František Žák

Representation form of de Finetti theorem and application to convexity

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 52 (2011), No. 2, 61--66

Persistent URL: http://dml.cz/dmlcz/143679

Representation Form of De Finetti Theorem and Application to Convexity

FRANTIŠEK ŽÁK

Praha

Received August 24, 2010
Revised February 26, 2011

The classical de Finetti theorem states that every symmetric measure on $\mathbb{R}^{\mathbb{N}}$ is an integral convex combination of probability distributions of i.i.d sequences of random variables. In this article we reprove the theorem as a corollary to what one may call the Conditional de Finetti theorem: Given a symmetric sequence of random variables, there exists a σ-algebra, on which are random variables conditionally independent and conditionally identically distributed.

Once this result is established, we use general results on integral representation of invariant measures to get related convexity statements. Especially, we prove that the extremal points of convex set of symmetric measures are exactly the distributions of the sequences of i.i.d random variables.

1. Introduction

The integral representation form of the de Finetti theorem and related convexity results were first established in general by Hewitt and Savage [3] in 1955. First they characterize extremal points of symmetric measures as being the distributions of sequences of i.i.d random variables. After this the Krein - Milman theorem is applied to prove the desired integral representation. Technically, the above procedure may be characterized as a functional analytic one.

We suggest a more probabilistic proof of de Finetti theorem. We start with a conditional form of the theorem (see Theorem 1.1) and apply it to get de Finetti representation Theorem 2.1 directly. Finally, in Section 3, we are able to prove that the

---

Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic

2000 Mathematics Subject Classification. 60G09

Key words and phrases. De Finetti theorem, integral representation

E-mail address: kotrny@centrum.cz

set of probability distributions of i.i.d sequences is exactly the set of extremal measures in the set of symmetric measures on $\mathbb{R}^{\mathbb{N}}$. In addition the integral representation in Theorem 2.1 is proved to be determined uniquely. The proof is supported by the non-compact Choquet theory for convex sets of invariant measures.

Recall that symmetric measure is a probability measure on $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}})$ invariant under all finite permutations of coordinates. A symmetric sequence of random variables is a sequence whose distribution is a symmetric measure.

Theorem 1.1 (Conditional de Finetti theorem) *Let $X_1, X_2, \ldots$ be a symmetric sequence of random variables. Then there exists a σ-algebra $\mathscr{F}^\infty$, on which are $X_1, X_2, \ldots$ conditionally independent and conditionally identically disributed, i.e.*

$$P[X_j \in A_j, 1 \leq j \leq n | \mathscr{F}^\infty] = \prod_{j=1}^{n} P[X_j \in A_j | \mathscr{F}^\infty] \; a.s.,$$

for $A_j \in \mathscr{B}(\mathbb{R}), 1 \leq j \leq n, n \in \mathbb{N}$ and

$$P[X_n \in A | \mathscr{F}^\infty] = P[X_1 \in A | \mathscr{F}^\infty] \; a.\ s.\ , A \in \mathscr{B}(\mathbb{R}),\ n \in \mathbb{N}.$$

Proof. See [5]. □

Our goal is to use this theorem to produce a representation form of the de Finetti theorem. We do so in Section 2, in which we also remind some used notations and facts.

2. Representation form of de Finetti theorem

Through the paper we shall denote $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}})$ the space of all real sequences with corresponding Borel σ-algebra and consider $\mathbb{R}^{\mathbb{N}}$ in the standart product topology. The Borel σ-algebra on real line will be identified as $\mathscr{B}$. If X_n converges in distribution to X, we write $X_n \xrightarrow{\mathscr{D}} X$. Given a metric space X, $\mathscr{P}(X)$ will mean the space of all probability measures on X. On such space we work with the usual weak topology and weak convergence of measures. The Borel σ-algebra generated by weak topology will simply be denoted as $\mathscr{B}(\mathscr{P}(X))$.

For the sake of simplicity we denote $\mathscr{I}$ the set of all probability measures on $\mathscr{B}^{\mathbb{N}}$ of the form $P^{\mathbb{N}}$, for some probability measure P on $(\mathbb{R}, \mathscr{B})$.

Theorem 2.1 (de Finetti theorem) *Let μ be a symmetric measure on $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}})$. Then there exists a Borel probability measure R on $\mathscr{B}(\mathscr{P}(\mathbb{R}^{\mathbb{N}}))$ that for every $A \in \mathscr{B}^{\mathbb{N}}$ holds*

$$\mu(A) = \int_{\mathscr{I}} r(A)\, R(dr).$$

Our proof is in a need of some prerequisities.

Lemma 2.2 *a)* $\mathscr{I} = \left\{Q \in \mathscr{P}(\mathbb{R}^{\mathbb{N}}) : Q = F^{\mathbb{N}}, F \in \mathscr{P}(\mathbb{R})\right\}$ *is a weakly closed, hence Borel set in* $\mathscr{P}(\mathbb{R}^{\mathbb{N}})$.

b) Mapping $g : \mathscr{P}(\mathbb{R}) \to \mathscr{P}(\mathbb{R}^{\mathbb{N}})$ *defined by* $g(Q) = Q^{\mathbb{N}}$ *is continuous with respect to weak convergence, hence Borel measurable as*

$$g : (\mathscr{P}(\mathbb{R}), \mathscr{B}(\mathscr{P}(\mathbb{R}))) \to (\mathscr{P}(\mathbb{R}^{\mathbb{N}}), \mathscr{B}(\mathscr{P}(\mathbb{R}^{\mathbb{N}}))).$$

Proof. a) Because $\mathbb{R}^{\mathbb{N}}$ is a separable metric space, the weak topology is metrizable (see [1] for example). So let a sequence of i.i.d random variables $X_n = (X_n^1, X_n,^2, \dots)$ converge to a sequence $X = (X^1, X^2, \dots)$ in distribution. Because

$$X_n \overset{\mathscr{D}}{\to} X \text{ on } \mathbb{R}^{\mathbb{N}} \iff (X_n^1, \dots, X_n^k) \overset{\mathscr{D}}{\to} (X^1, \dots, X^k) \text{ on } \mathbb{R}^k \ \forall k \in \mathbb{N} \tag{2.1}$$

(see Theorem 4.29 in [4], for example) we have to prove only that $X_k = (X^1, \dots, X^k)$ is a vector of i.i.d random variables for arbitrary $k \in \mathbb{N}$. As X_n^k converges in distribution to X^k as $n \to \infty$ for all $k \in \mathbb{N}$, we verify that random variables X^k are identically distributed. To prove that they are independent, simply compute the characteristic function of a vector $(X_n^1, \dots, X_n^k)$ and let $n \to \infty$ to see that $X^1, \dots, X^k$ are independent random variables.

b) Consider measures Q_n and Q in $\mathscr{P}(\mathbb{R})$ such that $Q_n \to Q$ weakly. It follows by (2.1) applying the characteristic functions calculus again that $Q_n^{\mathbb{N}} \to Q^{\mathbb{N}}$ weakly in $\mathscr{P}(\mathbb{R}^{\mathbb{N}})$, which proves the statement (b). □

Lemma 2.3 *Let* X *be a separable metric space. The mapping* $h : \mathscr{P}(X) \to \mathbb{R}$, $h(\mu) = \mu(B)$ *is Borel measurable for every* $B \in \mathscr{B}(X)$.

Proof. We define a collection of sets

$$\mathscr{D} = \{B \in \mathscr{B}(X) \ : \ h(\mu) = \mu(B) \text{ is a measurable mapping}\}.$$

If $\mu_n \in \mathscr{B}(X)$ converges weakly to $\mu \in \mathscr{B}(X)$, Portmanteau theorem gets $\limsup \mu_n(F) \leq \mu(F)$ for every closed set F. h is therefore an upper semi-continuous function for closed sets and thus Borel measurable. Routine arguments reveal that $\mathscr{D}$ is a Dynkin system and because closed sets generate $\mathscr{B}(\mathbb{P}(X))$, the lemma is proved. □

Lemma 2.4 *Let* $(\Omega, \mathscr{A}, P)$ *be a probability space and* $\omega \to \nu_\omega$ *a map* $\Omega \to \mathscr{P}(\mathbb{R})$ *such that* $\omega \to \nu_\omega(B)$ *is a* $(\mathscr{A}, \mathscr{B}(\mathbb{R}))$ *– measurable map for all* $B \in \mathscr{B}(\mathbb{R})$. *Then* $\omega \to \nu_\omega$ *is a map measurable w.r.t the* $\mathscr{A}$ *and Borel* σ*-algebra of* $\mathscr{P}(\mathbb{R})$.

The lemma states a known fact, we add a brief proof for the sake of completeness.

Proof. By assumption it follows that a mapping $\omega \to \int f \, d\nu_\omega$ is measurable for indicator functions. Uniform approximation by simple functions ensures measurability for all continuous bounded functions. $\mathscr{P}(\mathbb{R})$ is a separable space (see [1]). Hence, we have to check the measurability only on a base of weak topology on $\mathscr{P}(\mathbb{R})$. The base is formed by all finite intersections of the sets $\left\{\nu : a < \int f \, d\nu < b\right\}$ with $b > a > 0$ and f is a bounded continous function, which closes the proof together with the above observation. □

Proof of Theorem 2.1. Let μ be a symmetric measure and $X = (X_1, X_2, \ldots)$ a sequence whose distribution is μ. Further let D be a measurable rectangle in $\mathscr{B}^{\mathbb{N}}$, i.e. $D = \prod_{k=1}^{n} A_i \times \mathbb{R}^{\mathbb{N}}$, where $A_i \in \mathscr{B}$. Theorem 1.1 implies that

$$\mu(D) = P\left[X_1 \in A_1, \ldots, X_n \in A_n\right] = \int_\Omega \prod_{k=1}^{n} P\left[X_1 \in A_k | \mathscr{F}^\infty\right] dP.$$

According to Theorem 6.3 in [4] there is a regular version of the conditional probability distibution $\nu(.) = P[X_1 \in .|\mathscr{F}^\infty]$ that is a probability measure on $(\mathbb{R}, \mathscr{B})$ for each $\omega \in \Omega$, the map $\omega \to \nu_\omega(A)$ being a measurable map for all $A \in \mathscr{B}$. Hence,

$$\int_\Omega \prod_{k=1}^{n} P\left[X_1 \in A_k | \mathscr{F}^\infty\right](\omega)\, dP(\omega) = \int_\Omega \prod_{k=1}^{n} \nu_\omega(A_k)\, P(d\omega),$$

holds almost surely. Now, it follows by Lemma 2.3 and Lemma 2.4, denoting by p the image of P w.r.t the measurable mapping $\omega \to \nu_\omega$, that

$$\int_\Omega \prod_{k=1}^{n} \nu_\omega(A_k)\, P(d\omega) = \int_{\mathscr{P}(\mathbb{R})} \prod_{k=1}^{n} \nu(A_k)\, p(d\nu) = \int_{\mathscr{P}(\mathbb{R})} \nu^{\mathbb{N}}(D)\, p(d\nu). \tag{2.2}$$

As follows from Lemma 2.2 the map $\nu \to \nu^{\mathbb{N}}$ is measurable, let R be an image of p w.r.t this mapping, i. e. a Borel probability measure on $\mathscr{I}$. By using Lemma 2.3, Lemma 2.2 and (2.2) we conclude

$$\mu(D) = \int_{\mathscr{P}(\mathbb{R})} \nu^{\mathbb{N}}(D)\, p(d\nu) = \int_{\mathscr{I}} r(D)\, R(dr).$$

All we have to do now is to define collection of sets

$$\mathscr{K} = \left\{ B \in \mathscr{B}^{\mathbb{N}} \ : \ \mu(B) = \int_{\mathscr{I}} r(B)\, R(dr) \right\}.$$

The above proof ensures that $\mathscr{K}$ includes measurable rectangles. They generate $\mathscr{B}^{\mathbb{N}}$ and are closed under the finite intersection. It is easy to verify that $\mathscr{K}$ is a Dynkin system, so the theorem is established. □

3. Extremal points of symmetric measures

Consider a vector space of all signed finite measures on $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}})$ and denote by $\mathscr{M}$ the set of all symmetric measures. Obviously $\mathscr{M}$ is a convex set in this space. We plan to apply Theorem 2.1 to prove

Theorem 3.1 *The extremal points of the set of symmetric measures are exactly the distributions of sequences of iid. random variables. In symbols, ext* $\mathscr{M} = \mathscr{I}$.

Given a measurable space $(T, \mathscr{T})$ and a collection of measurable mappings $\mathscr{R} = \{r : r : (T, \mathscr{T}) \to (T, \mathscr{T})\}$ we can define the concept of ergodic measure. In fact there are two possible definitions of ergodicity and they are not equivalent in general (see [6], page 84). An invariant measure μ on $\mathscr{T}$ is called strongly ergodic if the following implication holds:

$$A \in \mathscr{T},\ \mu(r^{-1}A\ \Delta\ A) = 0\ \forall r \in \mathscr{R}\ \Rightarrow\ \mu(A) = 1 \text{ or } \mu(A) = 1.$$

Measure μ is ergodic with respect to $\mathscr{R}$, if it is $\mathscr{R}$-invariant and trivial on all invariant events $A \in \mathscr{T}$. However, additional requirements imposed on $\mathscr{R}$ may force the definitions to coincide.

Lemma 3.2 *Let $(T, \mathscr{T})$ be a measurable space and $\mathscr{R}$ a collection of measurable mappings on this space. If $\mathscr{R}$ forms a countable group with respect to the composition of mappings then a measure μ on $\mathscr{T}$ is ergodic, if and only if it is strongly ergodic.*

Proof. Obviously strong ergodicity implies ergodicity. To prove the converse, let $A \in \mathscr{T}$ be such an event that $\mu(A\ \Delta\ r^{-1}A) = 0$ for all $r \in \mathscr{R}$. We want to show $\mu(A) = 0$ or $\mu(A) = 1$. By assumption, it suffices to find an invariant event $B \in \mathscr{T}$ such as $\mu(A\ \Delta\ B) = 0$. We put $B = \cup_{k=1}^{\infty} r_k^{-1} A$. A simple calculation reveals that $\mu(A\ \Delta\ B) = 0$. $\mathscr{R}$ is a countable group, therefore we also get that

$$r_n^{-1} B = r_n^{-1} \cup_{k=1}^{\infty} r_k^{-1} A = \cup_{k=1}^{\infty} (r_k \circ r_n)^{-1} A = B$$

holds for arbitrary $n \in \mathbb{N}$. This means that B is invariant and the proof is completed. □

To prove theorem 3.1 we apply two general results concerned with extremal points of invariant measures and integral representation of invariant measures. Note that every collection of invariant probability measures forms a convex set in the space of finite signed measures.

Theorem 3.3 *Let X be a collection of $\mathscr{R}$-invariant probability measures on $(T, \mathscr{T})$. Then*

$$\mu \in ext\ X \Longleftrightarrow \mu \text{ is strongly ergodic w. r. t. } \mathscr{R}.$$

Proof. See [6], theorem 10.4. □

The first general theorems dealing with integral representation of invariant measures are probably due to Farell in [2]. Here we use for our purposes the version of Choquet theorem for invariant measures proved in 1984 by Štěpán in [7].

Theorem 3.4 *Let X be a complete and separable metric space, $\mathscr{R}$ a set of continous mappings $r : X \to X$. Then for all μ $\mathscr{R}$-invariant measures there is a unique probability measure V on $\mathscr{B}(\mathscr{P}(X))$ such that*

$$\mu(A) = \int_{\mathscr{E}} \nu(A)\ V(d\nu),\ A \in \mathscr{B}(X),$$

where $\mathscr{E}$ denotes the set of all strongly ergodic measures with respect to $\mathscr{R}$.

Proof. Apply Theorem 10 in [7] and recall that every finite measure on Polish space is a Radon measure. □

Proof of Theorem 3.1. As finite premutations on $\mathbb{R}^{\mathbb{N}}$ obviously form a countable group, it follows by Theorem 3.3 and Lemma 3.2 that extremal points of the set of symmetric measures are ergodic measures w. r. t. finite permutations. Classical Hewitt-Savage 0-1 law states that the distribution of the sequence of i.i.d random variables is an ergodic measure, i. e. $\mu \in \mathscr{I} \Rightarrow \mu$ is ergodic. Every ergodic measure ν has trivial integral representation by Dirac measure concetrated on ν. De Finetti theorem 2.1 implies though that every ergodic measure may be expressed as an integral average of the distributions from $\mathscr{I}$. It is easy to see that a finite permutation is a continous mapping, hence from the uniqueness contained in theorem 3.4 we get the desired equality $ext\, \mathscr{M} = \mathscr{I}$. □

Acknowledgement. The author would like to express his immense gratitude to Professor Josef Štěpán for his help and many important suggestions regarding the article.

References

[1] Billingsley, P.: Convergence of probability measures, Wiley-Interscience Publication, second edition (1999).

[2] Farell, R. H.: *Respresentation of invariant measures*, Illinois J. Math., **6** (1962), 447–467.

[3] Hewitt, E., Savage, L. J.: *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc., **80** (1955), 470–501.

[4] Kallenberg, O.: Foundations of Modern Probability, Springer-Verlag, New York – Berlin – Heidelberg second edition (2002).

[5] Kingman, J. F. C.: *Uses of exchangeability*, Annals of probability, **6** (1978), 183–197.

[6] Phelps, R. R.: Lectures on Choquet's theorem, Van Nostrand, London (1966).

[7] Štěpán, J.: *A noncompact Choquet theorem*, Commentationes Mathematicae Universitatis Carolinae, **25** (1984), 73–89.