Sublinear expectations: On large sample behaviours, Monte Carlo method, and coherent upper previsions

Pedro Terán

Abstract Shige Peng's sublinear expectations generalize ordinary linear expectation operators. It is shown that the behaviour of sample averages of Peng i.i.d. variables may be very different from the probabilistic intuition. In particular, Peng's generalization of the Monte Carlo method is shown to be wrong. It is also observed that sublinear expectations coincide with Peter Walley's coherent upper previsions on a linear space.

1 A tale of (improbably) three Pedros

It is my great pleasure to contribute to this volume honouring professor Pedro Gil. I met Pedro after participating in the local phase of the Mathematical Olympiad (which he organized) and he played a heavy role in my decision to become a Math student, a possibility I had not considered before. Some years later, I got a fellowship associated to his research project, which made it possible for me to get a doctoral degree and thus eventually led me to my current job. On that count alone, he has been one of the few most influential people in my life.

When I started as a fellow, the Department of Statistics at Oviedo had just spinned off from the Department of Mathematics. Both entities would continue to share equipment (like the copier machine) for some time but action was taken almost immediately to divide the post-graduates' room. So I found myself arriving there and being assigned no space until the works were done.

About one month later, I went to Pedro's office to inform him (in his capacity as head of the department) that I finally had a desk. 'But I have nothing to put on it anyway', I added nonchalantly. To my surprise, he smiled,

Escuela Politécnica de Ingeniería. Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, E-33071 Gijón, Spain, `teranpedro@uniovi.es`

emptied his pencil holder, and handed it to me. He went on to explain that it had been given to him decades earlier by Pedro Abellanas after a remarkably parallel conversation, and had been with him for all of his career. That token of appreciation I have cherished ever since.

This contribution is connected to Pedro's work on imprecise probability, a topic on which he was the advisor of Inés Couso and Enrique Miranda. It makes some observations about Shige Peng's theory of sublinear expectations. One is that a sublinear expectation is in fact the same thing as a coherent upper prevision on a linear space of gambles, in Walley's terminology. Another is that, although Peng's approach and particularly Peng's restrictive definition of independence allow one to obtain analogs of classical limit theorems, unfamiliar behaviours are possible too.

Finally, Peng's Monte Carlo method for computing sublinear expectations is wrong. That is a (rare, one should hope) instance of a counterexample to a result presented at a plenary lecture of the International Congress of Mathematicians, therefore something I'm happy to present in return for a pencil holder.

2 Introduction and preliminaries

Shige Peng devised a purely analytical approach to stochastic analysis. While that provides a path to learning stochastic calculus without any knowledge of probability theory, we are interested here in a different side effect. Peng's technical machinery, which in essence is based on viscosity solutions of nonlinear partial differential equations, allows one to extend the reach of probability theory in such a way that the usual expectation is replaced by more general functionals which are sublinear.

Extended results include fundamental limit theorems like the law of large numbers [6] (to be discussed below), the central limit theorem, and the large deviation principle. Peng also constructed a non-additive generalization of the Brownian motion with its associated stochastic calculus. For an introduction to Peng's approach, the reader is referred to the lecture notes [6], survey [4], and 2010 ICM plenary lecture [5].

A *sublinear expectation space* is a triple $(\Omega, \mathcal{H}, \mathbb{E})$ where Ω is a set, $\mathcal{H}$ is a linear space of real functions on Ω (including the constants and closed under taking the absolute value), and $\mathbb{E} : \mathcal{H} \to \mathbb{R}$ is a mapping such that

(SE.i) $\mathbb{E}[X] \leq \mathbb{E}[Y]$ if $X \leq Y$,
(SE.ii) $\mathbb{E}[c] = c$ for all constant c.
(SE.iii) $\mathbb{E}[X + Y] \leq \mathbb{E}[X] + \mathbb{E}[Y]$,
(SE.iv) $\mathbb{E}[cX] = c \cdot \mathbb{E}[X]$ for $c \geq 0$.

Elements of $\mathcal{H}$ will be called *random variables*. A subset $A \subset \Omega$ whose indicator function I_A is in $\mathcal{H}$ will be called an *event* (although, events are

conspicuously absent from Peng's approach). Its complement will be denoted A^c.

It turns out, as a consequence of the Hahn–Banach theorem, that every sublinear expectation admits the representation

$$\mathbb{E}[X] = \sup_{E \in \mathfrak{E}} E[X]$$

for some family of linear functionals $\mathfrak{E}$ (whose restrictions to events are thus finitely additive probabilities).

Set

$$C_{l.Lip}(\mathbb{R}^n) = \{\varphi : \mathbb{R}^n \to \mathbb{R} \mid \exists C, m \geq 0 \mid \forall x, y \in \mathbb{R}^n,$$

$$|\varphi(x) - \varphi(y)| \leq C(1 + |x|^m + |y|^m)|x - y|\}.$$

In order to consider independence or identical distribution in Peng's sense, one must also assume that $X \in \mathcal{H}$ implies $\varphi(X) \in \mathcal{H}$ for each $\varphi \in C_{l.Lip}(\mathbb{R})$. A random variable X will be called *Peng independent* of a random variable Y if

$$\mathbb{E}[\varphi(X, Y)] = \mathbb{E}[\mathbb{E}[\varphi(x, Y)]_{x=X}] \text{ for all } \varphi \in C_{l.Lip}(\mathbb{R}^2),$$

where $\mathbb{E}[\varphi(x, Y)]_{x=X}$ denotes the mapping $\omega \in \Omega \mapsto \mathbb{E}[\varphi(X(\omega), Y)]$. Observe that, under this definition, 'X independent of Y' does not imply 'Y independent of X' and, in fact, these are seldom equivalent [2].

Random variables X, Y will be called *Peng identically distributed* if

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)] \text{ for all } \varphi \in C_{l.Lip}(\mathbb{R}).$$

They will be called *Peng i.i.d.* if they are Peng identically distributed and X is Peng independent of Y.

A *random vector* is an element $(X_1, \ldots, X_n)$ of $\mathcal{H}^n$ such that $\varphi(X_1, \ldots, X_n) \in \mathcal{H}$ for each $\varphi \in C_{l.Lip}(\mathbb{R}^n)$. A random variable Y will be called *Peng independent* of a random vector $(X_1, \ldots, X_n)$ if

$$\mathbb{E}[\varphi(X_1, \ldots, X_n, Y)] = \mathbb{E}[\mathbb{E}[\varphi(x_1, \ldots, x_n, Y)]_{x_1=X_1, \ldots, x_2=X_2}]$$

for every $\varphi \in C_{l.Lip}(\mathbb{R}^{n+1})$. A sequence $\{X_n\}_n$ will be called *Peng independent and identically distributed* (Peng i.i.d.) if X_n is identically distributed as X_m for all n, m, and X_{n+1} is Peng independent of $(X_1, \ldots, X_n)$ for all $n \in \mathbb{N}$.

Dual to a sublinear expectation $\mathbb{E}$ is the superlinear functional $\mathcal{E}$ given by $\mathcal{E}[X] = -\mathbb{E}[-X]$. Unlike in ordinary probability theory, in general $\mathcal{E} \neq \mathbb{E}$. If $\mathbb{E}$ is linear and $\mathbb{E}[X_n] \to 0$ whenever X_n is a decreasing sequence converging to 0, then $\mathbb{E}$ is the expectation with respect to some probability measure. In that case one can check that, taking $\mathcal{H}$ to be a linear space of measurable functions, the notions above coincide with the usual ones.

Peng's PDE background and methods explain the idiosincratic choice of disregarding events and defining the distribution of a random variable only

in terms of expectations of the locally Lipschitz functions in $C_{l.Lip}(\mathbb{R})$. To handle events, it is convenient to measure them with

$$\nu(A) = \mathcal{E}[I_A], \quad \overline{\nu}(A) = \mathbb{E}[I_A] = 1 - \nu(A^c).$$

The notation $\nu, \overline{\nu}$ reflects the fact that, from a probabilistic point of view, it seems to be more intuitive to use ν since $\overline{\nu}(A) = \overline{\nu}(B) = 1$ can happen for disjoint A, B; hence, a set of $\overline{\nu}$-measure 1 is not necessarily 'big' in the sense of probability theory. But a set of ν-measure 1 is 'big' because, dually, ν is the infimum of a family of finitely additive probabilities.

We will say that a sequence of random variables $\{X_n\}_n$ converges to a random variable X *almost surely* if $\nu(X_n \to X) = 1$, *in probability* if $\nu(|X_n - X| < \varepsilon) \to 1$ for each $\varepsilon > 0$, and *in law* if $\mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)]$ for each $\varphi \in C_{l.Lip}(\mathbb{R})$.

Of these types of convergence, only convergence in law was considered by Peng. To ensure $\{X_n \to X\}$ and $\{|X_n - X| < \varepsilon\}$ are events, Peng's assumptions on $\mathcal{H}$ are not enough. That will not be a problem in this paper.

3 Counterexample to Monte Carlo method, and some unexpected behaviours

Peng proved the following law of large numbers for sublinear expectations.

Proposition 1. [6, Theorem II.3.1] *Let $(\Omega, \mathcal{H}, \mathbb{E})$ be a sublinear expectation space. Let $\{X_n\}_n$ be a sequence Peng i.i.d. as a random variable X. Then,*

$$\mathbb{E}[\varphi(S_n)] \to \sup_{x \in [\mathcal{E}[X], \mathbb{E}[X]]} \varphi(x)$$

for each $\varphi \in C_{l.Lip}(\mathbb{R})$.

It is interesting to note that, when $\mathbb{E}$ is the ordinary expectation against a probability measure, Peng's law is equivalent to a weak law of large numbers.

Corollary 1. (Weak law of large numbers) *Let X be an integrable random variable on a probability space $(\Omega, \mathcal{A}, P)$. Let $\{X_n\}_n$ be a sequence i.i.d. as X. Then*

$$S_n \to E(X)$$

in probability.

Proof. It suffices to prove it in the case when X is bounded, since standard truncation techniques yield then the more general result (see e.g. the proof of [7, Theorem 7.12]). Thus we assume without loss of generality that X takes on values in a compact interval $[a, b]$.

Let f be an arbitrary real continuous bounded function, and fix $\varepsilon > 0$. By the Weierstrass approximation theorem, there exists a polynomial φ such that $\sup_{x \in [a,b]} |f(x) - \varphi(x)| < \varepsilon$. With the triangle inequality,

$$|E[f(S_n)] - f(E[X])| \leq |E[f(S_n)] - E[\varphi(S_n)]|$$

$$+ |E[\varphi(S_n)] - \varphi(E[X])| + |\varphi(E[X]) - f(E[X])|.$$

The first and third terms are bounded above by ε.

We claim now $\varphi \in C_{l.Lip}(\mathbb{R})$. Since $C_{l.Lip}(\mathbb{R})$ is a linear space, it is enough to show that the mappings $x \mapsto x^i$ $(i \geq 1)$ are in $C_{l.Lip}(\mathbb{R})$. But

$$|x^i - y^i| = |\sum_{j=0}^{i-1} x^j y^{i-1-j}| \cdot |x - y| \leq (\sum_{j=0}^{i-1} |x|^j |y|^{i-1-j})|x - y|$$

$$\leq i \max\{|x|, |y|\}^{i-1} |x - y| \leq C(1 + |x|^m + |y|^m)|x - y|$$

for the choices $C = i, m = i - 1$.

Having established that claim, and observing that the X_n are Peng i.i.d., Proposition 1 gives

$$E[\varphi(S_n)] \to \sup_{x \in [\mathcal{E}[X], \mathbb{E}[X]]} \varphi(x) = \varphi(E[X]),$$

proving that the second term above goes to 0. In summary,

$$\limsup_n |E[f(S_n)] - f(E[X])| \leq 2\varepsilon,$$

whence the arbitrariness of ε implies $E[f(S_n)] \to f(E[X])$. By the portmanteau lemma [1, Theorem 2.1], the arbitrariness of f yields $S_n \to E[X]$ in distribution, which is equivalent to convergence in probability since the limit is a constant. $\square$

Based on this result, he proposed in [4, Section 5.2] and [5] a Monte Carlo method relying on the following formula to obtain an almost sure sample approximation to the expectation:

$$\mathbb{E}[\varphi(X)] = \limsup_n n^{-1} \sum_{i=1}^n \varphi(X_i), \tag{1}$$

for any $\varphi \in C_{l.Lip}(\mathbb{R})$ and Peng i.i.d. X_n. If one can generate a simulated sequence of X_n, that provides a way of computing the expectation of X (and, more generally, of nice functions of X) which would be quite more comfortable than the representation as a supremum of linear functionals.

We present now a sequence that satisfies Peng's law of large numbers and fails (1), showing that the latter does not follow from the former and thus cannot be used to compute the expectation.

For a given sequence $\{X_n\}_n$ of random variables, we write

$$S_n = n^{-1} \sum_{i=1}^{n} X_i.$$

Theorem 1. *There exist a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$ and a sequence $\{X_n\}_n \subset \mathcal{H}$ such that*

(a) *The X_n are Peng i.i.d. as some $X \in \mathcal{H}$.*
(b) *$\mathcal{E}[X] = 0, \mathbb{E}[X] = 1$.*
(c) *For every continuous $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(S_n)] \to \max_{x \in [\mathcal{E}[X], \mathbb{E}[X]]} \varphi(x).$$

(d) *$S_n \to \mathcal{E}[X]$ almost surely.*
(e) *$\nu(\mathcal{E}[X] \leq S_n \leq \mathbb{E}[X]) = 1$.*
(f) *$\nu(S_n < \mathbb{E}[X]) = 0$.*

Proof. Set $\Omega = \mathbb{N} \cup \{0\}$, take $\mathcal{H}$ to be the set of all bounded (necessarily Borel measurable) functions on Ω, and define $\mathbb{E}[X] = \sup_{\omega \in \Omega} X(\omega)$ for all $X \in \mathcal{H}$. It is readily checked that $\mathbb{E}$ is a sublinear expectation.

Let $X_n(\omega)$ be the nth bit in the binary representation of ω, i.e.

$$\omega = \sum_{n=1}^{\infty} 2^{n-1} X_n(\omega)$$

with $X_n(\omega) \in \{0, 1\}$.

Proof of part (a). In this space, Peng identical distribution of two random variables X, Y amounts to the identity

$$\sup_{\omega \in \Omega} \varphi(X(\omega)) = \sup_{\omega \in \Omega} \varphi(Y(\omega)) \quad \forall \varphi \in C_{l.Lip}(\mathbb{R}).$$

Therefore, any two variables with the same range $X(\Omega) = Y(\Omega)$ are Peng identically distributed. Since the range of each X_n is $\{0, 1\}$, they are all Peng identically distributed.

To establish Peng independence, we need to show

$$\sup_{\omega \in \Omega} \varphi(X_1(\omega), \dots, X_{n+1}(\omega)) = \sup_{\omega \in \Omega} [\sup_{\omega' \in \Omega} \varphi(x_1, \dots, x_n, X_{n+1}(\omega))]_{x_i = X_i(\omega')}$$

$$= \sup_{\omega, \omega' \in \Omega} \varphi(X_1(\omega), \dots, X_n(\omega), X_{n+1}(\omega')).$$

The '$\leq$' part is clear. Moreover, from the definition of the X_n, the set $\{(X_1(\omega), \dots, X_{n+1}(\omega))\}_{\omega \in \Omega}$ exhausts $\{0, 1\}^{n+1}$. Hence

$$\sup_{\omega \in \Omega} \varphi(X_1(\omega), \dots, X_{n+1}(\omega)) = \sup_{(x_1, \dots, x_{n+1}) \in \{0,1\}^{n+1}} \varphi(x_1, \dots, x_{n+1})$$

$$\geq \sup_{\omega,\omega' \in \Omega} \varphi(X_1(\omega), \ldots, X_n(\omega), X_{n+1}(\omega'))$$

as well.

Proof of part (b). Setting $X = X_1$, we have

$$\mathbb{E}[X] = \sup_{\omega \in \Omega} X(\omega) = 1$$

and

$$\mathcal{E}[X] = -\mathbb{E}[-X] = \inf_{\omega \in \Omega} X(\omega) = 0.$$

Proof of part (c). As ω ranges over Ω, the $(X_1(\omega), \ldots, X_n(\omega))$ exhaust $\{0,1\}^n$. Thus the range of S_n is $\{0, 1/n, \ldots, (n-1)/n, 1\}$. From the continuity of φ and part (b),

$$\mathbb{E}[\varphi(S_n)] = \sup_{\omega \in \Omega} \varphi(S_n(\omega)) = \sup_{0 \leq k \leq n} \varphi(k/n) \to \max_{x \in [0,1]} \varphi(x) = \max_{x \in [\mathcal{E}[X], \mathbb{E}[X]]} \varphi(x).$$

Proof of part (d). In view of part (b), we have to show $S_n(\omega) \to 0$ for each $\omega \in \Omega$. But that is clear, since the binary representation of each $\omega \in \Omega$ has finitely many non-zero terms.

Proof of part (e). Since X_n can only take on 0 and 1 as values, $0 \leq S_n \leq 1$ and then, from part (b),

$$\nu(\mathcal{E}[X] \leq S_n \leq \mathbb{E}[X]) = \nu(0 \leq S_n \leq 1) = \nu(\Omega) = \mathbb{E}[1] = 1.$$

Proof of part (f). Since the first n bits of the binary representation of $2^n - 1$ are 1,

$$\nu(S_n < 1) = \mathcal{E}[I_{\{S_n < 1\}}] = \inf_{\omega \in \Omega} I_{\{S_n < 1\}}(\omega) \leq I_{\{S_n < 1\}}(2^n - 1) = 0$$

because $S_n(2^n - 1) = n/n = 1$. □

Peng's formula (1) predicts

$$\limsup_n S_n = \sup_{x \in [\mathcal{E}[X], \mathbb{E}[X]]} x = \mathbb{E}[X] = 1$$

almost surely, but Theorem 1.(d) shows on the contrary that

$$S_n(\omega) \to \mathcal{E}[X] = 0 \quad \forall \omega \in \Omega.$$

Thus Peng's formula is wrong. It might have been the case that it actually followed from his law of large numbers and the latter were wrong, but Theorem 1.(c) implies that $\{X_n\}_n$ fulfils Peng's law of large numbers. Therefore the problem is that (1) does not follow from the law of large numbers.

Parts (e) and (f) suggest that the behaviour of S_n in this sublinear expectation space is quite different from the usual probability intuition. That is confirmed by the following proposition.

Proposition 2. *The sequence in Theorem 1 has the following properties:*

(a) $n^{-1} \sum_{i=1}^{n} a_i X_i \to 0$ *almost surely for any sequence* $a_n \geq 0$, *in particular* $S_n \to 0$ *almost surely.*
(b) S_n *does not converge in probability.*
(c) $S_n \to Y$ *in law, with the distribution of* Y *being given by* $\mathbb{E}_Y[\varphi]$ $= \max_{x \in [0,1]} \varphi(x)$.

Proof. Proof of part (a). For each $\omega \in \Omega$, it holds that $X_n(\omega) = 0$ for $n > \log_2 \omega + 1$. Thus, since X_n takes on values 0 and 1,

$$0 \leq n^{-1} \sum_{i=1}^{n} a_n X_n(\omega) \leq n^{-1} \sum_{i \leq \log_2 \omega + 1} a_i \to 0.$$

Proof of part (b). Fix $\varepsilon \in (0, 1/4)$. Reasoning by contradiction, assume $S_n \to Y$ in probability for some random variable Y. Accordingly, $\nu(|S_n - Y| < \varepsilon) \to 1$. For each event A one has

$$\nu(A) = \mathcal{E}[I_A] = \inf_{\omega \in \Omega} I_A(\omega);$$

the right-hand side is 0 unless A contains all $\omega \in \Omega$. Therefore, for all sufficiently large n,

$$\{|S_n - Y| < \varepsilon\} = \Omega$$

and then

$$\{|S_n - S_{2n}| < 2\varepsilon\} = \Omega.$$

Whatever n may be, since the first n bits of the binary representation of $2^n - 1$ are ones and the next n bits are zeros, we have

$$S_n(2^n - 1) = \frac{n}{n} = 1, \qquad S_{2n}(2^n - 1) = \frac{n}{2n} = 1/2.$$

But $\varepsilon < 1/4$, whence $|S_n(2^n - 1) - S_{2n}(2^n - 1)| > 2\varepsilon$, a contradiction.
Proof of part (c). It follows from Theorem 1.(c) or Peng's law. $\square$

Thus the behaviours of S_n with respect to the three types of convergence are all different, in particular

1) Almost sure convergence does not imply convergence in probability.
2) The limit of the sequence of laws exists but is not the law of the almost sure limit.
3) Almost sure convergence does not imply convergence in law.

It must be emphasized that a.s. convergence *is not* a weak notion in this context. In fact, it implies *uniform* a.s. convergence over a family of finitely additive probabilities (and, under appropriate conditions, see e.g. [3], σ-additive probabilities).

To make things even more interesting, for this sublinear expectation convergence in probability actually implies almost sure convergence; that is an instance of a more general phenomenon (see e.g. [8, Proposition 5.1]).

4 Sublinear expectations and coherent upper previsions

The purpose of this section is to observe that Peng's sublinear expectations are the same thing as the coherent upper previsions in e.g. Walley's book [10]. While Peng cites Walley, he seems to have been unaware of the equivalence.

A *coherent upper prevision* is a functional $\mathbb{E}$ on a linear space $\mathcal{H}$ of functions (called *gambles*) on a set Ω, such that, for any $n \geq 0$, $m \geq 1$, and $X_0, X_1, \ldots, X_n \in \mathcal{H}$,

$$\inf_{\omega \in \Omega} \Big[\sum_{k=1}^{n} (X_k(\omega) - \mathbb{E}[X_k]) - m(X_0(\omega) - \mathbb{E}[X_0]) \Big] \leq 0.$$

Proposition 3. *Let $\mathbb{E}$ be a functional on a linear space $\mathcal{H}$ of bounded functions (closed under taking the absolute value) on a set Ω. Then $\mathbb{E}$ is a sublinear expectation if and only if $\mathbb{E}$ is a coherent upper prevision.*

Proof. Walley [10, Section 2.6] showed that $\mathbb{E}$ is a coherent upper prevision if and only if the following three properties are met for any $X, Y \in \mathcal{H}$:

(CUP.i) $\mathbb{E}[X] \leq \sup_{\omega \in \Omega} X(\omega)$,
(CUP.ii) $\mathbb{E}[X + Y] \leq \mathbb{E}[X] + \mathbb{E}[Y]$,
(CUP.iii) $\mathbb{E}[cX] = c \cdot \mathbb{E}[X]$ for $c \geq 0$.

Properties (CUP.ii) and (CUP.iii) are the same as properties (SE.iii) and (SE.iv) of a sublinear expectation. Thus the proof splits into two parts.
1. A sublinear expectation satisfies (CUP.i).
Let $X \in \mathcal{H}$. Then, by (SE.i) and (SE.ii),

$$\mathbb{E}[X] \leq \mathbb{E}[\sup_{\omega \in \Omega} X(\omega)] = \sup_{\omega \in \Omega} X(\omega).$$

2. A coherent upper prevision satisfies (SE.i) and (SE.ii).
(SE.i): If $X \leq Y$, then, using (CUP.ii) and (CUP.i),

$$\mathbb{E}[X] = \mathbb{E}[Y + (X - Y)] \leq \mathbb{E}[Y] + \mathbb{E}[X - Y] \leq \mathbb{E}[Y] + \sup_{\omega \in \Omega}(X(\omega) - Y(\omega)) \leq \mathbb{E}[Y].$$

(SE.ii): From (CUP.iii) and (CUP.ii),

$$0 = \mathbb{E}[0] = \mathbb{E}[c - c] \leq \mathbb{E}[c] + \mathbb{E}[-c].$$

By (CUP.i), $\mathbb{E}[c] \leq c$ and $\mathbb{E}[-c] \leq -c$. Combining the three inequalities,

$$c \leq -\mathbb{E}[-c] \leq \mathbb{E}[c] \leq c,$$

which proves $\mathbb{E}[c] = c$. □

5 Concluding remarks

The sequence $\{X_n\}_n$ in Theorem 1 is interesting when contemplated from the probabilistic perspective. Each X_n should have, intuitively, a Bernoulli $\mathcal{B}(1/2)$ distribution. The product space $\{0,1\}^{\mathbb{N}}$ admits then a unique probability measure (the product measure $\mathbb{P}$) such that all those distributions are independent. But the subset of $\{0,1\}^{\mathbb{N}}$ which corresponds to Ω, namely the 0-1 sequences with finitely many ones, is $\mathbb{P}$-null, whence a probability for Ω cannot be retrieved from $\mathbb{P}$.

Proposition 5.3 in [8] presents an example of behaviours similar to Proposition 2, but the independence in the sense of Peng is not satisfied there (only a rather weaker form of independence holds), and a different definition of convergence in law is used as well. Theorem 1.(a) is similar to Proposition 6.1.(a) in [9], but both the notions of independence and identical distribution used are different.

References

1. Billingsley P (1968) Convergence of probability measures. Wiley, New York
2. Hu M, Li X (2014) Independence under the G-expectation framework. J Theoret Probab 27:1011–1020
3. Huber PJ, Strassen V (1973) Minimax tests and the Neyman–Pearson lemma for capacities. Ann Statist 1:251-263
4. Peng S (2009) Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations. Sci China A 52:1391-1411
5. Peng S (2010) Backward stochastic differential equation, nonlinear expectation and their applications. In: Bhatia R (ed) Proc Int Congress Math Hyderabad 2010, Vol I: Plenary lectures and ceremonies. Hindustan, New Delhi (Video available at `http://videos.icm2010.in/player.php?search=&video_id=plenary_lecture &date=&highres=1&play_from_clip=8&play_to_clip=8#player`)
6. Peng S (2010) Nonlinear expectations and stochastic calculus under uncertainty. Lecture notes available at arXiv:1002.4546
7. Proschan MA, Shaw PA (2016) Essentials of probability theory for statisticians. CRC, Boca Raton
8. Terán P (2014) Laws of large numbers without additivity. Trans Amer Math Soc 366:5431–5451
9. Terán P (2015) On independence and the law of large numbers for upper probabilities. Unpublished manuscript
10. Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman & Hall, London