Hausdorff moment sequence m . We know that these are not unique, so following standard practice, we could restrict our attention to the set of m -integrable bounded functions, i.e., those bounded functions on which all linear extensions of the moment sequence coincide. We could then study the properties of the resulting linear functional E_m on this set. Instead, what we do further on, is to study the non-linear functional \underline{E}_m , for which we can show easily that it is completely monotone as well (see Theorem 13). It can therefore again be written as a Choquet functional (and ultimately therefore as a Riemann integral). This allows us find simple and elegant expressions for \underline{E}_m (Theorems 13 and 14). It also allows us to characterise in a straightforward manner those bounded functions that are m -integrable (Theorem 13). We believe this would have been much harder to do by limiting our attention to the linear restriction E_m of \underline{E}_m alone.

2. A SHORT INTRODUCTION TO COHERENT 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 [23] 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 $\mathcal{L}(\Omega)$.

A *lower prevision* \underline{P} is a real-valued map defined on some subset \mathcal{H} of $\mathcal{L}(\Omega)$. If the domain \mathcal{H} 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 $-\mathcal{H}$ by $\overline{P}(f) := -\underline{P}(-f)$ for every $-f$ in \mathcal{H} . 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 $\mathcal{L}(\Omega)$ of all gambles is called *coherent* if 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: $\underline{P}(f) \geq \inf f$; here f and g are any gambles on Ω . A lower prevision

\underline{P} on an arbitrary domain \mathcal{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 $\sup[\sum_{i=1}^n f_i - mf_0] \geq \sum_{i=1}^n \underline{P}(f_i) - m\underline{P}(f_0)$ for any $n, m \geq 0$ and f_0, f_1, \dots, f_n in \mathcal{K} . 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 $\underline{P}(f_n) \rightarrow \underline{P}(f)$.

A *linear prevision* P on $\mathcal{L}(\Omega)$ is a coherent lower prevision that is moreover self-conjugate: $P(-f) = -P(f)$. 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 \mathcal{K} of $\mathcal{L}(\Omega)$ is called a *linear prevision* if it can be extended to a linear prevision on $\mathcal{L}(\Omega)$. This is the case if and only if $\sup[\sum_{i=1}^n f_i - \sum_{j=1}^m g_j] \geq \sum_{i=1}^n P(f_i) - \sum_{j=1}^m P(g_j)$ for any $n, m \geq 0$ and $f_1, \dots, f_n, g_1, \dots, g_m$ in \mathcal{K} . We let $\mathbb{P}(\Omega)$ denote the set of all linear previsions on $\mathcal{L}(\Omega)$.

The restriction Q of a linear prevision P on $\mathcal{L}(\Omega)$ to the set $\wp(\Omega)$ 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, [1]): for any gamble h in $\mathcal{L}(\Omega)$ we have $P(h) = (D) \int h dQ$.

The *natural extension* \underline{E}_P to $\mathcal{L}(\Omega)$ of a coherent lower prevision \underline{P} defined on \mathcal{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 $\mathcal{M}(\underline{P})$ of all linear previsions that point-wise dominate \underline{P} on its domain \mathcal{K} : for any gamble f in $\mathcal{L}(\Omega)$

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

Moreover, $\mathcal{M}(\underline{E}_P) = \mathcal{M}(\underline{P})$. Indeed, if \underline{P} is a coherent lower prevision on $\mathcal{L}(\Omega)$ and \bar{P} is its conjugate upper prevision, then for any gamble f and for any $a \in [\underline{P}(f), \bar{P}(f)]$ there exists a linear prevision $P \in \mathcal{M}(\underline{P})$ 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 $\mathcal{K}_1 \supseteq \mathcal{K}$ that dominates \underline{P} on \mathcal{K} (i.e., the natural extension of \underline{P} to \mathcal{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 $\mathcal{L}(\Omega)$. Moreover, $\mathcal{M}(\underline{E}_2) = \mathcal{M}(\underline{E}_1) = \mathcal{M}(\underline{P})$. In particular, if P is a linear prevision on a negation invariant \mathcal{K} that has a unique extension P_1 to some larger negation invariant domain \mathcal{K}_1 , then a linear prevision on all gambles will dominate (agree with) P on \mathcal{K} if and only if it dominates (agrees with) P_1 on \mathcal{K}_1 .

Next, we turn to the notion of n -monotonicity. A thorough study of the properties of n -monotone coherent lower previsions can be found in earlier papers [5, 6, 7]. A lower prevision defined on a *lattice* \mathcal{K} of gambles (a set of gambles closed under pointwise 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 \mathcal{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.

We can easily characterise the natural extension of a completely monotone coherent lower prevision \underline{P} . If it is defined on a lattice of events \mathcal{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

$$\underline{P}_*(A) = \sup\{\underline{P}(B) : B \in \mathcal{A}, B \subseteq A\}.$$

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

$$\underline{P}_*(f) = \sup \{ \underline{P}(g) : g \in \mathcal{K}, g \leq f \},$$

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 if $f(\omega) < f(\varpi)$ then also $g(\omega) \leq g(\varpi)$ for all ω and ϖ in Ω . Secondly, it is completely determined by the values it assumes on events: it is actually 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. And 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 $\{f \geq t\} := \{x \in [0, 1] : f(x) \geq t\}$ are *\underline{P} -integrable* for all but a countable number of t .

3. FORMULATION AND INITIAL SOLUTION OF THE PROBLEM

We are now ready to formulate the Hausdorff moment problem using the language established in the previous section.

Consider a sequence m of real numbers $m_k, k \geq 0$, and the subset $\mathcal{V}_p([0, 1])$ of the set $\mathcal{V}([0, 1])$ of all polynomials on the unit interval given by

$$\mathcal{V}_p([0, 1]) := \{ p^k : k \geq 0 \},$$

where $p^k(x) = x^k$ for all $x \in [0, 1]$ and $k \geq 0$. We define a functional P_m on this set by letting $P_m(p^k) := m_k$ for $k \geq 0$. This functional can be *uniquely* extended to a linear functional \hat{P}_m on the set $\mathcal{V}([0, 1])$ of all polynomials. This is done as follows:

$$\hat{P}_m\left(\sum_{k=0}^n a_k p^k\right) = \sum_{k=0}^n a_k P_m(p^k) = \sum_{k=0}^n a_k m_k$$

for all $n \geq 0$ and $a_k \in \mathbb{R}$.

We can then ask whether there is some finitely additive probability, or equivalently, some linear prevision Q with moment sequence m , i.e., such that $Q(p^k) = m_k = P_m(p^k)$ for all $k \geq 0$. This Q would extend the functional P_m (or equivalently, the functional \hat{P}_m) to the set $\mathcal{L}([0, 1])$ of all gambles on $[0, 1]$.

The following theorem provides an answer to this question by summarising a number of results from the literature. We give the proof, which is quite simple, in order to make the paper more self-contained.

Theorem 1. *The following statements are equivalent.*

1. *The functional P_m on $\mathcal{V}_p([0, 1])$ is a linear prevision.*
2. *The functional P_m can be extended to a linear prevision Q on $\mathcal{L}([0, 1])$.*
3. *The functional P_m can be extended uniquely to a linear prevision on $\mathcal{V}([0, 1])$, or equivalently, on the set $\mathcal{C}([0, 1])$ of all continuous³ gambles on $[0, 1]$. This unique extension*

³We shall always assume that $[0, 1]$ is endowed with the Euclidean topology, i.e., with the relativisation to $[0, 1]$ of the Euclidean topology on the set of real numbers \mathbb{R} .

to $\mathcal{V}([0, 1])$ is then actually \hat{P}_m . We shall also denote by \hat{P}_m the unique extension to $\mathcal{C}([0, 1])$.

4. For all polynomials $p = \sum_{k=0}^n a_k p^k$ in $\mathcal{V}([0, 1])$, it holds that

$$\min p \leq \hat{P}_m(p) = \sum_{k=0}^n a_k m_k \leq \max p.$$

5. $\hat{P}_m(1) = m_0 = 1$ and for all non-negative polynomials $p = \sum_{k=0}^n a_k p^k \geq 0$ in $\mathcal{V}([0, 1])$, it holds that

$$\hat{P}_m(p) = \sum_{k=0}^n a_k m_k \geq 0,$$

or, in other words, the linear functional \hat{P}_m is positive and normed on $\mathcal{V}([0, 1])$ ([22]).

6. The moment sequence m satisfies the Hausdorff moment condition [13, 14]: $m_0 = 1$ and m is completely monotone, meaning that $(-1)^n \Delta^n m_k \geq 0$ for all $k \geq 0$ and $n \geq 0$.

Proof. A real functional on some domain is a linear prevision if and only if it can be extended to a linear prevision on all gambles. By also observing the following, we see at once that the first three statements are equivalent. Let Q be any extension of P_m to a linear prevision on all gambles on $[0, 1]$. Then for any polynomial $p = \sum_{k=0}^n a_k p^k$ in $\mathcal{V}([0, 1])$, it follows from the linearity of Q that $Q(p) = \sum_{k=0}^n a_k Q(p^k) = \sum_{k=0}^n a_k m_k = \hat{P}_m(p)$, since Q and P_m coincide on $\mathcal{V}_p([0, 1])$. So Q coincides with \hat{P}_m on $\mathcal{V}([0, 1])$. By virtue of the Stone–Weierstraß Theorem, every continuous function h on $[0, 1]$ is a uniform limit of some sequence of polynomials p_n on $[0, 1]$. Then, since any linear prevision Q on $\mathcal{L}([0, 1])$ is in particular uniformly continuous, we see that $Q(h) = \lim_{n \rightarrow \infty} Q(p_n) = \lim_{n \rightarrow \infty} \hat{P}_m(p_n)$, where the last equality follows from the previously established fact that Q and \hat{P}_m must coincide on polynomials. This shows that (i) the value of $Q(h)$ must be independent of the actual choice of the sequence $p_n \rightarrow h$, and (ii) that all Q have the same value on h .

We now give a circular proof for the equivalence of these statements with the rest. To see that the third statement implies the fourth, recall that \hat{P}_m is a linear prevision on $\mathcal{V}([0, 1])$, and therefore satisfies $\min p \leq \hat{P}_m(p) \leq \max p$ for any p in $\mathcal{V}([0, 1])$. We see at once that the fourth statement implies the fifth. We now prove that the fifth statement implies the sixth. We already know that $m_0 = 1$, so consider, for any non-negative integers k and n , the non-negative polynomial $q(x) = x^k(1-x)^n$. It follows from the Binomial Theorem that $q(x) = \sum_{\ell=0}^n (-1)^\ell \binom{n}{\ell} x^{k+\ell}$ and therefore the assumption implies that $\sum_{\ell=0}^n (-1)^\ell \binom{n}{\ell} m_{k+\ell} = \hat{P}_m(q) \geq 0$. Now observe that the left-hand side of the above equality is equal to $(-1)^n \Delta^n m_k$. We complete the proof by showing that the last statement implies the third. Recall that \hat{P}_m is the only linear functional that extends P_m from $\mathcal{V}_p([0, 1])$ to $\mathcal{V}([0, 1])$. So in order to prove that \hat{P}_m is a linear prevision, it remains to show that \hat{P}_m satisfies $\hat{P}_m(p) \geq \min p$ for all polynomials p in $\mathcal{V}([0, 1])$. But we have by assumption that $\hat{P}_m(\min p) = \hat{P}_m(1) \min p = m_0 \min p = \min p$, so $\hat{P}_m(p) = \hat{P}_m(p - \min p) + \hat{P}_m(\min p) = \hat{P}_m(p - \min p) + \min p$ so we only have to prove that \hat{P}_m is positive.⁴ Consider any non-negative polynomial p , and the associated Bernstein polynomial $B_{n,p}$ of degree n , given by $B_{n,p}(x) = \sum_{k=0}^n p \binom{n}{k} \binom{n}{k} x^k (1-x)^{n-k}$ for all $x \in [0, 1]$. Since we have seen above that $\hat{P}_m(x^k(1-x)^{n-k}) = (-1)^{n-k} \Delta^{n-k} m_k$, it follows from the Hausdorff moment condition

⁴An alternative proof runs as follows. In his 1921 paper [13, Satz III on pp. 98–99], Hausdorff shows using a beautifully simple argument that any non-negative polynomial on $[0, 1]$ can be written as a non-negative linear combination of Bernstein basis polynomials $x^k(1-x)^\ell$. Since by assumption, \hat{P}_m assumes a non-negative value on these basis polynomials (see above), it follows by linearity that \hat{P}_m assumes a non-negative value on all non-negative polynomials.

that $\hat{P}_m(B_{n,p}) = \sum_{k=0}^n p\left(\frac{k}{n}\right) \binom{n}{k} (-1)^{n-k} \Delta^{n-k} m_k \geq 0$, since also $p(k/n) \geq 0$. By Lemma 2, $\hat{P}_m(p) = \lim_{n \rightarrow \infty} \hat{P}_m(B_{n,p}) \geq 0$, which completes the proof. \square

Lemma 2. *For any polynomial p in $\mathcal{V}([0, 1])$, the sequence $\hat{P}_m(B_{n,p})$ converges to $\hat{P}_m(p)$.⁵*

Proof. Consider any polynomial $p = \sum_{j=0}^r a_j x^j$ of degree $r \geq 0$. Let us define the difference operator $\Delta_h p(x) = [p(x+h) - p(x)]/h$, and let Δ_h^k be the composition of Δ_h with itself k times. Then $\Delta_h^k p(x) := h^{-k} \sum_{\ell=0}^k \binom{k}{\ell} (-1)^{k-\ell} p(x+h\ell)$ for any $x \in [0, 1]$ and $h > 0$ such that $x+hk \leq 1$, and in particular $\Delta_h^0 p(x) = p(x)$ for any $x \in [0, 1]$.

Using a general reciprocity relation involving differences (see Feller's book [12, Section VII.1, Eq. (1.8)]), we can write

$$B_{n,p}(x) = \sum_{k=0}^n p\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k} \Delta_{\frac{1}{n}}^k p(0) x^k = \sum_{k=0}^n \frac{\binom{n}{k}}{n^k} \Delta_{\frac{1}{n}}^k p(0) \frac{x^k}{k!},$$

where $(n)_k := n!/(n-k)! = n(n-1)\dots(n-k+1)$.

Now use induction to show that

$$\sum_{\ell=0}^k \ell^j \binom{k}{\ell} (-1)^{k-\ell} = \begin{cases} 0 & \text{if } j < k \\ k! & \text{if } j = k. \end{cases}$$

From this, we can deduce that

$$\Delta_{\frac{1}{n}}^k p(0) = \begin{cases} 0 & \text{if } r < k \\ a_k k! + \sum_{j=k+1}^r \frac{a_j}{n^{j-k}} \sum_{\ell=0}^k \binom{k}{\ell} (-1)^{k-\ell} \ell^j & \text{if } r \geq k, \end{cases} \quad (1)$$

and consequently

$$\hat{P}_m(B_{n,p}) = \sum_{k=0}^n \frac{\binom{n}{k}}{n^k} \Delta_{\frac{1}{n}}^k p(0) \frac{m_k}{k!} = \sum_{k=0}^r \frac{\binom{n}{k}}{n^k} \frac{m_k}{k!} \Delta_{\frac{1}{n}}^k p(0).$$

Now, using Equation (1) we can prove that for $0 \leq k \leq r$, $\Delta_{\frac{1}{n}}^k p(0)$ converges uniformly to $k!a_k$. Since $\binom{n}{k}/n^k$ converges uniformly to 1, we find $\lim_{n \rightarrow \infty} \hat{P}_m(B_{n,p}) = \sum_{k=0}^r a_k m_k = \hat{P}_m(p)$. This implies our claim. \square

It is trivial that if there exists a σ -additive probability P_m^σ with a given sequence of moments m , there is also a linear prevision with these moments, because P_m^σ has a linear extension to all gambles. Interestingly, this proposition tells us that the converse is also true: the Hausdorff moment problem has a solution (as a linear prevision) if and only if the Hausdorff moment condition is satisfied.

Let us assume from now on that the moment sequence m satisfies the Hausdorff moment condition, implying that there are linear previsions with moment sequence m . If we invoke the F. Riesz Representation Theorem in the form mentioned by Feller (see for instance [12, Section V.1]), we see that the linear prevision \hat{P}_m on the set $\mathcal{C}([0, 1])$ of all continuous gambles on $[0, 1]$ can be extended *uniquely* to a σ -additive probability measure P_m^σ on the Borel sets of $[0, 1]$, and that there is a *unique* (right-continuous) distribution function F_m^σ on $[0, 1]$ such that for all continuous gambles h

$$E_m^\sigma(h) = (L) \int h dP_m^\sigma = (LS) \int h dF_m^\sigma = \hat{P}_m(h).$$

In this expression, the first integral is the Lebesgue integral associated with the probability measure P_m^σ , and the second integral the Lebesgue–Stieltjes integral associated with

⁵The essence of this lemma can be found in Feller's book [12, Section VII.3].

the distribution function F_m^σ . Also, $F_m^\sigma(x) = P_m^\sigma([0, x])$ for all x in $[0, 1]$, and the expectation operator E_m^σ , as well as both integrals are actually defined for all Borel-measurable functions on $[0, 1]$.

In this sense, the moments determine a unique σ -additive probability measure on the Borel sets. But the solution is not as clear-cut if we look for the finitely additive probabilities on all events, or, equivalently, the linear previsions on all gambles, that correspond to the given moments. These are given by the set

$$\mathcal{M}(m) := \mathcal{M}(\hat{P}_m) = \{Q \in \mathbb{P}([0, 1]) : (\forall h \in \mathcal{C}([0, 1]))(Q(h) \geq \hat{P}_m(h))\}$$

of all linear previsions Q that dominate, or equivalently, coincide with, \hat{P}_m on continuous gambles.

For any gamble h on $[0, 1]$, it follows that the linear previsions that solve the moment problem can (and will) assume any value in the real interval $[\underline{E}_m(h), \bar{E}_m(h)]$, where

$$\underline{E}_m(h) := \inf \{Q(h) : Q \in \mathcal{M}(m)\} \text{ and } \bar{E}_m(h) := \sup \{Q(h) : Q \in \mathcal{M}(m)\}.$$

The functional \underline{E}_m on $\mathcal{L}([0, 1])$ is called the *natural extension* of \hat{P}_m , or, because of the transitivity of natural extension, of P_m . We shall also call it the *natural extension of the moment sequence m* . It is the point-wise smallest coherent lower prevision that coincides with \hat{P}_m on $\mathcal{C}([0, 1])$, or equivalently that coincides with P_m on $\mathcal{V}_p([0, 1])$. The functional \bar{E}_m is its *conjugate upper prevision* and satisfies $\bar{E}_m(h) = -\underline{E}_m(-h)$ for all gambles h on $[0, 1]$.

By definition, \underline{E}_m is the smallest coherent lower prevision for which $\underline{E}_m(p^k) = \bar{E}_m(p^k) = m_k$, $k \geq 0$, i.e., that has (precise) moment sequence m , and a linear prevision dominates \underline{E}_m if and only if it dominates, and therefore coincides with, \hat{P}_m (or P_m). In other words, we get

$$\mathcal{M}(\underline{E}_m) = \mathcal{M}(\hat{P}_m) = \mathcal{M}(P_m) = \mathcal{M}(m).$$

So we see that the coherent lower prevision \underline{E}_m *completely determines* the solution to the Hausdorff moment problem for linear previsions: a linear prevision P has a sequence of moments m if and only if it dominates \underline{E}_m .

In particular, the gambles h on $[0, 1]$ where the lower and upper natural extensions coincide, i.e., for which $\underline{E}_m(h) = \bar{E}_m(h)$, are precisely those gambles to which P_m has a unique extension as a linear prevision. We shall call such gambles *m -integrable*. One of the goals in this paper is precisely to study these m -integrable gambles. Another, closely related, goal is to study the functional \underline{E}_m .

In particular, we can associate with \underline{E}_m and its conjugate \bar{E}_m a *lower distribution function* \underline{F}_m and an *upper distribution function* \bar{F}_m on $[0, 1]$, given by

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

for all $x \in [0, 1]$. Then we can ask ourselves: what are the properties of these functions, what is their relationship to F_m^σ , and to what extent do they determine the functional \underline{E}_m ? We give the answer to these questions in the following section.

In order to be able to place the discussion of later sections in the right perspective, it is useful, at this point, to make a small historical digression, and to try and reconstruct exactly how Hausdorff and F. Riesz contributed to the solution of what we now call the Hausdorff moment problem. In a famous series of papers [13, 14] published in 1921, Hausdorff showed that for a sequence m of real numbers m_k , $k \geq 0$, there is a non-decreasing function α on $[0, 1]$ such that for the associated Riemann–Stieltjes integral

$$(RS) \int_0^1 x^k d\alpha(x) = m_k, \quad k \geq 0,$$

if and only if the sequence m is completely monotone. Some twelve years before that, F. Riesz had shown in his Representation Theorem ([19], see also [20, Section 50]) that a functional L on the linear space $\mathcal{C}([0, 1])$ of continuous functions on $[0, 1]$ is a positive continuous linear functional⁶ (with respect to the supremum-norm topology) if and only if it can be represented as a Riemann–Stieltjes integral $(RS) \int_0^1 \cdot d\alpha(x)$ on $\mathcal{C}([0, 1])$ with respect to some non-decreasing function α , meaning that

$$L(h) = (RS) \int_0^1 h(x) d\alpha(x) \quad (2)$$

for all continuous gambles h on $[0, 1]$.

Since the Stone–Weierstraß theorem tells us that the set $\mathcal{V}([0, 1])$ of polynomials on $[0, 1]$ is uniformly dense in the set $\mathcal{C}([0, 1])$ of continuous functions, the combination of both results, taking into account issues of normalisation, shows that there is another statement that is equivalent to each of the six statements in Theorem 1, namely:

7. There is a non-decreasing function α on $[0, 1]$ satisfying $\alpha(0) = 0$ and $\alpha(1) = 1$ such that $\hat{P}_m(h) = (RS) \int_0^1 h(x) d\alpha(x)$ for all continuous gambles h on $[0, 1]$.

Since it is a property of Riemann–Stieltjes integrals that an identity such as Equation (2) determines α uniquely, except in its points of discontinuity,⁷ we see that there will be a unique right-continuous (distribution function) α that corresponds to a given Hausdorff moment sequence m , and this unique α can be used to construct the unique σ -additive probability measure with moment sequence m , using standard measure-theoretic methods. But we already observe at this point that there will usually be an infinity of distribution functions α that are not right-continuous and still generate \hat{P}_m , and that the Riemann–Stieltjes integral somehow will have an important part in describing the finitely additive probabilities with a given moment sequence. Indeed, with each solution α of Equation (2) there will correspond an extension of \hat{P}_m as a linear prevision on the set of all gambles that are Riemann–Stieltjes integrable with respect to α . But this set of integrable gambles will usually only be a proper subset of $\mathcal{L}([0, 1])$, and we want to look at extensions of the moment sequence to all gambles. That is why we do not take up the subject of Riemann–Stieltjes integrals and distribution functions here,⁸ but take a completely different route.

4. THE NATURAL EXTENSION \underline{E}_m AND m -INTEGRABLE GAMBLES

Since $\mathcal{C}([0, 1])$ is a linear subspace of $\mathcal{L}([0, 1])$ that contains all constant gambles, we may apply a basic result [23, Corollary 3.1.8] in the theory of coherent lower previsions to obtain the following expressions: for any gamble h on $[0, 1]$,⁹

$$\begin{aligned} \underline{E}_m(h) &= \sup \{ \hat{P}_m(g) : g \in \mathcal{C}([0, 1]), g \leq h \} \\ \bar{E}_m(h) &= \inf \{ \hat{P}_m(g) : g \in \mathcal{C}([0, 1]), h \leq g \}. \end{aligned} \quad (3)$$

⁶Actually, Riesz proved his Representation Theorem for continuous linear functionals L and functions α of bounded variation, which can generally be written as a differences of two non-decreasing functions. Similarly, Hausdorff proved his result for what he calls a “C-Folge”, which is the difference of completely monotone sequences, and functions α of bounded variation. Our formulation here is equivalent.

⁷And up to an additive (integration) constant, which is fixed here because of the normalisation constraints $\alpha(1) = 1$ and $\alpha(0) = 0$. See, for instance, [20, Section 51].

⁸We address the relation between finitely additive measures, distribution functions and moment sequences in a companion paper [18].

⁹We can also consider similar expressions where the approximation is made by polynomials, i.e., elements of $\mathcal{V}([0, 1])$. This is because \hat{P}_m has a unique extension as a linear prevision from $\mathcal{V}([0, 1])$ to $\mathcal{C}([0, 1])$, as we have proven in Theorem 1.

We use these expressions to prove a number of interesting properties of \underline{E}_m and \bar{E}_m , and the lower and upper distribution functions \underline{F}_m and \bar{F}_m . We collect these properties in Propositions 3 through 11.

4.1. Basic properties of \underline{E}_m , \bar{E}_m and \bar{F}_m .

Proposition 3. *Consider a Hausdorff moment sequence m . Then the following statements hold.*

1. $0 \leq \underline{E}_m \leq F_m^\sigma \leq \bar{F}_m \leq 1$.
2. \underline{E}_m and \bar{F}_m are non-decreasing functions.
3. $\underline{E}_m(0) = 0$ and $\underline{E}_m(1) = \bar{F}_m(1) = 1$.

Proof. The first statement follows from the definition of \underline{E}_m and \bar{F}_m , and from (i) $0 \leq I_{[0,x]} \leq 1$; (ii) the fact that the lower prevision \underline{E}_m is coherent, which implies that it is monotone, satisfies $\underline{E}_m(0) = \bar{E}_m(0) = 0$ and $\underline{E}_m(1) = \bar{E}_m(1) = 1$ and $\underline{E}_m \leq \bar{E}_m$; and (iii) that $P_m^\sigma \in \mathcal{M}(m) = \mathcal{M}(\underline{E}_m)$. The second statement follows from the fact that \underline{E}_m and \bar{E}_m are coherent and therefore monotone. The third statement can be proven as follows. First of all $I_{[0,1]} = 1$ and $\underline{E}_m(1) = \bar{E}_m(1) = 1$ by coherence of \underline{E}_m . Secondly, consider that $\underline{E}_m(0) = \underline{E}_m(I_{\{0\}})$ and that for any continuous h on $[0, 1]$, $h \leq I_{\{0\}}$ implies that $h \leq 0$. Since \hat{P}_m is a linear prevision, it follows that $\hat{P}_m(h) \leq 0$. Equation (3) then tells us that $\underline{E}_m(I_{\{0\}}) = 0$. \square

Much more interesting properties can be proven for \underline{F}_m and \bar{F}_m . But first, let us introduce some additional notation. For any function h on $[0, 1]$ and any $x \in [0, 1]$ let $h(x-) := \lim_{t \rightarrow x, t < x} h(t)$ denote the *left limit* of h in x (if it exists) when $x > 0$, and let $h(0-) := h(0)$. Similarly, let $h(x+) := \lim_{t \rightarrow x, t > x} h(t)$ denote the *right limit* of h in x (if it exists) when $x < 1$, and let $h(1+) := h(1)$.

It follows from the first two statements in Proposition 3 that the left and right limits of \underline{F}_m and \bar{F}_m exist everywhere.¹⁰ Let us denote the set of all points of discontinuity of \underline{F}_m by $\mathcal{D}_{\underline{F}_m} := \{x \in [0, 1] : \underline{F}_m(x+) \neq \underline{F}_m(x-)\}$ and the set of points where \bar{F}_m is not continuous by $\mathcal{D}_{\bar{F}_m} := \{x \in [0, 1] : \bar{F}_m(x-) \neq \bar{F}_m(x+)\}$. Let $\mathcal{D}_m := \mathcal{D}_{\underline{F}_m} \cup \mathcal{D}_{\bar{F}_m}$ denote their union. It follows from the non-decreasing character of \underline{F}_m and \bar{F}_m [Proposition 3] that $\mathcal{D}_{\underline{F}_m}$, $\mathcal{D}_{\bar{F}_m}$ and \mathcal{D}_m are countable subsets of $[0, 1]$.

Lemma 4. *Consider a Hausdorff moment sequence m . Then for all $x \in [0, 1]$, $\underline{F}_m(x+) \geq \bar{F}_m(x)$. As a consequence, if $x \notin \mathcal{D}_m$ then $\underline{F}_m(x) = \bar{F}_m(x)$.*

Proof. The inequality holds trivially for $x = 1$, so let us now consider $x \in [0, 1)$. Let $0 < \varepsilon < 1 - x$, and define the gamble f_ε on $[0, 1]$ by

$$f_\varepsilon(t) = \begin{cases} 1 + \frac{\varepsilon}{2} & \text{if } t \in [0, x] \\ 1 + \frac{\varepsilon}{2} + \frac{x-t}{\varepsilon} & \text{if } t \in [x, x + \varepsilon] \\ \frac{\varepsilon}{2} & \text{if } t \in [x + \varepsilon, 1]. \end{cases}$$

This is a continuous gamble satisfying $f_\varepsilon > I_{[0,x]}$ and $f_\varepsilon - \varepsilon < I_{[0,x+\varepsilon]}$. Hence, it follows from Equation (3) and the fact that \hat{P}_m is a linear prevision on $\mathcal{C}([0, 1])$ that

$$\bar{F}_m(x) = \bar{E}_m([0, x]) \leq \hat{P}_m(f_\varepsilon) = \hat{P}_m(f_\varepsilon - \varepsilon) + \varepsilon \leq \underline{E}_m([0, x + \varepsilon]) + \varepsilon = \underline{F}_m(x + \varepsilon) + \varepsilon.$$

If we now let $\varepsilon \rightarrow 0$, we obtain the desired inequality.

The second part follows from the definition of \mathcal{D}_m and Proposition 3. \square

¹⁰This is because the limits of bounded non-decreasing or non-increasing sequences always exist.

Proposition 5. *Consider a Hausdorff moment sequence m . Then the following statements hold.*

1. For any $x \in [0, 1]$, $\underline{F}_m(x+) = \overline{F}_m(x) = \overline{F}_m(x+)$, so \overline{F}_m is right-continuous on $[0, 1]$.
2. For any $x \in (0, 1)$, $\underline{F}_m(x-) = \underline{F}_m(x) = \overline{F}_m(x-)$, so \underline{F}_m is left-continuous on $[0, 1]$.
3. $\underline{F}_m(0-) = \underline{F}_m(0) = 0 \leq \overline{F}_m(0-) = \overline{F}_m(0)$.
4. $\underline{F}_m(1-) = \overline{F}_m(1-) \leq \underline{F}_m(1) = \overline{F}_m(1) = 1$.
5. $F_m^\sigma = \overline{F}_m$.
6. $\mathcal{D}_{\underline{F}_m} \cap (0, 1] = \mathcal{D}_{\overline{F}_m} \cap (0, 1]$, whence $\mathcal{D}_m = \mathcal{D}_{\underline{F}_m}$.
7. The following statements are equivalent for all $x \in (0, 1)$: (i) $x \notin \mathcal{D}_m$; (ii) $\underline{F}_m(x) = \overline{F}_m(x)$; and (iii) F_m^σ is continuous in x .
8. $\underline{F}_m(0) = \overline{F}_m(0) \Leftrightarrow 0 \notin \mathcal{D}_m \Rightarrow F_m^\sigma$ is (right-)continuous in 0.
9. F_m^σ is (left-)continuous in 1 $\Leftrightarrow 1 \notin \mathcal{D}_m \Rightarrow \underline{F}_m(1) = \overline{F}_m(1)$.

Proof. Let us begin with the first statement. Consider $x \in [0, 1]$. By virtue of Proposition 3 and Lemma 4, it suffices to prove that $\overline{F}_m(x) = \overline{F}_m(x+)$. The equality holds by definition for $x = 1$. Consider then $x \in [0, 1)$. Then by Equation (3), for any $\varepsilon > 0$ there is some continuous gamble $h_\varepsilon \geq I_{[0, x]}$ such that $\hat{P}_m(h_\varepsilon) < \overline{F}_m(x) + \varepsilon$. Given $\eta > 0$, we have $h_\varepsilon(x) + \eta \geq 1 + \eta$, so by continuity of h_ε there is some $0 < \delta_{\varepsilon\eta} < 1 - x$ such that $h_\varepsilon(t) + \eta \geq 1$ for all $x \leq t \leq x + \delta_{\varepsilon\eta}$. This tells us that $h_\varepsilon + \eta \geq I_{[0, x + \delta_{\varepsilon\eta}]}$ and therefore $\overline{F}_m(x + \delta_{\varepsilon\eta}) \leq \hat{P}_m(h_\varepsilon + \eta) = \hat{P}_m(h_\varepsilon) + \eta < \overline{F}_m(x) + \varepsilon + \eta$. From the non-decreasing character of \overline{F}_m [Proposition 3] it then follows that $\overline{F}_m(x+) \leq \overline{F}_m(x + \delta_{\varepsilon\eta}) < \overline{F}_m(x) + \varepsilon + \eta$. Since this holds for any $\varepsilon, \eta > 0$, we deduce that $\overline{F}_m(x+) \leq \overline{F}_m(x)$. The converse inequality follows from the non-decreasing character of \overline{F}_m .

We now turn to the second statement. Take $x \in (0, 1]$. Then, $\underline{F}_m(x-) = \sup_{t < x} \underline{F}_m(t) = \sup_{t < x, t \notin \mathcal{D}_m} \underline{F}_m(t) = \sup_{t < x, t \notin \mathcal{D}_m} \overline{F}_m(t) = \overline{F}_m(x-)$, where the second and fourth equalities hold because \mathcal{D}_m is countable and the third follows from Lemma 4. On the other hand, a reasoning similar to the one in the first statement allows us to deduce that $\underline{F}_m(x-) = \underline{F}_m(x)$ when $x \in (0, 1)$.

The third statement follows from the definition of $\underline{F}_m(0-)$ and $\overline{F}_m(0-)$ and Proposition 3.

For the fourth statement, taking into account Proposition 3, it suffices to show that $\underline{F}_m(1-) = \overline{F}_m(1-)$. The proof of this fact has already been given above as an extra in the proof of statement 2.

The fifth statement follows from the first, Proposition 3 and the right-continuity of F_m^σ .

The first part of the sixth statement is a consequence of the first, second and fourth statements. For the second part, note that $\overline{F}_m(0-) = \overline{F}_m(0) = \overline{F}_m(0+)$ from the first and third statements, so 0 does not belong to $\mathcal{D}_{\overline{F}_m}$. Hence, $\mathcal{D}_m = (\mathcal{D}_m \cap (0, 1]) \cup (\mathcal{D}_m \cap \{0\}) = (\mathcal{D}_{\underline{F}_m} \cap (0, 1]) \cup (\mathcal{D}_{\underline{F}_m} \cap \{0\}) = \mathcal{D}_{\underline{F}_m}$.

Let us now prove the seventh statement. The equivalence between (i) and (ii) follows from the first two statements. On the other hand, the equivalence between (ii) and (iii) follows from the second and the fifth statements.

The equivalence in the eighth statement follows from the first and third statements. The implication is trivial: F_m^σ is always continuous in 0 because it is right-continuous everywhere.

Finally, the equivalence in the ninth statement follows from the fourth and fifth statements. For the implication, note that we always have $\underline{F}_m(1) = \overline{F}_m(1) = 1$. \square

We see that, if m is a sequence satisfying the Hausdorff moment condition, the distribution function F_m^σ of the unique σ -additive probability P_m^σ with these moments is equal to

the upper distribution function \bar{F}_m . This also allows us to determine \underline{E}_m and \bar{F}_m directly from the moment sequence m . Indeed, it is established in the literature [12, Section VII.3, Eq. (3.11)] that

$$F_m^\sigma(x) = \lim_{n \rightarrow \infty} \sum_{k \leq nx} \binom{n}{k} (-1)^{n-k} \Delta^{n-k} m_k$$

in all points of (left-)continuity x of F_m^σ . Since we know that F_m^σ is right-continuous, this completely determines F_m^σ , and therefore also \bar{F}_m and \underline{E}_m through Proposition 5. The following example adds further detail to the picture drawn in this proposition.

Example 1. Let us consider the moment sequence m given by $m_k = 1$ for $k \geq 0$. It satisfies the Hausdorff moment conditions since all non-trivial differences are zero. Observe that the σ -additive probability measure all of whose probability mass is concentrated in 1 is the unique σ -additive probability measure with this moment sequence. As a result, $\bar{E}_m(\{1\}) = 1$. It follows from conjugacy that $\underline{E}_m([0, 1]) = 1 - \bar{E}_m(\{1\}) = 0$, and as a consequence $\underline{E}_m(1-) = \sup_{t < 1} \underline{E}_m(t) = 0$. Hence, the inequality in the fourth statement of Proposition 5 may be strict. Moreover, 1 belongs to \mathcal{D}_m even though $\underline{E}_m(1) = \bar{F}_m(1) = 1$. Hence, the converse implication in the ninth statement of Proposition 5 need not hold.

Similarly, let us consider the moment sequence m given by $m_0 = 1$ and $m_k = 0$ for $k > 0$. It also satisfies the Hausdorff moment condition, and the σ -additive probability measure all of whose probability mass is concentrated in 0 is the unique σ -additive probability measure with this moment sequence. As a consequence, we see that $\bar{F}_m(0) = 1 > \underline{E}_m(0) = 0$. Hence, the inequality in the third statement of Proposition 5 may be strict, and the sixth statement cannot be extended to $\mathcal{D}_{\underline{E}_m} = \mathcal{D}_{\bar{F}_m}$. On the other hand, the distribution function F_m^σ of the σ -additive probability measure with these moments is constant on $[0, 1]$ (equal to one), and therefore continuous on $[0, 1]$, but still $0 \in \mathcal{D}_m$. This shows that the converse implication in the eighth statement of Proposition 5 need not hold. \blacklozenge

Let us now draw inspiration from the expressions in Equation (3) and define, for any gamble h on $[0, 1]$, the gambles h^\downarrow and h^\uparrow on $[0, 1]$ as follows: for all x in $[0, 1]$,

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

Observe that in particular for any $A \subseteq [0, 1]$, $(I_A)^\uparrow = I_{\text{int}(A)}$ and $(I_A)^\downarrow = I_{\text{cl}(A)}$, where $\text{int}(A)$ is the topological interior of A , and $\text{cl}(A)$ its topological closure. The gamble h^\uparrow is lower semi-continuous¹¹ (it is a point-wise supremum of continuous functions), and h^\downarrow is upper semi-continuous. Moreover, these gambles satisfy the following:

Proposition 6. *Consider a Hausdorff moment sequence m . Then for any gamble h on $[0, 1]$ we have that $\underline{E}_m(h) = \underline{E}_m(h^\uparrow)$ and $\bar{E}_m(h) = \bar{E}_m(h^\downarrow)$. In particular, for any event $A \subseteq [0, 1]$, $\underline{E}_m(A) = \underline{E}_m(\text{int}(A))$ and $\bar{E}_m(A) = \bar{E}_m(\text{cl}(A))$.*

Proof. Simply observe that for any $g \in \mathcal{C}([0, 1])$ such that $g \leq h$, it holds that $g \leq h^\uparrow \leq h$. The monotonicity [coherence] of \underline{E}_m then implies that $\hat{P}_m(g) = \underline{E}_m(g) \leq \underline{E}_m(h^\uparrow) \leq \underline{E}_m(h)$. Taking the supremum of the left-hand side of this inequality yields $\underline{E}_m(h) \leq \underline{E}_m(h^\uparrow) \leq \underline{E}_m(h)$, taking into account Equation (3). Then also $\underline{E}_m(A) = \underline{E}_m(I_A) = \underline{E}_m((I_A)^\uparrow) =$

¹¹A gamble f on $[0, 1]$ is lower semi-continuous if its strict cut sets $\{f > t\}$ are open for all real t . It is upper semi-continuous if its cut sets $\{f \geq t\}$ are closed for all real t , or equivalently, if $-f$ is lower semi-continuous. A gamble is continuous if and only if it is both lower and upper semi-continuous. For more details, see, for instance, [21, Section 15.22].

$\underline{E}_m(I_{\text{int}(A)}) = \underline{E}_m(\text{int}(A))$. The equalities involving \overline{E}_m can be proven in a completely similar way, or using conjugacy. \square

So in order to know the values of \underline{E}_m on events, it is sufficient to know its values on all open subsets of $[0, 1]$. Let us denote by \mathcal{S} the class of these open subsets.¹²

Now consider, for any set A its topological interior $\text{int}(A)$. Let, for any $x \in \text{int}(A)$, $I(x)$ be the maximal interval that contains x and is included in $\text{int}(A)$. Since $\text{int}(A)$ is open, we know that $I(x) \in \mathcal{S}$ and that it has (strictly) positive length. Moreover, let us define $\mathcal{S}(A) = \{I(x) : x \in \text{int}(A)\}$ as the collection of all maximal intervals of $\text{int}(A)$. All these maximal intervals are furthermore pairwise-disjoint (otherwise they would not be maximal), so they can only be countable in number. Moreover, $\text{int}(A) = \bigcup_{I \in \mathcal{S}(A)} I$, so we see that if we know the values that \underline{E}_m takes on *countable* unions of disjoint open intervals, we know the value of \underline{E}_m for any event.

The important result in Proposition 9 below tells us that it even suffices to know the values of \underline{E}_m on *finite* unions of disjoint open intervals. In order to prove this, we first need to establish a couple of simple lemmas.

Lemma 7. *Let D be a subset of $[0, 1]$ and let f be a continuous gamble on $[0, 1]$ such that $f \leq I_D$. Then there is some non-negative continuous gamble g on $[0, 1]$ such that $f \leq g \leq I_D$.*

Proof. Indeed, let the function g on $[0, 1]$ be such that $g(x) = \max\{f(x), 0\}$ for all $x \in [0, 1]$, then g is continuous, non-negative and it dominates f . Moreover, since I_D dominates both 0 and f , we deduce that $g \leq I_D$. \square

Lemma 8. *Let C and D be two disjoint open subsets of $[0, 1]$ and let f be a continuous gamble on $[0, 1]$ such that $0 \leq f \leq I_{C \cup D}$. Then $f_C := fI_C$ and $f_D := fI_D$ are continuous gambles on $[0, 1]$ and $f_C + f_D = f$.*

Proof. First of all, it is clear that $I_{C \cup D} = I_C + I_D$ since C and D are disjoint. Therefore $f_C + f_D = fI_{C \cup D}$. If $x \in (C \cup D)^c$ [superscript c denotes complement] then $f(x) = 0$ by assumption, so $f = fI_{C \cup D}$. This tells us that $f_C + f_D = f$. Now, $f_C = \min\{f, I_C\}$ is the minimum of two lower semi-continuous functions, which is again lower semi-continuous. In a similar way, we see that f_D is lower semi-continuous, so $-f_D$ is upper semi-continuous, and because the sum of two upper semi-continuous functions is upper semi-continuous, we see that $f_C = f + (-f_D)$ is upper semi-continuous as well, and therefore continuous. Similarly for f_D . \square

Proposition 9. *Let m be a Hausdorff moment sequence. Let B be a countable union of disjoint open subsets of $[0, 1]$, $B_n, n \geq 0$. Then $\underline{E}_m(B) = \sup_{n \geq 0} \underline{E}_m(\bigcup_{k=0}^n B_k)$.*

Proof. Consider $\varepsilon > 0$, then it follows from Equation (3) that there is some continuous gamble g_ε on $[0, 1]$ satisfying $g_\varepsilon \leq I_B$ and $\underline{E}_m(B) - \hat{P}_m(g_\varepsilon) < \frac{\varepsilon}{2}$. By Lemma 7, there is some continuous non-negative gamble f_ε on $[0, 1]$ satisfying $g_\varepsilon \leq f_\varepsilon \leq I_B$ and therefore, using the monotonicity of \hat{P}_m , $\underline{E}_m(B) - \hat{P}_m(f_\varepsilon) < \frac{\varepsilon}{2}$. Let us define $f_{\varepsilon, n} := f_\varepsilon I_{\bigcup_{k=0}^n B_k}$ for any $n \geq 0$. The sequence $f_{\varepsilon, n}, n \geq 0$ satisfies the following properties.

- (i) It converges point-wise to f_ε : if $x \in B$, there is some n such that $x \in B_n$, whence $f_{\varepsilon, \ell}(x) = f_\varepsilon(x)$ for all $\ell \geq n$. On the other hand, if $x \notin B$, we have $f_\varepsilon(x) = 0$, and then $f_{\varepsilon, \ell}(x) = f_\varepsilon(x) = 0$ for any $\ell \geq 0$.
- (ii) It is non-decreasing: $f_{\varepsilon, n} \leq f_{\varepsilon, n+1}$ for any $n \geq 0$.

¹²By an *open interval* we shall mean a subinterval of $[0, 1]$ that is open in the relativisation to $[0, 1]$ of the Euclidean topology on the reals \mathbb{R} , 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 $(y, 1]$.

- (iii) $f_{\varepsilon,n}$ is continuous for any n . This follows at once from Lemma 8 if we let $f = f_\varepsilon$, $C = \bigcup_{k=0}^n B_k$ (an open set), $D = \bigcup_{k>n} B_k$ (an open set), so $C \cup D = B$, $0 \leq f \leq I_B$ and $f_{\varepsilon,n} = f_C$.

Applying Dini's Convergence Theorem (see for instance [21, Section 17.7]), we conclude that the sequence $f_{\varepsilon,n}$, $n \geq 0$ converges uniformly to f_ε . Since the linear prevision \hat{P}_m is uniformly continuous and monotone, it follows that $\lim_{n \rightarrow \infty} \hat{P}_m(f_{\varepsilon,n}) = \sup_n \hat{P}_m(f_{\varepsilon,n}) = \hat{P}_m(f_\varepsilon)$, and as a consequence there is some $n_\varepsilon \geq 0$ such that $\hat{P}_m(f_\varepsilon) - \hat{P}_m(f_{\varepsilon,n}) < \frac{\varepsilon}{2}$ for all $n \geq n_\varepsilon$. Since $f_{\varepsilon,n} \leq I_{\bigcup_{k=0}^n B_k}$ for all $n \geq 0$, we deduce from the monotonicity [due to coherence] of \underline{E}_m that $\hat{P}_m(f_{\varepsilon,n}) = \underline{E}_m(f_{\varepsilon,n}) \leq \underline{E}_m(\bigcup_{k=0}^n B_k)$ for all $n \geq 0$, and so for all $n \geq n_\varepsilon$ we have that

$$\begin{aligned} \underline{E}_m(B) - \underline{E}_m\left(\bigcup_{k=0}^n B_k\right) &= \underline{E}_m(B) - \hat{P}_m(f_\varepsilon) + \hat{P}_m(f_\varepsilon) - \hat{P}_m(f_{\varepsilon,n}) \\ &\quad + \hat{P}_m(f_{\varepsilon,n}) - \underline{E}_m\left(\bigcup_{k=0}^n B_k\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} + 0 = \varepsilon. \end{aligned}$$

This means that $\underline{E}_m(B) = \sup_{n \geq 0} \underline{E}_m(\bigcup_{k=0}^n B_k)$. \square

We can go still one step further if we recognise that \underline{E}_m is additive on finite unions of disjoint open sets, so that \underline{E}_m is actually 'σ-additive' on countable unions of disjoint open sets.

Proposition 10. *Consider a Hausdorff moment sequence m . Then for any disjoint B_1, \dots, B_n in \mathcal{I} , $\underline{E}_m(\bigcup_{k=1}^n B_k) = \sum_{k=1}^n \underline{E}_m(B_k)$.*

Proof. First of all, $\underline{E}_m(\bigcup_{k=1}^n B_k) \geq \sum_{k=1}^n \underline{E}_m(B_k)$ because \underline{E}_m is super-additive [since it is coherent]. To prove the converse inequality, consider any continuous gamble $g \leq I_{\bigcup_{k=1}^n B_k}$ on $[0, 1]$. By Lemma 7 there is some non-negative continuous gamble f on $[0, 1]$ such that $g \leq f \leq I_{\bigcup_{k=1}^n B_k}$. Then by repeatedly applying Lemma 8, we find that $f = \sum_{k=1}^n f I_{B_k}$ and that $f I_{B_k}$ is continuous and dominated by I_{B_k} for $k = 1, \dots, n$. Therefore $\hat{P}_m(g) \leq \hat{P}_m(f) = \sum_{k=1}^n \hat{P}_m(f I_{B_k}) = \sum_{k=1}^n \underline{E}_m(f I_{B_k}) \leq \sum_{k=1}^n \underline{E}_m(B_k)$ [since \hat{P}_m is monotone and additive on $\mathcal{C}([0, 1])$, and \underline{E}_m is monotone], and consequently $\underline{E}_m(\bigcup_{k=1}^n B_k) \leq \sum_{k=1}^n \underline{E}_m(B_k)$, also using Equation (3). \square

Summarising, we find that for any subset A of $[0, 1]$,

$$\underline{E}_m(A) = \underline{E}_m(\text{int}(A)) = \sum_{I \in \mathcal{I}(A)} \underline{E}_m(I),$$

so \underline{E}_m is completely determined on events if we know its values on all open intervals. The following proposition establishes amongst other things that these values on open intervals are determined by the lower and upper distribution functions \underline{F}_m and \overline{F}_m .

Proposition 11. *Consider a Hausdorff moment sequence m . Then the following statements hold.*

1. $\underline{E}_m([0, x]) = \underline{E}_m((0, x)) = \underline{F}_m(x)$ for all $x \in [0, 1)$, and $\underline{E}_m([0, 1]) = \underline{F}_m(1)$.
2. $\underline{E}_m((0, x]) = \underline{E}_m((0, x)) = \underline{F}_m(x) - \overline{F}_m(0)$ for all $x \in (0, 1)$.
3. $\overline{E}_m([x, y]) = \overline{F}_m(y) - \underline{F}_m(x)$ for any $0 \leq x \leq y \leq 1$, and $\overline{E}_m([x, y]) = \overline{E}_m((x, y)) = \overline{F}_m(y) - \underline{F}_m(x)$ for any $0 \leq x < y \leq 1$.
4. $\underline{E}_m((x, y]) = \underline{F}_m(y) - \overline{F}_m(x)$ for any $0 \leq x < y \leq 1$, and $\underline{E}_m((x, y]) = \underline{E}_m((x, y)) = \underline{F}_m(y) - \overline{F}_m(x)$ for all $0 < x < y < 1$.
5. $\underline{E}_m((x, 1]) = 1 - \overline{F}_m(x)$ for any $x \in [0, 1)$, and $\underline{E}_m((x, 1]) = \underline{E}_m([x, 1])$ for any $x \in (0, 1]$.

6. $\underline{E}_m((x, 1)) = \underline{E}_m(1-) - \overline{F}_m(x)$ for any $x \in [0, 1]$, and
 $\underline{E}_m([x, 1)) = \underline{E}_m((x, 1))$ for any $x \in (0, 1]$.
7. $\underline{E}_m(\{x\}) = 0$ for any $x \in [0, 1]$.
8. $\underline{E}_m([0, 1]) = 1$.

Proof. We begin with the first statement. The first equality is a consequence of Proposition 6, and the fact that $\text{int}([0, x]) = [0, x)$ for any $x \in [0, 1]$. The second equality follows from the definition of \underline{F}_m . For the second part, note that $\underline{E}_m([0, 1]) \geq \sup_{x < 1} \underline{E}_m(x) = \underline{E}_m(1-)$. Conversely consider a continuous gamble g that satisfies $g \leq I_{[0, 1]}$. Since $g(1) \leq 0$, we have for any $\varepsilon > 0$ that $g(1) - \varepsilon < 0$, and since the gamble $g - \varepsilon$ is continuous, there is some $\delta > 0$ such that $g(x) - \varepsilon \leq 0$ for any $1 - \delta \leq x \leq 1$. Hence $g - \varepsilon \leq I_{[0, 1 - \delta]}$ and therefore

$$\hat{P}_m(g) - \varepsilon = \hat{P}_m(g - \varepsilon) \leq \underline{E}_m([0, 1 - \delta]) = \underline{E}_m(1 - \delta) \leq \underline{E}_m(1-).$$

It follows that $\hat{P}_m(g) \leq \underline{E}_m(1-)$, and Equation (3) now implies that $\underline{E}_m([0, 1]) \leq \underline{E}_m(1-)$.

Let us now prove the third statement. Consider $x \leq y$ in $[0, 1]$. Then

$$\begin{aligned} \overline{E}_m([x, y]) &= 1 - \underline{E}_m([0, x] \cup (y, 1]) = 1 - \underline{E}_m([0, x]) - \underline{E}_m((y, 1]) \\ &= \overline{E}_m([0, y]) - \underline{E}_m([0, x]) = \overline{F}_m(y) - \underline{E}_m(x-), \end{aligned}$$

where we used conjugacy, Propositions 5 and 10 and the first statement. The other equalities in this statement follow from Proposition 6.

We turn to the fourth statement. From Proposition 6 we have the equalities $\underline{E}_m([x, y]) = \underline{E}_m((x, y]) = \underline{E}_m([x, y)) = \underline{E}_m((x, y))$ for $x \leq y$ in $(0, 1)$. Consider then $x < y$ in $[0, 1]$. Since $I_{(x, y)} = I_{[0, y]} - I_{[0, x]}$ if follows from the super-additivity [due to coherence] of \underline{E}_m that

$$\begin{aligned} \underline{E}_m((x, y)) &= \underline{E}_m(I_{(x, y)}) \geq \underline{E}_m(I_{[0, y]}) + \underline{E}_m(-I_{[0, x]}) \\ &= \underline{E}_m(I_{[0, y]}) - \overline{E}_m(I_{[0, x]}) \\ &= \underline{E}_m([0, y]) - \overline{E}_m([0, x]) = \underline{E}_m(y-) - \overline{F}_m(x), \end{aligned} \quad (4)$$

where the last equality follows from the first statement. Now let f be any continuous gamble on $[0, 1]$ dominated by $I_{(x, y)}$. Then since $f(x) \leq 0$ and $f(y) \leq 0$, we have for any $\varepsilon > 0$ that $f(x) - \varepsilon < 0$ and $f(y) - \varepsilon < 0$. Since the gamble $f - \varepsilon$ is continuous, there is some $0 < \delta < \frac{y-x}{2}$ such that $f(t) - \varepsilon \leq 0$ for any t outside the non-empty open interval $(x + \delta, y - \delta)$, so $f - \varepsilon \leq I_{(x + \delta, y - \delta)}$. So we get

$$\begin{aligned} \hat{P}_m(f) - \varepsilon &= \hat{P}_m(f - \varepsilon) \leq \underline{E}_m((x + \delta, y - \delta)) \\ &\leq \overline{E}_m((x + \delta, y - \delta)) \\ &= \overline{F}_m(y - \delta) - \underline{E}_m(x + \delta) \\ &\leq \overline{F}_m(y-) - \underline{E}_m(x+) = \underline{E}_m(y-) - \overline{F}_m(x), \end{aligned}$$

where we have used Equation (3), the third statement, the non-decreasing character of \underline{E}_m and \overline{F}_m [Proposition 3], and Proposition 5. This implies that $\hat{P}_m(f) \leq \underline{E}_m(y-) - \overline{F}_m(x)$, so $\underline{E}_m((x, y)) \leq \underline{E}_m(y-) - \overline{F}_m(x)$ by Equation (3). Combined with Equation (4), this yields the desired equality.

The first equality in the second statement follows from Proposition 6. The second follows from the fourth statement.

The first part of the fifth statement follows by conjugacy, and the second part from Proposition 6.

The sixth statement follows from the fourth statement and Proposition 6.

The seventh statement follows from Proposition 6, and the last statement has been proven in Proposition 3. \square

In what follows, it will be useful to be able to characterise which intervals I are m -integrable, meaning that $\underline{E}_m(I) = \overline{E}_m(I)$. The following proposition gives a surprisingly simple characterisation: the m -integrability of an interval I is completely determined by whether its endpoints belong to \mathcal{D}_m .

Proposition 12. *Consider a Hausdorff moment sequence m . Let I be a non-empty interval in $[0, 1]$. We have the following possibilities.*

1. *If $0 \in I$ and $1 \in I$ (whence $I = [0, 1]$), then I is m -integrable.*
2. *If $0 \in I$ and $1 \notin I$, then I is m -integrable if and only if $\sup I \notin \mathcal{D}_m$.*
3. *If $0 \notin I$ and $1 \in I$, then I is m -integrable if and only if $\inf I \notin \mathcal{D}_m$.*
4. *If $0 \notin I$ and $1 \notin I$, then I is m -integrable if and only if $\inf I, \sup I \notin \mathcal{D}_m$.*

Proof. The first statement is obvious. For the second, assume that $I = (0, x)$ for $x \in (0, 1]$, or $I = [0, x]$ for $x \in [0, 1)$. Then the first and third statements in Proposition 11 imply that I is m -integrable if and only if $\overline{F}_m(x) = \underline{F}_m(x-)$, which by virtue of Proposition 5 holds if and only if $x \notin \mathcal{D}_m$.

Let us now turn to the third statement. Assume that $I = (x, 1]$ for $x \in [0, 1)$, or $I = [x, 1]$ for $x \in (0, 1]$. Then the third and fifth statements of Proposition 11 imply that I is m -integrable if and only if $\overline{F}_m(x) = \underline{F}_m(x-)$, which again is equivalent to $x \notin \mathcal{D}_m$.

Finally, if $0, 1 \notin I$, the second, third, fourth, sixth and seventh statements in Proposition 11 allow us to deduce that I is m -integrable if and only if $\overline{F}_m(\sup I) = \underline{F}_m(\sup I-)$ and $\overline{F}_m(\inf I) = \underline{F}_m(\inf I-)$. This again holds if and only if $\inf I, \sup I \notin \mathcal{D}_m$. \square

We have seen in Example 1 that the set \mathcal{D}_m may be non-empty (or, equivalently, that not all intervals are always m -integrable). Hence, the linear prevision that solves the Hausdorff moment problem may not be unique. Let us show next that there is actually *never* a unique linear prevision that solves the Hausdorff moment problem.

Remark 1 (The non-uniqueness of finitely additive probability measures with a given moment sequence). Consider the set $\mathbb{Q} \cap [0, 1]$ of all rational numbers between zero and one, then $\text{int}(\mathbb{Q} \cap [0, 1]) = \emptyset$ and $\text{cl}(\mathbb{Q} \cap [0, 1]) = [0, 1]$, so we infer from Proposition 6 that $\underline{E}_m(\mathbb{Q} \cap [0, 1]) = \underline{E}_m(\emptyset) = 0$ and $\overline{E}_m(\mathbb{Q} \cap [0, 1]) = \overline{E}_m([0, 1]) = 1$. So for any $0 \leq a \leq 1$ there is some linear prevision Q that solves the Hausdorff moment problem for which $Q(\mathbb{Q} \cap [0, 1]) = a$, and this holds for any moment sequence m that satisfies the Hausdorff moment condition! This shows that there is always either none or an uncountable infinity of linear previsions that produce a given sequence of moments $m_k, k \geq 0$.

We have said before that whenever there is a linear prevision with a particular sequence of moments m_k , there also is a unique σ -additive probability P_m^σ with these moments. We now see that, although the σ -additive probability will be unique, this will never be the case for probabilities that are only finitely additive. To put it differently, a linear prevision on $\mathcal{L}([0, 1])$ is *never* completely determined by its moments. We will comment further on this fact in the Conclusion. \blacklozenge

4.2. The complete monotonicity of \underline{E}_m and its immediate consequences. The values of \underline{E}_m on events are completely determined by the values that \underline{E}_m assumes on open intervals, and therefore, by Proposition 11, by its lower and upper distribution functions. There are two further questions we should still like to look at in the rest of this section.

1. Are the values of \underline{E}_m on events also completely determined by \underline{F}_m and \overline{F}_m in their points of continuity, or in other words, by F_m^σ in its points of continuity? By virtue of Propositions 11 and 12, this comes down to \underline{E}_m being determined by its values on *m-integrable open intervals*.
2. Can we say something similar about the values that \underline{E}_m assumes on gambles, and not just events?

We shall answer both questions positively in Theorem 14 further on.

But before we can address these issues, we need to prepare ourselves a bit better. In order to answer the first question, it will help us to consider the set of all *m-integrable open intervals*. By Proposition 12 this is the set

$$\{[0, 1]\} \cup \{[0, x) : x \notin \mathcal{D}_m\} \cup \{(y, 1] : y \notin \mathcal{D}_m\} \cup \{(x, y) : x, y \notin \mathcal{D}_m\}.$$

This set is closed under intersections, so the lattice of events \mathcal{O}_m generated by all *m-integrable open intervals* is the set made up of all finite unions of disjoint *m-integrable open intervals*. Let us denote by \tilde{P}_m the restriction of \underline{E}_m to \mathcal{O}_m . Then we know by Proposition 10 that \tilde{P}_m is additive on this lattice of events. Moreover, any non-trivial element O of \mathcal{O}_m has the form

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

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

$$\begin{aligned} \tilde{P}_m(O) &= \underline{E}_m(x_1) + \sum_{k=1}^{n-1} [\underline{E}_m(x_{2k+1}) - \overline{F}_m(x_{2k})] + 1 - \overline{F}_m(x_{2n}) \\ &= \underline{E}_m(O) = \overline{E}_m(O). \end{aligned} \quad (5)$$

Note that since $x_k \notin \mathcal{D}_m$ for all k , we have $\underline{E}_m(x_{2k-1}) = \underline{E}_m(x_{2k-1})$, and we could also replace \underline{E}_m by \overline{F}_m or vice versa in this equation.

In order to answer the second question, we point out that there exists, besides Equation (3), another representation of the natural extension \underline{E}_m to gambles, that is more informative, and provides additional insight about its nature. This representation derives from the fact that \underline{E}_m is the natural extension of the linear prevision \tilde{P}_m , which is in particular a completely monotone lower prevision. Moreover, the domain $\mathcal{C}([0, 1])$ of \tilde{P}_m is a linear lattice of gambles on $[0, 1]$ that contains all constant gambles. Hence, we can apply the results mentioned in Section 2 about this type of previsions. Let \mathcal{L}_m denote the class of *m-integrable gambles*

$$\mathcal{L}_m := \{h \in \mathcal{L}([0, 1]) : \underline{E}_m(h) = \overline{E}_m(h)\},$$

and let

$$\mathcal{A}_m := \{A \subseteq [0, 1] : \underline{E}_m(A) = \overline{E}_m(A)\} = \{A \subseteq [0, 1] : I_A \in \mathcal{L}_m\}$$

denote the class of *m-integrable events*, i.e., those events with *m-integrable indicators*. Then we have the following:

Theorem 13. *Consider a Hausdorff moment sequence m . Then the following statements hold.*

1. \underline{E}_m is the natural extension of its restriction to events, which is a completely monotone coherent lower probability.

2. \underline{E}_m is a completely monotone and comonotone additive coherent lower prevision on $\mathcal{L}([0, 1])$, and for all gambles h on $[0, 1]$,

$$\begin{aligned} \underline{E}_m(h) &= (C) \int h d\underline{E}_m := \inf h + (R) \int_{\inf h}^{\sup h} \underline{E}_m(\{h \geq t\}) dt \\ &= \inf h + (R) \int_{\inf h}^{\sup h} \underline{E}_m(\{h > t\}) dt, \end{aligned}$$

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.

3. \mathcal{L}_m is a uniformly closed linear lattice that contains all constant gambles.
4. \mathcal{A}_m is a field of subsets of $[0, 1]$ that includes \mathcal{O}_m .
5. A gamble f is m -integrable if and only if its cut sets $\{f \geq t\}$, or equivalently, its strict cut sets $\{f > t\}$, are m -integrable for all but a countable number of real numbers t .

We show next that \underline{E}_m is completely determined by the values it assumes on \mathcal{O}_m .

Theorem 14. Consider a Hausdorff moment sequence m . Let \tilde{P}_m denote the restriction of \underline{E}_m to the lattice of events \mathcal{O}_m generated by the m -integrable open intervals. Then the following statements hold.

1. The natural extension of \tilde{P}_m to all events is the inner set function $\tilde{P}_{m,*}$ of \tilde{P}_m . It is a completely monotone coherent lower probability.
2. For any $A \subseteq [0, 1]$, $\underline{E}_m(A) = \tilde{P}_{m,*}(A)$.
3. \underline{E}_m is the natural extension of \tilde{P}_m to all gambles and consequently $\underline{E}_m(h) = (C) \int h d\tilde{P}_{m,*}$ for any gamble h on $[0, 1]$.

Proof. Observe that \tilde{P}_m is a completely monotone coherent lower probability defined on a lattice of events \mathcal{O}_m that contains \emptyset and $[0, 1]$; this is because it is a restriction of the completely monotone and coherent lower prevision \underline{E}_m . Hence, the natural extension of \tilde{P}_m to all events is its inner set function $\tilde{P}_{m,*}$, and this inner set function is completely monotone as well. This proves the first statement.

To prove the second statement, first recall that $\tilde{P}_{m,*}$ is the natural extension of \tilde{P}_m , so it is the smallest coherent lower probability that coincides with \tilde{P}_m on \mathcal{O}_m . Since \underline{E}_m coincides with \tilde{P}_m on \mathcal{O}_m and is coherent, it must therefore dominate $\tilde{P}_{m,*}$ on all events. To prove that, conversely, $\tilde{P}_{m,*}$ dominates \underline{E}_m on events, consider any $A \subseteq [0, 1]$. Then by definition

$$\tilde{P}_{m,*}(A) = \sup \{ \tilde{P}_m(O) : O \in \mathcal{O}_m, O \subseteq A \}.$$

But since for any open subset O of $[0, 1]$, we have $O \subseteq A$ if and only if $O \subseteq \text{int}(A)$, we find that $\tilde{P}_{m,*}(A) = \tilde{P}_{m,*}(\text{int}(A))$. Since the same holds for \underline{E}_m [Proposition 6], we may assume without loss of generality that A is open. We have already argued that this implies that A is a countable union $\bigcup_{k \geq 0} O_k$ of disjoint open intervals O_k . We now show that $\tilde{P}_{m,*}$ dominates \underline{E}_m on open intervals. Indeed, given $x < y$ in $(0, 1)$,

$$\begin{aligned} \underline{E}_m((x, y)) &= \underline{E}_m(y-) - \bar{F}_m(x) = \sup_{x < t < z < y; t, z \notin \mathcal{D}_m} \underline{E}_m(z) - \bar{F}_m(t) \\ &= \sup_{x < t < z < y; t, z \notin \mathcal{D}_m} \tilde{P}_m((t, z)) \leq \tilde{P}_{m,*}((x, y)), \end{aligned}$$

using Propositions 5 and 11 and Equation (5), as well as the fact that \mathcal{D}_m is countable. Similarly, we can show that $\underline{E}_m([0, x]) \leq \tilde{P}_{m,*}([0, x])$ and that $\underline{E}_m((y, 1]) \leq \tilde{P}_{m,*}((y, 1])$ for any $x, y \in [0, 1]$. Note also that $\underline{E}_m([0, 1]) = 1 = \tilde{P}_m([0, 1])$ trivially.

Hence, by Propositions 9 and 10,

$$\begin{aligned} \underline{E}_m(A) &= \sup_{n \geq 0} E_m\left(\bigcup_{k=0}^n O_k\right) = \sup_{n \geq 0} \sum_{k=0}^n \underline{E}_m(O_k) \\ &\leq \sup_{n \geq 0} \sum_{k=0}^n \tilde{P}_{m,*}(O_k) \leq \sup_{n \geq 0} \tilde{P}_{m,*}\left(\bigcup_{k=0}^n O_k\right) \leq \tilde{P}_{m,*}(A), \end{aligned}$$

also taking into account that $\tilde{P}_{m,*}$ is super-additive and monotone [because it is a coherent lower probability, as a restriction of the coherent lower prevision \underline{E}_m]. This completes the proof of the second statement. The third statement now follows from Theorem 13 and the transitivity of natural extension (see Section 2). \square

5. SOME REMARKS ON THE HAUSDORFF LOWER MOMENT PROBLEM

Before ending this paper, we devote some attention to a more general type of ‘lower’ moment problem. Indeed, it may be interesting to consider the moment problem for some particular types of coherent lower previsions that are not necessarily linear previsions. This means that we consider a sequence m of real numbers, and study under which conditions there is some coherent lower prevision \underline{P} on $\mathcal{L}([0, 1])$ with *lower moments* m_k , i.e., satisfying $\underline{P}(p^k) = m_k$ for all $k \geq 0$. We should warn the reader here that in the present case the moments will no longer be precise, that is, the *upper moments* $\overline{P}(p^k)$ need not be equal to (may strictly dominate) the *lower moments* $\underline{P}(p^k) = m_k$.

Of course, if the lower moment sequence m satisfies the Hausdorff moment condition, then we know from Theorem 1 that there is a linear prevision P with this moment sequence, and therefore in particular also a coherent lower prevision. However, the converse is not true, as the following simple counterexample shows.

Example 2. Let P_1 be the linear prevision given by $P_1(\{0.5\}) = P_1(\{0.9\}) = 0.5$. Then its first four moments are given by $m_0^1 = 1$, $m_1^1 = 0.7$, $m_2^1 = 0.53$ and $m_3^1 = 0.427$. Consider on the other hand the linear prevision P_2 determined by $P_2(\{0.25\}) = P_2(\{1\}) = 0.5$. Its first four moments are $m_0^2 = 1$, $m_1^2 = 0.625$, $m_2^2 = 0.53125$ and $m_3^2 = 0.5078125$. Now let \underline{P} be the lower envelope of $\{P_1, P_2\}$. This is a coherent lower prevision, and its first four lower moments are $m_0 = 1$, $m_1 = 0.625$, $m_2 = 0.53$ and $m_3 = 0.427$. To see that there is no linear prevision with these first four moments, it suffices to check that there is no completely monotone sequence starting with $\{m_0, m_1, m_2, m_3\}$. This follows at once if we note that $(-1)^2 \Delta^2 m_1 = m_1 - 2m_2 + m_3 = 0.625 - 1.06 + 0.427 = -0.08 < 0$. \blacklozenge

Hence, the moment problem for coherent lower previsions is more general than the one for linear previsions. It is therefore quite remarkable that both problems are equivalent when we consider the additional assumption of 2-monotonicity.

Proposition 15. *Given a lower moment sequence m , there is a 2-monotone coherent lower prevision \underline{P} on $\mathcal{L}([0, 1])$ satisfying $\underline{P}(p^k) = m_k$ for all $k \geq 0$ if and only if there is a linear prevision on $\mathcal{L}([0, 1])$ with these moments, i.e., if and only if m satisfies the Hausdorff moment condition.*

Proof. The ‘if’ part is trivial, since any linear prevision is in particular a 2-monotone coherent lower prevision. To prove the ‘only if’ part, assume that \underline{P} is coherent, 2-monotone and satisfies $\underline{P}(p^k) = m_k$ for all k . We already know from coherence that $m_0 = \underline{P}(p^0) = \underline{P}(1) = 1$. It therefore remains to show that m is completely monotone. To this end, we first show that given $k_1, k_2 \geq 0$, there is some linear prevision Q in $\mathcal{M}(\underline{P})$ such that $Q(p^k) =$

$\underline{P}(p^k) = m_k$ for all $k_1 \leq k \leq k_1 + k_2$. Indeed, the gambles in $\{p^k: k_1 \leq k \leq k_1 + k_2\}$ are all comonotone with each other. Moreover, since a coherent 2-monotone lower prevision is also comonotone additive, we get

$$\underline{P}\left(\sum_{k=k_1}^{k_1+k_2} p^k\right) = \sum_{k=k_1}^{k_1+k_2} \underline{P}(p^k).$$

On the other hand, since \underline{P} is coherent, there is some linear prevision Q in $\mathcal{M}(\underline{P})$ such that

$$\underline{P}\left(\sum_{k=k_1}^{k_1+k_2} p^k\right) = Q\left(\sum_{k=k_1}^{k_1+k_2} p^k\right) = \sum_{k=k_1}^{k_1+k_2} Q(p^k),$$

where the second equality follows from the additivity of Q . Now each of the terms $Q(p^k)$ in the second sum dominates the corresponding term $\underline{P}(p^k)$ in the first sum, and since the sums are equal, so are all the corresponding terms: $Q(p^k) = \underline{P}(p^k) = m_k$, which proves our claim.

Now, since Q is a linear prevision, its moments $Q(p^k)$ must satisfy the Hausdorff moment condition because of Theorem 1. Hence in particular $(-1)^{k_2} \Delta^{k_2} Q(p^{k_1}) \geq 0$. But since in this difference there only appear moments $Q(p^k)$ with $k_1 \leq k \leq k_1 + k_2$, it follows from the considerations above that $(-1)^{k_2} \Delta^{k_2} m_{k_1} \geq 0$. Since we can do this for any $k_1, k_2 \geq 0$, we deduce that the moment sequence m is completely monotone. \square

6. CONCLUSION

The sequence of moments of a σ -additive probability measure P^σ on the Borel sets of $[0, 1]$ completely determines this measure: we cannot find two different σ -additive probabilities with the same sequence of moments. We have shown in this paper that this is not the case when we consider probabilities which are merely finitely additive: indeed, the sequence of moments of a finitely additive probability P will be shared by an infinity of different finitely additive probabilities. Among their restrictions to the Borel sets of $[0, 1]$, only one will be σ -additive, namely P^σ .

We have also proven that this class of finitely additive probabilities on $[0, 1]$ with a given sequence of moments is characterised in particular by the values they take on events, or even more precisely, by the values they take on integrable open intervals. The integrability of these intervals can be easily determined by looking at the discontinuity points of the associated upper and lower distribution functions, or equivalently, of F^σ .

The fact that the lower and upper envelopes of the finitely additive probabilities with a given sequence of moments are respectively completely monotone and completely alternating allow us to relate this set of linear previsions to specific types of (lower and upper) integrals: for instance, the lower and upper previsions it induces are the Choquet integrals with respect to their restrictions to events.

There is an interesting connection between our results and de Finetti's [8] work on exchangeable sequences. In fact, his Representation Theorem, stripped to its bare essentials,¹³ states that given an exchangeable sequence of events, there is a *representing* positive linear functional R (i.e., linear or coherent prevision) on the set $\mathcal{V}([0, 1])$ of polynomial gambles on $[0, 1]$ (actually, on the set of Bernstein basis polynomials $B_{n,p}$) such that the probability of p successes in n observations is

$$P(S_n = p) = R(B_{n,p})$$

¹³When we reduce it to this simple form, the result becomes very simple and fairly easy to prove; see also [4] for a more general treatment, also allowing for exchangeable *lower* previsions.

for all natural numbers n and $0 \leq p \leq n$. Of course, the linear prevision R is completely determined by its raw moments $m_n = R(B_{n,0})$, and it can be extended to a unique linear prevision on the set $\mathcal{C}([0, 1])$ of continuous gambles on $[0, 1]$. In later discussions of this result, several authors (see, for instance, [12, 15, 16]) use the F. Riesz representation theorem in some form to extend this linear prevision uniquely to a σ -additive probability measure on the Borel sets of $[0, 1]$, which allows them to claim that exchangeability is equivalent to being conditionally independent and identically distributed; see for instance [16, Theorem 1.1], and to use this extension to prove various limit theorems (such as the strong law of large numbers, or the Hewitt–Savage zero-one law). Of course, de Finetti would have strongly disagreed with this claim, because he refused to consider σ -additivity as normative, or as anything more than a mathematical convenience.¹⁴ This is why the (weak, but especially the strong) limit laws he proves for exchangeable sequences, are expressed only in finitary terms, using ‘strong convergence’ rather than ‘almost sure convergence’.¹⁵ As a consequence, the prevision of any gamble h that is not m -integrable (and therefore *a fortiori* discontinuous) on $[0, 1]$ is essentially undetermined, and the most we can say, in accordance with de Finetti’s Fundamental Theorem of Probability, is that it can assume any value between $\underline{E}_m(h)$ and $\bar{E}_m(h)$. In this sense, if we let go of the σ -additivity requirement, the natural extension \underline{E}_m is actually the representing lower prevision.

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).

REFERENCES

- [1] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.
- [2] D. M. Cifarelli and E. Regazzini. De Finetti’s contributions to probability and statistics. *Statistical Science*, 11:253–282, 1996.
- [3] G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.
- [4] G. de Cooman, E. Miranda, and E. Quaeghebeur. Exchangeable lower previsions. 2006. In preparation.
- [5] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n -Monotone lower previsions. *Journal of Intelligent and Fuzzy Systems*, 16:253–263, 2005.
- [6] 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.
- [7] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n -Monotone exact functionals. 2006. Submitted for publication.
- [8] B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l’Institut Henri Poincaré*, 7:1–68, 1937. English translation in [17].

¹⁴For instance, see Section 6.3 of his book [10], suggestively titled “The parting of the ways”.

¹⁵See [8] and the rather nice overview paper of his work [2] for the formulation of these laws, and [10, Section 6.8] and [11, 3] for details on strong convergence and finitary formulations of strong laws of large numbers.

- [9] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.
- [10] B. de Finetti. *Theory of Probability*, volume 1. John Wiley & Sons, Chichester, 1974. English translation of [9].
- [11] L. E. Dubins. On Lebesgue-like extensions of finitely additive measures. *The Annals of Probability*, 2:456–463, 1974.
- [12] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley and Sons, New York, 1971.
- [13] F. Hausdorff. Summationmethoden und Momentfolgen I. *Mathematische Zeitschrift*, 9:74–109, 1921.
- [14] F. Hausdorff. Summationmethoden und Momentfolgen II. *Mathematische Zeitschrift*, 9:280–299, 1921.
- [15] E. Hewitt and L. J. Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80:470–501, 1955.
- [16] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer, New York, 2005.
- [17] H. E. Kyburg Jr. and H. E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.
- [18] E. Miranda, G. de Cooman, and E. Quaeghebeur. Finitely additive extensions of distribution functions and moment sequences: the coherent lower prevision approach. 2006. Submitted for publication.
- [19] F. Riesz. Sur les opérations fonctionelles linéaires. *Comptes Rendus de l'Académie des Sciences de Paris*, 149:974–977, 1909.
- [20] 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 fonctionnelle”. Reprinted by Dover Publications in 1990.
- [21] E. Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, CA, 1997.
- [22] J. A. Shohat and J. D. Tamarkin. *The Problem of Moments*. Number 1 in Mathematical Surveys. American Mathematical Society, Providence, Rhode Island, 1943.
- [23] 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