WEAK AND STRONG LAWS OF LARGE NUMBERS FOR COHERENT LOWER PREVISIONS

GERT DE COOMAN AND ENRIQUE MIRANDA

ABSTRACT. We prove weak and strong laws of large numbers for coherent lower previsions, where the lower prevision of a random variable is given a behavioural interpretation as a subject's supremum acceptable price for buying it. Our laws are a consequence of the rationality criterion of coherence, and they can be proven under assumptions that are surprisingly weak when compared to the standard formulation of the laws in more classical approaches to probability theory.

1. INTRODUCTION

In order to set the stage for this paper, let us briefly recall a simple derivation for Bernoulli's weak law of large numbers. Consider N successive tosses of the same coin. The outcome for the k-th toss is denoted by X_k, $k = 1,\dots,N$. This is a random variable, taking values in the set $\{-1,1\}$, where -1 stands for 'tails' and $+1$ for 'heads'. We denote by p the probability for any toss to result in 'heads'. The common expected value μ of the outcomes X_k is then given by $\mu = 2p-1$, and their common variance σ^2 by $\sigma^2 = 4p(1-p) \leq 1$. We are interested in the *sample mean*, which is the random variable $S_N = \frac{1}{N}\sum_{k=1}^N X_k$ whose expectation is μ. If we make the *extra assumption that the successive outcomes X_k are independent*, then the variance σ_N^2 of S_N is given by $\sigma_N^2 = \sigma^2/N \leq 1/N$, and if we use Chebychev's inequality, we find for any $\varepsilon > 0$ that the probability that $|S_N - \mu| > \varepsilon$ is bounded as follows

$$P(\{|S_N-\mu|>\varepsilon\}) \leq \frac{\sigma_N^2}{\varepsilon^2} \leq \frac{1}{N\varepsilon^2}. \quad (1)$$

This tells us that for any $\varepsilon > 0$, the probability $P(\{|S_N-\mu|>\varepsilon\})$ tends to zero as the number of observations N goes to infinity, and we say that the sample mean S_N *converges in probability* to the expectation μ. If we let $Y_k = \frac{1+X_k}{2}$, then the random variable $\frac{1}{N}\sum_{k=1}^N Y_k$ represents the frequency of 'heads' in N tosses. We may rewrite Eq. (1) as $P(\{|\frac{1}{N}\sum_{k=1}^N Y_k - p| > \varepsilon\}) \leq 1/4N\varepsilon^2$, and this tells us that the frequency of 'heads' converges in probability to the probability p of 'heads'.

This convergence result is the *weak law of large numbers* in the context of a binomial process as originally envisaged by Bernoulli (1713). It can be generalised in a number of ways. We can look at random variables that may assume more than two (and possibly an infinite number of) values. We can also try and replace the *convergence in probability* by the stronger *almost sure convergence*. In standard, measure-theoretic probability theory, this

Date: January 10, 2008.

2000 *Mathematics Subject Classification.* 60A99, 60F05,60F15.

Key words and phrases. Imprecise probabilities, coherent lower previsions, law of large numbers, epistemic irrelevance, 2-monotone capacities.

Supported by research grant G.0139.01 of the Flemish Fund for Scientific Research (FWO), and by MCYT, projects MTM2004-01269, TSI2004-06801-C04-01.

has led to the so-called *strong law of large numbers*, due to Borel (1909) and Kolmogorov (1930). In essence, this law states that if we look at an infinite sequence of random variables X_k with bounded variance, then the sample mean S_k will converge to μ almost surely, i.e., with probability one. Finally, we can weaken or modify the independence assumption, and this has led to the well-known martingale convergence theorems, due to Ville (1939) and Doob (1953), and the ergodic theorems, going back to Birkhoff (1932) and von Neumann (1932). Clearly, Bernoulli's law has been the starting point for many important developments in modern, measure-theoretic probability theory (Kallenberg, 2002). At the same time, because it so obviously connects frequencies and probabilities, it has been a source of inspiration for the frequentist interpretation of probability theory.

But Bernoulli's law is perhaps easier to interpret on the subjective, behavioural account of probability theory, championed by de Finetti (1974–1975). Here, a subject's probability for an event is his fair rate for betting on the event, and the law (1) has the following interpretation: if a subject's probability for 'heads' is p, and he judges the tosses of the coin to be independent, then the rationality criterion of *coherence* requires him to bet on the event $\{|S_N - \mu| > \varepsilon\}$ at rates smaller than $\frac{1}{N\varepsilon^2}$. Specifying a higher betting rate would make him subject to a sure loss.[1]

In the case of coin tossing, specifying the expectation μ completely determines the probability distribution of the random variable X_k, because it can only assume two values. This is no longer true for more general random variables, and this observation points to a possible generalisation of Bernoulli's law that so far seems to have received little explicit attention in the literature. What can be said about the probability that $|S_N - \mu| > \varepsilon$ when the probability distributions of the random variables X_k aren't fully specified, but when, as in the Bernoulli case, only the expectation μ of these random variables is given?

In the present paper, we go even further than this. As with the more standard versions of the laws of large numbers, it seems easier to interpret our results in terms of the rational behaviour of a subject. A bounded random variable X_k can be interpreted as a random (or unknown) reward, expressed in terms of some predetermined linear utility. A subject's *lower prevision* m for X_k is the supremum price for which he is willing to buy the random reward X_k, and his *upper prevision* M for X_k is his infimum price for selling X_k. In contradistinction with the Bayesian approach to probability theory, we don't assume that lower and upper previsions coincide, leading to a *prevision* or *fair price* for X_k (de Finetti, 1974–1975), but we do require, as in the Bayesian theory, that lower and upper previsions satisfy some basic rationality, or *coherence*, criteria. In Section 2, we present the basic ideas behind the behavioural theory of coherent lower previsions, which goes back to Smith (1961) and Williams (1975), and was brought to a recent synthesis by Walley (1991).

We shall prove laws of large numbers that make precise the following loosely formulated statement:[2] *if a subject gives a lower prevision m for the bounded random variables X_k, $k = 1, \dots, N$, and assesses that he can't learn from the past, in the sense that observations of the variables X_1, ..., X_{k-1} don't affect the lower prevision for X_k, $k = 2, \dots, N$, then the rationality requirement of coherence implies that he should bet on the event that the sample mean S_N dominates the lower prevision m at rates that increase to one as the number of observations N increases to infinity.* So if a subject doesn't learn from past

[1] Because it represents an upper bound that needn't be tight, the inequality gives a necessary condition for avoiding a sure loss that needn't be sufficient.

[2] Our subsequent treatment is more general in that we don't assume that all variables have the same lower prevision m, but the resulting laws are more difficult to summarise in an intuitive manner.

observations, and specifies a lower prevision for a single observation, then, loosely formulated, coherence implies that he should also believe that the sample mean will eventually dominate this lower prevision. Our law therefore provides a connection between lower previsions and sample means. A similar (dual) statement can be given for upper previsions.

By our analysis, we establish that laws of large numbers can be formulated under conditions that are much weaker than what is usually assumed. This will be explained in much more detail in the following sections, but it behoves us here to at least indicate in what way our assumptions are indeed much weaker.

Above, we have summarised our results using the behavioural interpretation of lower previsions as supremum buying prices. But they can also be given a Bayesian sensitivity analysis interpretation, which makes it easier to compare them to the standard probabilistic results. On the sensitivity analysis view, the uncertainty about a random variable X_k is ideally described by some probability distribution, which may not be well known. Specifying the lower prevision m for the X_k amounts to providing a lower bound for the expectation of X_k under this ideal distribution, or equivalently, it amounts to specifying the set $\mathscr{M}$ of those probability distributions for which the associated expectation of X_k dominates m. So by specifying m, we state that the ideal probability distribution belongs to the set $\mathscr{M}$, but nothing more. And secondly, and more importantly, we model the assessment that the subject can't learn from past observations by stating *only* that the *lower prevision* for X_k doesn't change (remains equal to m) after we observe the outcomes $X_1, \dots, X_{k-1}$. In the standard, precise probabilistic approach, independence implies that the *entire probability distribution* for X_k doesn't change after observing $X_1, \dots, X_{k-1}$. This is a much stronger assumption than ours, at least if the random variables X_k can assume more than two values! To put it differently, after observing $X_1, \dots, X_{k-1}$ our model for the uncertainty about X_k will still be the set of distributions $\mathscr{M}$. So all that is known is that the ideal updated distribution still belongs to $\mathscr{M}$. But we make no claim that this ideal distribution will be the same for all the possible values of $X_1, \dots, X_{k-1}$, nor that it should be the same for all times k! So, in the end, we have sets $\mathscr{M}$ of possible values for the ideal marginal and updated distributions, and we can combine them by applying Bayes rule to all possible combinations, in the usual Bayesian sensitivity analysis fashion. In this way, we end up with a set of candidates for the ideal joint probability distribution of all the variables $X_1, \dots, X_N$. We prove (amongst other things) that the probability of the event $\{S_N \geq m - \varepsilon\}$ goes to one as N increases, in a uniform way for all candidate joint distributions.

How do we proceed to derive our results? In Section 2 we give a brief introduction to the basic ideas behind the theory of coherent lower previsions, and we explain how these lower previsions can be identified with sets of (finitely additive) probability measures.

In Section 3, we prove our very general version of the weak law of large numbers for coherent lower previsions that satisfy a so-called *forward factorisation* condition. We want to stress here that this law is a quite general mathematical result, which holds regardless of the interpretation given to coherent lower previsions.

We discuss a number of possible interpretations of our weak law, as well as specific special cases in Section 4, where we show that our results subsume much of the previous work in the field. Also, in Theorem 4, we give an alternative general formulation of our weak law in terms of (sets of) precise probabilities. This should be easily understandable by, and possibly relevant to, anyone interested in probability theory, on any interpretation.

Interestingly, we can use our weak law to prove versions of the strong law, which is what we do in Section 5. In the last section, we once again draw attention to the more salient features of our approach, and point to possible further generalisations.

2. Coherent lower and upper previsions

In this section, we present a succinct overview of the relevant main ideas underlying the behavioural theory of imprecise probabilities, in order to make it easier for the reader to understand the main ideas of the paper. We refer to (Walley, 1991) for extensive discussion and motivation, and for many of the results and formulae that we shall use below.

2.1. **Basic notation and behavioural interpretation.** Consider a subject who is uncertain about something, say, the value that a random variable X assumes in a set of possible values $\mathscr{X}$. Then a bounded real-valued function on $\mathscr{X}$ is called a *gamble* on $\mathscr{X}$, and the set of all gambles on $\mathscr{X}$ is denoted by $\mathscr{L}(\mathscr{X})$. We interpret a gamble as an uncertain reward: if the value of the random variable X turns out to be $x \in \mathscr{X}$, then the corresponding reward will be $f(x)$ (positive or negative), expressed in units of some (predetermined) linear utility.

The subject's *lower prevision* $\underline{P}(f)$ for a gamble f is defined as his supremum acceptable price for buying f, i.e., it is the highest price μ such that the subject will accept to buy f for all prices strictly smaller than μ (buying f for a price α is the same thing as accepting the uncertain reward $f-\alpha$). Similarly, a subject's *upper prevision* $\overline{P}(f)$ for f is his infimum acceptable selling price for f. Clearly, $\overline{P}(f) = -\underline{P}(-f)$ since selling f for a price α is the same thing as buying $-f$ for the price $-\alpha$. This *conjugacy relation* implies that we can limit our attention to lower previsions: any result for lower previsions can immediately be reformulated in terms of upper previsions.

A subset A of $\mathscr{X}$ is called an *event*, and it can be identified with its indicator (function) I_A, which is a gamble on $\mathscr{X}$. The *lower probability* $\underline{P}(A)$ of A is nothing but the lower prevision $\underline{P}(I_A)$ of its indicator, and it represents the supremum acceptable price for buying A. Similarly, the *upper probability* $\overline{P}(A)$ of A is the infimum acceptable price for selling A. In the case of events it is perhaps more intuitive to regard their lower and upper probabilities as betting rates: the lower probability of A, which is the supremum value of α such that $I_A - \alpha$ is an acceptable gamble for our subject, can also be seen as his supremum acceptable betting rate *on* the event A. Similarly, the upper probability of A can also be seen as one minus our subject's supremum acceptable betting rate *against* A. Note that in this case the conjugacy relation between upper and lower previsions becomes $\overline{P}(A) = 1 - \underline{P}(A^c)$ for any $A \subseteq \mathscr{X}$. In what follows, we don't distinguish between events A and their indicators I_A, and we shall freely move from one notation to another. We shall also, whenever we deem it convenient, switch between the equivalent notations $\underline{P}(A)$ and $\underline{P}(I_A)$: lower/upper probabilities are just special lower/upper previsions.

2.2. **Rationality requirements.** Assume that the subject has given lower prevision assessments $\underline{P}(f)$ for all gambles f in some set of gambles $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$, which needn't have any predefined structure. We can then consider $\underline{P}$ as a real-valued function with domain $\mathscr{K}$, and we call this function a *lower prevision* on $\mathscr{K}$. Since the assessments present in $\underline{P}$ represent commitments of the subject to act in certain ways, they are subject to a number of rationality requirements. The strongest such requirement is that the lower prevision $\underline{P}$ should be *coherent*. Coherence means first of all that the subject's assessments *avoid sure loss*: for any n in the set of positive natural numbers $\mathbb{N}$ and for any $f_1, \ldots, f_n$ in $\mathscr{K}$ we require that

$$\sup_{x \in \mathscr{X}} \left[\sum_{k=1}^{n} [f_k(x) - \underline{P}(f_k)] \right] \geq 0.$$

Otherwise, there would be some $\varepsilon > 0$ such that for all x in $\mathscr{X}$, $\sum_{k=1}^n [f_k(x) - \underline{P}(f_k) + \varepsilon] \leq -\varepsilon$, i.e., the net reward of buying the gambles f_k for the acceptable prices $\underline{P}(f_k) - \varepsilon$ is sure to lead to a loss of at least ε, whatever the value of the random variable X.

But coherence also means that if we consider any f in $\mathscr{K}$, we can't force the subject to accept f for a price strictly higher than his specified supremum buying price $\underline{P}(f)$, by exploiting buying transactions implicit in his lower previsions $\underline{P}(f_k)$ for a finite number of gambles f_k in $\mathscr{K}$, which he is committed to accept. More explicitly, we also require that for any natural numbers $n \geq 0$ and $m \geq 1$, and $f_0, \dots, f_n$ in $\mathscr{K}$:

$$\sup_{x \in \mathscr{X}} \left[\sum_{k=1}^n [f_k(x) - \underline{P}(f_k)] - m[f_0(x) - \underline{P}(f_0)] \right] \geq 0.$$

Otherwise, there would exist $\varepsilon > 0$ such that $m[f_0 - [\underline{P}(f_0) + \varepsilon]]$ pointwise dominates the acceptable combination of buying transactions $\sum_{k=1}^n [f_k - \underline{P}(f_k) + \varepsilon]$, and is therefore acceptable as well. This would mean that by combining these acceptable transactions derived from his assessments, the subject can be effectively forced to buy f_0 at the price $\underline{P}(f_0) + \varepsilon$, which is strictly higher than the supremum acceptable buying price $\underline{P}(f_0)$ that he has specified for it. This is an inconsistency that is to be avoided.

Coherent lower previsions $\underline{P}$ satisfy a number of basic properties. For instance, given gambles f and g in $\mathscr{K}$, real μ and non-negative real λ, coherence implies that the following properties hold, whenever the gambles that appear are in the domain $\mathscr{K}$ of $\underline{P}$:

(C1) $\underline{P}(f) \geq \inf_{x \in \mathscr{X}} f(x)$;
(C2) $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ [super-additivity];
(C3) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [positive homogeneity];
(C4) $\underline{P}(\lambda f + \mu) = \lambda \underline{P}(f) + \mu$.

Other properties can be found in (Walley, 1991, Section 2.6). It is important to mention here that when $\mathscr{K}$ is a linear space, coherence is equivalent to (C1)–(C3).

2.3. **Natural extension.** We can always extend a coherent lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K}$ to a coherent lower prevision $\underline{E}$ on the set of all gambles $\mathscr{L}(\mathscr{X})$, through a procedure called *natural extension*. The natural extension $\underline{E}$ of $\underline{P}$ is defined as the pointwise smallest coherent lower prevision on $\mathscr{L}(\mathscr{X})$ that coincides on $\mathscr{K}$ with $\underline{P}$. It is given for all $f \in \mathscr{L}(\mathscr{X})$ by

$$\underline{E}(f) = \sup_{\substack{f_1,\dots,f_n \in \mathscr{K} \\ \mu_1,\dots,\mu_n \geq 0, n \geq 0}} \inf_{x \in \mathscr{X}} \left[f(x) - \sum_{k=1}^n \mu_k [f_k(x) - \underline{P}(f_k)] \right],$$

where the $\mu_1, \dots, \mu_n$ in the suprema are non-negative real numbers. The natural extension summarises the behavioural implications of $\underline{P}$: $\underline{E}(f)$ is the supremum buying price for f that can be derived from the lower prevision $\underline{P}$ by arguments of coherence alone. We see from its definition that it is the supremum of all prices that the subject can be effectively forced to buy the gamble f for, by combining finite numbers of buying transactions implicit in his lower prevision assessments $\underline{P}$. Note that $\underline{E}$ will not be in general the unique coherent extension of $\underline{P}$ to $\mathscr{L}(\mathscr{X})$; but any other coherent extension will pointwise dominate $\underline{E}$ and will therefore model behavioural dispositions not present in $\underline{P}$.

2.4. **Relation to precise probability theory.** When $\underline{P}(f) = \overline{P}(f)$, the subject's supremum buying price coincides with his infimum selling price, and this common value is a *prevision* or *fair price* for the gamble f, in the sense of de Finetti (1974–1975). This means that our subject is disposed to buy the gamble f for any price $\mu < P(f)$, and to sell it for any price $\mu' > P(f)$ (but he may be undecided about his behaviour for $\mu = P(f)$). A prevision P

defined on a set of gambles $\mathscr{K}$ is called a *linear prevision* if it is coherent both as a lower and as an upper prevision.

A linear prevision P on the set $\mathscr{L}(\mathscr{X})$ can also be characterised as a linear functional that is positive (if $f \geq 0$ then $P(f) \geq 0$) and has unit norm ($P(I_{\mathscr{X}}) = 1$). Its restriction to events is a finitely additive probability. Moreover, any finitely additive probability defined on the set $\wp(\mathscr{X})$ of all events can be uniquely extended to a linear prevision on $\mathscr{L}(\mathscr{X})$. For this reason, we shall identify linear previsions on $\mathscr{L}(\mathscr{X})$ with finitely additive probabilities on $\wp(\mathscr{X})$. We denote by $\mathbb{P}(\mathscr{X})$ the set of all linear previsions on $\mathscr{L}(\mathscr{X})$, or equivalently, of all finitely additive probabilities on $\wp(\mathscr{X})$.

Linear previsions are the *precise* probability models, and we call coherent lower and upper previsions *imprecise* probability models. That linear previsions are only required to be *finitely* additive, and not σ-additive, derives from the finitary character of the coherence requirement. Throughout the paper, we shall work with finitely additive probabilities, and only bring in σ-additivity when we think it's absolutely necessary.

The notions of avoiding sure loss, coherence, and natural extension can be characterised in terms of sets of linear previsions. Consider a lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K}$. Its *set of dominating linear previsions* $\mathscr{M}(\underline{P})$ is given by

$$\mathscr{M}(\underline{P}) = \{P \in \mathbb{P}(\mathscr{X}) \colon (\forall f \in \mathscr{K}) P(f) \geq \underline{P}(f)\}. \tag{2}$$

Then $\underline{P}$ avoids sure loss if and only if $\mathscr{M}(\underline{P}) \neq \emptyset$, i.e., if it has a dominating linear prevision. $\underline{P}$ is coherent if and only if $\underline{P}(f) = \min\{P(f) \colon P \in \mathscr{M}(\underline{P})\}$ for all f in $\mathscr{K}$, i.e., if it is the *lower envelope* of $\mathscr{M}(\underline{P})$. And the natural extension $\underline{E}$ of $\underline{P}$ is given by $\underline{E}(f) = \min\{P(f) \colon P \in \mathscr{M}(\underline{P})\}$ for all f in $\mathscr{L}(\mathscr{X})$. This means that we have the important equality $\mathscr{M}(\underline{E}) = \mathscr{M}(\underline{P})$, another way of expressing that the natural extension $\underline{E}$ carries essentially the same information as the coherent lower prevision $\underline{P}$. Moreover, the lower envelope of any set of linear previsions is always a coherent lower prevision.

We can use these relationships to formulate the results (limit laws) for coherent lower previsions in the rest of the paper in terms of their dominating linear previsions. They provide coherent lower previsions with a *Bayesian sensitivity analysis* interpretation, as opposed to the more direct behavioural one given above: we may assume the existence of an ideal (but unknown) precise probability model P_T on $\mathscr{L}(\mathscr{X})$, and represent our imperfect knowledge about P_T by means of a set of possible candidates $\mathscr{M}$ for P_T. The information given by this set is equivalent to the one provided by its *lower envelope* $\underline{P}$, which is given by $\underline{P}(f) = \min_{P \in \mathscr{M}} P(f)$ for all $f \in \mathscr{L}(\mathscr{X})$. This lower envelope $\underline{P}$ is a coherent lower prevision; and indeed, $P_T \in \mathscr{M}$ is equivalent to $P_T \geq \underline{P}$.

2.5. **Joint and marginal lower previsions.** Now consider a number of random variables $X_1, X_2, \dots, X_N$ that may assume values in the respective sets $\mathscr{X}_1, \mathscr{X}_2, \dots, \mathscr{X}_N$. We assume that these variables are *logically independent*: the joint random variable $(X_1, \dots, X_N)$ may assume all values in the product set $\mathscr{X}^N := \mathscr{X}_1 \times \mathscr{X}_2 \times \dots \mathscr{X}_N$. A subject's coherent lower prevision $\underline{P}^N$ on a subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X}^N)$ is a model for his uncertainty about the value that the joint random variable $(X_1, \dots, X_N)$ assumes in $\mathscr{X}^N$, and we call it a *joint lower prevision*.

For $k = 1, \dots, N$, we can associate with $\underline{P}^N$ its so-called $\mathscr{X}_k$*-marginal* (lower prevision) $\underline{P}_k$, defined by $\underline{P}_k(g) = \underline{P}^N(g')$ for all gambles g on $\mathscr{X}_k$, such that the corresponding gamble g' on $\mathscr{X}^N$, defined by $g'(x_1, \dots, x_N) = g(x_k)$ for all $(x_1, \dots, x_N)$ in $\mathscr{X}^N$, belongs to $\mathscr{K}$. The gamble g' is constant on the sets $\mathscr{X}_1 \times \dots \times \{x_k\} \dots \times \mathscr{X}_N$, and we call it $\mathscr{X}_k$*-measurable*. In what follows, we shall identify g and g', and simply write $\underline{P}^N(g)$ rather than $\underline{P}^N(g')$.

The marginal $\underline{P}_k$ is the corresponding model for the subject's uncertainty about the value that X_k assumes in $\mathscr{X}_k$, irrespective of what values the remaining $N-1$ random variables assume. The coherence of the joint lower prevision $\underline{P}^N$ clearly implies the coherence of its marginals $\underline{P}_k$. If $\underline{P}^N$ is in particular a linear prevision on $\mathscr{L}(\mathscr{X}^N)$, its marginals are linear previsions too.

Conversely, assume we start with N coherent marginal lower previsions $\underline{P}_k$, defined on the respective domains $\mathscr{K}_k \subseteq \mathscr{L}(\mathscr{X}_k)$. We can interpret $\mathscr{K}_k$ as a set of gambles on $\mathscr{X}^N$ that are $\mathscr{X}_k$-measurable. Any coherent joint lower prevision defined on a set $\mathscr{K}$ of gambles on $\mathscr{X}^N$ that includes the $\mathscr{K}_k$ and that coincides with the $\underline{P}_k$ on their respective domains, i.e., has marginals $\underline{P}_k$, will be called a *product* of the lower previsions $\underline{P}_k$. We shall come across various ways of defining such products further on in the paper. It should be stressed here that, in contradistinction with Walley (1991, Section 9.3.1), we don't intend the mere term 'product' to imply that the variables X_k are assumed to be independent in any way. On our approach, there may be many types of products, some of which may be associated with certain types of interdependence between the random variables X_k.

3. A WEAK LAW OF LARGE NUMBERS

3.1. **Formulation.** We are now ready to turn to the most general formulation of our weak law of large numbers. We consider N random variables X_k taking values in respective sets $\mathscr{X}_k$. As before, we assume these random values to be logically independent, meaning that the joint random variable $(X_1,\dots,X_N)$ may assume all values in the product set $\mathscr{X}^N := \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_N$. We also consider a coherent joint lower prevision $\underline{P}^N$ on $\mathscr{L}(\mathscr{X}^N)$.

Definition 1. A joint lower prevision $\underline{P}^N$ on $\mathscr{L}(\mathscr{X}^N)$ is called *forward factorising* if $\underline{P}^N(g[h-\underline{P}^N(h)]) \geq 0$ for all $k \in \{1,\dots,N\}$, all $g \in \mathscr{L}_+(\mathscr{X}^{k-1})$ and all $h \in \mathscr{L}(\mathscr{X}_k)$.

We have used the notations $\mathscr{X}^k$ for the product set $\times_{\ell=1}^k \mathscr{X}_\ell$ and $\mathscr{L}_+(\mathscr{X}^k)$ for the set of non-negative gambles on $\mathscr{X}^k$. For $k=0$, there is some abuse of notation: we let $\mathscr{X}^0 := \emptyset$, and we identify $\mathscr{L}_+(\mathscr{X}^0) = \mathscr{L}_+(\emptyset)$ with the set $\mathbb{R}_+$ of non-negative real numbers. The corresponding inequality for $k=0$ is implied by the coherence of $\underline{P}^N$.

Why do we use the term 'forward factorising'? It is easy to see that for joint linear previsions P^N, the condition is equivalent to $P^N(gh) = P^N(g)P^N(h)$ for all g in $\mathscr{L}(\mathscr{X}^{k-1})$ and all h in $\mathscr{L}(\mathscr{X}_k)$, where $k \in \{1,\dots,N\}$: for the direct implication, apply the condition to $g - \inf g$ and h and use the linearity of P^N to deduce that $P^N(gh) \geq P^N(g)P^N(h)$ for all g in $\mathscr{L}(\mathscr{X}^{k-1})$ and all h in $\mathscr{L}(\mathscr{X}_k)$, and then use this with $-g$, h to deduce the equality; the converse implication is trivial. This means that the linear prevision P^N factorises on products of gambles, where one of the factors refers to the 'present time k', and the other factor refers to the 'entire past $1,\dots,k-1$'. Our condition will turn out to be the appropriate generalisation of this idea to coherent joint lower previsions.

It is for forward factorising coherent joint lower previsions that we shall formulate our weak law of large numbers. The following theorem is instrumental in proving it, but as we shall see in Section 4, it is of some interest in itself as well.

Theorem 1. *Let $\underline{P}^N$ be a coherent joint lower prevision on $\mathscr{L}(\mathscr{X}^N)$. Then $\underline{P}^N$ is forward factorising if and only if*

$$\underline{P}^N(f) \geq \inf_{x \in \mathscr{X}^N} \left[f(x) - \sum_{k=1}^{N} \sum_{j_k=1}^{n_k} g_{kj_k}(x_1,\dots,x_{k-1})[h_{kj_k}(x_k) - m_{kj_k}] \right] \tag{3}$$

for all gambles f on $\mathscr{X}^N$, all $n_k \geq 0$, all $h_{kj_k} \in \mathscr{L}(\mathscr{X}_k)$, all $g_{kj_k} \in \mathscr{L}_+(\mathscr{X}^{k-1})$, and all $m_{kj_k} \leq \underline{P}^N(h_{kj_k})$, where $j_k \in \{1\dots,n_k\}$ and $k \in \{1,\dots,N\}$.

Remark 1. We can see from the proof of Theorem 1 (see Section A.1 in the Appendix) that if we only require the forward factorising property to hold for strictly positive g and arbitrary h, then it is equivalent to condition (3), but now restricted to $h_{kj_k} \in \mathscr{L}(\mathscr{X}_k)$ and strictly positive $g_{kj_k} \in \mathscr{L}_+(\mathscr{X}^{k-1})$.

Now consider, for each random variable X_k a gamble h_k on its set of possible values $\mathscr{X}_k$. Let B be a common bound for the ranges of these gambles, i.e., $\sup h_k - \inf h_k \leq B$ for all $k \in \{1,\dots,N\}$. Then the '*sample mean*' $\frac{1}{N}\sum_{k=1}^N h_k$ is a gamble whose range is also bounded by B. Given $\varepsilon > 0$, we are interested in the lower probability of the event

$$\left\{\frac{1}{N}\sum_{k=1}^N \underline{P}^N(h_k) - \varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k \leq \frac{1}{N}\sum_{k=1}^N \overline{P}(h_k) + \varepsilon\right\}$$
$$:= \left\{x \in \mathscr{X}^N \colon \frac{1}{N}\sum_{k=1}^N \underline{P}^N(h_k) - \varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k(x_k) \leq \frac{1}{N}\sum_{k=1}^N \overline{P}(h_k) + \varepsilon\right\}$$

that the sample mean lies, up to ε, between the average of the lower previsions $\underline{P}^N(h_k)$ and the average of the upper previsions $\overline{P}^N(h_k)$ of these gambles. If the coherent lower prevision $\underline{P}^N$ is forward factorising, then the lower probability of this event goes to one as N increases to infinity: in fact, we have the following result.

Theorem 2 (Weak law of large numbers – general version)**.** *Let $\underline{P}^N$ be a lower prevision on $\mathscr{L}(\mathscr{X}^N)$ that is coherent and forward factorising. Let $\varepsilon > 0$ and consider arbitrary gambles h_k on $\mathscr{X}_k$. Let B be a common bound for the ranges of these gambles and let $\inf h_k \leq m_k \leq \underline{P}^N(h_k) \leq \overline{P}^N(h_k) \leq M_k \leq \sup h_k$. Then*

$$\underline{P}^N\left(\left\{\frac{1}{N}\sum_{k=1}^N m_k - \varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k \leq \frac{1}{N}\sum_{k=1}^N M_k + \varepsilon\right\}\right) \geq 1 - 2\exp\left(-\frac{N\varepsilon^2}{4B^2}\right).$$

This is a general mathematical result, valid on any interpretation that might be given to a lower prevision. It holds for all functionals $\underline{P}^N$ on $\mathscr{L}(\mathscr{X}^N)$ that are coherent [in the sense that they satisfy the mathematical conditions (C1)–(C3)] and forward factorising.

Remark 2. We can infer from the proof of this theorem (in Section A.2 of the Appendix) that we actually have two limit laws. If we only specify upper bounds M_k for the upper previsions $\overline{P}(h_k)$ then we can prove that

$$\underline{P}^N\left(\left\{\frac{1}{N}\sum_{k=1}^N (h_k(x_k) - M_k) \leq \varepsilon\right\}\right) \geq 1 - \exp\left(-\frac{N\varepsilon^2}{4B^2}\right),$$

and if we only specify lower bounds m_k for the lower previsions $\underline{P}^N(h_k)$ then we can prove that

$$\underline{P}^N\left(\left\{\frac{1}{N}\sum_{k=1}^N (h_k(x_k) - m_k) \geq -\varepsilon\right\}\right) \geq 1 - \exp\left(-\frac{N\varepsilon^2}{4B^2}\right),$$

for all $\varepsilon > 0$. If we specify both, we get Theorem 2.

Remark 3. In our definition of a forward factorising lower prevision, we require that the 'factorisation' inequality $\underline{P}^N(g[h - \underline{P}(h)]) \geq 0$ should be satisfied for *all* g in $\mathscr{L}_+(\mathscr{X}^{k-1})$ and *all* gambles h in $\mathscr{L}(\mathscr{X}_k)$. But when we want to prove a weak law of large numbers in more specific situations, we can sometimes weaken the factorisation requirement. For instance, when the sets $\mathscr{X}_k$ are bounded subsets of $\mathbb{R}$, and we want to prove the weak law for a restricted choice of the gambles h_k, e.g., Borel measurable ones, then we may deduce from our method of proof in Section A.2, that we only need the factorisation property to hold for Borel measurable g and h. When, even more restrictively, we only want a weak law for the case that the gambles h_k are identity maps, we can do with the identity map for h and continuous, or even polynomial, g in the forward factorisation inequality.

3.2. **Special cases.** Let us look at the specific formulation of our weak law in a number of particular special cases: (i) the classical case of independent variables; and (ii) when the random variables X_k are related to the occurrence of events. We begin with the first case.

Consider the case of independent and identically distributed bounded random variables X_k: all random variables are real, and have the same distribution P, which in this classical case is assumed to be a σ-additive probability measure defined on the Borel σ-field $\mathscr{B}$ on $\mathbb{R}$. The distribution of the joint random variable is the usual product measure P^N on the product algebra $\mathscr{B}^N$. We have then that $m_k = M_k = \mu = E_P(X_k)$, where E_P is the expectation operator associated with P. We denote the common variance of the X_k by $\sigma^2 = E_P((X_k-\mu)^2)$. Since there exists a common bound B for the ranges of the random variables X_k, B^2 is then a bound for σ^2. Let us denote

$$D_N = \frac{1}{N}\sum_{k=1}^{N}(X_k-\mu).$$

Then $E_{P^N}(D_N) = 0$, where E_{P^N} is the expectation operator associated with the (independent) product measure P^N. Also $E_{P^N}(D_N^2) \leq \frac{B^2}{N}$, so we infer from Chebychev's inequality that

$$E_{P^N}(\{|D_N(x)| \leq \varepsilon\}) \geq 1-\frac{B^2}{N\varepsilon^2},$$

and we deduce that $E_{P^N}(\{|D_N(x)| \leq \varepsilon\})$ goes to one as N goes to infinity. This is the usual Chebychev bound found in many probability textbooks; see (Ash and Doléans-Dade, 2000) and also (Shafer and Vovk, 2001), where this same bound is derived in a way that is similar to our derivation of Theorem 2 in the Section A.2 of the Appendix.

But since the expectation operator E_{P^N} is forward factorising,[3] we can also use our formulation of the weak law, which provides a different upper bound, $1-2\exp(-\frac{N\varepsilon^2}{4B^2})$. In Figure 1, we compare the functions $2\exp(-x/4)$ and $\frac{1}{x}$ (where $x = \frac{N\varepsilon^2}{B^2}$) in a loglog plot. It is seen that our new bound is far superior to the one given by Chebychev's inequality for large enough $N\varepsilon^2/B^2$ (more than 10). This will be important in the next section, where it will ultimately allow us to derive a finitary version of the strong law of large numbers directly from the weak one. Curiously, perhaps, the form of our bound corresponds much better (up to a factor in the exponential) to Hoeffding's (1963) inequality for N *independent* bounded random variables $X_1,\dots,X_N$, which can be written as

$$P\left(\left\{\frac{1}{N}\sum_{k=1}^{N}(X_k-\mu_k) \leq -\varepsilon\right\}\right) \leq \exp\left(-\frac{2N\varepsilon^2}{B^2}\right),$$

where μ_k is the expected value of X_k.

Next, consider N logically independent events A_k for which a subject has specified lower and upper probabilities $\underline{P}(A_k)$ and $\overline{P}(A_k)$, $k = 1,\dots,N$. Consider the random variables $X_k = I_{A_k}$, then each X_k assumes values in $\mathscr{X}_k = \{0,1\}$. Let $\underline{P}^N$ be any coherent and forward factorising lower prevision on $\mathscr{L}(\mathscr{X}^N)$ that extends these lower and upper probability

[3] Actually, since this operator is only defined on measurable functions, it is forward factorising only on measurable gambles. But that is enough to apply our weak law: we use only a specific type of measurable gamble in its proof; see Remark 3 and Section A.2 of the Appendix.

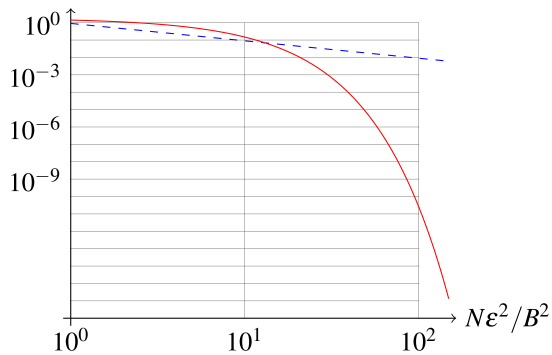

FIGURE 1. Comparison of the bounds $2\exp(-\frac{N\varepsilon^2}{4B^2})$ (full line) and $B^2/N\varepsilon^2$ (dashed line) as a function of $N\varepsilon^2/B^2$ in a loglog plot.

assessments. Then $\underline{P}^N(X_k) = \underline{P}(A_k)$, and similarly $\overline{P}^N(X_k) = \overline{P}(A_k)$. Since $B = 1$ in this case, our weak law tells us in particular that

$$\underline{P}^N\left(\left\{\frac{1}{N}\sum_{k=1}^{N}\underline{P}(A_k) - \varepsilon \leq \frac{1}{N}\sum_{k=1}^{N} I_{A_k} \leq \frac{1}{N}\sum_{k=1}^{N}\overline{P}(A_k) + \varepsilon\right\}\right) \geq 1 - 2\exp\left(-\frac{N\varepsilon^2}{4}\right).$$

This version of the weak law therefore relates the lower and upper probabilities of the events A_k to the 'frequency of occurrence' $\frac{1}{N}\sum_{k=1}^{N} I_{A_k}$.

4. INTERPRETATION

We now turn to a discussion of the significance of our weak law. We present various ways of interpreting it by considering a diversity of situations where we are naturally led to consider joint lower previsions that are forward factorising.

We consider N random variables X_k, taking values in the respective sets $\mathscr{X}_k$, and gambles h_k on $\mathscr{X}_k$, $k = 1,\dots,N$. A subject specifies a lower prevision m_k and an upper prevision M_k for each gamble h_k, which only depends on the value of the k-th random variable. In addition, he assesses that he isn't learning from previous observations by expressing that his lower and upper previsions for the gamble h_k won't change after observing the values of the previous variables $X_1, \dots, X_{k-1}$; and this for all $k = 1,\dots,N$.

On a behavioural interpretation, this means that our subject has specified N marginal lower previsions $\underline{P}_k$ on the domains $\mathscr{K}_k := \{h_k, -h_k\} \subseteq \mathscr{L}(\mathscr{X}_k)$, given by

$$m_k = \underline{P}_k(h_k) \text{ and } M_k = \overline{P}_k(h_k) = -\underline{P}_k(-h_k). \tag{4}$$

That the lower and upper previsions of h_k depending on the value of X_k don't change after learning the values of previous variables $X_1, \dots, X_{k-1}$ can be expressed using so-called *forward epistemic irrelevance assessments*

$$\underline{P}(h_k|x_1,\dots,x_{k-1}) = m_k \text{ and } \underline{P}(-h_k|x_1,\dots,x_{k-1}) = -\overline{P}(h_k|x_1,\dots,x_{k-1}) = -M_k, \tag{5}$$

for $2 \leq k \leq N$ and for all $x_1,\dots,x_{k-1}$ in $\mathscr{X}^{k-1}$. The left hand sides of these expressions represent conditional lower previsions, i.e., the subject's supremum buying prices for the relevant gambles h_k and $-h_k$ conditional on the observation of the values $(x_1,\dots,x_{k-1})$ of the previous random variables.

We shall now consider various joint lower previsions $\underline{P}^N$ on $\mathscr{L}(\mathscr{X}^N)$ that are compatible with these assessments (4) and (5), in the sense that (i) they are 'products' of the marginal lower previsions $\underline{P}_k$, meaning that they coincide with them: $\underline{P}^N(h_k) = \underline{P}_k(h_k) = m_k$ and $\underline{P}^N(-h_k) = \underline{P}_k(-h_k) = -M_k$; and (ii) they reproduce, in some specific sense, the epistemic irrelevance assessments (5).

4.1. **The forward irrelevant natural extension.** We have shown elsewhere (De Cooman and Miranda, 2006) that the point-wise smallest (behaviourally most conservative) joint lower prevision on $\mathscr{L}(\mathscr{X}^N)$ that is coherent[4] with the assessments (4) and (5), is given by the so-called *forward irrelevant natural extension* $\underline{E}^N$ of the marginals $\underline{P}_k$. An immediate application of the general results in (De Cooman and Miranda, 2006, Proposition 4) also allows us to conclude that this $\underline{E}^N$ is actually forward factorising, and given by

$$\underline{E}^N(f) = \sup_{\substack{g_k,h_k\in\mathscr{L}_+(\mathscr{X}^{k-1})\\ k=1,\dots,N}} \inf_{x\in\mathscr{X}^N}\Bigg[f(x) - \sum_{k=1}^N [g_k(x_1,\dots,x_{k-1})(x_k-m_k)+h_k(x_1,\dots,x_{k-1})(M_k-x_k)]\Bigg],$$

for all gambles f on $\mathscr{X}^N$. A comparison with Eq. (3) in Theorem 1 tells us that $\underline{E}^N$ *is actually the point-wise smallest (most conservative) product of the marginal lower previsions* $\underline{P}_k$ *that is still forward factorising.* An immediate application of Theorem 2 then tells us that

$$\underline{E}^N\left(\left\{\frac{1}{N}\sum_{k=1}^N m_k - \varepsilon \le \frac{1}{N}\sum_{k=1}^N h_k \le \frac{1}{N}\sum_{k=1}^N M_k + \varepsilon\right\}\right) \ge 1 - 2\exp\left(-\frac{N\varepsilon^2}{4B^2}\right).$$

where B is any common bound for the ranges of the gambles h_k. Summarising, we find the following result.

Theorem 3 (Weak law of large numbers – behavioural version)**.** *Consider any gambles h_k on $\mathscr{X}_k$ with a common bound B for their ranges, and assume that a subject (i) assesses lower previsions m_k and upper previsions M_k for these gambles h_k, where* $\inf h_k \le m_k \le M_k \le \sup h_k$*, and (ii) assesses that these lower and upper previsions won't change upon learning the values of previous random variables $X_1, \dots, X_{k-1}$, $k = 1,\dots,N$. Then coherence requires him to bet on the event* $\{\frac{1}{N}\sum_{k=1}^N m_k - \varepsilon \le \frac{1}{N}\sum_{k=1}^N h_k \le \frac{1}{N}\sum_{k=1}^N M_k + \varepsilon\}$ *at rates that are at least* $1-2\exp(-N\varepsilon^2/4B^2)$*, for all* $\varepsilon > 0$.

The conditions for applying this version of the weak law are very weak: our subject need only give (coherent) lower and upper prevision assessments for the gambles h_k that lie in $[m_k, M_k]$; *he needn't give assessments for any other gambles*. Or he may give lower and upper previsions for a number of other gambles, and the only requirement is then that the *implied* lower and upper previsions (by natural extension) for h_k should lie in $[m_k, M_k]$. Moreover, he need only assess that his lower and upper previsions for the gambles h_k, *and these two numbers alone*, are not affected by observing the values of the previous random variables $X_1, \dots, X_{k-1}$.

Of course, any reader who doesn't really care for lower previsions, or the associated coherence requirements, or more generally for Walley's behavioural approach to probability, may at this point rightfully wonder why he or she should bother about this version of

[4]We refer to Walley's notion of coherence of conditional and unconditional lower previsions, which is introduced and studied in Walley (1991, Chapter 7).

our law. We now proceed to show that our result can be given another, sensitivity analysis interpretation. This shows our law to be of potential interest for anyone dealing with probability theory, on any interpretation.

4.2. **The forward irrelevant product and the sensitivity analysis version.** Any product lower prevision $\underline{P}^N$ of marginal lower previsions $\underline{P}_k$ can be seen as some specific way to combine these marginals into a joint lower prevision. One such product is the forward irrelevant natural extension $\underline{E}_N$ discussed above. It combines these marginals in such a way that it is as conservative as possible, while still remaining coherent with the forward epistemic irrelevance assessments (5).

But any product lower prevision $\underline{P}^N$ that dominates $\underline{E}^N$ automatically also satisfies $\underline{P}^N(\{\frac{1}{N}\sum_{k=1}^N m_k - \varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k \leq \frac{1}{N}\sum_{k=1}^N M_k + \varepsilon\}) \geq 1 - 2\exp(-N\varepsilon^2/4B^2)$, and therefore leads to a version of the weak law of large numbers. We now give an interesting procedure for constructing such a product, with a nice sensitivity analysis interpretation.

Consider the set $\mathscr{M}_k$ of all (marginal) linear previsions, or equivalently, finitely additive probability measures, P_k on $\mathscr{L}(\mathscr{X}_k)$ that are compatible with the given lower and upper previsions assessments, in the sense that

$$\mathscr{M}_k := \{P_k \in \mathbb{P}(\mathscr{X}_k)\colon m_k \leq P_k(h_k) \leq M_k\}. \tag{6}$$

We denote the lower envelope of the set of linear previsions $\mathscr{M}_k$ by $\underline{E}_k$, i.e., $\underline{E}_k(f_k) = \min\{P_k(f)\colon P_k \in \mathscr{M}_k\}$ for all gambles f_k on $\mathscr{X}_k$.

Let P_1 be any element of $\mathscr{M}(\underline{P}_1)$, and for $2 \leq k \leq N$ and any $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$, let $P_k(\cdot|x_1,\dots,x_{k-1})$ be any element of $\mathscr{M}(\underline{P}_k)$, where these sets are defined by Eq. (2). This leads to our considering a marginal linear prevision P_1 on $\mathscr{L}(\mathscr{X}_1)$ and conditional linear previsions $P_k(\cdot|X^{k-1})$ on $\mathscr{L}(\mathscr{X}^k)$, defined as follows: for any gamble g_k on $\mathscr{X}^k$, $P_k(g_k|X^{k-1})$ is the gamble on $\mathscr{X}^{k-1}$ that assumes the value

$$P_k(g_k(x_1,\dots,x_{k-1},\cdot)|x_1,\dots,x_{k-1}) \tag{7}$$

in the element $(x_1,\dots,x_{k-1})$ of $\mathscr{X}^{k-1}$, for $2 \leq k \leq N$. Let, for any gamble f on $\mathscr{X}^N$,

$$P(f) = P_1(P_2(\dots(P_N(f|X^{N-1}))\dots|X^1)), \tag{8}$$

i.e., apply, in the usual fashion, Bayes's rule to combine the marginal linear prevision P_1 and the conditional linear previsions $P_2(\cdot|X^1)$, …, $P_N(\cdot|X^{N-1})$ into the joint linear prevision P^N on $\mathscr{L}(\mathscr{X}^N)$. See (Miranda and de Cooman, 2007, Section 4) for more details on this construction procedure.

We can repeat this procedure to end up with a joint linear prevision P^N *for any possible choice* of the linear previsions P_1 and $P_k(\cdot|X^{k-1}), k = 2,\dots,N$.[5] In this way, we end up with a set of joint linear previsions on $\mathscr{L}(\mathscr{X}^N)$ that is completely characterised by its lower envelope, which we shall denote by $\underline{M}^N$. In another paper (De Cooman and Miranda, 2006, Section 3.4), we have called this coherent joint lower prevision $\underline{M}^N$ the *forward irrelevant product* of the marginals $\underline{P}_k$. We have shown there that this lower prevision can also be found directly in terms of the marginal lower envelopes $\underline{E}_k$ as follows:

$$\underline{M}^N(h) = \underline{E}_1(\underline{E}_2(\dots(\underline{E}_N(h))\dots)),$$

for any gamble h on $\mathscr{X}^N$. Here we are using the general convention that for any gamble g on $\mathscr{X}^k$, $\underline{E}_k(g)$ denotes the gamble on $\mathscr{X}^{k-1}$, whose value in an element $(x_1,\dots,x_{k-1})$ of $\mathscr{X}^{k-1}$ is given by $\underline{E}_k(g(x_1,\dots,x_{k-1},\cdot))$. In other words, $\underline{E}_N(h)$ is the gamble on $\mathscr{X}^{N-1}$

[5] This is similar to the definition of integrals from strategies by Dubins and Savage (1965, Chapter 2).

that is obtained by 'integrating out' the last variable, i.e., the gamble that assumes the value $\underline{E}^N(h(x_1,\dots,x_{N-1},\cdot))$ in the element $(x_1,\dots,x_{N-1})$ of $\mathscr{X}^{N-1}$, and so on.

It is easy to see (De Cooman and Miranda, 2006, Proposition 4) that $\underline{M}^N$ is really a product of the marginals $\underline{P}_k$ (coincides with the $\underline{P}_k$ on the gambles h_k and $-h_k$), and that it is *forward factorising*. This implies that it generally dominates the forward irrelevant natural extension $\underline{E}^N$.[6] So does therefore any joint linear prevision P^N constructed in the manner described above. This leads at once to the following interesting result.

Theorem 4 (Weak law of large numbers – sensitivity analysis version)**.** *Consider any gambles h_k on $\mathscr{X}_k$ with a common bound B on their ranges, and assume that a subject assesses lower previsions m_k and upper previsions M_k for these gambles h_k, where* $\inf h_k \leq m_k \leq M_k \leq \sup h_k$. *Then any joint linear prevision P^N on $\mathscr{L}(\mathscr{X}^N)$ that is constructed from the marginal linear previsions in the sets $\mathscr{M}(\underline{P}_k)=\{P_k\colon m_k \leq P_k(h_k) \leq M_k\}$ in the manner described above, using Eqs.* (7)–(9)*, will satisfy, for all $\varepsilon > 0$,*

$$P^N\left(\left\{\frac{1}{N}\sum_{k=1}^N m_k-\varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k \leq \frac{1}{N}\sum_{k=1}^N M_k+\varepsilon\right\}\right) \geq 1-2\exp\left(-\frac{N\varepsilon^2}{4B^2}\right).$$

Of course, a similar result holds if we combine marginal linear previsions from sets $\mathscr{M}_k'$ that are subsets of the $\mathscr{M}_k$ obtained from Eq. (7), meaning that we base ourselves on assessments which imply that for the corresponding lower envelopes $\underline{E}_k'(h_k) \geq m_k$ and $\underline{E}_k'(-h_k) \geq -M_k$.

4.3. **The strong product.** Consider again the procedure, described in the previous section, for obtaining joint linear previsions P^N from the marginal linear previsions P_k in the sets $\mathscr{M}_k$. If we now consistently take the *same* linear prevision $P(\cdot|x_1,\dots,x_{k-1})=P_k$ for all $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$, or in other words, let P^N be the forward irrelevant product of the marginal linear previsions P_k, then we end up with a set of joint linear previsions that is only a subset of the one considered previously. Its lower envelope is a product of the marginal lower previsions $\underline{P}_k$, and is usually called their *strong*, or *type-I product*; see Walley (1991, Section 9.3.5) and Couso et al. (2000). It therefore dominates both their forward irrelevant product and their forward irrelevant natural extension. Indeed, it is easy to see that it is also forward factorising. For this strong product (and any of the P^N constructed above) we therefore have the same upper bound $1-2\exp(-N\varepsilon^2/4B^2)$ for the corresponding (lower) probability of the event $\{\frac{1}{N}\sum_{k=1}^N m_k-\varepsilon \leq \frac{1}{N}\sum_{k=1}^N h_k \leq \frac{1}{N}\sum_{k=1}^N M_k+\varepsilon\}$.

4.4. **The Kuznetsov product.** In a quite interesting, but relatively unknown book (in Russian) on interval probability, Kuznetsov (1991) proves a number of limit laws, and in particular a strong law of large numbers, for interval probabilities. The assumption he needs for proving these laws, is essentially that the random variables X_k satisfy a special kind of independence condition, which we shall call *Kuznetsov independence*. In order to relate this condition to the forward factorising property considered in this paper, let us focus on the case of two random variables. The extension to the case of more than two variables is then more or less straightforward, using induction. Consider a coherent joint lower prevision $\underline{P}^2$ on $\mathscr{L}(\mathscr{X}_1\times\mathscr{X}_2)$ that models the available knowledge about the value of the joint random variable (X_1,X_2) in $\mathscr{X}_1\times\mathscr{X}_2$. We say that the random variables X_1 and X_2 are *Kuznetsov independent* if for all gambles f on $\mathscr{X}_1$ and g on $\mathscr{X}_2$:

$$\underline{P}^2(fg)=\min\{\underline{P}^2(f)\underline{P}^2(g),\underline{P}^2(f)\overline{P}^2(g),\overline{P}^2(f)\underline{P}^2(g),\overline{P}^2(f)\overline{P}^2(g)\},$$

[6] In De Cooman and Miranda (2006, Theorem 3), we show that $\underline{M}^N$ actually coincides with $\underline{E}^N$ whenever the sets $\mathscr{X}_k$ are finite.

i.e., if $[\underline{P}^2(fg), \overline{P}^2(fg)]$ is the interval product of $[\underline{P}^2(f), \overline{P}^2(f)]$ and $[\underline{P}^2(g), \overline{P}^2(g)]$. A detailed discussion of Kuznetsov independence can be found in (Cozman, 2000, 2003).

We show in the Appendix (see Section A.3) that if X_1 and X_2 are Kuznetsov independent, then the coherent joint lower prevision $\underline{P}^2$ is forward factorising. Therefore, Kuznetsov's results are implied by our law of large numbers.[7]

5. A STRONG LAW OF LARGE NUMBERS

5.1. **Preliminary work.** We turn now towards a strong law of large numbers, where we are dealing with a sequence of random variables X_k taking values in sets $\mathscr{X}_k$, $k \geq 1$. We also consider a corresponding sequence of gambles h_k on $\mathscr{X}_k$ that is *uniformly bounded*, in the sense that there should exist some real number B such that $\sup h_k - \inf h_k \leq B$ for all $k \geq 1$. We also consider two sequences of real numbers m_k and M_k such that $\inf h_k \leq m_k \leq M_k \leq \sup h_k$ for all $k \geq 1$.

We shall be concentrating on an arbitrary sequence of coherent joint lower previsions $\underline{P}^N$ on $\mathscr{L}(\mathscr{X}^N)$, $N \geq 1$ such that

- (S1) $m_k \leq \underline{P}^N(h_k) \leq \overline{P}^N(h_k) \leq M_k$ for all $N \geq 1$ and $k = 1, \dots, N$, where $\overline{P}^N$ is the conjugate upper prevision of $\underline{P}^N$.
- (S2) For each $N \geq 1$, $\underline{P}^N$ is forward factorising. This implies that for these joint lower previsions $\underline{P}^N$ our weak law of large numbers in the form of Theorem 2 holds.
- (S3) This sequence of lower previsions is *consistent*: $\underline{P}^N$ is the $\mathscr{X}^N$-marginal of its successors $\underline{P}^{N+\ell}$, $\ell \geq 1$, $N \geq 1$.

As discussed in the previous section, these $\underline{P}^N$ could for instance be forward irrelevant natural extensions $\underline{E}^N$ of marginal assessments $\underline{P}_k(h_k) = m_k$ and $\overline{P}_k(h_k) = M_k$, $k \geq 1$, or their forward irrelevant products $\underline{M}^N$, (epistemically) independent products, type-1 products, products that are Kuznetsov independent, or even linear previsions in $\mathscr{M}(\underline{E}^N)$ or $\mathscr{M}(\underline{M}^N)$, … The conclusions that we shall reach further on will be valid for any such choice.

The first step in our study is necessarily the investigation of the behavioural consequences of the sequence of lower previsions $\underline{P}^N$. We shall summarise the information present in the $\underline{P}^N$, $N \geq 1$ by means of a coherent lower prevision $\underline{P}^{\mathbb{N}}$ defined on some set of gambles $\mathscr{K}^{\mathbb{N}} \subseteq \mathscr{L}(\mathscr{X}^{\mathbb{N}})$, where of course, $\mathscr{X}^{\mathbb{N}} = \times_{k \in \mathbb{N}} \mathscr{X}_k$ is the set of all maps (sequences) x from $\mathbb{N}$ to $\bigcup_{k \in \mathbb{N}} \mathscr{X}_k$ satisfying $x_k = x(k) \in \mathscr{X}_k$ for all $k \in \mathbb{N}$.

To this end, we must first define the following projection and extension operators. For any natural numbers $N_1 \leq N_2$, we define the *projection* proj_{N_2,N_1} by

$$\mathrm{proj}_{N_2,N_1} \colon \mathscr{X}^{N_2} \to \mathscr{X}^{N_1} : x \mapsto x|_{N_1},$$

where $x|_{N_1}$ is the element of $\mathscr{X}^{N_1}$ whose components are the first N_1 components of the N_2-tuple x. Similarly, we define the *cylindrical extension* $\mathrm{ext}_{N_1,N_2} \colon \mathscr{L}(\mathscr{X}^{N_1}) \to \mathscr{L}(\mathscr{X}^{N_2})$ as follows: for any $f \in \mathscr{L}(\mathscr{X}^{N_1})$, $\mathrm{ext}_{N_1,N_2}(f)$ is a gamble on $\mathscr{X}^{N_2}$, such that for any $x \in \mathscr{X}^{N_2}$, $\mathrm{ext}_{N_1,N_2}(f) \cdot x = f(\mathrm{proj}_{N_2,N_1}(x)) = f(x|_{N_1})$, or in other words, $\mathrm{ext}_{N_1,N_2}(f) = f \circ \mathrm{proj}_{N_2,N_1}$. Observe that $\mathrm{ext}_{N_1,N_2}(f)$ is essentially the same gamble as f, but defined on the larger space $\mathscr{X}^{N_2}$. In a completely similar way, we can define the operators $\mathrm{proj}_{\mathbb{N},N}$ and $\mathrm{ext}_{N,\mathbb{N}}$ for any natural number N. For instance, $\mathrm{proj}_{\mathbb{N},N}$ maps any sequence x in $\mathscr{X}^{\mathbb{N}}$ to the N-tuple containing its N first elements.

We now define a set of gambles $\mathscr{K}^{\mathbb{N}}$ on $\mathscr{X}^{\mathbb{N}}$ and a lower prevision $P^{\mathbb{N}}$ on $\mathscr{X}^{\mathbb{N}}$ as follows: a gamble f belongs to $\mathscr{K}^{\mathbb{N}}$ if and only if $f = \mathrm{ext}_{N,\mathbb{N}}(g)$ for some $g \in \mathscr{L}(\mathscr{X}^N)$

[7] The 'weak law' version of his result follows from our weak law, and the 'strong law' version from our discussion of the strong law in Section 5, and in particular the considerations about σ-additivity in Section 5.3.

and some $N \geq 1$, and then we define $\underline{P}^{\mathbb{N}}(f) = \underline{P}^N(g)$. This definition of $\underline{P}^{\mathbb{N}}$ is consistent, because we assumed that the sequence of lower previsions $\underline{P}^N$ is consistent. It is clear that $\underline{P}^{\mathbb{N}}$ represents exactly the same information as all the lower previsions $\underline{P}^N$ taken together. Observe that $\underline{P}^{\mathbb{N}}$ is coherent, because the $\underline{P}^N$ are. In order to find the behavioural consequences of all these lower previsions $\underline{P}^N$, we need to consider the natural extension $\underline{E}^{\mathbb{N}}$ of $\underline{P}^{\mathbb{N}}$ to the set $\mathscr{L}(\mathscr{X}^{\mathbb{N}})$ of all gambles on $\mathscr{X}^{\mathbb{N}}$. We show in the Appendix (Section A.4) that it is given by:

$$\underline{E}^{\mathbb{N}}(f) = \sup_{N\in\mathbb{N}} \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(f)) = \lim_{N\to\infty} \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(f)), \tag{9}$$

for all gambles f on $\mathscr{X}^{\mathbb{N}}$, and similarly,

$$\overline{E}^{\mathbb{N}}(f) = \inf_{N\in\mathbb{N}} \overline{P}^N(\overline{\mathrm{proj}}_{\mathbb{N},N}(f)) = \lim_{N\to\infty} \overline{P}^N(\overline{\mathrm{proj}}_{\mathbb{N},N}(f)), \tag{10}$$

where for any y in $\mathscr{X}^N$,

$$\underline{\mathrm{proj}}_{\mathbb{N},N}(f)(y) = \inf_{\mathrm{proj}_{\mathbb{N},N}(x)=y} f(x) \quad \text{and} \quad \overline{\mathrm{proj}}_{\mathbb{N},N}(f)(y) = \sup_{\mathrm{proj}_{\mathbb{N},N}(x)=y} f(x).$$

This natural extension can also be formulated in terms of linear previsions. Consider, for each $N \in \mathbb{N}$, the set

$$\mathscr{M}(\underline{P}^N) = \{P^N \in \mathbb{P}(\mathscr{X}^N) \colon (\forall f \in \mathscr{L}(\mathscr{X}^N))\underline{P}^N(f) \leq P^N(f)\}$$

of those linear previsions P^N on $\mathscr{L}(\mathscr{X}^N)$ that dominate the lower prevision $\underline{P}^N$ on its domain $\mathscr{L}(\mathscr{X}^N)$. The set $\mathscr{M}(\underline{P}^{\mathbb{N}}) = \mathscr{M}(\underline{E}^{\mathbb{N}})$ can be easily expressed in terms of the $\mathscr{M}(\underline{P}^N)$, by means of the following *marginalisation operators*: for any $N \in \mathbb{N}$, let the map $\mathrm{mar}_{\mathbb{N},N} \colon \mathbb{P}(\mathscr{X}^{\mathbb{N}}) \to \mathbb{P}(\mathscr{X}^N)$ be defined by $\mathrm{mar}_{\mathbb{N},N}(P^{\mathbb{N}}) = P^{\mathbb{N}} \circ \mathrm{ext}_{N,\mathbb{N}}$, for all $P^{\mathbb{N}}$ in $\mathbb{P}(\mathscr{X}^{\mathbb{N}})$, or in other words, the linear prevision $\mathrm{mar}_{\mathbb{N},N}(P^{\mathbb{N}})$ is the $\mathscr{X}^N$-marginal of the linear prevision $P^{\mathbb{N}}$.

Theorem 5. *Let $\underline{P}^N$, $N \geq 1$, be the sequence of coherent lower previsions considered above, and let $\underline{P}^{\mathbb{N}}$ be the equivalent lower prevision defined on the set $\mathscr{K}^{\mathbb{N}}$ of gambles on $\mathscr{X}^{\mathbb{N}}$. Then $\mathscr{M}(\underline{E}^{\mathbb{N}}) = \mathscr{M}(\underline{P}^{\mathbb{N}}) = \bigcap_{N\in\mathbb{N}} \mathrm{mar}_{\mathbb{N},N}^{-1}(\mathscr{M}(\underline{P}^N))$, where $\mathrm{mar}_{\mathbb{N},N}^{-1}(\mathscr{M}(\underline{P}^N))$ is the set of linear previsions on $\mathscr{L}(\mathscr{X}^{\mathbb{N}})$ whose $\mathscr{X}^N$-marginals belong to $\mathscr{M}(\underline{P}^N)$.*

In other words, the natural extension $\underline{E}^{\mathbb{N}}$ is the lower envelope of all the linear previsions on $\mathscr{L}(\mathscr{X}^{\mathbb{N}})$ whose $\mathscr{X}^N$-marginals belong to $\mathscr{M}(\underline{P}^N)$, i.e., dominate $\underline{P}^N$, for all $N \geq 1$. See the Appendix (section A.5) for a proof.

5.2. **Derivation of the strong law.**

5.2.1. *The classical case.* In order to get some idea about what it is we want to achieve, let us consider the classical law of large numbers for the uniformly bounded Borel measurable gambles h_k. Here it is assumed that each marginal lower prevision $\underline{P}_k$ is a σ-additive probability measure P_k defined on the Borel σ-field $\mathscr{B}_k$ on the bounded subset $\mathscr{X}_k$ of $\mathbb{R}$. The random variables are assumed to be stochastically independent, which means that for each $N \geq 1$, the behaviour of the random variable X^N is described by the product P^N of the marginals $P_1, \dots, P_N$, which is a σ-additive probability measure defined on the product σ-field $\mathscr{B}^N$ of the Borel σ-fields $\mathscr{B}_1, \dots, \mathscr{B}_N$. By the Daniell-Kolmogorov Theorem, these products P^N have a unique σ-additive extension P_σ to the σ-field $\mathscr{B}^{\mathbb{N}}$ generated by the field $\mathscr{F} = \bigcup_{N\geq 1} \mathrm{proj}_{\mathbb{N},N}^{-1}(\mathscr{B}^N)$ of all measurable cylinders.

Let $\mu_k := \int_{\mathscr{X}_k} h_k \,\mathrm{d}P_k$, and consider, for $\varepsilon > 0$, the following subsets of $\mathscr{X}^{\mathbb{N}}$:

$$\overline{\Delta}_{r,\varepsilon,p} := \left\{ \limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k - \mu_k) > \varepsilon \right\} = \left\{ x \in \mathscr{X}^{\mathbb{N}} \colon \limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k(x_k) - \mu_k) > \varepsilon \right\}$$

and

$$\overline{\Delta}_{\ell,\varepsilon,p} := \left\{ \liminf_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k - \mu_k) < -\varepsilon \right\} = \left\{ x \in \mathscr{X}^{\mathbb{N}} \colon \liminf_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k(x_k) - \mu_k) < -\varepsilon \right\}.$$

These are the sets of sequences whose averages are at a distance greater than ε from the averages of the means μ_k infinitely often. The classical strong law of large numbers for σ-additive probabilities establishes the almost sure (P_σ) convergence of $\frac{1}{N}\sum_{k=1}^{N}(h_k - \mu_k)$ to zero, or, equivalently, that for all $\varepsilon > 0$, $P_\sigma(\overline{\Delta}_{r,\varepsilon,p}) = P_\sigma(\overline{\Delta}_{\ell,\varepsilon,p}) = 0$.

5.2.2. *A first attempt.* Let us now return to the context established in Section 5.1. It has been established already some time ago (Walley and Fine, 1982) that in the case of imprecise marginals, we can't expect almost sure convergence of the sequence $\frac{1}{N}\sum_{k=1}^{N} h_k$, even under much stronger independence conditions than the forward factorisation (related to forward epistemic irrelevance) we are imposing here. For this reason, we shall look at the limit inferior and the limit superior of such sequences separately, and consider the sets

$$\overline{\Delta}_{r,\varepsilon} := \left\{ \limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k - M_k) > \varepsilon \right\} \text{ and } \overline{\Delta}_{\ell,\varepsilon} := \left\{ \liminf_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k - m_k) < -\varepsilon \right\},$$

where $\varepsilon > 0$. These are the sets of sequences whose averages are at a distance greater than ε from the averages of the bounds infinitely often. At first sight, our goal would then be to prove that for all ε:

$$\overline{E}^{\mathbb{N}}(\overline{\Delta}_{\ell,\varepsilon}) = \overline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon}) = 0. \tag{11}$$

However, the following result tells us that, unless in trivial cases, natural extension is too weak for a strong law of large numbers to be formulated in terms of it. An earlier hint about such a result can be found in (Dubins, 1974). This result holds, regardless of the coherent lower previsions $\underline{P}^N$; they may even be linear previsions.

Theorem 6. *Let $\underline{P}^N$, $N \geq 1$, be a sequence of coherent lower previsions satisfying the conditions (S1)–(S3) in Section 5.1, and let $\underline{E}^{\mathbb{N}}$ be their natural extension to all gambles on $\mathscr{X}^{\mathbb{N}}$. Then*

(i) if $\overline{\Delta}_{r,\varepsilon} \neq \mathscr{X}^{\mathbb{N}}$ then $\underline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon}) = 0$ and if $\overline{\Delta}_{r,\varepsilon} \neq \emptyset$ then $\overline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon}) = 1$; and similarly

(ii) if $\overline{\Delta}_{\ell,\varepsilon} \neq \mathscr{X}^{\mathbb{N}}$ then $\underline{E}^{\mathbb{N}}(\overline{\Delta}_{\ell,\varepsilon}) = 0$ and if $\overline{\Delta}_{\ell,\varepsilon} \neq \emptyset$ then $\overline{E}^{\mathbb{N}}(\overline{\Delta}_{\ell,\varepsilon}) = 1$.

In other words, if $\overline{\Delta}_{\ell,\varepsilon}$ and $\overline{\Delta}_{r,\varepsilon}$ are proper subsets of $\mathscr{X}^{\mathbb{N}}$, then the natural extension $\underline{E}^{\mathbb{N}}$ is vacuous (completely uninformative) on these sets.

The behavioural interpretation of this result is the following: if a subject has lower previsions $\underline{P}^N$, $N \geq 1$ as described in Section 5.1, then coherence alone imposes no restrictions on his betting rates for the (proper) events $\overline{\Delta}_{\ell,\varepsilon}$ and $\overline{\Delta}_{r,\varepsilon}$: they can lie anywhere between zero and one. This obliterates all hope of formulating a strong law of large numbers involving the natural extension $\underline{E}^{\mathbb{N}}$ in the form (12). This happens even if the lower previsions $\underline{P}^N$ are linear. The formulation (12) of the strong law of large numbers will only hold if the sets $\overline{\Delta}_{r,\varepsilon}$ and $\overline{\Delta}_{\ell,\varepsilon}$ eventually become empty for ε small enough.[8]

[8]But as soon as $\sup h_k > M_k$ and $\inf h_k < m_k$ for all k the sets $\overline{\Delta}_{r,\varepsilon}$ and $\overline{\Delta}_{\ell,\varepsilon}$ will be non-empty for all $\varepsilon > 0$.

5.2.3. *Precise marginals.* It is instructive to interpret Theorem 6, in the classical case that all coherent joint lower previsions $\underline{P}^N$ are forward irrelevant products of σ-additive probability measures P_k. More precisely, we assume that we have marginal σ-additive probability measures P_k defined on a σ-field $\mathscr{B}_k$ on $\mathscr{X}_k$. Consider their product measure P^N defined on the product $\mathscr{B}^N$ of the σ-fields $\mathscr{B}_1, \ldots, \mathscr{B}_N$, then it is easy to see [use the results in (De Cooman and Miranda, 2006)] that the forward irrelevant product $\underline{M}^N$ of these probability measures will satisfy

$$\underline{M}^N(f) = \int_{\mathscr{X}_k} f \, \mathrm{d}P^N,$$

for all $\mathscr{B}^N$-measurable gambles f. In other words, the forward irrelevant product $\underline{M}^N$ of these marginals is a linear prevision on the set of all $\mathscr{B}^N$-measurable gambles. Observe that $\mathscr{M}(\underline{M}^N) \subseteq \mathscr{M}(P^N)$, where we denote by $\mathscr{M}(P^N)$ the set of all linear previsions on $\mathscr{L}(\mathscr{X}^N)$, or equivalently all *finitely additive* probabilities on $\wp(\mathscr{X}^N)$, that coincide with the probability measure P^N on $\mathscr{B}^N$. We denote by P_σ the unique σ-additive extension of the products P^N to the σ-field $\mathscr{B}^{\mathbb{N}}$. Because in this case $m_k = M_k = \mu_k = \int h_k \, \mathrm{d}P_k$, we find that $\overline{\Delta}_{r,\varepsilon,p} = \overline{\Delta}_{r,\varepsilon}$, and $\overline{\Delta}_{\ell,\varepsilon,p} = \overline{\Delta}_{\ell,\varepsilon}$. The classical strong law of large numbers therefore tells us that $P_\sigma(\overline{\Delta}_{r,\varepsilon}) = P_\sigma(\overline{\Delta}_{\ell,\varepsilon}) = 0$. We shall prove a more general result, see Eq. (13).

But we know by Theorem 5 that the natural extension $\underline{E}^{\mathbb{N}}$ of the forward irrelevant products $\underline{M}^N$ is the lower envelope of all *finitely additive extensions* to $\mathscr{X}^{\mathbb{N}}$ of the linear previsions in $\mathscr{M}(\underline{M}^N)$, and is therefore a lower envelope of finitely additive extensions of P^N to $\mathscr{X}^{\mathbb{N}}$. Among these extensions, P_σ is the only one that is actually σ-additive on $\mathscr{B}^{\mathbb{N}}$, and this P_σ is *zero* on the sets $\overline{\Delta}_{r,\varepsilon}$ and $\overline{\Delta}_{\ell,\varepsilon}$. It is an essential consequence of Theorem 6 that the other extensions in $\mathscr{M}(\underline{E}^{\mathbb{N}})$, which are only finitely additive and not σ-additive, can assume *any value between zero and one* on the sets $\overline{\Delta}_{r,\varepsilon}$ and $\overline{\Delta}_{\ell,\varepsilon}$. For discussions of similar phenomena in more general contexts, we refer to (Walley, 1991, Corollary 3.4.3) and (Bhaskara Rao and Bhaskara Rao, 1983, Section 3.3).

So we can surmise that the reason why we aren't able to prove a strong law involving the natural extension $\underline{E}^{\mathbb{N}}$, is that it is an extension based on the essentially *finitary* notion of coherence, and therefore a lower envelope of probabilities that are only guaranteed to be finitely additive. This leads us to the formulation of a weaker, finitary formulation of the strong law of large numbers. It was suggested by Dubins (1974) (see also de Finetti (1974–1975, Section 7.5) and Gnedenko (1975, Section 34)), as follows: For all $\varepsilon > 0$, there is an integer N such that, for all positive integers k,

$$P_\sigma\left(\left\{x \in \mathscr{X}^{\mathbb{N}} \colon (\exists n \in [N, N+k]) |\frac{1}{n}\sum_{\ell=1}^{n} x_\ell| > \varepsilon\right\}\right) < \varepsilon.$$

This finitary version also holds for all *finitely* additive extensions of the P^N, and it is therefore not really surprising that in our more general context, we can prove a generalisation.

5.2.4. *Finitary formulation of the strong law.*

Theorem 7 (Strong law – finitary version)**.** *Consider the sequence of coherent lower previsions $\underline{P}^N$, $N \geq 1$, introduced in Section 5.1 and satisfying (S1)–(S3), and let $\underline{E}^{\mathbb{N}}$ be their natural extension to all gambles on $\mathscr{X}^{\mathbb{N}}$. Then for every $\varepsilon > 0$, there is some $N(\varepsilon)$ such that for all integer $N \geq N(\varepsilon)$ and for all positive integer k,*

$$\overline{E}^{\mathbb{N}}\left(\bigcup_{n=N}^{N+k}\left\{\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell - M_\ell) > \varepsilon\right\}\right) < \varepsilon \text{ and } \overline{E}^{\mathbb{N}}\left(\bigcup_{n=N}^{N+k}\left\{\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell - m_\ell) < -\varepsilon\right\}\right) < \varepsilon.$$

The behavioural interpretation of this law, loosely speaking, is the following: given a sequence of uniformly bounded gambles h_k on random variables X_k with associated upper previsions M_k, the requirements of coherence and epistemic irrelevance together imply that we should bet at rates greater than $1-\varepsilon$ against the N-th average of these variables exceeding the N-th average of their upper previsions by more than ε, if this average is considered for a 'sufficiently large' number of variables.

As we can see from the proof in Section A.7 of the Appendix, the $N(\varepsilon)$ we found essentially increases as ε^{-2} for small enough ε. Perhaps surprisingly, this strong law follows from the weak one: In the proof we use a method that resembles the (Borel-)Cantelli lemma, which involves the summation of the bounds $\exp(-k\varepsilon^2/4B^2)$ of Theorem 2 for all k greater than some N. A similar course of reasoning using the Chebychev bound $1/k\varepsilon^2$ wouldn't allow us to establish this result, as the corresponding sum is infinite.

5.3. **Further comments on the strong law.** Let us show next how Theorem 7 subsumes a number of strong laws in the literature: (i) the classical strong law for σ-additive probabilities; (ii) the strong law for independent and indistinguishable distributions by Epstein and Schneider (2003); and (iii) the strong law for capacities established by Maccheroni and Marinacci (2005).

Consider the case that there are elements P_σ of $\mathscr{M}(\underline{E}^{\mathbb{N}})$ that are σ-additive on the σ-field $\mathscr{B}^{\mathbb{N}}$. Then Theorem 7 implies that:

$$P_\sigma\left(\left\{\limsup_n \frac{1}{n}\sum_{\ell=1}^n (h_\ell - M_\ell) > 0\right\}\right) = P_\sigma\left(\left\{\liminf_n \frac{1}{n}\sum_{\ell=1}^n (h_\ell - m_\ell) < 0\right\}\right) = 0; \quad (12)$$

see the Appendix, Section A.8 for a proof. Consequently, if we denote by $\overline{E}_\sigma^{\mathbb{N}}$ the upper envelope of all the elements of $\mathscr{M}(\underline{E}^{\mathbb{N}})$ that are σ-additive on the σ-field $\mathscr{B}^{\mathbb{N}}$, we get

$$\overline{E}_\sigma^{\mathbb{N}}\left(\left\{\limsup_n \frac{1}{n}\sum_{\ell=1}^n (h_\ell - M_\ell) > 0\right\}\right) = \overline{E}_\sigma^{\mathbb{N}}\left(\left\{\liminf_n \frac{1}{n}\sum_{\ell=1}^n (h_\ell - m_\ell) < 0\right\}\right) = 0. \quad (13)$$

So we can get to a form of the strong law in the spirit of the discussion in Section 5.2.2 (see Eq. (12)) if we restrict ourselves to the linear previsions in $\mathscr{M}(\underline{E}^{\mathbb{N}}) = \mathscr{M}(\underline{P}^{\mathbb{N}})$ that are σ-additive on $\mathscr{B}^{\mathbb{N}}$.

If in particular $m_k = M_k = m$ for all $k \geq 1$, this tells us that the sequence $\frac{1}{n}\sum_{k=1}^n X_k$ converges almost surely (P_σ) to m, for all σ-additive extensions P_σ to $\mathscr{B}^{\mathbb{N}}$ of the linear previsions P^N, $N \geq 1$, constructed from the marginal linear previsions in the sets

$$\mathscr{M}(\underline{P}_k) = \{P_k \colon P_k(X_k) = m\}, \quad k = 1, \dots, N$$

using the construction mentioned in Section 4.2. This shows that the finitary version established in our last theorem is indeed a significant extension of the classical strong law of large numbers for bounded random variables.

This discussion also shows that the results in this section are more general than the strong law of large numbers for so-called 'independent and indistinguishably distributed variables' proven by Epstein and Schneider (2003). These authors have essentially proven that Eq. (14) holds in the special case where the gambles h_k are identity maps on random variables X_k that may assume only a *finite* number of values, and the joint lower previsions $\underline{P}^N$ are forward irrelevant natural extensions/products of *identical* coherent marginal lower previsions $\underline{P}_k$ that satisfy $\underline{P}_k(X_k) \geq m$ and $\overline{P}_k(X_k) \leq M$, and that moreover are 2-*monotone* (or super-modular). Our analysis shows that none of these extra (italicised) requirements is essential.

In a similar vein, this analysis allows us to prove in a simple way a significantly more general version of a law of large numbers for completely (or totally) monotone capacities that was first proven by Maccheroni and Marinacci (2005). [9]

Indeed, consider a sequence of random variables X_k taking values in the respective sets $\mathscr{X}_k$. Each set $\mathscr{X}^k$ is provided with a topology, and with a Borel σ-field $\mathscr{B}_k$ generated by the open sets in that topology. The product σ-field $\mathscr{B}^{\mathbb{N}}$ of the Borel σ-fields $\mathscr{B}_k$ is actually the Borel σ-field generated by the open sets in the product topology on $\mathscr{X}^{\mathbb{N}}$. Consider uniformly bounded and Borel measurable gambles h_k on $\mathscr{X}_k$, and a lower probability $\underline{P}^{\mathbb{N}}$ defined on the product σ-field $\mathscr{B}^{\mathbb{N}}$, that satisfies the following properties.

(M1) $\underline{P}^{\mathbb{N}}(\emptyset)=0$ and $\underline{P}^{\mathbb{N}}(\mathscr{X}^{\mathbb{N}})=1$.

(M2) $\underline{P}^{\mathbb{N}}$ is 2-monotone, meaning that for any A and B in $\mathscr{B}^{\mathbb{N}}$, (i) $\underline{P}^{\mathbb{N}}(A)\leq\underline{P}^{\mathbb{N}}(B)$ if $A\subseteq B$; and (ii) $\underline{P}^{\mathbb{N}}(A\cup B)+\underline{P}^{\mathbb{N}}(A\cap B)\geq\underline{P}^{\mathbb{N}}(A)+\underline{P}^{\mathbb{N}}(B)$.

(M3) $\underline{P}^{\mathbb{N}}(A\times B)\geq\underline{P}^{\mathbb{N}}(A)\times\underline{P}^{\mathbb{N}}(B)$ for all $\mathscr{B}^{k-1}$-measurable subsets A of $\mathscr{X}^{k-1}$ and all $\mathscr{B}_k$-measurable subsets of $\mathscr{X}_k$, and all $k\geq 1$.

(M4) $\underline{P}^{\mathbb{N}}(B_n)\uparrow\underline{P}^{\mathbb{N}}(\mathscr{X}^{\mathbb{N}})$ when B_n is an increasing sequence of elements of $\mathscr{B}^{\mathbb{N}}$ that converges to $\mathscr{X}^{\mathbb{N}}$.

It follows from a result by Walley (1981) that (M1) and (M2) imply that $\underline{P}^{\mathbb{N}}$ is a coherent lower probability on $\mathscr{B}^{\mathbb{N}}$, and that its natural extension $\underline{E}^{\mathbb{N}}$ to gambles is given by

$$\underline{E}^{\mathbb{N}}(f):=\inf f+\int_{\inf f}^{\sup f}\underline{P}_*^{\mathbb{N}}(\{f>\alpha\})\,\mathrm{d}\alpha$$

for all $f\in\mathscr{L}(\mathscr{X}^{\mathbb{N}})$, where the lower probability $\underline{P}_*^{\mathbb{N}}$ is defined on all subsets of $\mathscr{X}^{\mathbb{N}}$ as the inner set function of $\underline{P}^{\mathbb{N}}$, meaning that

$$\underline{P}_*^{\mathbb{N}}(A):=\sup\{\underline{P}^{\mathbb{N}}(B)\colon B\subseteq A, B\in\mathscr{B}^{\mathbb{N}}\}$$

for all $A\subseteq\mathscr{X}^{\mathbb{N}}$. In particular, we find for the Borel measurable gambles h_k on $\mathscr{X}_k$ that

$$m_k:=\underline{E}^{\mathbb{N}}(h_k)=\inf h_k+\int_{\inf h_k}^{\sup h_k}\underline{P}_*^{\mathbb{N}}(\{h_k>\alpha\})\,\mathrm{d}\alpha=\inf h_k+\int_{\inf h_k}^{\sup h_k}\underline{P}^{\mathbb{N}}(\{h_k>\alpha\})\,\mathrm{d}\alpha,$$

and similarly

$$M_k:=\overline{E}^{\mathbb{N}}(h_k)=\inf h_k+\int_{\inf h_k}^{\sup h_k}\overline{P}^{\mathbb{N}}(\{h_k>\alpha\})\,\mathrm{d}\alpha.$$

Theorem 8 (Strong law of large numbers for capacities)**.** *Consider any lower probability $\underline{P}^{\mathbb{N}}$ on $\mathscr{B}^{\mathbb{N}}$ that satisfies (M1)–(M4). Then the $\mathscr{X}^N$-marginals $\underline{E}^N$ of its natural extension $\underline{E}^{\mathbb{N}}$ to all gambles on $\mathscr{X}^{\mathbb{N}}$ satisfy (S1)–(S3). Hence, for every $\varepsilon>0$, there is some $N(\varepsilon)$ such that for all $N\geq N(\varepsilon)$ and all $k>0$ in $\mathbb{N}$, $\overline{P}^{\mathbb{N}}(\bigcup_{n=N}^{N+k}\{\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell-M_\ell)>\varepsilon\})<\varepsilon$, and $\overline{P}^{\mathbb{N}}(\bigcup_{n=N}^{N+k}\{\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell-m_\ell)<-\varepsilon\})<\varepsilon$. Moreover, for any σ-additive probability P_σ in $\mathscr{M}(\underline{P}^{\mathbb{N}})$ we see that Eq.* (13) *holds.*

Under some additional conditions, $\underline{P}^{\mathbb{N}}$ will actually be the lower envelope of all σ-additive probabilities P_σ in $\mathscr{M}(\underline{P}^{\mathbb{N}})$, which means that $\overline{P}^{\mathbb{N}}(\{\limsup_n\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell-M_\ell)>0\})=0$, and similarly $\overline{P}^{\mathbb{N}}(\{\liminf_n\frac{1}{n}\sum_{\ell=1}^{n}(h_\ell-m_\ell)<0\})=0$. The additional conditions listed by Maccheroni and Marinacci are sufficient to prove just that lower envelope result.

---

[9] These authors require, in addition to complete monotonicity, some additional continuity properties that we don't need here. See Maccheroni and Marinacci (2005, Section 2) for more details.

6. Discussion

Why do we believe that the results presented here merit attention?

First of all, our sufficient condition for the existence of a weak law (forward factorisation) is really weak. We only require, loosely speaking, that if we consider a non-negative mixture of a gamble with lower prevision 0, this mixture has non-negative lower prevision. We have shown that this holds in particular when we make marginal assessments about a number of variables and consider an assumption of forward epistemic irrelevance.

But this (stronger) assumption of forward epistemic irrelevance, is by itself already quite weak. It only requires that the lower and upper previsions for the variables X_k (and these two *numbers* alone) don't change after observing values of the previous variables. We can of course deduce weak laws of large numbers from it under stronger conditions, such as epistemic independence, or independence on the Bayesian sensitivity analysis interpretation, which leads to type-1 (or strong) products: any of these assessments will provide us with a more specific, or equivalently less conservative, upper and lower probability model. This implies that our formulation of the laws of large numbers is more general, or is based on weaker assumptions, than a number of other formulations in the literature, in particular work by Walley and Fine (1982), Kuznetsov (1991), and Epstein and Schneider (2003).

The classical formulation of the laws of large numbers requires the measurability of the involved random variables with respect to some σ-field on the initial space, and the σ-additivity of the probability measure defined on that space. The approach followed in this paper weakens also these two requirements by working within the behavioural theory of imprecise probabilities. We consider lower and upper previsions defined on sets of gambles which needn't have any predefined structure, and in particular show that measurability isn't necessary in order to derive laws of large numbers. Nevertheless, we show when deriving the strong law that it is not difficult to define a suitable σ-field if we want to relate our results to the more classical ones.

The suppression of the hypothesis of σ-additivity in favour of the finite additivity is, however, more involved. Although the weak law can be derived without any continuity assumption, in the case of the strong law, where we study the behaviour of infinite sequences, we find that the lower and upper previsions modelling our behavioural assessments are in general too conservative. It turns out that we must replace the classical version of the strong law by a finitary formulation which, in the case of σ-additive probabilities, is equivalent to it. With hindsight, it is only logical that, under the general coherentist (finitary) setting considered in this paper, a strong law of large numbers for upper and lower previsions, or sets of finitely additive probabilities, can only be formulated in terms of the behaviour of finite sequences (albeit arbitrarily large), because it is this behaviour that is used to determine the natural extension on $\mathscr{X}^{\mathbb{N}}$. It is also interesting to remark that this finitary version can be proven because we have tightened, in our formulation of the weak law, the bounds derived in the classical approach using Chebychev's inequality.

Finally, we want to point out that there is one important limitation for our laws of large numbers: the variables X_k should be *bounded*. The main reason for this is that, so far, the theory of coherent lower previsions is only able to deal with gambles, i.e., bounded functions of a random variable. However, there are some recent developments (Troffaes and De Cooman, 2002a,b) that deal with extending the theory of lower previsions in order to deal with unbounded functions of random variables, so there is some hope at least that the limitation of boundedness will eventually be overcome.

APPENDIX A. PROOFS OF THEOREMS

In this appendix, we have gathered all of the more technical developments. We believe the actual proofs of results related to the weak and strong laws are fairly straightforward.

A.1. **Proof of Theorem 1.** We first prove that the condition (3) is necessary. Consider any $f \in \mathscr{L}(\mathscr{X}^N)$, $n_k \geq 0$, $h_{kj_k} \in \mathscr{L}(\mathscr{X}_k)$, $g_{kj_k} \in \mathscr{L}_+(\mathscr{X}^{k-1})$, and $m_{kj_k} \leq \underline{P}^N(h_{kj_k})$, where $j_k \in \{1\dots,n_k\}$ and $k \in \{1,\dots,N\}$. Define the gambles g and h on $\mathscr{X}^N$ by $g(x) := \sum_{k=1}^N \sum_{j_k=1}^{n_k} g_{kj_k}(x_1,\dots,x_{k-1})[h_{kj_k}(x_k) - m_{kj_k}]$ for all $x \in \mathscr{X}^N$ and $h := f - g$. We have to prove that $\underline{P}^N(f) \geq \inf h$. Indeed:

$$\underline{P}^N(f) = \underline{P}^N(h+g) \geq \underline{P}^N(h) + \sum_{k=1}^N \sum_{j_k=1}^{n_k} \underline{P}^N(g_{kj_k}[h_{kj_k} - m_{kj_k}])$$

$$\geq \inf h + \sum_{k=1}^N \sum_{j_k=1}^{n_k} \underline{P}^N(g_{kj_k}[h_{kj_k} - \underline{P}^N(h_{kj_k})]) \geq \inf h,$$

where the first inequality follows from the coherence [use (C2)] of $\underline{P}^N$, the second also from the coherence of $\underline{P}^N$ [use (C1) to see that $\underline{P}^N(h) \geq \inf h$, and use (C1) and (C2) to show that $\underline{P}^N$ is monotone, whence $\underline{P}^N(g_{kj_k}[h_{kj_k} - m_{kj_k}]) \geq \underline{P}^N(g_{kj_k}[h_{kj_k} - \underline{P}^N(h_{kj_k})])$], and the last inequality from the fact that $\underline{P}^N$ is forward factorising.

Next, we show that condition (3) is sufficient. Take k in $\{1,\dots,N\}$, g in $\mathscr{L}_+(\mathscr{X}^{k-1})$ and h in $\mathscr{L}(\mathscr{X}_k)$. In (3), let $f := g[h - \underline{P}^N(h)]$, let all $n_\ell := 0$ for all $\ell \neq k$, let $n_k := 1$, $h_{k1} := h$, $g_{k1} := g$ and $m_{k1} = \underline{P}^N(g_{k1}) = \underline{P}^N(g)$, then we find that indeed

$$\underline{P}^N(g[h - \underline{P}^N(h)]) \geq \inf \left[g[h - \underline{P}^N(h)] - g[h - \underline{P}^N(h)]\right] = 0.$$

A.2. **Proof of Theorem 2.** Fix $N \geq 1$ and $\varepsilon > 0$. Observe that

$$-B \leq h_k(x_k) - m_k \leq B \text{ and } -B \leq h_k(x_k) - M_k \leq B. \tag{14}$$

This implies that we may restrict ourselves without loss of generality to the case $\varepsilon < B$, since for other values the proof is trivial. It is easier to work with the conjugate upper prevision $\overline{P}$ of $\underline{P}$. Consider the sets

$$\Delta_{r,\varepsilon,N} := \left\{ \frac{1}{N} \sum_{k=1}^N h_k \leq \frac{1}{N} \sum_{k=1}^N M_k + \varepsilon \right\} \quad \text{and} \quad \Delta_{\ell,\varepsilon,N} := \left\{ \frac{1}{N} \sum_{k=1}^N m_k - \varepsilon \leq \frac{1}{N} \sum_{k=1}^N h_k \right\}.$$

Let $\Delta_{\varepsilon,N} = \Delta_{r,\varepsilon,N} \cap \Delta_{\ell,\varepsilon,N}$. Consider $\overline{P}(\Delta_{r,\varepsilon,N}^c) = 1 - \underline{P}(\Delta_{r,\varepsilon,N})$, for which we have, by applying Theorem 1 in its conjugate form to this special case, that

$$\overline{P}(\Delta_{r,\varepsilon,N}^c) \leq \sup_{x \in \mathscr{X}^N} \left[I_{\Delta_{r,\varepsilon,N}^c}(x) + \sum_{k=1}^N g_k(x_1,\dots,x_{k-1})(M_k - h_k(x_k)) \right],$$

for any choice of the non-negative gambles g_k on $\mathscr{X}^{k-1}$. We construct an upper bound for $\overline{P}(\Delta_{r,\varepsilon,N}^c)$ by judiciously choosing the functions g_k. Our choice is inspired by a combination of ideas discussed in (Shafer and Vovk, 2001, Lemma 3.3). Fix $\beta \geq 0$ and $\delta > 0$, let $g_1 := \delta\beta$ and let, for all $2 \leq k \leq N$ and $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$, $g_k(x_1,\dots,x_{k-1}) := \delta\beta \prod_{i=1}^{k-1}[1 + \delta(h_i(x_i) - M_i)]$. Then it follows after some manipulations that

$$\sum_{k=1}^N g_k(x_1,\dots,x_{k-1})(M_k - h_k(x_k)) = \beta - \beta \prod_{k=1}^N [1 + \delta(h_k(x_k) - M_k)].$$

Recalling (15), we see that if we let $0 < \delta < \frac{1}{2B}$, then all the g_k are guaranteed to be non-negative (as well as bounded), and they can be used to calculate an upper bound for $\overline{P}(\Delta^c_{r,\varepsilon,N})$. We then find that

$$\overline{P}(\Delta^c_{r,\varepsilon,N}) \leq \beta + \sup_{x\in\mathscr{X}^N}\left[I_{\Delta^c_{r,\varepsilon,N}}(x) - \beta\prod_{k=1}^{N}[1+\delta(h_k(x_k)-M_k)]\right].$$

If for $\beta \geq 0$ the supremum on the right hand side is non-positive, we have that $\overline{P}(\Delta^c_{r,\varepsilon,N}) \leq \beta$; hence, we can use such β to provide an upper bound for $\overline{P}(\Delta^c_{r,\varepsilon,N})$. The only condition to be imposed on β is therefore that if $x \in \Delta^c_{r,\varepsilon,N}$, or in other words if

$$\frac{1}{N}\sum_{k=1}^{N}(h_k(x_k)-M_k) > \varepsilon, \tag{15}$$

then we must have $\beta\prod_{k=1}^{N}[1+\delta(h_k(x_k)-M_k)] \geq 1$, or equivalently,

$$\ln\beta + \sum_{k=1}^{N}\ln[1+\delta(h_k(x_k)-M_k)] \geq 0. \tag{16}$$

Since $\ln(1+x) \geq x - x^2$ for $x > -\frac{1}{2}$, and since $\delta(h_k(x_k)-M_k) \geq -\delta B > -\frac{1}{2}$ by our previous choice for δ, we find, also using Eq. (15),

$$\begin{aligned}\sum_{k=1}^{N}\ln[1+\delta(h_k(x_k)-M_k)] &\geq \sum_{k=1}^{N}\delta(h_k(x_k)-M_k) - \sum_{k=1}^{N}[\delta(h_k(x_k)-M_k)]^2\\ &\geq \delta\sum_{k=1}^{N}(h_k(x_k)-M_k) - \delta^2NB^2\\ &= N\delta\left[\frac{1}{N}\sum_{k=1}^{N}(h_k(x_k)-M_k) - B^2\delta\right].\end{aligned}$$

Recalling (16), we find that for all $x \in \Delta^c_{r,\varepsilon,N}$, $\sum_{k=1}^{N}\ln[1+\delta(h_k(x_k)-M_k)] > N\delta(\varepsilon - B^2\delta)$. Choose β such that $\ln\beta + N\delta(\varepsilon - B^2\delta) \geq 0$, or equivalently $\beta \geq \exp(-N\delta(\varepsilon - B^2\delta))$, then requirement (17) will indeed be satisfied for $x \in \Delta^c_{r,\varepsilon,N}$. The tightest bound is achieved for $\delta = \frac{\varepsilon}{2B^2}$, so we see that $\overline{P}(\Delta^c_{r,\varepsilon,N}) \leq \exp(-\frac{N\varepsilon^2}{4B^2})$. We previously required that $\delta < \frac{1}{2B}$, so if we use this value for δ, we find that we have indeed proven this inequality for $\varepsilon < B$. In a similar way, we can prove that $\overline{P}(\Delta^c_{\ell,\varepsilon,N}) \leq \exp(-\frac{N\varepsilon^2}{4B^2})$, and therefore, using the coherence (sub-additivity) of $\overline{P}$,

$$\overline{P}(\Delta^c_{\varepsilon,N}) = \overline{P}(\Delta^c_{r,\varepsilon,N}\cup\Delta^c_{\varepsilon,N}) \leq \overline{P}(\Delta^c_{\ell,\varepsilon,N}) + \overline{P}(\Delta^c_{r,\varepsilon,N}) \leq 2\exp\left(-\frac{N\varepsilon^2}{4B^2}\right).$$

By passing back to the lower prevision $\underline{P}$ we find that indeed $\underline{P}(\Delta_{\varepsilon,N}) \geq 1 - 2\exp(-\frac{N\varepsilon^2}{4B^2})$.

A.3. **Proof of the claims about Kuznetsov independence in Section 4.4.** Consider any non-negative gamble g on $\mathscr{X}_1$ and any gamble h on $\mathscr{X}_2$. Then $0 \leq \underline{P}^2(g) \leq \overline{P}^2(g)$ by coherence, and also $\underline{P}^2(h - \underline{P}(h)) = \underline{P}^2(h) - \underline{P}^2(h) = 0$ and $\overline{P}^2(h - \underline{P}^2(h)) = \overline{P}^2(h) - \underline{P}^2(h) \geq 0$. Consequently it follows from the Kuznetsov independence of X_1 and X_2 that $\underline{P}^2(g[h - \underline{P}(h)])$ is equal to

$$\begin{aligned}&\min\{\underline{P}^2(g)\underline{P}^2(h-\underline{P}(h)),\underline{P}^2(g)\overline{P}^2(h-\underline{P}(h)),\overline{P}^2(g)\underline{P}^2(h-\underline{P}(h)g),\overline{P}^2(g)\overline{P}^2(h-\underline{P}(h))\}\\ &\qquad= \min\{0,\underline{P}^2(g)\overline{P}^2(h-\underline{P}^2(h),\overline{P}^2(g)\overline{P}^2(h-\underline{P}(h)) = 0,\end{aligned}$$

so $\underline{P}^2$ is forward factorising.

A.4. **Proof of Eq.** (10)**.** We first prove the formula for $\underline{E}^{\mathbb{N}}(h)$. Since $\underline{E}^{\mathbb{N}}$ is the natural extension of $\underline{P}^{\mathbb{N}}$, we have by definition that

$$\underline{E}^{\mathbb{N}}(h) = \sup_{\substack{f_k\in\mathscr{K}^{\mathbb{N}},\lambda_k\geq 0\\ k=1,\dots,n,n\geq 0}} \inf_{x\in\mathscr{X}^{\mathbb{N}}} \left[h(x) - \sum_{k=1}^{n} \lambda_k [f_k(x) - \underline{P}^{\mathbb{N}}(f_k)]\right].$$

Now it follows from the definition of $\underline{P}^{\mathbb{N}}$ and $\mathscr{K}^{\mathbb{N}}$ that for any choice of the natural number $n\geq 0$ and gambles f_1, …, f_n in $\mathscr{K}^{\mathbb{N}}$, there are natural numbers m_1, …, m_n and gambles g_1 on $\mathscr{X}^{m_1}$, …, g_n on $\mathscr{X}^{m_n}$ such that $f_k = \mathrm{ext}_{m_k,\mathbb{N}}(g_k)$ and $\underline{P}^{\mathbb{N}}(f_k) = \underline{P}^{m_k}(g_k)$, for $k = 1,\dots,n$. Let $M = \max\{m_1,\dots,m_n\}$, and consider the gambles $h_1 = \mathrm{ext}_{m_1,M}(g_1),\dots,h_n = \mathrm{ext}_{m_n,M}(g_n)$ on $\mathscr{X}^M$. Then obviously

$$f_k = \mathrm{ext}_{m_k,\mathbb{N}}(g_k) = \mathrm{ext}_{M,\mathbb{N}}(\mathrm{ext}_{m_k,M}(g_k)) = \mathrm{ext}_{M,\mathbb{N}}(h_k) = h_k\circ \mathrm{proj}_{\mathbb{N},M},$$

and the consistency of the joint lower previsions $\underline{P}^N$, $N\geq 1$, implies the equality $\underline{P}^M(h_k) = \underline{P}^{m_k}(g_k) = \underline{P}^{\mathbb{N}}(f_k)$, for $k = 1,\dots,n$. As a result, we find that

$$\begin{aligned}
&\inf_{x\in\mathscr{X}^{\mathbb{N}}}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[f_k(x) - \underline{P}^{\mathbb{N}}(f_k)\right]\right]\\
&\qquad = \inf_{x\in\mathscr{X}^{\mathbb{N}}}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[h_k(\mathrm{proj}_{\mathbb{N},M}(x)) - \underline{P}^M(h_k)\right]\right]\\
&\qquad = \inf_{y\in\mathscr{X}^M}\inf_{x\in\mathrm{proj}^{-1}_{\mathbb{N},M}(\{y\})}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[h_k(\mathrm{proj}_{\mathbb{N},M}(x)) - \underline{P}^M(h_k)\right]\right]\\
&\qquad = \inf_{y\in\mathscr{X}^M}\left[\inf_{x\in\mathrm{proj}^{-1}_{\mathbb{N},M}(\{y\})} h(x) - \sum_{k=1}^{n}\lambda_k\left[h_k(y) - \underline{P}^M(h_k)\right]\right]\\
&\qquad \leq \underline{P}^M(\underline{\mathrm{proj}}_{\mathbb{N},M}(h)) \leq \sup_{N\in\mathbb{N}}\underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(h)),
\end{aligned}$$

where the next to last inequality follows from the coherence of $\underline{P}^M$ and the definition of $\underline{\mathrm{proj}}_{\mathbb{N},M}(h)$. Consequently $\underline{E}^{\mathbb{N}}(h)\leq \sup_{N\in\mathbb{N}}\underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(h))$. Let us prove the converse inequality. From the definition of $\underline{P}^{\mathbb{N}}$ and $\mathscr{K}^{\mathbb{N}}$, we have, for any natural number $N\geq 1$,

$$\begin{aligned}
\underline{E}^{\mathbb{N}}(h) &= \sup_{\substack{f_k\in\mathscr{K}^{\mathbb{N}},\lambda_k\geq 0\\ k=1,\dots,n,n\geq 0}} \inf_{x\in\mathscr{X}^{\mathbb{N}}}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[f_k(x) - \underline{P}^{\mathbb{N}}(f_k)\right]\right]\\
&\geq \sup_{\substack{f_k\in\mathrm{ext}_{N,\mathbb{N}}(\mathscr{L}(\mathscr{X}^N)),\lambda_k\geq 0\\ k=1,\dots,n,n\geq 0}} \inf_{x\in\mathscr{X}^{\mathbb{N}}}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[f_k(x) - \underline{P}^{\mathbb{N}}(f_k)\right]\right]\\
&= \sup_{\substack{g_k\in\mathscr{L}(\mathscr{X}^N),\lambda_k\geq 0\\ k=1,\dots,n,n\geq 0}} \inf_{x\in\mathscr{X}^{\mathbb{N}}}\left[h(x) - \sum_{k=1}^{n}\lambda_k\left[g_k(\mathrm{proj}_{\mathbb{N},N}(x)) - \underline{P}^N(g_k)\right]\right]\\
&= \sup_{\substack{g_k\in\mathscr{L}(\mathscr{X}^N),\lambda_k\geq 0\\ k=1,\dots,n,n\geq 0}} \inf_{y\in\mathscr{X}^N}\left[\underline{\mathrm{proj}}_{\mathbb{N},N}(h)(y) - \sum_{k=1}^{n}\lambda_k\left[g_k(y) - \underline{P}^N(g_k)\right]\right]\\
&= \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(h)),
\end{aligned}$$

where the last equality follows from the fact that the lower prevision $\underline{P}^N$ is coherent on $\mathscr{L}(\mathscr{X}^N)$, and therefore coincides with its natural extension on this domain. Consequently, $\underline{E}^{\mathbb{N}}(h) \geq \sup_{N\in\mathbb{N}} \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(h))$. This completes the proof for $\underline{E}^{\mathbb{N}}$. The proof for $\overline{E}^{\mathbb{N}}$ follows immediately by conjugacy.

A.5. **Proof of Theorem 5.** A crucial step in the proof lies in the observation that since $\underline{P}^{\mathbb{N}}$ is a coherent lower prevision, we have that $\mathscr{M}(\underline{P}^{\mathbb{N}}) = \mathscr{M}(\underline{E}^{\mathbb{N}})$. Consider a linear prevision $P^{\mathbb{N}}$ on $\mathscr{L}(\mathscr{X}^{\mathbb{N}})$. Then $P^{\mathbb{N}} \in \mathscr{M}(\underline{P}^{\mathbb{N}}) = \mathscr{M}(\underline{E}^{\mathbb{N}})$ if and only if

$$(\forall f \in \mathscr{L}(\mathscr{X}^{\mathbb{N}}))P^{\mathbb{N}}(f) \geq \underline{E}^{\mathbb{N}}(f) \tag{17}$$

and this is equivalent to $(\forall f \in \mathscr{L}(\mathscr{X}^{\mathbb{N}}))(\forall N \in \mathbb{N})P^{\mathbb{N}}(f) \geq \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(f))$. Observe that for any $N \in \mathbb{N}$, for any gamble f on $\mathscr{X}^{\mathbb{N}}$ and for any gamble g on $\mathscr{X}^N$, we also have that $f \geq \mathrm{ext}_{N,\mathbb{N}}(\underline{\mathrm{proj}}_{\mathbb{N},N}(f))$ and $g = \underline{\mathrm{proj}}_{\mathbb{N},N}(\mathrm{ext}_{N,\mathbb{N}}(g))$. So $(\forall f \in \mathscr{L}(\mathscr{X}^{\mathbb{N}}))P^{\mathbb{N}}(f) \geq \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(f))$ is equivalent to $(\forall g \in \mathscr{L}(\mathscr{X}^N))P^{\mathbb{N}}(\mathrm{ext}_{N,\mathbb{N}}(g)) \geq \underline{P}^N(g)$. Since, by definition, $\mathrm{mar}_{\mathbb{N},N}(P^{\mathbb{N}}) = P^{\mathbb{N}} \circ \mathrm{ext}_{N,\mathbb{N}}$, we see that (18) is equivalent to $(\forall N \in \mathbb{N})\,\mathrm{mar}_{\mathbb{N},N}(P^{\mathbb{N}}) \in \mathscr{M}(\underline{P}^N)$, and this completes the proof.

A.6. **Proof of Theorem 6.** We give the proof for $\overline{\Delta}_{r,\varepsilon}$ and for lower probabilities. The rest of the proof is analogous. Observe that, because of Eq. (10),

$$\underline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon}) = \sup_{N\in\mathbb{N}} \underline{P}^N(\underline{\mathrm{proj}}_{\mathbb{N},N}(\overline{\Delta}_{r,\varepsilon})),$$

where for all $N \geq 1$, $\underline{\mathrm{proj}}_{\mathbb{N},N}(\overline{\Delta}_{r,\varepsilon}) = \{y \in \mathscr{X}^N \colon \overline{\Delta}^c_{r,\varepsilon} \cap \mathrm{proj}^{-1}_{\mathbb{N},N}(\{y\}) = \emptyset\}$. First, assume that $\overline{\Delta}_{r,\varepsilon} \neq \mathscr{X}^{\mathbb{N}}$, or equivalently, that $\overline{\Delta}^c_{r,\varepsilon} \neq \emptyset$, and consider any x in $\overline{\Delta}^c_{r,\varepsilon}$, then

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (h_k(x_k) - M_k) \leq \varepsilon.$$

Fix $N \geq 1$ and consider any $y \in \mathscr{X}^N$. If we replace the first N components of x by y, then this does not affect the lim sup in the above expression. If we denote the sequence thus obtained from x by x', this means that $x' \in \overline{\Delta}^c_{r,\varepsilon}$ and moreover $\mathrm{proj}_{\mathbb{N},N}(x') = y$. This tells us that $y \notin \underline{\mathrm{proj}}_{\mathbb{N},N}(\overline{\Delta}_{r,\varepsilon})$, and therefore $\underline{\mathrm{proj}}_{\mathbb{N},N}(\overline{\Delta}_{r,\varepsilon}) = \emptyset$, whence indeed $\underline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon}) = 0$.

A.7. **Proof of Theorem 7.** We only prove the first inequality. The proof of the second inequality is then completely analogous. Define, for $N \geq 1$ and $k \geq 0$, the set

$$\overline{\Delta}_{r,\varepsilon,N,k} = \bigcup_{n=N}^{N+k} \left\{ \frac{1}{n} \sum_{\ell=1}^{n} (h_\ell - M_\ell) > \varepsilon \right\}.$$

Then clearly, using the notations established in our proof of the weak law (see Section A.2), we get $\overline{\Delta}_{r,\varepsilon,N,k} = \bigcup_{n=N}^{N+k} \mathrm{proj}^{-1}_{\mathbb{N},n}(\Delta^c_{r,\varepsilon,n})$, and, using the coherence (sub-additivity) of the upper prevision $\overline{E}^{\mathbb{N}}$, we derive that

$$\overline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon,N,k}) \leq \sum_{n=N}^{N+k} \overline{E}^{\mathbb{N}}(\mathrm{proj}^{-1}_{\mathbb{N},n}(\Delta^c_{r,\varepsilon,n})) = \sum_{n=N}^{N+k} \overline{P}^n(\Delta^c_{r,\varepsilon,n}),$$

where the equality follows from Eq. (11), the consistency [(S3)] of the joint lower previsions $\underline{P}^k$ (and hence the consistency of the conjugate upper previsions $\overline{P}^k$) and the fact that for $m \geq n$, $\overline{\mathrm{proj}}_{\mathbb{N},m}(\mathrm{proj}^{-1}_{\mathbb{N},n}(\Delta^c_{r,\varepsilon,n})) = \mathrm{proj}^{-1}_{m,n}(\Delta^c_{r,\varepsilon,n})$. Since the joint lower previsions

$\underline{P}^n$ satisfy our weak law [by (S1) and (S2)], we get that $\overline{P}^n(\Delta^c_{r,\varepsilon,n}) \leq \exp(-n\lambda)$, where $\lambda = \varepsilon^2/4B^2$. So we find that

$$\overline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon,N,k}) \leq \sum_{n=N}^{N+k} e^{-n\lambda} = e^{-N\lambda} \sum_{n=0}^{k} e^{-n\lambda} = e^{-N\lambda} \frac{1-e^{-(k+1)\lambda}}{1-e^{-\lambda}} < \frac{e^{-N\lambda}}{1-e^{-\lambda}}.$$

This means that if we let $N = N(\varepsilon)$ be determined by

$$e^{-N(\varepsilon)\frac{\varepsilon^2}{4B^2}} \frac{1}{1-e^{-\frac{\varepsilon^2}{4B^2}}} = \varepsilon \text{ or equivalently } N(\varepsilon) = -\frac{4B^2}{\varepsilon^2} \ln \varepsilon (1-e^{-\frac{\varepsilon^2}{4B^2}}),$$

then we have indeed that $\overline{E}^{\mathbb{N}}(\overline{\Delta}_{r,\varepsilon,N,k}) < \varepsilon$ for $N \geq N(\varepsilon)$ and all $k \geq 0$. Observe that $N(\varepsilon)$ increases as ε^{-2} for small enough ε.

A.8. **Proof of Eq.** (13). Consider the sets $\overline{\Delta}_r = \{x \in \mathscr{X}^{\mathbb{N}} \colon \limsup_N \frac{1}{N} \sum_{\ell=1}^N (x_\ell - M_\ell) > 0\}$ and $\overline{\Delta}_{r,\varepsilon,n} = \{x \in \mathscr{X}^{\mathbb{N}} \colon \frac{1}{n} \sum_{\ell=1}^n (x_\ell - M_\ell) > \varepsilon\}$. Check that $\overline{\Delta}_r = \bigcap_{m\geq 1} \bigcap_{N\geq 1} \bigcup_{n\geq N} \overline{\Delta}_{r,\frac{1}{m},n}$. Since P_σ is lower continuous, we find that $P_\sigma(\overline{\Delta}_r) = \inf_{m\geq 1} \inf_{N\geq 1} P_\sigma(\bigcup_{n\geq N} \overline{\Delta}_{r,\frac{1}{m},n})$. On the other hand, we infer from Theorem 7 and the upper continuity of P_σ that for every $m \geq 1$ there is some $N(m) \geq 1$ such that $\inf_{N\geq 1} P_\sigma(\bigcup_{n\geq N} \overline{\Delta}_{r,\frac{1}{m},n}) \leq P_\sigma(\bigcup_{n\geq N(m)} \overline{\Delta}_{r,\frac{1}{m},n}) \leq \frac{1}{m}$, whence indeed $P_\sigma(\overline{\Delta}_r) = 0$.

A.9. **Proof of Theorem 8.** The consistency requirement (S3) holds trivially, by construction. Consider any sequence of $\mathscr{B}_k$ measurable gambles h_k on $\mathscr{X}_k$, then

$$m_k = \inf h_k + \int_{\inf h_k}^{\sup h_k} \underline{P}^{\mathbb{N}}(\{h_k > \alpha\})\,\mathrm{d}\alpha = \inf h_k + \int_{\inf h_k}^{\sup h_k} \underline{P}^N(\{h_k > \alpha\})\,\mathrm{d}\alpha = \underline{E}^N(h_k)$$

and similarly $M_k = \overline{E}^N(h_k)$ for all $N \geq k$. This shows that (S1) holds. We now show that (S2) holds in the sense that all the $\underline{E}^N$ are forward factorising when we restrict ourselves to Borel measurable gambles. Consider a strictly positive gamble g_k on $\mathscr{X}^{k-1}$ that is measurable with respect to the product σ-field $\mathscr{B}^{k-1}$ and any gamble h_k on $\mathscr{X}_k$ that is $\mathscr{B}_k$-measurable. Then we know that for any $n \geq 1$ and any real α,

$$\{g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] > \alpha\} \supseteq \left\{g_k > \frac{1}{n}\right\} \cap \{h_k > \underline{E}^{\mathbb{N}}(h_k) + n\alpha\}$$

and therefore, using the coherence (monotonicity) of $\underline{P}^{\mathbb{N}}$ and (M3), we find that

$$\begin{aligned}\underline{E}^{\mathbb{N}}\left(\{g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] > \alpha\}\right) &\geq \underline{E}^{\mathbb{N}}\left(\left\{g_k > \frac{1}{n}\right\} \cap \{h_k > \underline{E}^{\mathbb{N}}(h_k) + n\alpha\}\right) \\ &\geq \underline{P}^{\mathbb{N}}\left(\left\{g_k > \frac{1}{n}\right\}\right) \underline{P}^{\mathbb{N}}\left(\{h_k > \underline{E}^{\mathbb{N}}(h_k) + n\alpha\}\right)\end{aligned}$$

Now it follows from (M1) and (M4) that, taking into account that g_k is strictly positive, $\lim_{n\to\infty} \underline{P}^{\mathbb{N}}(\{g_k > \frac{1}{n}\}) = 1$, and since h_k is bounded, we also find that, using (M1) again,

$$\lim_{n\to\infty} \underline{P}^{\mathbb{N}}(\{h_k > \underline{E}^{\mathbb{N}}(h_k) + n\alpha\}) = \begin{cases} 0 & \text{if } \alpha > 0 \\ 1 & \text{if } \alpha < 0. \end{cases}$$

As a result

$$\underline{E}^{\mathbb{N}}(\{g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] > \alpha\}) \geq \begin{cases} 0 & \text{if } \alpha > 0 \\ 1 & \text{if } \alpha < 0. \end{cases}$$

Now since g_k is positive, we find that $\inf g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] \leq g_k[\inf h_k - \underline{E}^{\mathbb{N}}(h_k)] \leq 0$, where the last inequality follows from the coherence of $\underline{E}^{\mathbb{N}}$. Similarly, $\sup g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] \geq g_k[\sup h_k - \underline{E}^{\mathbb{N}}(h_k)] \geq 0$. Hence

$$\underline{E}^{\mathbb{N}}(g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)]) \geq \inf g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)] + \int_{\inf g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)]}^{0} 1 \,\mathrm{d}\alpha + \int_{0}^{\sup g_k[h_k - \underline{E}^{\mathbb{N}}(h_k)]} 0 \,\mathrm{d}\alpha = 0.$$

This shows that (S2) holds if we restrict ourselves to Borel measurable gambles g_k and h_k, where the g_k are strictly positive everywhere. But this is enough for our proof of the weak law to work, see Remark 1 and the fact that in the proof of our weak law (Section A.2), we only use strictly positive and measurable g_k. The rest of the proof is now immediate.

Acknowledgements

We would like to thank Matthias Troffaes for many provocative discussions, and for his patience when our enthusiasm must have been hard to bear. We are also indebted to two anonymous referees, who waded through progressively shorter versions of this paper.

References

R. B. Ash and C. A. Doléans-Dade. *Probability and Measure Theory*. Academic Press, London, second edition, 2000.

J. Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713.

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

G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the USA*, 17:656–660, 1932.

E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–270, 1909.

I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

F. G. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. In *Proceedings of ISIPTA '01*, pages 104–111. Shaker Publishing, Maastricht, 2000.

F. G. Cozman. Computing lower expectations with Kuznetsov's independence condition. In *Proceedings of ISIPTA '03*, pages 177–187. Carleton Scientific, 2003.

G. de Cooman and E. Miranda. Forward irrelevance. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.

B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975.

J. L. Doob. *Stochastic Processes*. John Wiley & Sons, New York, 1953.

L. E. Dubins. On Lebesgue-like extensions of finitely additive measures. *The Annals of Probability*, 2:456–463, 1974.

L. E. Dubins and L. J. Savage. *Inequalities for Stochastic Processes – How to Gamble If You Must*. McGraw-Hill, New York, 1965. Reprinted by Dover Publications in 1976.

L. G. Epstein and M. Schneider. IID: independently and indistinguishably distributed. *Journal of Economic Theory*, 113:32–50, 2003.

B. V. Gnedenko. *The Theory of Probability*. MIR Publishers, Moscow, 1975.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the Americal Statistical Association*, 58:13–30, 1963.

O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, second edition, 2002.

A. N. Kolmogorov. Sur la loi forte des grands nombres. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 191:910–912, 1930. Translated into English on pp. 60–61 of Kolmogorov (1992).

A. N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume II: Probability Theory and Mathematical Statistics*. Kluwer, Dordrecht, 1992.

V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., 1991. (in Russian).

F. Maccheroni and M. Marinacci. A strong law of large numbers for capacities. *Annals of Probability*, 33(3):1171–1178, 2005.

E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 46(1):188–225, 2007.

G. Shafer and V. Vovk. *Probability and Finance: It's Only a Game*. Wiley, New York, 2001.

C. A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series A*, 23:1–37, 1961.

M. C. M. Troffaes and G. de Cooman. Extension of coherent lower previsions to unbounded random variables. In *Proceedings of IPMU 2002*, pages 735–42. ESIA – Université de Savoie, 2002a.

M. C. M. Troffaes and G. de Cooman. Lower previsions for unbounded random variables. In P. Grzegorzewski, O. Hryniewicz, and M. Ángeles Gil, editors, *Soft Methods in Probability, Statistics and Data Analysis*, pages 146–155. Physica-Verlag, New York, 2002b.

J. Ville. *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.

J. von Neumann. Proof of the quasi-ergodic theorem. *Proceedings of the National Academy of Sciences of the USA*, 18:70–82, 1932.

P. Walley. Coherent lower (and upper) probabilities. Technical Report Statistics Research Report 22, University of Warwick, Coventry, 1981.

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

P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10:741–761, 1982.

P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.

GHENT UNIVERSITY, SYSTEMS RESEARCH GROUP, TECHNOLOGIEPARK – ZWIJNAARDE 914, 9052 ZWIJNAARDE, BELGIUM

E-mail address: gert.decooman@ugent.be

REY JUAN CARLOS UNIVERSITY, DEPT. OF STATISTICS AND OPERATIONS RESEARCH – C-TULIPÁN, S/N, 28933 MÓSTOLES, SPAIN

E-mail address: enrique.miranda@urjc.es