

An elementary proof of de Finetti’s Theorem

Werner Kirsch
Fakultät für Mathematik und Informatik
FernUniversität in Hagen, Germany

September 5, 2018

Abstract

A sequence of random variables is called exchangeable if the joint distribution of the sequence is unchanged by any permutation of the indices. De Finetti’s theorem characterizes all \(\{0, 1\}\)-valued exchangeable sequences as a ‘mixture’ of sequences of independent random variables.

We present an new, elementary proof of de Finetti’s Theorem. The purpose of this paper is to make this theorem accessible to a broader community through an essentially self-contained proof.

1 Introduction

Definition 1 A finite sequence of (real valued) random variables \(X_1, X_2, \ldots, X_N\) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) is called exchangeable, if for any permutation \(\pi\) of \(\{1, 2, \ldots, N\}\) the distributions of \(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(N)}\) and \(X_1, X_2, \ldots, X_N\) agree, i.e. if for any Borel sets \(A_1, A_2, \ldots, A_N\)

\[
\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \ldots, X_N \in A_N) = \mathbb{P}(X_{\pi(1)} \in A_1, X_{\pi(2)} \in A_2, \ldots, X_{\pi(N)} \in A_N)
\]

An infinite sequence \(\{X_i\}_{i \in \mathbb{N}}\) is called exchangeable, if the finite sequences \(X_1, X_2, \ldots, X_N\) are exchangeable for any \(N \in \mathbb{N}\).

Obviously, independent, identically distributed random variables are exchangeable, but there are many more examples of exchangeable sequences.

Let us denote by \(\pi_p\) the (Bernoulli) probability measure on \(\{0, 1\}\) given by \(\pi_p(1) = p\) and \(\pi_p(0) = 1 - p\). If the random variables \(X_i\) are independent and distributed according to \(\pi_p\), i.e. \(\mathbb{P}(X_i = 1) = \pi_p(1) = p\) and \(\mathbb{P}(X_i = 0) = 1 - p\).
\( \pi_p(0) = 1 - p \), then the probability distribution of the sequence \( X_1, \ldots, X_N \) is the product measure

\[
\mathcal{P}_p = \bigotimes_{i=1}^{N} \pi_p \quad \text{on} \quad \{0, 1\}^N
\]  

(2)

In 1931 B. de Finetti proved the following remarkable theorem which now bears his name:

**Theorem 2 (de Finetti’s Representation Theorem)** Let \( X_i \) be an infinite sequence of \( \{0, 1\} \)-valued exchangeable random variables then there exists a probability measure \( \mu \) on \( [0, 1] \) such that for any \( N \) and any sequence \( (x_1, \ldots, x_N) \in \{0, 1\}^N \)

\[
\mathbb{P}(X_1 = x_1, \ldots, X_N = x_N) = \int \mathcal{P}_p(x_1, \ldots, x_N) d\mu(p) = \int \prod_{i=1}^{N} \pi_p(x_i) d\mu(p)
\]  

(3)  

(4)

Loosely speaking: An exchangeable sequence with values in \( \{0, 1\} \) is a ‘mixture’ of independent sequences with respect to a measure \( \mu \) on \( [0, 1] \).

De Finetti’s Theorem was extended in various directions, most notably to random variables with values in rather general spaces [4]. For reviews on the theorem see e. g. [1], see also the textbook [6] for a proof.

The proof of Theorem 2 we present here is very elementary. It is based on the method of moments which allows us to prove weak convergence of measures.

**Acknowledgement** It is a pleasure to thank Michael Fleermann for careful proof-reading and many helpful suggestions.

2 **Preliminaries**

For a probability measure \( \mu \) on \( \mathbb{R} \) we define the \( k^{th} \) moments by \( m_k(\mu) := \int x^k \, d\mu(x) \) whenever the latter integral exists (in the sense that \( \int |x|^k \, d\mu(x) < \infty \)). In the following we will be dealing with measures with compact support so that all moments exist (and are finite). The following theorem is a light version of the method of moments which is nevertheless sufficient for our purpose.

**Proposition 3**

1. Let \( \mu_n \ (n \in \mathbb{N}) \) be probability measures with support contained in a (fixed) interval \([a, b]\). If for all \( k \) the moments \( m_k(\mu_n) \) converge to some \( m_k \) then the sequence \( \mu_n \) converges weakly to a measure \( \mu \) with moments \( m_k(\mu) = m_k \) and with support contained in \([a, b]\).
2. If \( \mu \) is a probability measure with support contained in \([a, b]\) and \( \nu \) is a probability measures on \( \mathbb{R} \) such that \( m_k(\mu) = m_k(\nu) \) then \( \mu = \nu \).

**Remark 4** Let \( \mu_n \) and \( \mu \) be probability measures on \( \mathbb{R} \). Recall that weak convergence of the measures \( \mu_n \) to \( \mu \) means that

\[
\int f(x) \, d\mu_n(x) \to \int f(x) \, d\mu(x)
\]

for all bounded, continuous functions \( f \) on \( \mathbb{R} \).

The above theorem is true and, in fact, well known if the support condition is replaced by the much weaker assumption that the moments \( m_k(\mu) \) (resp. the numbers \( m_k \)) do not grow too fast as \( k \to \infty \) (see [6] or [5] for details).

**Proof.** We sketch the proof, for details see the literature cited above. By Weierstrass approximation theorem the polynomials on \( I = [a - 1, b + 1] \) are uniformly dense in the space of continuous functions on \( I \). Hence the integral \( \int f(x) \, d\mu(x) \) for continuous \( f \) can be computed from the knowledge of the moments of \( \mu \). From this part 2 of the theorem follows.

Moreover, we get that the integrals \( \int f(x) \, d\mu_n(x) \) converge for any continuous \( f \). The limit is a positive linear functional. Thus the probability measures \( \mu_n \) converge weakly to a measure \( \mu \) with

\[
\int f(x) \, d\mu_n(x) \to \int f(x) \, d\mu(x)
\]

which implies part 1. \( \blacksquare \)

## 3 Proof of de Finetti’s Theorem

The following theorem is a substitute for a (very weak) law of large numbers.

**Theorem 5** Let \( X_i \) be an infinite sequence of \( \{0, 1\} \)-valued exchangeable random variables then \( S_N := \frac{1}{N} \sum_{i=1}^{N} X_i \) converges in distribution to a probability measure \( \mu \). 

\( \mu \) is concentrated on \([0, 1]\) and its moments are given by

\[
m_k(\mu) = \mathbb{E}\left(X_1 \cdot X_2 \cdot \ldots \cdot X_k\right)
\]

where \( \mathbb{E} \) denotes expectation with respect to \( \mathbb{P} \).

**Definition 6** We call the measure \( \mu \) associated with \( X_i \) according to Theorem 5 the de Finetti measure of \( X_i \).
Proof. (Theorem 5)
To express the moments of \( S_N \) we compute
\[
\left( \sum_{i=1}^{N} X_i \right)^k = \sum_{(i_1, \ldots, i_k) \in \{1, \ldots, N\}^k} X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k} \quad (7)
\]
To simplify the evaluation of the above sum we introduce the number of different indices in \((i_1, \ldots, i_k)\) as
\[
\rho(i_1, i_2, \ldots, i_k) = \#\{i_1, i_2, \ldots, i_k\} \quad (8)
\]
Consequently
\[
(7) = \sum_{r=1}^{k} \sum_{(i_1, \ldots, i_k) \in \{1, \ldots, N\}^k \atop \rho(i_1, \ldots, i_k) = r} X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k} \quad (9)
\]
Thus we may write
\[
\mathbb{E}\left( \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right)^k \right) = \frac{1}{N^k} \sum_{i_1, \ldots, i_k = 1}^{N} \mathbb{E}\left( X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k} \right) \nonumber
\]
\[
+ \frac{1}{N^k} \sum_{i_1, \ldots, i_k = 1}^{N} \mathbb{E}\left( X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k} \right) \quad (10)
\]
There are at most \((k - 1)^k N^{k-1}\) index tuples \((i_1, \ldots, i_k)\) with \(\rho(i_1, \ldots, i_k) < k\). Indeed, we have \(N^{k-1}\) possibilities to chose the possible indices (‘candidates’) for \((i_1, \ldots, i_k)\). Then for each of the \(k\) positions in the \(k\)-tuple we may chose one of the \(k - 1\) candidates which gives \((k - 1)^k\) possibilities. This covers also tuples with less than \(k - 1\) different indices as some of the candidates may finally not appear in the tuple. It follows that the second term in (10) goes to zero. So
\[
\mathbb{E}\left( \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right)^k \right) \approx \frac{1}{N^k} \sum_{i_1, \ldots, i_k = 1}^{N} \mathbb{E}\left( X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k} \right),
\]
so using exchangeability:
\[
= \frac{1}{N^k} \sum_{i_1, \ldots, i_k = 1}^{N} \mathbb{E}\left( X_1 \cdot X_2 \cdot \ldots \cdot X_k \right) \nonumber
\]
\[
\approx \mathbb{E}\left( X_1 \cdot X_2 \cdot \ldots \cdot X_k \right) \quad (11)
\]
An application of Proposition 3 gives the desired result.

We note a Corollary to the Theorem 5 or better to its proof.

**Corollary 7** If \( \{X_i\} \) is an exchangeable sequence of \( \{0, 1\}\)-valued random variables and

\[
r = \rho(i_1, i_2, \ldots, i_k) = \#\{i_1, i_2, \ldots, i_k\}
\]

then

\[
\mathbb{E}\left(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_k}\right) = \mathbb{E}\left(X_1 \cdot X_2 \cdot \ldots \cdot X_r\right)
\]  

(12)

**Proof.** Since \( X_i \in \{0, 1\} \) we have \( X_i^\ell = X_i \) for all \( \ell \in \mathbb{N}, \ell \geq 1 \). Hence, the product in the left hand side is actually a product of \( r \) different \( X_j \), the expectation of which equals the right hand side due to exchangeability.

For the proof of Theorem 2 we will use the following simple lemma.

**Lemma 8** Suppose \( \{X_i\}_{i \in \mathbb{N}} \) is a \( \{0, 1\}\)-valued exchangeable sequence. Then for pairwise distinct \( i_1, i_2, \ldots, i_k \in \mathbb{N} \) and \( x_1, \ldots, x_k \in \{0, 1\} \) with \( \sum x_i = m \)

\[
P\left(X_{i_1} = x_1, \ldots, X_{i_k} = x_k\right) = \frac{1}{\binom{k}{m}} P\left(\sum_{i=1}^{k} X_i = m\right)
\]

(13)

**Proof.** There are \( \binom{k}{m} \) tuples \( x_1, \ldots, x_k \) with \( \sum x_i = m \). Due to exchangeability they all lead to the same probability.

We now prove Theorem 2.

**Proof.** (Theorem 2)

Let \( \mu \) be the de Finetti measure of \( X_i \) (see Definition 6) and define a \( \{0, 1\}\)-valued process \( \{Y_i\}_i \) by

\[
P\left(Y_1 = y_1, \ldots, Y_k = y_k\right) = \int \prod_{i=1}^{k} \pi_p(y_i) \, d\mu(p)
\]

(14)

The process \( Y_i \) is obviously exchangeable.

We’ll prove that \( X_i \) and \( Y_i \) have the same finite dimensional distributions.

According to Lemma 8 it suffices to show that \( S_N = \sum_{i=1}^{N} X_i \) and \( T_N = \sum_{i=1}^{N} Y_i \) have the same distributions for all \( N \) and for this it is enough by Propo-
position 3 to prove that their moments agree.

\[
\mathbb{E}\left(S_N^k\right) = \sum_{r=1}^{k} \sum_{i_1, \ldots, i_k=1}^{N} \mathbb{E}\left(X_{i_1} \cdot \ldots \cdot X_{i_k}\right)
= \sum_{r=1}^{k} \sum_{i_1, \ldots, i_k=1}^{N} \mathbb{E}\left(X_1 \cdot \ldots \cdot X_r\right) \quad \text{(by Corollary 7)}
= \sum_{r=1}^{k} \sum_{i_1, \ldots, i_k=1}^{N} \int \prod_{i=1}^{r} \pi_p(x_i) \, d\mu(p) \quad \text{(by Theorem 5)}
= \sum_{r=1}^{k} \sum_{i_1, \ldots, i_k=1}^{N} \mathbb{E}\left(Y_1 \cdot \ldots \cdot Y_r\right) \quad \text{(by (14))}
= \sum_{r=1}^{k} \sum_{i_1, \ldots, i_k=1}^{N} \mathbb{E}\left(Y_{i_1} \cdot \ldots \cdot Y_{i_k}\right) \quad \text{(Corollary 7)}
= \mathbb{E}\left(T_N^k\right) \quad \text{(15)}
\]

References

[1] D. Aldous: Exchangeability and related topics, pp. 1-198 in: Lecture Notes in Mathematics 117, Springer (1985).

[2] B. de Finetti: Funzione caratteristica di un fenomeno aleatorio, Atti della R. Accademia Nazionale dei Lincei, Ser. 6, Memorie, Classe di Scienze Fisiche, Matematiche e Naturali 4, 251–299 (1931).

[3] B. de Finetti: La prevision: ses lois logiques, ses sources subjectives, Annales de l’Institut Henri Poincare, 7, 1–68 (1937).

[4] Hewitt, E. and Savage, L. J.: Symmetric measures on Cartesian products. Transactions of the American Mathematical Society, 80, 470–501 (1955).

[5] W. Kirsch: Moments in Probability, book in preparation, to appear at DeGruyter.
[6] A. Klenke: Probability Theory, Springer (2013)

Werner Kirsch  werner.kirsch@fernuni-hagen.de