The Annals of Probability
1983, Vol. 11, No. 4, 1051–1054

STRONG LAW OF LARGE NUMBERS WITH RESPECT TO A SET-VALUED PROBABILITY MEASURE[1]

By Madan L. Puri and Dan A. Ralescu

Indiana University and University of Cincinnati

In this paper we define the expected value of a random vector with respect to a set-valued probability measure. The concepts of independent and identically distributed random vectors are appropriately defined, and a strong law of large numbers is derived in this setting. Finally, an example of a set-valued probability useful in Bayesian inference is provided.

1. Introduction. This research is motivated by the following consideration: there are instances in Bayesian estimation when the prior probability is not known precisely. In such situations DeRobertis and Hartigan (1981) suggest using an interval of measures rather than a single prior, and extend the Bayes theorem in this setting. This idea is reminiscent of upper and lower probabilities (see Koopman, 1940, and Dempster, 1967). The risk $R(\theta, \delta)$ associated with a decision function δ is a random variable in the Bayes setting (since the unknown parameter θ is assumed to be a random variable). The main question then is: how one can evaluate the average risk when the prior measure is not known precisely. The concept which seems to be useful in such situations is that of a set-valued measure (see Debreu and Schmeidler, 1970, and Artstein, 1972) with respect to which the expectation of a random variable is evaluated.

In Section 2 we give some preliminaries on set-valued measures, and we define the expected value. In Section 3 we prove a strong law of large numbers with respect to a set-valued probability. In Section 4 we give an example of a set-valued probability measure.

2. Expectation with respect to a set-valued probability measure. The concept of a set-valued measure was defined in connection with the integral of a set-valued function (see Debreu and Schmeidler, 1970).

Let Ω be a set, $\mathscr{A}$ a σ-algebra of subsets of Ω, and $\mathscr{P}(\mathbb{R}^n)$ the collection of all subsets of $\mathbb{R}^n$. A set-valued measure is a function $\Pi: \mathscr{A} \to \mathscr{P}(\mathbb{R}^n)$, such that (i) $\Pi(A) \neq \phi$ for every $A \in \mathscr{A}$, (ii) $\Pi(\cup_{j=1}^\infty A_j) = \sum_{j=1}^\infty \Pi(A_j)$ for every disjoint family $\{A_j\}_j$, $A_j \in \mathscr{A}$.

Here the sum $\sum_{j=1}^\infty B_j$ of subsets of $\mathbb{R}^n$ is defined as the collection of all vectors $b = \sum_{j=1}^\infty b_j$ where $b_j \in B_j$ and $\sum_{j=1}^\infty \| b_j \| < \infty$.

In what follows we consider only bounded set-valued measures (such that $\Pi(\Omega)$ is bounded). It follows that for such measures, $\Pi(\phi) = \{0\}$.

A *selection* μ of Π is a vector-valued measure $\mu: \mathscr{A} \to \mathbb{R}^n$, such that $\mu(A) \in \Pi(A)$ for every $A \in \mathscr{A}$.

An atom of the set-valued measure Π is an event $A \in \mathscr{A}$ with $\Pi(A) \neq \{0\}$ and such that $A_1 \subset A$ implies $\Pi(A_1) = \{0\}$ or $\Pi(A \backslash A_1) = \{0\}$. A set-valued measure with no atoms is called nonatomic.

The following theorem due to Artstein (1972) will be used in the sequel:

Theorem 2.1. (a) *If* Π *is bounded, nonatomic set-valued measure, then* $\Pi(A)$ *is convex for every* $A \in \mathscr{A}$.

(b) *If* Π *is bounded set-valued measure, then, for every* $A \in \mathscr{A}$ *and* $x \in \Pi(A)$, *there exists a selection* μ *of* Π *such that* $\mu(A) = x$.

Received October 1982.

[1] Research supported by the National Science Foundation Grant IST-7918468.

AMS 1980 *subject classification*. Primary, 60B12; secondary, 60F15.

Key words and phrases. Set-valued measure, strong law of large numbers, interval of measures.

A *set-valued probability* on Ω is a set-valued measure $\Pi: \mathscr{A} \to \mathscr{P}([0, 1])$ such that $1 \in \Pi(\Omega)$.

A *set-valued probability space* is a triple $(\Omega, \mathscr{A}, \Pi)$ where Π is a set-valued probability.

Without loss of generality, one can assume that Π is absolutely continuous with respect to a probability measure P on Ω; $\Pi \ll P$, that is, for a set $A \in \mathscr{A}$ for which $P(A) = 0$, we have $\Pi(A) = \{0\}$. (see Artstein, 1972).

Let $X: \Omega \to \mathbb{R}^n$ be a random vector such that $E_P(\|X\|) = \int_\Omega \|X\| \, dP < \infty$. The *expected value* of X with respect to Π is defined as $\int_\Omega X \, d\Pi = \{\int_\Omega X \, d\mu : \mu \text{ is a selection of } \Pi\}$. According to Theorem 2.1 (b) it is clear that $\int_\Omega X d\Pi \neq \phi$ if $E_P(\|X\|) < \infty$.

3. Strong law of large numbers. Let $(\Omega, \mathscr{A}, \Pi)$ be a set-valued probability space, and let $X: \Omega \to \mathbb{R}^n$ be a random vector. Then X induces a set-valued probability on the Borel sets in $\mathbb{R}^n$ (denoted by $\mathscr{B}_n$) in the following way:

$$B \in \mathscr{B}_n, \quad \Pi_X(B) = \Pi(X \in B).$$

The random vectors X_i, $i \geq 1$ defined on $(\Omega, \mathscr{A}, \Pi)$ are *independent* if $\Pi(X_1 \in B_1, X_2 \in B_2, \cdots, X_i \in B_i) = \Pi(X_1 \in B_1) \cdots \Pi(X_i \in B_i)$ where the product of subsets M and N of $[0, 1]$ is defined by $MN = \{mn : m \in M, n \in N\}$. They are identically distributed if $\Pi_{X_1} = \cdots = \Pi_{X_i} = \cdots$. Clearly these concepts generalize the classical concepts of independent and identically distributed random vectors (with respect to an ordinary probability measure).

Finally, we need another notation: if $x \in \mathbb{R}^n$, and $A \subset \mathbb{R}^n$, then

$$d(x, A) = \inf_{a \in A} \|x - a\|.$$

We now prove our main theorem.

THEOREM 3.1. *Let X_i, $i \geq 1$ be independent and identically distributed random vectors defined on a set-valued probability space $(\Omega, \mathscr{A}, \Pi)$ such that $\Pi \ll P$ where P is a probability measure. If $E_P(\|X_1\|) < \infty$, then $d((1/n) \sum_{j=1}^n X_j, \int_\Omega X_1 \, d\Pi) \to 0$ almost everywhere with respect to Π.*

PROOF. Clearly, if μ is a selection of Π (which exists according to Theorem 2.1 (b)), it is not true in general that X_i, $i \geq 1$ are independent and identically distributed with respect to μ.

To prove the theorem, we will show the following:

There exists a probability measure Q on Ω which is a selection of Π and such that X_i, $i \geq 1$ are independent and identically distributed with respect to Q, and $E_Q(\|X_1\|) < \infty$.

Let $Q(A) = \sup \Pi(A)$ for every $A \in \mathscr{A}$. The fact that Q is a probability measure follows from Proposition 3.1 of Artstein (1972). Also it is clear that $Q(X_i \in B) = Q(X_1 \in B)$ for every $B \in \mathscr{B}$, and $i \geq 1$ i.e., X_i, $i \geq 1$ are identically distributed.

To prove that X_i, $i \geq 1$ are independent, it suffices to show that $Q(X_1 \in B_1, X_2 \in B_2) = Q(X_1 \in B_1) Q(X_2 \in B_2)$ for every $B_1, B_2 \in \mathscr{B}_n$. By the definition of independence, it suffices to show that $\sup(MN) = \sup M \sup N$ where $M, N \subset [0, 1]$. This being easy to establish, the desired independence follows.

Since $\Pi \ll P$, it follows from classical results that $E_P(\|X_1\|) < \infty$ implies $\int_\Omega \|X_1\| \, dQ < \infty$.

We now prove that Q is a selection of Π. From Theorem 2.1(b) there exists a probability measure Q_1 which is a selection of Π. Clearly $Q_1(A) \leq Q(A)$ for every $A \in \mathscr{A}$, but this implies that $Q_1 = Q$.

Now from the classical law of large numbers, it follows that $(1/n) \sum_{j=1}^n X_j \to E_Q(X_1)$ almost everywhere with respect to Q. Thus

$$d\left(\frac{1}{n} \sum_{j=1}^n X_j, \int_\Omega X_1 \, d\Pi\right) \leq \left\| \frac{1}{n} \sum_{j=1}^n X_j - E_Q(X_1) \right\| \to 0,$$

and so $d((1/n) \sum_{j=1}^n X_j, \int_\Omega X_1 \, d\Pi) \to 0$ almost everywhere (Q).

The definition of Q implies that the above convergence actually holds almost everywhere with respect to Π, that is, $\Pi(\sum_{j=1}^{n} X_j/n \not\to \int_\Omega X_1 \, d\Pi) = \{0\}$. This completes the proof.

4. An example. The strong law of large numbers proved in Section 3 is a generalization of the classical law of large numbers.

The simplest example of a set-valued probability measure is provided by an interval of measures (as studied by DeRobertis and Hartigan, 1981). More precisely let P_1 and P_2 be two finite measures on $(\Omega, \mathscr{A})$ such that $P_1(A) \le P_2(A)$ for every $A \in \mathscr{A}$, and let P_2 be a probability measure. Let $\Pi : \mathscr{A} \to \mathscr{P}([0, 1])$ be defined as

$$\Pi(A) = [P_1(A), P_2(A)], \quad A \in \mathscr{A}. \tag{4.1}$$

Clearly $\Pi(\phi) = \{0\}$ and $1 \in \Pi(\Omega)$.

Let $\{A_j\}_j$ be a disjoint family, $A_j \in \mathscr{A}$. We must show that $\Pi(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \Pi(A_j)$. This is equivalent to

$$\textstyle\sum_{j=1}^{\infty} [P_1(A_j), P_2(A_j)] = [\sum_{j=1}^{\infty} P_1(A_j), \sum_{j=1}^{\infty} P_2(A_j)].$$

The above equality follows from the formulas

$$\inf \textstyle\sum_{j=1}^{\infty} [P_1(A_j), P_2(A_j)] = \sum_{j=1}^{\infty} P_1(A_j),$$

$$\sup \textstyle\sum_{j=1}^{\infty} [P_1(A_j), P_2(A_j)] = \sum_{j=1}^{\infty} P_2(A_j),$$

and from the convexity of $\sum_{j=1}^{\infty} [P_1(A_j), P_2(A_j)]$.

Thus Π defined by (4.1) is a set-valued probability. Also Π is absolutely continuous with respect to P_2.

If $X : \Omega \to \mathbb{R}^n$ is a random vector such that $E_{P_2}(\|X\|) < \infty$, then the expected value of X (as defined in Section 2) is given by $\int_\Omega X \, d\Pi = \{E_P(X) : P_1 \le P \le P_2\}$ where P is a finite measure.

If X_i, $i \ge 1$ are independent and identically distributed with respect to Π (given by (4.1)) and note that the latter condition is equivalent to the fact that X_i, $i \ge 1$ are identically distributed with respect to P_1 and P_2, then the law of large numbers given by Theorem 3.1 implies

$$\inf_{P_1 \le P \le P_2} \left\| \frac{1}{n} \textstyle\sum_{j=1}^{n} X_j - E_P(X_1) \right\| \to 0 \quad \text{almost everywhere with respect to } P_2.$$

It is interesting to note that, under certain hypotheses, every set-valued probability is of the form (4.1).

THEOREM 4.1. *Let $\Pi : \mathscr{A} \to \mathscr{P}([0, 1])$ be a nonatomic set-valued probability measure such that $\Pi(\Omega)$ is closed. Then $\Pi(A) = [P_1(A), P_2(A)]$ for every $A \in \mathscr{A}$, where P_1 is a measure and P_2 is a probability measure such that $P_1(A) \le P_2(A)$, $A \in \mathscr{A}$.*

PROOF. Denote $P_1(A) = \inf \Pi(A)$ and $P_2(A) = \sup \Pi(A)$, $A \in \mathscr{A}$. We show that P_1 and P_2 are measures. Let $\{A_j\}_j$ be a disjoint family of sets in $\mathscr{A}$. Then, clearly

$$P_1(\cup_{j=1}^{\infty} A_j) = \inf(\textstyle\sum_{j=1}^{\infty} \Pi(A_j)) \ge \sum_{j=1}^{\infty} \inf \Pi(A_j) = \sum_{j=1}^{\infty} P_1(A_j). \tag{4.2}$$

Let $\varepsilon > 0$. Then there exists $x_j \in \Pi(A_j)$ such that $x_j < \inf \Pi(A_j) + \varepsilon/2^j$, $j \ge 1$. Thus $\sum_{j=1}^{\infty} x_j \le \sum_{j=1}^{\infty} \inf \Pi(A_j) + \varepsilon$. From (4.2), the series $\sum_{j=1}^{\infty} \inf \Pi(A_j)$ is convergent. So $\sum_{j=1}^{\infty} x_j \in \sum_{j=1}^{\infty} \Pi(A_j)$. Consequently

$$\inf(\textstyle\sum_{j=1}^{\infty} \Pi(A_j)) \le \sum_{j=1}^{\infty} \inf \Pi(A_j) + \varepsilon. \tag{4.3}$$

Since $\varepsilon > 0$ is arbitrary, it follows from (4.2) and (4.3) that

$$P_1(\cup_{j=1}^{\infty} A_j) = \textstyle\sum_{j=1}^{\infty} P_1(A_j).$$

Now $P_1(\Omega) = \inf \Pi(\Omega) \in \Pi(\Omega)$ since $\Pi(\Omega)$ is closed. From Theorem 2.1(b), there exists a selection Q_1 of Π such that $Q_1(\Omega) = P_1(\Omega)$. Since $P_1 \le Q_1$, we have $P_1 = Q_1$, so P_1 is a selection of Π. Similarly P_2 is a selection of Π, and obviously $P_2(\Omega) = 1$. Finally, since (from Theorem 2.1 (a)) $\Pi(A)$ is convex for every $A \in \mathscr{A}$, it follows that $\Pi(A) = [P_1(A), P_2(A)]$, $A \in \mathscr{A}$, which was to be proved.

REMARK. It may be noted that the intervals of probability measures (DeRobertis and Hartigan, op. cit.) are restandardized when computing expectations and posterior probabilities, so that the ranges of expectations for set-valued probabilities and for intervals of probabilities do not coincide.

REFERENCES

[1] ARTSTEIN, Z. (1972). Set-valued measures. *Trans. Amer. Math. Soc.* **165** 103–125.
[2] DEBREU, G. and SCHMEIDLER, D. (1970). The Radon-Nikodym derivative of a correspondence. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, 41-56. Univ. of California Press.
[3] DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **36** 325–339.
[4] DEROBERTIS, L. and HARTIGAN, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.
[5] KOOPMAN, B. O. (1940). The axioms and algebra of intuitive probability. *Ann. Math.* **41** 269–278.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45221