Math-Net.Ru

Общероссийский математический портал

P. Berti, E. Regazzini, P. Rigo, Well-calibrated, coherent forecasting systems, *Теория вероятн. и ее примен.*, 1997, том 42, выпуск 1, 144–168

DOI: <https://doi.org/10.4213/tvp1717>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.209.158.208

26 апреля 2023 г., 17:08:58

© 1997 г.

BERTI P.*, REGAZZINI E.***, RIGO P.*

WELL-CALIBRATED,
COHERENT FORECASTING SYSTEMS¹⁾

В данной работе вводится определение предсказуемого вывода, основанного на последовательных наблюдениях, в свете принципа когерентности де Финетти. Даются характеристики когерентных предсказуемых выводов и стратегических предсказуемых выводов. Проводится тщательный анализ понятий правильно калиброванной (well-calibrated) и финитно правильно калиброванной систем прогнозирования. В связи с данной тематикой устанавливаются некоторые новые законы больших чисел относительно конечно-аддитивных распределений вероятностей. Эти результаты используются, чтобы решительно опровергнуть некоторые возражения, высказываемые против принципа когерентности.

Ключевые слова и фразы: калибрование, когерентность, условная вероятность, конгломеративность (conglomerability), продолжение, конечная аддитивность, предсказуемый вывод, стратегия.

According to a prevailing view, the goal of statistical inference is a statement about the probability law governing a given observable phenomenon. Consequently, statistical methods are conceived just as tools to single out that law. This point of view presupposes that the existence of probability laws, as well as the existence of laws governing observable phenomena, is viewed as an objective fact. Hence, the above mentioned idea of statistical inference might clash with the pragmatic position according to which the essential role of any scientific theory lies in making previsions about future possible facts. On the other hand, the process of inferring values for unknown observable facts, based on current observations and other information, is immune to the criticism from those who refuse to assume the existence of «true» probability laws (subjectivistic interpretation of probability). This kind of process, which is known as *predictive* statistical approach, was shared by almost all the pioneers of statistical induction; see [17]. It can be carried out directly, without passing through the mediation of a family of parametric probability laws (statistical model), under the sole guidance of de Finetti's

*Dipartimento di Statistica, «G. Parenti», viale Morgagni 59, 50134 Firenze, Italy.

**IMQ-Universita' «L. Bocconi», via R. Sarfatti 25, 20136 Milano, Italy.

¹⁾Research partially supported by: MURST (60% 1992, Inferenza statistica predittiva), MURST (40% 1992, Modelli probabilistici e statistica matematica) and Universita' «L. Bocconi» (1991-1992).

principle of coherence. Obviously, predictive inferences can also be obtained after assigning a statistical model and a prior distribution on parameters, according to the usual Bayesian procedures.

The present paper deals with *predictive inferences*, in the case of sequential observations. Consequently, we will have to suggest conditions on a sequence of conditional expectations in order that all terms of the sequence may be considered admissible as a whole. The resulting scheme and conditions have natural applications to concrete problems of forecasting. Moreover, we think that they could be used to set up a general theoretical basis for some modern non-conventional inferential approaches and, in particular, for Dawid's prequential approach; see [8], [9]. We will deal with the subject within de Finetti's theory of probability, because we think that this is the only one which does not surreptitiously introduce extra-statistical technical restrictions and that, consequently, it is the most suitable as a basis for any discussion about the logical foundation of statistical methodology.

The present article consists of three sections. Section 1 contains some basic statements of de Finetti's concept of *coherent prevision*. In particular, it includes criteria which enable one to decide on the coherence of real-valued functions defined on suitable classes of bounded conditional random quantities. Section 2 deals with the concept of predictive inference in the presence of observations made at instants $1, 2, \dots$ and when, at each instant n , one assesses inferences on future facts, on the basis of the first n observations. In the same section, one defines and characterizes (*via* the notion of conglomerability) *strategic predictive inferences* which, in a sense, are finitely additive versions of the Ionescu-Tulcea argument in order to assess probability measures on infinite dimensional spaces; cf. [21]. Finally, Section 3 analyses calibration as a check of empirical validity of a given forecasting system. It includes some new theorems about *well-calibration* and *finitary well-calibration* of strategic predictive inferences. Among other things, it is shown that coherent predictive inferences needn't be well-calibrated, and this statement is used to prove that de Finetti's coherence is exempt from the imperfections pointed out by Dawid in [7].

1. Preliminaries

1.1. Events and random quantities. Given the space Ω of elementary cases, each relevant *event* is viewed as a subset of Ω . In particular, Ω corresponds to the sure event and $\emptyset$, the empty set, to the impossible one. We will adopt the useful convention that the same symbol that designates an event also designates the indicator of that event. Likewise, the same symbol that designates a class of events also designates the class of the corresponding indicators. Consequently, if $\mathcal{H}$ designates a class of events and $\mathcal{L}$ a class of real-valued functions on Ω , $\mathcal{H} \subset \mathcal{L}$ will mean that the indicators of the elements of $\mathcal{H}$ belong to $\mathcal{L}$. Moreover, if $\mathcal{L}$ is the class of the indicators of

all the elements of an algebra of events, then we will also say that $\mathcal{L}$ is an algebra of events.

Any real-valued function g on Ω will be said to be a *random quantity* (r.q.). Given a r.q. g and an event $H \neq \emptyset$, the restriction of g to H , denoted by $g | H$, is said to be a *conditional r.q.* In particular, if $g = E$, where E is the indicator of some event, then $E | H$ is also said to be a *conditional event*. Note that, if $H = \Omega$, then $g | H = g$.

1.2. De Finetti's coherence principle. A real-valued function P , defined on a class $\mathcal{C}$ of *bounded* conditional r.q.'s, is considered as a candidate to represent an expectation about the true realization of each element of $\mathcal{C}$, if and only if P meets de Finetti's *coherence principle*. More precisely, according to de Finetti [12, vol. 1], we introduce the following

Definition 1. Given a class $\mathcal{C}$ of bounded conditional r.q.'s, $P: \mathcal{C} \rightarrow \mathbf{R}$ is said to be a *prevision* on $\mathcal{C}$ if it meets the coherence principle. In particular, if $\mathcal{C}$ is a class of conditional events, then a prevision on $\mathcal{C}$ is also said to be a *probability* on $\mathcal{C}$.

As far as the coherence principle is concerned, suppose that after signing P on $\mathcal{C}$, one is committed to accepting any bet whatsoever on each element of $\mathcal{C}$, with arbitrary (positive or negative) stakes, on the understanding that any bet on $g | H$ is called off if H does not occur. In this framework, $P(g | H)$ represents the price of every bet on $g | H$. Precisely, what one gains from a combination of bets on $g_1 | H_1, \dots, g_n | H_n$ with stakes $s_1, \dots, s_n$, respectively, is given by

$$G(g_k | H_k, s_k; k = 1, \dots, n) = \sum_{k=1}^n s_k H_k \{P(g_k | H_k) - g_k\} | H_0$$

with $H_0 = \cup_{k=1}^n H_k$. Then, P is said to be coherent if and only if the inequalities

$$\inf G \leq 0 \leq \sup G$$

hold for every choice of $n \in \mathbf{N}$, $\{g_1 | H_1, \dots, g_n | H_n\} \subset \mathcal{C}$, and $(s_1, \dots, s_n)$ in $\mathbf{R}^n$. In other words, one has to fix a coherent P , if one wants nobody to make Dutch book against him.

Definition 1 makes sense because, given any $\mathcal{C}$, there exists at least one prevision on $\mathcal{C}$. Such a statement is a direct consequence of the following extension theorem, whose proof — together with a concise treatment of the main consequences of the coherence principle — can be found in [24].

Theorem 1. Let $\mathcal{C}$ and $\mathcal{C}^*$ be classes of bounded conditional r.q.'s such that $\mathcal{C} \subset \mathcal{C}^*$ and let P be a prevision on $\mathcal{C}$. Then there exists a prevision (which needn't be unique) P^* on $\mathcal{C}^*$ for which

$$P^*(g | H) = P(g | H) \quad \text{for every } g | H \text{ in } \mathcal{C}.$$

Throughout the rest of this paper, we shall be basically concerned with

the case in which the domain $\mathcal{C}$ of P is given by

$$\mathcal{C} = \{g \mid H: g \in \mathcal{L}, H \in \mathcal{H}\} \quad (1)$$

where $\mathcal{H} = \cup_{n=0}^{\infty} \Pi_n$, (Π_n) being a sequence of partitions of Ω such that Π_{n+1} is a refinement of Π_n for every $n \geq 0$ and $\Pi_0 = \{\Omega\}$; $\mathcal{L}$ is any class of bounded r.q.'s such that $\mathcal{L} \supset \mathcal{H}$ and $gH \in \mathcal{L}$ whenever $g \in \mathcal{L}$ and $H \in \mathcal{H}$.

1.3. A few characterizations of probabilities and previsions.

We begin by mentioning a useful characterization of a prevision on $\mathcal{C}$ — whose proof can be found in [1] — whenever $\mathcal{C}$ coincides with (1).

Theorem 2. *Let $\mathcal{C}$ be defined according to (1). Then $P: \mathcal{C} \rightarrow \mathbf{R}$ is a prevision if and only if P meets the following conditions:*

(p₁) $P(\cdot \mid H)$ is a prevision on $\mathcal{L}$ for every H in $\mathcal{H}$;

(p₂) $\inf g \mid H \leq P(g \mid H) \leq \sup g \mid H$ for every H in $\mathcal{H}$ and g in $\mathcal{L}$;

(p₃) $P(gH_1 \mid H_2) = P(g \mid H_1 \cap H_2)P(H_1 \mid H_2)$ for every g, H_1 in $\mathcal{L}$ and H_2 in $\mathcal{H}$ such that $H_1 \subset \Omega$, $gH_1 \in \mathcal{L}$ and $H_1 \cap H_2 \in \mathcal{H}$.

If $\mathcal{L}$ is a class of events, then (p₁)–(p₃) can be restated as follows:

(π_1) $P(\cdot \mid H)$ is a probability on $\mathcal{L}$ for every H in $\mathcal{H}$;

(π_2) $P(A \mid H) = 1$ whenever $H \in \mathcal{H}$, $A \in \mathcal{L}$ and $H \subset A$;

(π_3) $P(A \cap H_1 \mid H_2) = P(A \mid H_1 \cap H_2)P(H_1 \mid H_2)$ provided that A, H_1 and $A \cap H_1$ belong to $\mathcal{L}$ and $H_1 \cap H_2, H_2$ are elements of $\mathcal{H}$.

The following propositions provide characterizations of unconditional previsions in case $\mathcal{L}$ has some suitable structure; for their proof, see [24].

Theorem 3. *Let $\mathcal{L}$ be a linear space of bounded r.q.'s including the constants. Then $Q: \mathcal{L} \rightarrow \mathbf{R}$ is a prevision if and only if it turns out to be a positive, linear functional on $\mathcal{L}$ such that $Q(\Omega) = 1$.*

Theorem 4. *Let $\mathcal{L}$ be an algebra of events. Then $Q: \mathcal{L} \rightarrow \mathbf{R}$ is a probability if and only if it turns out to be a non-negative-valued, additive function such that $Q(\Omega) = 1$.*

1.4. Basic differences between Kolmogorov's and de Finetti's theories. As already noted, in the present paper we essentially deal with classes of random elements of the type of (1), but it should be remembered that, thanks to Theorem 1, previsions can be assessed on *arbitrary* classes of conditional bounded r.q.'s. In particular, one can fix a conditional prevision without having preassigned any unconditional probability law. In other words, the coherence principle suffices to state whether a real-valued function, defined on any class of bounded *conditional* r.q.'s, can be considered as a prevision or not. Consequently, conditional probabilities, in de Finetti's theory, needn't be evaluated as derivatives of probability measures with respect to probability measures. Moreover, coherent probabilities (previsions, respectively) needn't be continuous with respect to monotone sequences of events (r.q.'s, respectively). Apart from the previous remarks which, in any case, stress some peculiarities of de Finetti's approach compared with Kolmogorov's, the main difference being that de Finetti explicitly considers

conditional r.q.'s and events, given a single event and, in many cases, such an event represents an isolated given hypothesis whose probability equals zero. In fact, such cases appear frequently enough in probability and statistical practice; cf. [12, p. 276 of the English translation] and [23, p. 66]. On the contrary, Kolmogorov [20, p. 51 of the English translation] asserts «the concept of conditional probability with regard to an isolated given hypothesis whose probability equals zero is inadmissible». Consequently, in Kolmogorov's theory, conditioning is considered with respect to specific classes of events.

2. Predictive inferences

2.1. Definition of predictive inference. Roughly speaking, predictive inference is any family of consistent conditional expectations on a class of r.q.'s, given some facts relating to the realizations of other random elements. As an example, we might consider a predictor (meteorologist, economist, bookmaker, etc.) who makes regular periodic forecasts on the basis of past data. In fact, forecasting is a significant case of predictive inference, which will be of help to us to explain our point of view in a concrete way.

We start by making precise the domain of a predictive inference, in the case of sequential observations. This means that observations are made at instants $1, 2, \dots$ and, at each instant n , one assesses inferences on future facts, on the basis of the first n observations.

Let $\mathfrak{X}$ denote the set of all possible outcomes of each element of a sequence of trials. According to Dubins and Savage [15] call the elements of $\Omega := \mathfrak{X}^\infty$ *histories*, and call the elements of $\mathfrak{X}^n$ *partial histories* ($n = 1, 2, \dots$). Hence, the set Π_n defined by

$$\Pi_n = \left\{ \{ (x_1, \dots, x_n) \} \times \mathfrak{X}^\infty : (x_1, \dots, x_n) \in \mathfrak{X}^n \right\} \quad (2)$$

is the partition of Ω whose elements can be identified with partial histories of length n . Then, the domain of a predictive inference in case of sequential observations will be defined to be the class $\mathcal{C}$, defined by (1), with $\Omega = \mathfrak{X}^\infty$ and Π_n as in (2).

In the framework just described, predictive inferences are assigned without any reference to some statistical model involving unknown parameters. Anyhow, it is a common statistical practice to derive previsions on $\mathcal{C}$ from joint assignments of a parametric statistical model and of a prior distribution on its unknown parameters. This situation will be investigated in a forthcoming paper. Here, we just take into account predictive inferences, free from parametric superstructures.

Definition 2. Given $\mathcal{C}$ as in (1), with $\Omega = \mathfrak{X}^\infty$ and Π_n as in (2), any prevision P on $\mathcal{C}$ is said to be a *predictive inference* (p.i.).

We now provide a characterization for p.i.'s based on Theorem 2.

After denoting the indicator of $\{(x_1, \dots, x_n)\} \times \mathfrak{X}^\infty$ with $I(x_1, \dots, x_n)$, $P: \mathcal{C} \rightarrow \mathbf{R}$ is a p.i. if and only if

- (i) $P(\cdot)$, $P(\cdot | x_1, \dots, x_n)$ are previsions on $\mathcal{L}$,
 - (ii) $P(I(x_1, \dots, x_n) | x_1, \dots, x_n) = 1$,
 - (iii) $P(g I(x_1, \dots, x_k, \dots, x_n) | x_1, \dots, x_k) = P(g | x_1, \dots, x_k, \dots, x_n) \times P(I(x_1, \dots, x_k, \dots, x_n) | x_1, \dots, x_k)$,
 - (iv) $P(g I(x_1, \dots, x_n)) = P(g | x_1, \dots, x_n) P(I(x_1, \dots, x_n))$
- for every g in $\mathcal{L}$, $(x_1, \dots, x_k, \dots, x_n)$ in $\mathfrak{X}^n$ and for all naturals k, n with $k \leq n$.

In the above proposition, as well as throughout the rest of this paper, $P(\cdot)$ stands for $P(\cdot | \Omega)$ and $P(\cdot | x_1, \dots, x_n)$ for $P(\cdot | \{(x_1, \dots, x_n)\} \times \mathfrak{X}^\infty)$.

2.2. Example of predictive inference within a Bayesian framework. Let $\mathcal{L} = \{A_1 \times \dots \times A_n \times \mathfrak{X}^\infty: A_1, \dots, A_n \in \mathcal{A}, n \in \mathbf{N}\}$ where $\mathcal{A}$ is some algebra of subsets of $\mathfrak{X}$ including the singletons. In compliance with the Bayesian standard procedure, suppose that for each θ belonging to a specific parameter space Θ , a prevision $P_\theta(\cdot)$ is assigned on $\mathcal{L}$ in such a way that

$$P_\theta(A_1 \times \dots \times A_n \times \mathfrak{X}^\infty) = \prod_{i=1}^n \int_{A_i} l(x, \theta) \lambda(dx)$$

where λ denotes a measure on $\mathcal{A}$ — finitely additive, possibly — and $l: \mathfrak{X} \times \Theta \rightarrow [0, +\infty)$ is such that $\int l(x, \theta) \lambda(dx) = 1$ for every $\theta \in \Theta$. Throughout the present paper, integrals are to be meant in the sense of Dunford and Schwartz; cf. [16] and [4]. Moreover, let us assume that

$$P_\theta(A_1 \times \dots \times A_n \times \mathfrak{X}^\infty | x_1, \dots, x_k) = P_\theta(\mathfrak{X}^k \times A_{k+1} \times \dots \times A_n \times \mathfrak{X}^\infty)$$

if $n > k$ and $x_i \in A_i$ for $i = 1, \dots, k$,

$$P_\theta(A_1 \times \dots \times A_n \times \mathfrak{X}^\infty | x_1, \dots, x_k) = 1$$

if $n \leq k$ and $x_i \in A_i$ for $i = 1, \dots, n$, and

$$P_\theta(A_1 \times \dots \times A_n \times \mathfrak{X}^\infty | x_1, \dots, x_k) = 0$$

otherwise.

These positions are consistent with the introduction of a sequence of $\mathfrak{X}$ -valued random elements which, under $P_\theta(\cdot)$, are independent and identically distributed. In view of (i)–(iv) of Subsection 2.1, $\mathcal{P}_\theta := \{P_\theta(\cdot), P_\theta(\cdot | x_1, \dots, x_n): (x_1, \dots, x_n) \in \mathfrak{X}^n, n \in \mathbf{N}\}$ is a p.i. for every θ . Let us now introduce a countably additive prior probability q on a σ -algebra $\mathcal{B}$ of subsets of Θ , and assume that $\theta \rightarrow l(x, \theta)$ is a $\mathcal{B}$ -measurable function for every fixed $x \in \mathfrak{X}$. Consequently, if

$$\int \prod_{i=1}^n l(x_i, \theta) q(d\theta) \in (0, +\infty)$$

for each partial history $(x_1, \dots, x_n)$, then

$$q(H \mid x_1, \dots, x_n) = \frac{\int_H \prod_{i=1}^n l(x_i, \theta) q(d\theta)}{\int_{\Theta} \prod_{i=1}^n l(x_i, \theta) q(d\theta)}, \quad H \in \mathcal{B},$$

represents a coherent posterior; cf. [2]. At this stage, a p.i. P can be assessed according to

$$P(g) = \int P_{\theta}(g) q(d\theta),$$

$$P(g \mid x_1, \dots, x_n) = \int P_{\theta}(g \mid x_1, \dots, x_n) q(d\theta \mid x_1, \dots, x_n)$$

for all g in $\mathcal{L}$, $(x_1, \dots, x_n)$ in $\mathfrak{X}^n$ and $n \in \mathbb{N}$.

Indeed, since P_{θ} is a p.i. for every θ , P satisfies (i) and (ii). Further, (iii) trivially holds whenever $P(I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k) = 0$, so that one can assume that

$$P(I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k)$$

$$= \prod_{i=k+1}^n \lambda(\{x_i\}) \int \prod_{i=k+1}^n l(x_i, \theta) q(d\theta \mid x_1, \dots, x_k)$$

is strictly positive. In that case, for any $g \in \mathcal{L}$

$$P(g I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k)$$

$$= \int P_{\theta}(g I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k) q(d\theta \mid x_1, \dots, x_k)$$

$$= \prod_{i=k+1}^n \lambda(\{x_i\}) \int P_{\theta}(g \mid x_1, \dots, x_k, \dots, x_n)$$

$$\times \prod_{i=k+1}^n l(x_i, \theta) q(d\theta \mid x_1, \dots, x_k)$$

$$= P(I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k) \int P_{\theta}(g \mid x_1, \dots, x_k, \dots, x_n)$$

$$\times \frac{\prod_{i=k+1}^n l(x_i, \theta) q(d\theta \mid x_1, \dots, x_k)}{\int \prod_{i=k+1}^n l(x_i, \theta) q(d\theta \mid x_1, \dots, x_k)}$$

$$= P(I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k)$$

$$\times \int P_{\theta}(g \mid x_1, \dots, x_k, \dots, x_n) q(d\theta \mid x_1, \dots, x_k, \dots, x_n).$$

Since the last integral represents $P(g \mid x_1, \dots, x_k, \dots, x_n)$, we have stated that (iii) comes true. Finally, since (iv) can be checked in the same way, P turns out to be a p.i.

2.3. Definition and existence of strategic predictive inferences.

This subsection deals with a special kind of p.i.'s, which satisfy the usual *disintegrability* condition characterizing Kolmogorovian conditional expectations.

According to Dubins and Savage [15], a *strategy* is a sequence $\sigma = (\sigma_0, \sigma_1, \dots)$ in which σ_0 is a probability on $\mathcal{P}(\mathfrak{X})$, the power set of $\mathfrak{X}$, and, for every n in $\mathbb{N}$, σ_n is a function on $\mathfrak{X}^n$ which associates a probability on $\mathcal{P}(\mathfrak{X})$ — denoted by $\sigma_n(x_1, \dots, x_n)$ — to every partial history $(x_1, \dots, x_n)$. For any $B \subset \mathfrak{X}$, $\sigma_n(x_1, \dots, x_n)(B)$ can be viewed as the conditional probability, under σ , of $\{x_{n+1} \in B\}$, given $\{(x_1, \dots, x_n)\} \times \mathfrak{X}^\infty$.

If P is a p.i. and $\mathcal{L}$ includes

$$\mathfrak{B} := \{B \times \mathfrak{X}^\infty, \mathfrak{X}^n \times B \times \mathfrak{X}^\infty : B \subset \mathfrak{X}, n \in \mathbb{N}\},$$

then the strategy σ given by

$$\sigma_0(B) = P(B \times \mathfrak{X}^\infty), \quad \sigma_n(x_1, \dots, x_n)(B) = P(\mathfrak{X}^n \times B \times \mathfrak{X}^\infty \mid x_1, \dots, x_n),$$

is said to be the *strategy induced* by P .

Definition 3. If $\mathcal{L} \supset \mathfrak{B}$, a p.i. P is said to be *strategic* if

$$\begin{aligned} P(g) &= \int P(g \mid x) \sigma_0(dx), \\ P(g \mid x_1, \dots, x_n) &= \int P(g \mid x_1, \dots, x_n, x) \sigma_n(x_1, \dots, x_n)(dx) \end{aligned} \quad (3)$$

for every g in $\mathcal{L}$, $(x_1, \dots, x_n)$ in $\mathfrak{X}^n$ and n in $\mathbb{N}$, where σ is the strategy induced by P .

As already noted, equations (3) look like the usual conditions requested for conditional expectations in Kolmogorov's theory. However, strategic p.i.'s needn't be continuous with respect to monotone sequences of r.q.'s.

We also note that, in order to check whether an arbitrary function $P: \mathcal{C} \rightarrow \mathbb{R}$ is a strategic p.i., it suffices to verify (3) together with (i) and (ii) of Subsection 2.1; indeed, under (i) and (ii), (iii) and (iv) follow from (3).

As far as the existence of strategic p.i.'s is concerned, we have the following proposition involving the class $\mathcal{L}^*$ of the *inductively integrable functions* on Ω ; see [15].

Theorem 5. *If $\mathfrak{B} \subset \mathcal{L} \subset \mathcal{L}^*$ and σ is a strategy, then there exists a unique strategic p.i. P , such that σ is the strategy induced by P .*

P r o o f. This proof heavily relies on pages 12–20 of [15]. In particular, we refer to [15] for the notation, the notion of *structure* and the existence and properties of the function $E(\cdot, \cdot)$. Define

$$P(g) := E(\sigma, g), \quad P(g \mid x_1, \dots, x_n) := E(\sigma[x_1, \dots, x_n], gx_1 \dots x_n)$$

for every g in $\mathcal{L}$, $(x_1, \dots, x_n)$ in $\mathfrak{X}^n$ and n in $\mathbb{N}$. Then, P is a strategic p.i. Indeed, according to Theorems 1 and 2, pages 17–18, of [15], P satisfies (3)

and (i)–(ii) of Subsection 2.1. Next, let P' be any strategic p.i. inducing σ . We are showing that $P' = P$. If g is constant, then $P(g) = P'(g)$. Likewise, for every $n \in \mathbb{N}$, $(x_1, \dots, x_n) \in \mathfrak{X}^n$ and $g \in \mathcal{L}$, if $gx_1 \dots x_n$ is constant, then $P(g | x_1, \dots, x_n) = P'(g | x_1, \dots, x_n)$. Fix an ordinal $\gamma > 0$, and suppose that $P(g) = P'(g)$ for every $g \in \mathcal{L}$ with structure less than γ , and $P(g | x_1, \dots, x_n) = P'(g | x_1, \dots, x_n)$ for every $n \in \mathbb{N}$, $(x_1, \dots, x_n) \in \mathfrak{X}^n$ and $g \in \mathcal{L}$ such that $gx_1 \dots x_n$ is of structure less than γ . Let $f \in \mathcal{L}$ be of structure γ . Then, for every $x \in \mathfrak{X}$, fx is of structure less than γ . Hence $P(f | x) = P'(f | x)$, so that strategicity implies

$$P(f) = \int P(f | x) \sigma_0(dx) = \int P'(f | x) \sigma_0(dx) = P'(f).$$

Similarly, if $f \in \mathcal{L}$ and $fx_1 \dots x_n$ is of structure γ , strategicity implies $P(f | x_1, \dots, x_n) = P'(f | x_1, \dots, x_n)$, and this concludes the proof.

One can notice that, with respect to significant statistical applications, $\mathcal{L}^*$ is large enough. For instance, if g is bounded and depends on a fixed finite number of coordinates, then g belongs to $\mathcal{L}^*$. Moreover, $\mathcal{L}^*$ includes bounded functions depending on a random number of coordinates, say $\tau: \Omega \rightarrow \mathbb{N}$, provided that τ is a stopping time.

Some significant characteristic aspects of strategic p.i.'s stand out of the following proposition, which involves a condition of *conglomerability*; see [10], [11] and [14].

Theorem 6. *If $\mathcal{L}$ is a linear space such that $\mathcal{L} \supset \mathfrak{B}$, then $P: \mathcal{C} \rightarrow \mathbb{R}$ is a strategic p.i. if and only if (i), (ii) in Subsection 2.1, and*

$$\begin{aligned} \inf_{x \in \mathfrak{X}} P(g | x) &\leq P(g) \leq \sup_{x \in \mathfrak{X}} P(g | x), \\ \inf_{x \in \mathfrak{X}} P(g | x_1, \dots, x_n, x) &\leq P(g | x_1, \dots, x_n) \leq \sup_{x \in \mathfrak{X}} P(g | x_1, \dots, x_n, x) \end{aligned} \quad (4)$$

hold for every $g \in \mathcal{L}$, $(x_1, \dots, x_n)$ in $\mathfrak{X}^n$ and n in $\mathbb{N}$.

P r o o f. The «only if» part is trivial, so that it suffices to prove that (i), (ii) and (4) imply (3). But, under (i) and (ii), the first of conditions (4) means that $P(\cdot)$ is conglomerable with respect to the partition $\{\{x\} \times \mathfrak{X}^\infty\}$, and the second means that $P(\cdot | x_1, \dots, x_n)$ is conglomerable with respect to the partition $\{\mathfrak{X}^n \times \{x\} \times \mathfrak{X}^\infty\}$. Thus, since $\mathcal{L}$ is a linear space and $\mathcal{L} \supset \mathfrak{B}$, Theorem 3.1 of [3] implies that (4) is equivalent to (3).

2.4. Example of strategic predictive inference. Let $\mathfrak{X} = \mathbb{R}$. According to [18], [19], in order to describe a situation of extremely vague a priori knowledge, a set of reasonable assumptions are:

- (a) $P(\{\omega: x_i \neq x_j \text{ for } i \neq j \text{ and } i, j \leq n\}) = 1$;
- (b) $P(B_{j_1} \times \dots \times B_{j_n} \times \mathbb{R}^\infty) = P(B_1 \times \dots \times B_n \times \mathbb{R}^\infty)$ for every permutation $(j_1, \dots, j_n)$ of $(1, \dots, n)$, and every n -tuple of intervals of the form $B_i = (-\infty, x_i]$;

(c) $P(\mathbf{R}^n \times I_i \times \mathbf{R}^\infty \mid x_1, \dots, x_n) = 1/(n+1)$ for every partial history $(x_1, \dots, x_n)$ without ties, I_i denoting the open interval $(x_{(i-1)}, x_{(i)})$, $i = 1, \dots, n+1$, with $x_{(0)} = -\infty$, $x_{(n+1)} = +\infty$ and $x_{(1)}, \dots, x_{(n)}$ the order statistics of $x_1, \dots, x_n$.

The set of conditions (a), (b) and (c), assessed for a particular n , represents *Hill's A_n -model*.

We now single out a strategic p.i. satisfying A_n for every n . Following [6] and [19], one assigns any strategy σ such that

$$\sigma_0((-\infty, x)) = \sigma_0((x, +\infty)) = \frac{1}{2} \quad \text{for every } x \text{ in } \mathbf{R},$$

$$\text{and } \sigma_n(x_1, \dots, x_n) = \frac{1}{n+1} \left(\sigma_0 + \sum_{i=1}^n d_{x_i} \right)$$

where d_x is a probability for which $d_x((x-\varepsilon, x)) = d_x((x, x+\varepsilon)) = \frac{1}{2}$ for every $\varepsilon > 0$. In view of Theorem 5, σ admits a unique extension P as a strategic p.i. and, moreover, it is easy to show that P meets A_n , for every n .

Clearly, the previous P is not countably additive. In fact, in [18] one shows that A_n is incompatible with countable additivity in the framework of Kolmogorov's theory of conditional expectations. We now show that, if strategicity is not required, then a countably additive p.i., agreeing with A_n , exists. By «a countably additive p.i.», we mean a p.i. P such that $P(\cdot)$ and every $P(\cdot \mid x_1, \dots, x_n)$ are countably additive on $\mathcal{L}$, or at least on some relevant sub-algebra of $\mathcal{L}$.

Indeed, let $\mathcal{L} = \{B \times \mathbf{R}^\infty : B \text{ Borel set in } \mathbf{R}^n, n \in \mathbf{N}\}$, and let $\{P'(\cdot), P'(\cdot \mid x_1, \dots, x_n)\}$ be any family of countably additive probabilities on $\mathcal{L}$ satisfying (ii) of Subsection 2.1 and

$$P'(I(x_1, \dots, x_k)) = P'(I(x_1, \dots, x_k, \dots, x_n) \mid x_1, \dots, x_k) = 0$$

for every $(x_1, \dots, x_k, \dots, x_n)$ in $\mathfrak{X}^n$ and every $k < n$. In view of (i)–(iv), P' is a p.i. Furthermore, since the single probabilities composing P' can be chosen independently of one another, it is plain that they can be chosen in such a way that A_n holds for every n .

3. Well-calibrated predictive inferences

3.1. Definition of well-calibrated predictive inference. After stating the definition of predictive inference, we now deal with one of the checks of empirical validity, which one usually applies to appreciate the «correctness» of the forecasting system. Indeed, the present section is concerned with the concept of well-calibrated inference.

Throughout this section one assumes that $\mathcal{L} = \{S \times \mathfrak{X}^\infty : S \subset \mathfrak{X}^n, n \in \mathbf{N}\}$, the set of all cylinders with finite-dimensional base, and that a p.i. P has been assessed on the corresponding class $\mathcal{C}$. Let us introduce sequences

(E_n) and (h_n) such that

$$E_1 = B_1 \times \mathfrak{X}^\infty, \quad E_2 = \mathfrak{X} \times B_2 \times \mathfrak{X}^\infty, \quad E_3 = \mathfrak{X}^2 \times B_3 \times \mathfrak{X}^\infty, \dots$$

with $B_n \subset \mathfrak{X}$ for each n , and

$h_n: \Omega \rightarrow \{0, 1\}$ depends on only the first $(n-1)$ coordinates of each history $\omega \in \Omega$, whenever $n > 1$, and h_1 is constant.

Then, for every $\omega = (x_1, x_2, \dots)$, let us put

$$\nu_n(\omega) = \sum_{i=1}^n h_i(\omega), \quad n = 1, 2, \dots,$$

$$p_n(\omega) = \begin{cases} \frac{1}{\nu_n(\omega)} \sum_{i=1}^n h_i(\omega) E_i(\omega) & \text{if } \nu_n(\omega) \neq 0, \\ 0 & \text{if } \nu_n(\omega) = 0, \end{cases}$$

$$\pi_n(\omega) = \begin{cases} \frac{1}{\nu_n(\omega)} \sum_{i=1}^n h_i(\omega) P(E_i | x_1, \dots, x_{i-1}) & \text{if } \nu_n(\omega) \neq 0, \\ 0 & \text{if } \nu_n(\omega) = 0, \end{cases}$$

where $P(E_i | x_1, \dots, x_{i-1}) := P(E_1)$ if $i = 1$.

Here is a real situation in which h_n , ν_n , p_n and π_n acquire a concrete meaning. If a meteorologist, every evening, assigns a probability to the event of precipitation within the next 24 hours, then E_n can be regarded as the event of precipitation on day n . Under these circumstances, the following is a plausible choice for h_n : after fixing (p, δ) in $[0, 1] \times (0, +\infty)$, set

$$h_n(\omega) = \begin{cases} 1 & \text{if } |P(E_n | x_1, \dots, x_{n-1}) - p| < \delta, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

In other words, h_n ($n \geq 2$) selects day n if and only if the conditional probability of precipitation on day n , given the partial history $(x_1, \dots, x_{n-1})$, is suitably close to p . Consequently, ν_n represents the number of days in $\{1, \dots, n\}$ in which the forecaster assigns an inference suitably close to p , π_n is the mean of the inferences assessed on those very same days and p_n the corresponding frequency of precipitation.

A possible requirement for a p.i. is that p_n and π_n have the same asymptotic behaviour. Clearly, this requires attaching a probability to the events

$$E = \{\omega: \nu_n(\omega) \rightarrow +\infty \text{ as } n \rightarrow +\infty\},$$

$$F = \{\omega \in E: p_n(\omega) - \pi_n(\omega) \not\rightarrow 0 \text{ as } n \rightarrow +\infty\}$$

which, usually, are not finite-dimensional cylinders. Consequently, the p.i. P , assessed on $\mathcal{C}$, must be extended.

Formally, according to [7]:

If P' is a p.i. on $\{g \mid H: g \in \mathcal{L} \cup \{E, F\}, H \in \mathcal{H}\}$, agreeing with P on $\mathcal{C}$, and if $P'(E) > 0$, then P' is said to be *well-calibrated* when

$$P'(F) = 0. \quad (6)$$

Note that (6) trivially holds whenever $P'(E) = 0$.

Even if the previous definition seems to be intuitively appealing, it cannot be considered as an *operational* one, since no human being can, in general, decide whether an inference is well-calibrated or not. This state of things depends on the fact that sets E and F involve limit conditions which, in general, cannot be observed. Hence, as a property of statistical procedures, well-calibration is concretely unimportant. At the very most, it expresses an unconfirmable belief which — as we will show shortly — is usually independent of the specification of P .

These remarks lead us to introduce a different notion of well-calibration, based on a finitary, observable condition, which will be denominated finitary well-calibration. More precisely:

A p.i. P on $\mathcal{C}$ is said to be *finitarily well-calibrated* if

$$\begin{aligned} \inf_{\varepsilon, c > 0} \liminf_n \inf_k P\left(\left\{\omega: \max_{n \leq j \leq n+k} |p_j(\omega) - \pi_j(\omega)| \leq \varepsilon, \nu_n(\omega) > c\right\}\right) \\ = \inf_{c > 0} \lim_n P\left(\left\{\omega: \nu_n(\omega) > c\right\}\right) \end{aligned} \quad (7)$$

whenever $\inf_{c > 0} \lim_n P(\{\omega: \nu_n(\omega) > c\}) > 0$.

Condition (7) is a bit cumbersome but, unlike (6), can be tested just by relying on the p.i. P actually assessed on $\mathcal{C}$, without selecting any of its extensions. Of course, the need for (7) is less pressing whenever one is forced to adopt a particular extension, for instance, if one is committed to fixing a countably additive P and, then, to extending it in a countably additive way. In this connection, note that, if P' is countably additive, then (6) and (7) are equivalent. Note also how the requirement that $\inf_{c > 0} \lim_n P(\{\omega: \nu_n(\omega) > c\}) > 0$, like its non-finitary counterpart $P'(E) > 0$, is logically compelling in order to make (7) meaningful. Examples 2 and 3 will show that neither well-calibration implies finitary well-calibration nor finitary well-calibration implies well-calibration.

3.2. Well-calibration and Kolmogorov's theory. In this subsection, P is assumed to be strategic and σ denotes the strategy induced by P . Moreover, $P'(\cdot)$ and $P'(\cdot \mid x_1, \dots, x_n)$ stand for the so called *Lebesgue-like-extensions* of $P(\cdot)$ and $P(\cdot \mid x_1, \dots, x_n)$, respectively. We send back to [13] and [22] for the definition and properties of Lebesgue-like-extensions. For our purposes it suffices to note that, setting $\mathcal{C}' = \{g \mid H: g \in \mathcal{L}', H \in \mathcal{H}\}$, where $\mathcal{L}'$ is the σ -algebra generated by $\mathcal{L}$, P' is a strategic p.i. on $\mathcal{C}'$. A further remark is that, under the usual assumptions of countable additivity and measurability, P' coincides with the countably additive extension of P . More precisely, let $\mathcal{D}$ be a σ -algebra of subsets of $\mathcal{X}$ and let us suppose that

σ_0 and every $\sigma_n(x_1, \dots, x_n)$ are countably additive when restricted to $\mathcal{D}$; $(x_1, \dots, x_n) \rightarrow \sigma_n(x_1, \dots, x_n)(B)$ is a $\mathcal{D}^n$ -measurable function on $\mathfrak{X}^n$ for every n and $B \in \mathcal{D}$.

Then, $P(\cdot)$ and every $P(\cdot | x_1, \dots, x_n)$ are countably additive on $\mathfrak{L}^{**} := \{S \times \mathfrak{X}^\infty : S \in \mathcal{D}^n, n \in \mathbb{N}\}$, and also, denoting by $\mathcal{D}^\infty$ the σ -algebra generated by $\mathfrak{L}^{**}$, $P'(\cdot)$ and $P'(\cdot | x_1, \dots, x_n)$ coincide on $\mathcal{D}^\infty$ with the countably additive extensions of $P(\cdot)$ and $P(\cdot | x_1, \dots, x_n)$ to $\mathcal{D}^\infty$.

To sum up, P' is a strategic extension of P and, in case P is assessed in line with the Kolmogorov theory, P' is just the Kolmogorovian extension of P . We now show that, in addition, well-calibration holds with respect to P' .

(In what follows, just to make the notation less cumbersome, whenever a function g on Ω depends only on the first n coordinates of each history $\omega = (x_1, \dots, x_n, \dots)$, $g(x_1, \dots, x_n)$ is sometimes used instead of $g(\omega)$.)

Theorem 7. *Let P be a strategic p.i. on $\mathcal{C}$, and let P' be its Lebesgue-like-extension. Then, $P'(F) = 0$.*

P r o o f. With reference to, and using the same notation as, Theorem 6.2 in [5, p. 353], let us define $Y_n := h_n E_n$ and

$$\begin{aligned} M_n(x_1, \dots, x_{n-1}) &:= P(Y_n | x_1, \dots, x_{n-1}) \\ &= h_n(x_1, \dots, x_{n-1}) P(E_n | x_1, \dots, x_{n-1}), \\ V_n(x_1, \dots, x_{n-1}) &:= P((Y_n - M_n)^2 | x_1, \dots, x_{n-1}). \end{aligned}$$

Let $A = \{\omega : \sum_{i=1}^\infty V_i(\omega) = +\infty\}$. Since $(Y_n - M_n)^2 \leq h_n$, one obtains $\sum_{i=1}^n V_i \leq \sum_{i=1}^n h_i = \nu_n$. Thus, setting

$$a_{n,\varepsilon}(\omega) = \left(\sum_{i=1}^n V_i(\omega) \right)^{1/2} \left(\log \sum_{i=1}^n V_i(\omega) \right)^{1/2+\varepsilon}, \quad \varepsilon > 0,$$

then $a_{n,\varepsilon}(\omega) \leq \nu_n(\omega)$ whenever $\omega \in A$ and n is sufficiently large. Now, Chen's theorem states that $\sum_{i=1}^n (Y_i - M_i)$ converges, $P'(\cdot)$ -a.s. on A^c , and $(a_{n,\varepsilon})^{-1} \sum_{i=1}^n (Y_i - M_i) \rightarrow 0$, $P'(\cdot)$ -a.s. on A . Hence, noting that $\sum_{i=1}^n (Y_i - M_i) = \nu_n(p_n - \pi_n)$, one obtains

$$\begin{aligned} P'(E) &= P' \left(\left\{ \omega \in E \cap A : (a_{n,\varepsilon}(\omega))^{-1} \nu_n(\omega) (p_n(\omega) - \pi_n(\omega)) \rightarrow 0 \right\} \right) \\ &\quad + P' \left(\left\{ \omega \in E \cap A^c : \nu_n(\omega) (p_n(\omega) - \pi_n(\omega)) \text{ converges} \right\} \right) \\ &\leq P' \left(\left\{ \omega \in E \cap A : p_n(\omega) - \pi_n(\omega) \rightarrow 0 \right\} \right) \\ &\quad + P' \left(\left\{ \omega \in E \cap A^c : p_n(\omega) - \pi_n(\omega) \rightarrow 0 \right\} \right) = P'(E \setminus F). \end{aligned}$$

In view of the previous remarks, Theorem 7 directly yields the following proposition, due to Dawid [7]:

$P'(F) = 0$, whenever P' is assessed according to the Kolmogorov theory of conditional probability.

In order to appreciate the contribution of de Finetti's theory to the analysis of the concept of well-calibrated inference, let us quote some of the comments made by Dawid on the above proposition.

Even if $P'(F) = 0$, «in practice, ..., it is rare for probability forecasts to be well-calibrated (so far as can be judged from finite experience) and no realistic forecaster would believe too strongly in his own calibration performance. We have a paradox: an event can be distinguished (easily, and indeed in many ways) that is given subjective probability one and yet is not regarded as «morally certain»», [7, p. 608].

Subsequently, Dawid concludes that the previous alleged paradox has «destructive implications for the theory of coherence», [7, abstract].

In fact, this line of reasoning rests on a misunderstanding: Dawid thinks that the only coherent p.i.'s are those assessed in conformity with Kolmogorov's notion of conditional probability. Hence, he considers the thesis of the corollary to Theorem 7 as the only one compatible with the coherence principle.

It is then about time that one carefully examined what really happens, as regards well-calibration, within the theory of coherence.

3.3. Well-calibration and the theory of coherence. We begin by showing that, under very weak assumptions, a coherent forecaster can either expect to be well-calibrated or, instead, give strictly positive probability to the event F of miscalibration.

Theorem 8. *Let P be a p.i. on $\mathcal{C}$, and let P' be any coherent extension of P to $\{g \mid H: g \in \mathcal{L} \cup \{E\}, H \in \mathcal{H}\}$. If*

(j) $F \cap A \neq \emptyset, (E \setminus F) \cap A \neq \emptyset$ for every A in $\mathcal{L}$ with $A \cap E \neq \emptyset$, then, given $\alpha \in [0, P'(E)]$, there is a further coherent extension of P', P'' , such that $P''(F) = \alpha$. Moreover, P' can be taken such that $P'(E) = 1$, whenever

(jj) $E \cap A \neq \emptyset$ for every A in $\mathcal{L} \setminus \{\emptyset\}$.

Proof. We need the following claim.

Claim. Let $\mathcal{A}$ be an algebra of subsets of Ω , $B \in \mathcal{A}$ and $V \subset B$. If $D \cap V \neq \emptyset$ whenever $D \in \mathcal{A}$ and $D \cap B \neq \emptyset$, then every probability μ on $\mathcal{A}$ can be extended to a probability μ' on $\mathcal{A} \cup \{V\}$ in such a way that $\mu'(V) = \mu(B)$.

In fact, for $C \subset \Omega$ let $\mu^*(C) = \inf\{\mu(H): C \subset H \in \mathcal{A}\}$. Then, since no proper subset of B , which belongs to $\mathcal{A}$, can cover V , it must be $\mu^*(V) = \mu(B)$. Hence, the claim follows from Theorem 3.3.3 of [4, p. 73].

Now, let $\mathcal{A}$ be the algebra generated by $\mathcal{L} \cup \{E\}$, $\mathcal{G}$ the algebra generated by $\mathcal{A} \cup \{F\}$, and Q any coherent extension of P' to $\{g \mid H: g \in \mathcal{A}, H \in \mathcal{H}\}$. We are showing that there are two p.i.'s Q_1 and Q_2 , extending Q to $\mathcal{C}^* = \{g \mid H: g \in \mathcal{G}, H \in \mathcal{H}\}$, such that $Q_1(F) = 0$ and $Q_2(F) = P'(E)$. After proving this, it is an easy consequence of the principle of coherence that, for every $\alpha \in [0, P'(E)]$ there is a p.i. P'' on $\mathcal{C}^*$, extending Q and such that $P''(F) = \alpha$.

Let us start with Q_1 . By (j) and the Claim (applied with $B = E$ and $V = E \setminus F$), there is a probability on $\mathcal{G}$, say $S(\cdot)$, extending $Q(\cdot)$ and such that $S(F) = 0$. Let $\mathcal{Z}$ be the set of partial histories $(x_1, \dots, x_n)$ such that $Q(I(x_1, \dots, x_n)) = 0$. Being a «part» of a p.i., $\{Q(\cdot | x_1, \dots, x_n): (x_1, \dots, x_n) \in \mathcal{Z}\}$ is a coherent family of probabilities on $\mathcal{A}$. Hence, by Theorem 1, it can be extended to a coherent family of probabilities on $\mathcal{G}$, say $\{S(\cdot | x_1, \dots, x_n): (x_1, \dots, x_n) \in \mathcal{Z}\}$. Let Q_1 be the function on $\mathcal{C}^*$ defined by $Q_1(\cdot) = S(\cdot)$, $Q_1(\cdot | x_1, \dots, x_n) = S(\cdot | x_1, \dots, x_n)$ for $(x_1, \dots, x_n) \in \mathcal{Z}$, and $Q_1(g | x_1, \dots, x_n) = S(gI(x_1, \dots, x_n))/Q(I(x_1, \dots, x_n))$ for $g \in \mathcal{G}$ and $(x_1, \dots, x_n) \notin \mathcal{Z}$. By (i)–(iv) of Subsection 2.1, Q_1 is easily seen to be a p.i. on $\mathcal{C}^*$. Moreover, the existence of Q_2 can be shown precisely as that of Q_1 , after interchanging F with $E \setminus F$. This proves the first part of the theorem. As to the remaining part, it can be argued following the same line of reasoning used in the first.

Condition (j) of Theorem 8 states that F and $E \setminus F$ are tail subsets of E , in the sense that a point of E belongs to F or to $E \setminus F$ independently of its first n coordinates, whatever n may be. Furthermore, Theorem 8 definitively dispels the doubts that Dawid raised about the value of the theory of coherence. As a matter of fact, coherence does not compel one to ignore the event of miscalibration. This situation, as the following example witnesses, can occur even within very common statistical contexts.

Example 1. Let $\mathcal{X} = \mathbf{R}$ and let σ be any strategy such that, when restricted to $\{(-\infty, x]: x \in \mathbf{R}\}$, σ_0 coincides with a normal law with parameter $(0, 2)$ and $\sigma_n(x_1, \dots, x_n)$ coincides with a normal law with parameter $((n/(n+1))\bar{x}_n, (n+2)/(n+1))$, where $\bar{x}_n = \sum_{i=1}^n x_i/n$. The above σ is consistent with the model of a sequence of independent $N(\theta, 1)$ random variables, θ being a $N(0, 1)$ random variable.

Let P be the strategic p.i. induced by σ on $\mathcal{C}$, let $B_n = (-\infty, 0]$, and let h_n be defined as in (5) with $p = \frac{1}{2}$ and δ any positive number. For $\omega = (x_1, \dots, x_n, \dots)$, let $\bar{x}_n(\omega) = \sum_{i=1}^n x_i/n$. If ω is such that $\bar{x}_n(\omega) \rightarrow 0$, then ω belongs to E . Moreover, if ω' coincides with ω up to a finite number of coordinates, then $\omega' \in E$. Hence condition (jj) holds. Fix $A \in \mathcal{L}$ with $A \cap E \neq \emptyset$, and let $(y_1, \dots, y_m)$ be such that the history $(y_1, \dots, y_m, \omega) := (y_1, \dots, y_m, x_1, x_2, \dots)$ is in A for all $\omega \in \mathcal{X}^\infty$. Taking ω such that $x_i > 0$ and $\bar{x}_n(\omega) \rightarrow 0$, $(y_1, \dots, y_m, \omega) \in A \cap F$. Likewise, choosing ω such that $x_i > 0$, $x_{i+1} < 0$ and $\bar{x}_n(\omega) \rightarrow 0$ sufficiently fast, $(y_1, \dots, y_m, \omega) \in A \cap (E \setminus F)$. Hence condition (j) holds, too. In view of Theorem 8, coherent extensions of P can be found, say P_1 and P_2 , such that $P_1(E) = P_2(E) = 1$, $P_1(F) > 0$ and $P_2(F) = 0$. Plainly, P_2 is well-calibrated and P_1 is not.

The following proposition provides conditions for a p.i. to be finitarily well-calibrated. In spite of the formal analogy with Theorem 7, such a proposition is definitely more consistent with the finitary nature of real statistical problems. In fact, it does not assume any hypothesis on possible infinite-dimensional extensions of P . Instead, Theorem 7 is proved under the

adoption of a particular infinite-dimensional extension of P . On the other hand, the proof of the new proposition involves some fine technical devices, which the proof of Theorem 7 does not actually require.

Theorem 9. *Let P be a strategic p.i. on $\mathcal{C}$, such that $\lim_n P(\{\omega: \nu_n(\omega) > c\}) = 1$ for every $c > 0$. Then, P is finitarily well-calibrated.*

P r o o f. Since the proof we are able to show is quite long, let us first state three claims whose proof will be anyway postponed. Moreover, since P is strategic, according to Theorem 5 it can be uniquely extended as a strategic p.i. on $\{g \mid H: g \text{ is a bounded r.q. depending only on a fixed finite number of coordinates, } H \in \mathcal{H}\}$. Let us denote such an extension again with P .

For $\omega = (x_1, x_2, \dots)$ and $i \in \mathbb{N}$, set $\psi_i(\omega) = E_i(\omega) - P(E_i \mid x_1, \dots, x_{i-1})$, $y_i(\omega) = h_i(\omega)/\nu_i(\omega)$ if $\nu_i(\omega) > 0$, and $y_i(\omega) = 0$ otherwise. Further, let

$$M_{jn}(\omega) = \begin{cases} \sum_{i=j}^n y_i(\omega) \psi_i(\omega) & \text{if } j \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Note that, for $r \leq q \leq s$, $M_{rs} = M_{rq} + M_{q+1,s}$.

Claim 1. For $1 \leq r < s \leq n$, $P(y_r \psi_r y_s \psi_s) = 0$.

Claim 2. For every $m, n \in \mathbb{N}$ with $m \leq n$ and every $a > 0$,

$$P\left(\bigcup_{j=1}^m \{|M_{jn}| > a\}\right) \leq \frac{m}{a^2} \sum_{i=1}^{\infty} \frac{1}{i^2}.$$

Claim 3. For every $\varepsilon, \delta > 0$ there is n_0 such that for each $m, n, k \in \mathbb{N}$ with $n \geq m \geq n_0$

$$P\left(\{|M_{j,n+i}| \leq \varepsilon, m \leq j \leq n+i, 0 \leq i \leq k\}\right) > 1 - \delta.$$

We now prove that Claims 2–3 imply Theorem 9. Let $\mathcal{U}_n(\omega) = p_n(\omega) - \pi_n(\omega)$ for all $n \in \mathbb{N}$ and $\omega \in \Omega$. Since $\nu_n \rightarrow +\infty$ in probability, it suffices to show that

$$\forall \varepsilon, \delta > 0 \exists n_1 \in \mathbb{N}: n \geq n_1 \implies P\left(\{|\mathcal{U}_n| \leq \varepsilon, \dots, |\mathcal{U}_{n+k}| \leq \varepsilon\}\right) > 1 - \delta,$$

$\forall k \in \mathbb{N}$. Note that $\mathcal{U}_n = 0$ whenever $\nu_n = 0$, and otherwise

$$\begin{aligned} \mathcal{U}_n &= \frac{1}{\nu_n} \sum_{i=1}^n \nu_i y_i \psi_i = \frac{1}{\nu_n} \sum_{i=1}^n \sum_{j=1}^i h_j y_i \psi_i \\ &= \frac{1}{\nu_n} \sum_{j=1}^n h_j \sum_{i=j}^n y_i \psi_i = \frac{1}{\nu_n} \sum_{j=1}^n h_j M_{jn}. \end{aligned}$$

Hence, after fixing $\varepsilon, c > 0$ and $n, m, k \in \mathbb{N}$ with $m \leq n$, one obtains

$$\begin{aligned}
 & P\left(\left\{|\mathcal{U}_n| \leq \varepsilon, \dots, |\mathcal{U}_{n+k}| \leq \varepsilon\right\}\right) \\
 &= P\left(\bigcap_{i=0}^k \left\{\frac{1}{\nu_{n+i}} \left| \sum_{j=1}^m h_j M_{j,n+i} + \sum_{j=m+1}^{n+i} h_j M_{j,n+i} \right| \leq \varepsilon\right\}\right) \\
 &\geq P\left(\bigcap_{i=0}^k \left\{\frac{1}{\nu_{n+i}} \left| \sum_{j=1}^m h_j M_{j,n+i} \right| \leq \frac{\varepsilon}{2}, \frac{1}{\nu_{n+i}} \left| \sum_{j=m+1}^{n+i} h_j M_{j,n+i} \right| \leq \frac{\varepsilon}{2}\right\}\right) \\
 &\geq P\left(\bigcap_{i=0}^k \left\{\frac{1}{\nu_{n+i}} \left| \sum_{j=1}^m h_j M_{j,n+i} \right| \leq \frac{\varepsilon}{2}\right\}\right) \\
 &\quad + P\left(\bigcap_{i=0}^k \left\{\frac{1}{\nu_{n+i}} \left| \sum_{j=m+1}^{n+i} h_j M_{j,n+i} \right| \leq \frac{\varepsilon}{2}\right\}\right) - 1 \\
 &\geq P\left(\bigcap_{i=0}^k \left\{\frac{1}{\nu_{n+i}} \sum_{j=1}^m |M_{j,n+i}| \leq \frac{\varepsilon}{2}, |M_{j,n+i}| \leq c, 1 \leq j \leq m\right\}\right) \\
 &\quad + P\left(\left\{|M_{j,n+i}| \leq \frac{\varepsilon}{2}, m+1 \leq j \leq n+i, 0 \leq i \leq k\right\}\right) - 1 \\
 &\geq P\left(\bigcap_{i=0}^k \left\{\frac{mc}{\nu_{n+i}} \leq \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{|M_{j,n+i}| \leq c, 1 \leq j \leq m, 0 \leq i \leq k\right\}\right) \\
 &\quad + P\left(\left\{|M_{j,n+i}| \leq \frac{\varepsilon}{2}, m \leq j \leq n+i, 0 \leq i \leq k\right\}\right) - 2 \\
 &\geq P\left(\left\{\frac{mc}{\nu_n} \leq \frac{\varepsilon}{2}\right\}\right) - P\left(\bigcup_{j=1}^m \left\{|M_{jn}| > \frac{c}{2}\right\}\right) \\
 &\quad - P\left(\bigcup_{i=1}^k \left\{|M_{n+1,n+i}| > \frac{c}{2}\right\}\right) \\
 &\quad + P\left(\left\{|M_{j,n+i}| \leq \frac{\varepsilon}{2}, m \leq j \leq n+i, 0 \leq i \leq k\right\}\right) - 1. \tag{8}
 \end{aligned}$$

Now, four probabilities are involved in (8). Let $\delta > 0$. By Claim 3, there is n_0 such that for every $m \geq n_0$ and every k and $n \geq m$, the third probability is less than $\delta/4$, whenever $c > \varepsilon$, while the fourth is greater than $1 - \delta/4$. Fix $m \geq n_0$. By Claim 2, there is $c > \varepsilon$ such that the second probability is less than $\delta/4$. Finally, since $\nu_n \rightarrow +\infty$ in probability, one can take n_1 sufficiently large so that the first probability is greater than $1 - \delta/4$, for all $n \geq n_1$.

In what follows, for $n \in \mathbb{N}$, γ_n denotes the probability on the power set of $\mathfrak{X}^n$ given by $\gamma_n(S) = P(S \times \mathfrak{X}^\infty)$ for all $S \subset \mathfrak{X}^n$.

Proof of Claim 1 of Theorem 9. By strategicity, and since $P(\psi_s | x_1, \dots, x_{s-1}) = 0$,

$$\begin{aligned} P(y_r y_s \psi_r \psi_s) &= \int P(y_r y_s \psi_r \psi_s | x_1, \dots, x_{s-1}) d\gamma_{s-1} \\ &= \int y_r(x_1, \dots, x_{r-1}) y_s(x_1, \dots, x_{s-1}) \psi_r(x_1, \dots, x_r) \\ &\quad \times P(\psi_s | x_1, \dots, x_{s-1}) d\gamma_{s-1} = 0. \end{aligned}$$

Proof of Claim 2 of Theorem 9. By Claim 1 and Chebyshev's inequality,

$$\begin{aligned} P\left(\bigcup_{j=1}^m \{|M_{jn}| > a\}\right) &\leq \sum_{j=1}^m \frac{1}{a^2} P(M_{jn}^2) = \sum_{j=1}^m \frac{1}{a^2} P\left(\sum_{i=j}^n y_i^2 \psi_i^2\right) \\ &\leq \frac{m}{a^2} \sum_{i=1}^{\infty} \frac{1}{i^2}. \end{aligned}$$

Proof of Claim 3 of Theorem 9. Fix $\varepsilon > 0$, and $m, n, k \in \mathbb{N}$ with $n \geq m$. Set

$$\begin{aligned} C_1 &= \{|M_{jn}| \leq \varepsilon, m \leq j \leq n\}, \quad C_2 = \{|M_{n+1, n+k}| \leq \varepsilon\}, \\ C_3 &= \{|M_{j, n+k}| \leq \varepsilon, n+2 \leq j \leq n+k\}. \end{aligned}$$

For $\omega \in C_1 \cap C_2 \cap C_3$, $m \leq j \leq n+i$ and $0 \leq i \leq k$, it must be $|M_{j, n+i}(\omega)| \leq 3\varepsilon$, since

$$M_{m+p, n+i} = M_{m+p, n} + M_{n+1, n+i} = M_{m+p, n} + M_{n+1, n+k} - M_{n+i+1, n+k}$$

for $p = 0, \dots, n-m$ and $i = 0, \dots, k$, and

$$M_{n+1+p, n+i} = M_{n+1+p, n+k} - M_{n+i+1, n+k}$$

for $p = 0, \dots, i-1$ and $i = 1, \dots, k$.

Hence, it suffices to show that

$$\begin{aligned} \forall \varepsilon, \delta > 0 \exists n_0 \text{ such that } \forall r, s \in \mathbb{N}, \text{ with } s \geq r \geq n_0, \\ P\left(\{|M_{js}| \leq \varepsilon, r \leq j \leq s\}\right) > 1 - \delta. \end{aligned} \quad (9)$$

We now prove (9). To this end, some new notation is needed:

$$\begin{aligned} A_p &= \{\omega: \nu_r(\omega) = p\}, \quad p = 0, \dots, r, \\ \rho(i)(\omega) &= \min \{q: \nu_q(\omega) = \nu_r(\omega) + i\}, \\ f(t)(\omega) &= \rho(t - \nu_r(\omega))(\omega) \text{ [roughly, } f(t)(\omega) \text{ is the first index } q \\ &\quad \text{such that } \nu_q(\omega) = t], \\ t(j)(\omega) &= \max \left\{ r, f\left(\left[\sqrt{\nu_j(\omega)}\right]^2\right) \right\}, \end{aligned}$$

$$\alpha(p) = [\sqrt{p}] + 1 \quad \text{for } p = 0, \dots, r \text{ and } \beta = [\sqrt{s}] - 1,$$

$$D_p = \left\{ \alpha^2(p), (\alpha(p) + 1)^2, \dots, (\beta + 1)^2 \right\}.$$

Then,

$$P\left(\bigcup_{j=r}^s \{|M_{js}| > \varepsilon\}\right) \leq P\left(\left\{\exists j = r, \dots, s \text{ with } |M_{t(j),s}| > \frac{\varepsilon}{2},\right.\right. \\ \left.\left.\text{or } \exists j = r, \dots, s \text{ with } |M_{js} - M_{t(j),s}| > \frac{\varepsilon}{2}\right\}\right) \\ \leq P\left(\left\{|M_{rs}| > \frac{\varepsilon}{2}\right\}\right) \quad (*) \\ + \sum_{p=0}^r P\left(\left\{\exists j = r+1, \dots, s \text{ with } A_p |M_{t(j),s}| > \frac{\varepsilon}{2}\right\}\right) \quad (**) \\ + \sum_{p=0}^r P\left(\left\{\exists j = r+1, \dots, s \text{ with } A_p |M_{js} - M_{t(j),s}| > \frac{\varepsilon}{2}\right\}\right). \quad (***)$$

We treat (*), (**) and (***) separately.

Let us start with (*). By Claim 1, and for every fixed $k \in \mathbb{N}$,

$$P\left(\left\{|M_{rs}| > \frac{\varepsilon}{2}\right\}\right) \leq \frac{4}{\varepsilon^2} P\left(\sum_{i=r}^s y_i^2\right) \\ \leq \frac{4}{\varepsilon^2} \left(\sum_{i=1}^{\infty} \frac{1}{i^2} P(\{\nu_r \leq k\}) + \sum_{i=k+1}^{\infty} \frac{1}{i^2} P(\{\nu_r > k\}) \right).$$

Thus, since $\nu_r \rightarrow +\infty$ in probability, it is easily verified that (*) converges to 0 as $r \rightarrow +\infty$, uniformly in s .

Let us turn now to (**).

$$(**) \leq \sum_{p=0}^r \sum_{q^2 \in D_p} P\left(\left\{A_p \left| \sum_{i=\rho(q^2-p)}^s y_i \psi_i \right| > \frac{\varepsilon}{2}\right\}\right) \\ \leq \sum_{p=0}^r \sum_{q^2 \in D_p} \frac{4}{\varepsilon^2} P\left(\left(A_p \sum_{i=\rho(q^2-p)}^s y_i \psi_i\right)^2\right) \\ = \sum_{p=0}^r \sum_{q^2 \in D_p} \frac{4}{\varepsilon^2} P\left(A_p \sum_{i=\rho(q^2-p)}^s y_i^2 \psi_i^2\right) \quad (10)$$

where the last equality will be proved later. Then,

$$\begin{aligned}
 (**) &\leq \sum_{p=0}^r \sum_{i=\alpha(p)}^{\infty} \frac{4}{\varepsilon^2} P(A_p) \sum_{j=0}^{\infty} \frac{1}{(i^2 + j)^2} \\
 &\leq \sum_{p=0}^r \sum_{i=\alpha(p)}^{\infty} \frac{4}{\varepsilon^2} \left(\frac{1}{i^4} + \frac{1}{i^2} \right) P(A_p) = \sum_{p=0}^r a(p) P(A_p),
 \end{aligned}$$

where $a(p) = \sum_{i=\alpha(p)}^{\infty} (4/\varepsilon^2) (1/i^4 + 1/i^2)$. Moreover, for $0 \leq M < r$,

$$\begin{aligned}
 \sum_{p=0}^r a(p) P(A_p) &= \sum_{p=0}^M a(p) P(A_p) + \sum_{p=M+1}^r a(p) P(A_p) \\
 &\leq \sum_{p=0}^M a(p) P(A_p) + a(M+1) \\
 &\leq \sup_p a(p) (1 - P(\{\nu_r > M\})) + a(M+1).
 \end{aligned}$$

Now, since $a(p) \downarrow 0$ and $\nu_r \rightarrow +\infty$ in probability, there is M such that $a(M+1) < \varepsilon$ and, fixed such M , there is r_0 such that $P(\{\nu_r > M\}) > 1 - \varepsilon$ for all $r \geq r_0$. To sum up, $(**)$ converges to 0 as $r \rightarrow +\infty$, uniformly in s .

We still have to prove (10). Let $B_t = \{\rho(q^2 - p) = t\}$ and $B = \cup_{t=p+1}^{s-1} B_t$. Then,

$$\begin{aligned}
 P\left(A_p B \sum_{\rho(q^2-p) \leq i < j \leq s} y_i \psi_i y_j \psi_j\right) &= \sum_{t=p+1}^{s-1} \sum_{t \leq i < j \leq s} P(A_p B_t y_i \psi_i y_j \psi_j) \\
 &= \sum_{t=p+1}^{s-1} \sum_{t \leq i < j \leq s} \int P(A_p B_t y_i \psi_i y_j \psi_j \mid x_1, \dots, x_{j-1}) d\gamma_{j-1} \\
 &= \sum_{t=p+1}^{s-1} \sum_{t \leq i < j \leq s} \int A_p B_t y_i \psi_i y_j P(\psi_j \mid x_1, \dots, x_{j-1}) d\gamma_{j-1} = 0,
 \end{aligned}$$

and this yields (10).

Finally, let's consider $(***)$.

$$\begin{aligned}
 (***) &= \sum_{p=0}^r P\left(\left\{\exists j = r+1, \dots, s \text{ with } A_p |M_{js} - M_{t(j),s}| > \frac{\varepsilon}{2}\right\}\right) \\
 &\leq \sum_{p=0}^r P\left(\left\{\exists j \text{ with } p \leq \nu_j \leq \alpha^2(p) \text{ and } A_p |M_{t(j),j-1}| > \frac{\varepsilon}{2}\right\}\right) \\
 &\quad + \sum_{p=0}^r P\left(\left\{\exists j \text{ with } \alpha^2(p) < \nu_j \leq (\alpha(p)+1)^2\right\}\right)
 \end{aligned}$$

$$\text{or } (\alpha(p) + 1)^2 < \nu_j \leq (\alpha(p) + 2)^2$$

$$\text{or } \dots \text{ or } \beta^2 < \nu_j \leq (\beta + 1)^2, \text{ and } A_p |M_{t(j), j-1}| > \frac{\varepsilon}{2} \Bigg\}$$

$$+ \sum_{p=0}^r P \left(\left\{ \exists j \text{ with } (\beta + 1)^2 < \nu_j \leq s \text{ and } A_p |M_{t(j), j-1}| > \frac{\varepsilon}{2} \right\} \right).$$

Let $\varphi_i = \psi_i$ if $i \leq s$ and $\varphi_i = 0$ otherwise. Then,

$$\begin{aligned} (***) &\leq P(A_0) + \sum_{p=1}^r P \left(\left\{ A_p \left| \frac{h_r \varphi_r}{p} \right| > \frac{\varepsilon}{2} \text{ or } A_p \left| \frac{h_r \varphi_r}{p} + \frac{\varphi_{f(p+1)}}{p+1} \right| > \frac{\varepsilon}{2} \right. \right. \\ &\quad \left. \left. \text{or } A_p \left| \frac{h_r \varphi_r}{p} \right| + \frac{\varphi_{f(p+1)}}{p+1} + \dots + \frac{\varphi_{f(\alpha^2(p))}}{\alpha^2(p)} > \frac{\varepsilon}{2} \right\} \right) \\ &\quad + \sum_{p=1}^r \sum_{q=\alpha(p)}^{\beta} P \left(\left\{ A_p \left| \frac{\varphi_{f(q^2)}}{q^2} \right| > \frac{\varepsilon}{2} \text{ or } A_p \left| \frac{\varphi_{f(q^2)}}{q^2} + \frac{\varphi_{f(q^2+1)}}{q^2+1} \right| > \frac{\varepsilon}{2} \right. \right. \\ &\quad \left. \left. \text{or } \dots \text{ or } A_p \left| \frac{\varphi_{f(q^2)}}{q^2} + \dots + \frac{\varphi_{f((q+1)^2)}}{(q+1)^2} \right| > \frac{\varepsilon}{2} \right\} \right) \\ &\quad + \sum_{p=1}^r P \left(\left\{ A_p \left| \frac{\varphi_{f((\beta+1)^2)}}{(\beta+1)^2} \right| > \frac{\varepsilon}{2} \right. \right. \\ &\quad \left. \left. \text{or } \dots \text{ or } A_p \left| \frac{\varphi_{f((\beta+1)^2)}}{(\beta+1)^2} + \dots + \frac{\varphi_{f(s)}}{s} \right| > \frac{\varepsilon}{2} \right\} \right) \\ &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{t=0}^{\alpha^2(p)-p} P \left(A_p \left(\frac{h_r \varphi_r}{p} + \frac{\varphi_{f(p+1)}}{p+1} + \dots + \frac{\varphi_{f(p+t)}}{p+t} \right)^2 \right) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q=\alpha(p)}^{\beta} \sum_{t=0}^{(q+1)^2-q^2} P \left(A_p \left(\frac{\varphi_{f(q^2)}}{q^2} + \dots + \frac{\varphi_{f(q^2+t)}}{q^2+t} \right)^2 \right) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{t=0}^{s-(\beta+1)^2} P \left(A_p \left(\frac{\varphi_{f((\beta+1)^2)}}{(\beta+1)^2} + \dots + \frac{\varphi_{f((\beta+1)^2+t)}}{(\beta+1)^2+t} \right)^2 \right) \\ &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{t=0}^{\alpha^2(p)-p} \left(\frac{1}{p^2} + \dots + \frac{1}{(p+t)^2} \right) P(A_p) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q=\alpha(p)}^{\beta} \sum_{t=0}^{(q+1)^2-q^2} \left(\frac{1}{q^4} + \dots + \frac{1}{(q^2+t)^2} \right) P(A_p) \end{aligned}$$

$$+ \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{t=0}^{s-(\beta+1)^2} \left(\frac{1}{(\beta+1)^4} + \cdots + \frac{1}{((\beta+1)^2+t)^2} \right) P(A_p) \quad (11)$$

where the last inequality will be proved later. Next,

$$\begin{aligned} (***) &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r (\alpha^2(p) - p + 1) \\ &\quad \times \left(\frac{1}{p^2} + \frac{1}{(p+1)^2} + \cdots + \frac{1}{\alpha^4(p)} \right) P(A_p) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q=\alpha(p)}^{\beta} ((q+1)^2 - q^2 + 1) \\ &\quad \times \left(\frac{1}{q^4} + \frac{1}{(q^2+1)^2} + \cdots + \frac{1}{(q+1)^4} \right) P(A_p) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r (s - (\beta+1)^2 + 1) \left(\frac{1}{(\beta+1)^4} + \cdots + \frac{1}{s^2} \right) P(A_p) \\ &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r \left(\alpha^2(p) - (\alpha(p) - 1)^2 + 1 \right) \\ &\quad \times \left(\frac{1}{(\alpha(p) - 1)^4} + \frac{1}{((\alpha(p) - 1)^2 + 1)^2} + \cdots + \frac{1}{\alpha^4(p)} \right) P(A_p) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q=\alpha(p)}^{\beta} ((q+1)^2 - q^2 + 1) \\ &\quad \times \left(\frac{1}{q^4} + \frac{1}{(q^2+1)^2} + \cdots + \frac{1}{(q+1)^4} \right) P(A_p) \\ &\quad + \frac{4}{\varepsilon^2} \sum_{p=1}^r ((\beta+2)^2 - (\beta+1)^2 + 1) \\ &\quad \times \left(\frac{1}{(\beta+1)^4} + \frac{1}{((\beta+1)^2 + 1)^2} + \cdots + \frac{1}{(\beta+2)^4} \right) P(A_p) \\ &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q=\alpha(p)-1}^{\beta+1} (2q+2) \\ &\quad \times \left(\frac{1}{q^4} + \frac{1}{(q^2+1)^2} + \cdots + \frac{1}{(q+1)^4} \right) P(A_p) \\ &\leq P(A_0) + \frac{4}{\varepsilon^2} \sum_{p=1}^r \sum_{q \geq \alpha(p)-1} (2q+2) \left(\frac{1}{q^4} + \frac{2q+1}{(q+1)^2 q^2} \right) P(A_p) \end{aligned}$$

$$= P(A_0) + \sum_{p=1}^r a(p) P(A_p),$$

where

$$a(p) = \frac{4}{\varepsilon^2} \sum_{q \geq \alpha(p)-1} (2q+2) \left(\frac{1}{q^4} + \frac{2q+1}{(q+1)^2 q^2} \right).$$

To sum up, $(***) \leq P(A_0) + \sum_{p=1}^r a(p) P(A_p)$, with $a(p) \downarrow 0$. Following the same argument used for $(**)$, one can show that $(***)$ converges to 0 as $r \rightarrow +\infty$, uniformly in s .

Let's now turn to (11). As is easily seen, it suffices to show that $P(A_p \varphi_{f(i)} \varphi_{f(j)}) = 0$ for $p < i < j$. To this purpose, let $D_z = \{f(j) = z\}$. Then,

$$\begin{aligned} P(A_p \varphi_{f(i)} \varphi_{f(j)}) &= \sum_{z=r}^s P(D_z A_p \psi_{f(i)} \psi_{f(j)}) \\ &= \sum_{z=r}^s \int P(D_z A_p \psi_{f(i)} \psi_{f(j)} \mid x_1, \dots, x_{z-1}) d\gamma_{z-1} \\ &= \sum_{z=r}^s \int D_z A_p \psi_{f(i)} P(\psi_z \mid x_1, \dots, x_{z-1}) d\gamma_{z-1} = 0. \end{aligned}$$

This completes the proof.

We conclude the present subsection with a couple of examples in order to show that neither well-calibration implies finitary well-calibration nor finitary well-calibration implies well-calibration.

E x a m p l e 2 [well-calibration does not imply finitary well-calibration]. Let $\mathfrak{X} = \mathbb{N}$, $h_n \equiv 1$ and $B_n = B$ for all n , where B is the set of even numbers. Let P be any p.i. such that $P(B \times \mathfrak{X}^\infty) = P(\mathfrak{X}^n \times B \times \mathfrak{X}^\infty) = b$, for some $b \in (0, 1)$ and all n . Setting $c_0 = P(B \times \mathfrak{X}^\infty)$ and $c_n(x_1, \dots, x_n, \dots) = P(\mathfrak{X}^n \times B \times \mathfrak{X}^\infty \mid x_1, \dots, x_n)$, one obtains

$$\int (p_n - \pi_n) dP = b - \frac{1}{n} \sum_{i=1}^n \int c_{i-1} dP.$$

Since $\int (p_n - \pi_n) dP \rightarrow 0$ is necessary for P to be finitarily well-calibrated, it suffices to choose c_n in such a way that $(1/n) \sum_{i=1}^n \int c_{i-1} dP$ does not converge to b and the resulting P satisfies condition (j) of Theorem 8 (so that there are well-calibrated coherent extensions of P).

For instance, set

$$c_n(x_1, \dots, x_n, \dots) = \alpha_n \frac{1}{n} \sum_{i=1}^n B(x_i) + \beta_n \frac{1}{n} \sum_{i=1}^n D(x_i) + \gamma_n b \quad (12)$$

where $\alpha_n, \beta_n, \gamma_n \geq 0$, $\alpha_n + \beta_n + \gamma_n = 1$ and D is some subset of $\mathfrak{X}$. Moreover, take $\alpha_n, \beta_n, \gamma_n$ and D in such a way that $\alpha_n \rightarrow \alpha$, $\beta_n \rightarrow \beta$, $\gamma_n \rightarrow 0$, $P(\mathfrak{X}^n \times$

$D \times \mathcal{X}^\infty) = d$ for each n , with $\beta > 0$ and $d \neq b$. Clearly, $(1/n) \sum_{i=1}^n \int c_{i-1} dP$ does not converge to b . Fix $A \in \mathcal{L} \setminus \{\emptyset\}$, and let $(y_1, \dots, y_m)$ be such that $(y_1, \dots, y_m, \omega) \in A$ for all ω . Taking ω_1 such that all its coordinates are in $B \setminus D$, and ω_2 with all its coordinates in $D \cap B$, then $(y_1, \dots, y_m, \omega_1) \in F \cap A$ and $(y_1, \dots, y_m, \omega_2) \in (E \setminus F) \cap A$. Hence, condition (j) of Theorem 8 holds.

We note that (12) is a rather common assessment in insurance field, in presence of the so called «collateral data». For instance, B could represent the event «occurrence of some accident», and D any other event different from, but strictly related to, B .

Example 3 [finitary well-calibration does not imply well-calibration]. Consider Example 1 again but, for the sake of simplicity, take $h_n \equiv 1$ instead of defining h_n as in (5). According to Theorem 9, P is finitarily well-calibrated. Hence, it suffices to find a coherent extension of P which is not well-calibrated. This follows by Theorem 8, after noting that $E = \Omega$ and condition (j) can be checked precisely as in Example 1.

3.4. Further comments. Contrary to what Dawid writes in [7], there is no conflict between calibration and coherence, since there are coherent p.i.'s which do not ignore the event F of miscalibration; cf. Theorem 8. In any case, the probability attached to F rests on a subjective judgement, just as is the case with any other probability assessment. Clearly, there are specific coherent assignments which may be conducive to ignoring F . For instance, the assumption that P is strategic enables one to choose a particular strategic extension of P , the Lebesgue-like-extension P' of P , and Theorem 7 implies that $P'(F) = 0$. Incidentally, we note how Theorem 7 involves one particular, important though it might be, strategic extension of P . However, the single probability laws $P'(\cdot)$ and $P'(\cdot | x_1, \dots, x_n)$ composing an arbitrary strategic extension P' of P are mutually linked in such a manner as to suggest that strategicity implies well-calibration. More precisely, it is plausible that Theorem 7, and perhaps some form of the martingale convergence theorem, holds for every strategic extension of P . On the other hand, Theorem 7 does not state that the adoption of a strategic P on $\mathcal{C}$ is conducive to ignoring F . Indeed, the p.i. in Example 1 is strategic on $\mathcal{C}$ and F can be given strictly positive probability. Finally, according to Theorem 9, the assumption that P is strategic is enough for finitary well-calibration to come true. In conclusion, it is the assumption that a p.i. is strategic, rather than its coherence, of concern to well-calibration.

REFERENCES

1. Berti P., Regazzini E., Rigo P. De Finetti's coherence and complete predictive inferences. — Quaderno I.A.M.I. 90. 5 Milano, 1990.
2. Berti P., Regazzini E., Rigo P. Coherent statistical inference and Bayes theorem. — Ann. Statist., 1991, v. 19, p. 366–381.

3. Berti P., Rigo P. Weak disintegrability as a form of preservation of coherence. — J. Italian Statist. Soc., 1992, v. 1, p. 161–181.
4. Bhaskara Rao K. P. S., Bhaskara Rao M. Theory of Charges. London: Academic Press, 1983.
5. Chen R. On almost sure convergence in a finitely additive setting. — Z. Wahrscheinlichkeitstheor. verw. Geb., 1977, B. 37, S. 341–356.
6. Cifarelli M. D., Regazzini E. Sopra una versione finitamente additiva del processo di Ferguson–Dirichlet. — In: Scritti in omaggio a L. Daboni. Trieste: Ed. Lint, 1990, p. 67–81.
7. Dawid P. The well-calibrated Bayesian (With discussion). — J. Amer. Statist. Assoc., 1982, v. 77, p. 605–613.
8. Dawid P. Present position and potential developments: some personal views, statistical theory, the prequential approach (With discussion). — J. Roy. Statist. Soc., Ser. A, 1984, v. 147, p. 278–292.
9. Dawid P. Fisherian inference in likelihood and prequential frames of reference (With discussion). — J. Roy. Statist. Soc., Ser. B, 1991, v. 53, p. 79–109.
10. De Finetti B. Sulla proprieta' conglomerativa delle probabilita' subordinate. — Rendiconti R. Istituto Lombardo di Scienze e Lett., 1930, v. 63, p. 414–418.
11. De Finetti B. Sull'impostazione assiomatica del calcolo delle probabilita'. — Annali Triestini dell'Universita' di Trieste, 1949, v. 19, p. 29–81. [English translation in: De Finetti B. Probability, Induction and Statistics. New York: Wiley, 1972.]
12. De Finetti B. Teoria delle probabilita'. Torino: Einaudi, 1970. [English translation: De Finetti B. Theory of Probability. Chichester: Wiley Classics Library Edition, 1990.]
13. Dubins L. E. On Lebesgue-like extensions of finitely additive measures. — Ann. Probab., 1974, v. 2, p. 456–463.
14. Dubins L. E. Finitely additive conditional probabilities, conglomerability and disintegrations. — Ann. Probab., 1975, v. 3, № 1, p. 89–99.
15. Dubins L. E., Savage L. J. How to gamble if you must (Inequalities for stochastic processes). New York: McGraw-Hill, 1965. [Dubins L. E., Savage L. J. Inequalities for stochastic processes (How to gamble if you must). New York: Dover, 1976.]
16. Dunford N., Schwartz J. T. Linear operators: Part I, General Theory. New York: Interscience, 1958.
17. Geisser S. Predictive analysis. — In: Encyclopedia of Statistical Sciences. New York: Wiley-Interscience, 1986, v. 7, p. 158–170.
18. Hill B. Posterior distribution of percentiles: Bayes theorem for sampling from a population. — J. Amer. Statist. Assoc., 1968, v. 63, p. 677–691.
19. Hill B. Parametric models for A_n : Splitting processes and mixtures. — J. Roy. Statist. Soc., Ser. B, 1993, v. 55, p. 423–433.
20. Kolmogorov A. N. Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin: Springer, Ergebnisse Mathematic, 1933. [English translation: Kolmogorov A. N. Foundations of the theory of probability. New York: Chelsea, 1950.]
21. Neveu J. Bases mathematiques du calcul des probabilités. 2ème ed. Paris: Masson, 1980.
22. Purves R. A., Sudderth W. D. Some finitely additive probability. — Ann. Probab., 1976, v. 4, p. 259–276.
23. Rao M. M. Conditional measures and applications. New York: Marcel Dekker, 1993.
24. Regazzini E. Finitely additive conditional probabilities. — Rend. Sem. Mat. Fis. Milano, 1985, v. 55, p. 69–89. Corrections in Rend. Sem. Mat. Fis. Milano, v. 57, p. 599.