A CONSTRUCTIVE PROBABILISTIC PROOF OF CHOQUET’S THEOREM

Wiebe R. Pestman

Report No. 9913 (March 1999)
A constructive probabilistic proof of
Choquet’s theorem

Wiebe R. Pestman

Let $E$ be an arbitrary locally convex space and let $K$ be a convex compact subset of $E$. Then for every Borel probability measure $\mu$ on $K$ there is a unique point $a$ in $K$ such that for all continuous linear forms $\ell$ on $E$ one has

$$\ell(a) = \int \ell(x) \, d\mu(x).$$

Such a point $a$ is called the resultant or barycenter of $\mu$; we shall denote it by $r(\mu)$. An element $a$ in $K$ is said to be an extreme point of $K$ if it is not possible to split it up as

$$a = \frac{1}{2}(x + y)$$

where $x$ and $y$ are elements in $K$, both different from $a$. We shall denote the set of all extreme points of $K$ by $\text{Ext}(K)$. A well-known theorem in functional analysis, the so-called Krein-Milman theorem (see [4], [5]), states that every point $a$ in $K$ is the resultant of some Borel probability measure on $K$ which is concentrated on the closure of $\text{Ext}(K)$. However, it is not at all a rare phenomenon (see [8]) that the closure of $\text{Ext}(K)$ is equal to $K$. In such cases the Krein-Milman theorem is, of course, trivial for one could choose $\mu$ to be the Dirac measure in the point $a$. One may therefore wonder whether it is possible to choose $\mu$ such as to be concentrated on $\text{Ext}(K)$ itself rather than on its closure. It turns out to be impossible in this general setting, as is testified by examples (see [3]). This negative phenomenon is closely related to the fact that the set $\text{Ext}(K)$ can be of an extremely wild (non-measurable) nature. However, if $K$ is metrisable, then this ‘catastrophe’ cannot occur. This can be understood by means of the following simple argument: Let $D$ be the diagonal in $K \times K$. Then the set $K \times K \setminus D$ is, if $K$ is metrisable, the union of a countable collection of compact sets. The set $K \setminus \text{Ext}(K)$ is the continuous image of $K \times K \setminus D$ under the map $(x, y) \mapsto \frac{1}{2}(x + y)$. It follows that $K \setminus \text{Ext}(K)$ is a countable union of compact sets in $K$ and therefore it is Borel. We conclude that $\text{Ext}(K)$ is Borel (even a $G_\delta$) if $K$ is metrisable. The famous Choquet theorem (see [4], [6]), a pearl in functional analysis, states that for such $K$ it is indeed possible to choose $\mu$ in such a way that it is concentrated on $\text{Ext}(K)$.

If $E$ is finite-dimensional, say of dimension $n$, then $\mu$ can be chosen such as to be concentrated on $n+1$ extreme points of $K$. In literature this special case of Choquet’s theorem is known as Carathéodory’s theorem; it can be proved by a simple induction argument on $n$ (see [7]). In this paper we deduce, in a probabilistic way, Choquet’s
result from Carathéodory’s. Contrary to [5], where also probabilistic methods are
exploited, we don’t use transfinite induction and there is no reliance on scalar or
vectorial martingale convergence. The proof given below is highly constructive.

To prepare the reader notationally, let the triple \((\Omega, \mathcal{F}, \mathbb{P})\) be some probability
space and let \(X : \Omega \rightarrow K\) be such that

\[ X^{-1}(A) \in \mathcal{F} \quad \text{for every Borel subset} \quad A \quad \text{in} \quad K. \]

We then say that \(X\) is a **stochastic variable** with values in \(K\). The image measure of
\(\mathbb{P}\) under \(X\) is, in this context, usually called the **probability distribution** of \(X\). It will
be denoted by \(\mathbb{P}_X\); its resultant will be called the **expectation value** of \(X\) and denoted
by \(E(X)\) rather than \(r(\mathbb{P}_X)\). In these notations we now prove:

**Theorem** (G. Choquet). Let \(K\) be a convex metrisable compact set in a locally
convex space \(E\). Then every point in \(K\) can be represented as the resultant of a Borel
probability measure on \(K\) which is concentrated on \(\text{ext}(K)\).

**Proof.** Let \(\ell_1, \ell_2, \ldots\) be an arbitrary sequence of continuous linear forms on \(E\) which
is separating for the points of \(K\). For every \(n = 1, 2, 3, \ldots\) we define an associated
continuous linear map \(T_n : E \rightarrow \mathbb{R}^n\) by

\[ T_n(x) = (\ell_1(x), \ell_2(x), \ldots, \ell_n(x)) \quad \text{for all} \quad x \in E. \]

A sequence of convex compact sets \(K_1, K_2, K_3, \ldots\) in respectively \(\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \ldots\) is
defined as

\[ K_n = T_n(K) \quad \text{for all} \quad n = 1, 2, 3, \ldots \]

Note that a sequence \(x_1, x_2, \ldots\) in \(K\) is convergent if and only if for every fixed \(n\)
the sequence \(T_n(x_1), T_n(x_2), \ldots\) converges in \(K_n\). Now let \(a\) be an arbitrarily chosen
element of \(K\). We have to prove that there is then a Borel probability measure on \(K\),
concentrated on \(\text{ext}(K)\), that has \(a\) as its resultant. To this end we set up a branching
process in the following way: First we choose points \(x(1)\) and \(x(2)\) in \(K\) and positive
scalars \(A_1\) and \(A_2\) such that

\[ f T_2(x(1)) = A_1 T_2(x(1) + A_2 T_2(x(2)), \text{where} \quad A_1 + A_2 = 1 \]

This is possible by Carathéodory’s theorem. Once the \(x(1), x(2)\) and \(A_1, A_2\) have been
chosen, we choose points \(x(1,1), x(1,2), x(1,3)\) in \(K\) and positive scalars \(A_{11}, A_{12}, A_{13}\)
such that

\[ T_2(x(1), x(2), x(1,3)) \in \text{ext}(K_2) \]

Again Carathéodory’s theorem assures us that the above is possible and it also assures
us that a couple of points \(x(2,1), x(2,2), x(2,3)\) in \(K\) and positive scalars \(A_{21}, A_{22}, A_{23}\)
can be chosen in such a way that

\[ T_2(x(2,1), x(2,2), x(2,3)) \in \text{ext}(K_2) \]

2
In this way we can go on: Once we have chosen the points \( x(i_1, \ldots, i_n) \) in \( K \) and the positive scalars \( \lambda_{i_1 \ldots i_n i_{n+1}} \) for all

\[
(i_1, i_2, \ldots, i_n) \in \{1,2\} \times \{1,2,3\} \times \cdots \times \{1,2,3,\ldots, n+1\},
\]

the ‘next generation’ of points \( x(i_1, \ldots, i_n, i_{n+1}) \) in \( K \) and positive scalars \( \lambda_{i_1 \ldots i_n i_{n+1}} \) can be chosen such as to satisfy

\[
\begin{align*}
T_{n+1} [x(i_1, \ldots, i_n)] &= \sum_{i_{n+1}=1}^{n+2} \lambda_{i_1 \ldots i_n i_{n+1}} T_{n+1} [x(i_1, \ldots, i_n, i_{n+1})] \\
T_{n+1} [x(i_1, \ldots, i_n, i_{n+1})] &\in \mathcal{E}xt(K_{n+1}) \quad \text{for all} \\
(i_1, \ldots, i_n, i_{n+1}) &\in \{1,2\} \times \cdots \times \{1,2,\ldots, n+2\}
\end{align*}
\]

By construction we now have

\[
T_n [x(i_1, \ldots, i_m)] = T_n [x(i_1, \ldots, i_n)] \in \mathcal{E}xt(K_n) \quad \text{if} \quad m \geq n. \quad \text{(\(\ast\))}
\]

Next we define for every \( n = 1,2,\ldots \) the finite probability space \( (\Omega_n, \mathcal{F}_n, \mathbb{P}_n) \) as

\[
\begin{align*}
\Omega_n &:= \{1,2\} \times \{1,2,3\} \times \cdots \times \{1,2,3,\ldots, n+1\} \\
\mathcal{F}_n &:= \sigma\text{-algebra of all subsets of } \Omega_n \\
\mathbb{P}_n \left[ (i_1, \ldots, i_n) \right] &:= \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_1 i_2 \ldots i_n}.
\end{align*}
\]

Moreover, we define \( \Omega \) to be the set

\[
\Omega := \{1,2\} \times \{1,2,3\} \times \cdots.
\]

Let \( p_n \) be the natural projection of \( \Omega \) on \( \Omega_n \) and let \( \mathcal{F} \) be the smallest \( \sigma \)-algebra that makes all the \( p_n : \Omega \to \Omega_n \) measurable. Now (see [1], [2]) there is on \( (\Omega, \mathcal{F}) \) a unique probability measure \( \mathbb{P} \) such that for all \( n \) the image measure \( p_n(\mathbb{P}) \) of \( \mathbb{P} \) under \( p_n \) equals \( \mathbb{P}_n \). Define on the probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) the sequence of stochastic variables \( X_1, X_2, \ldots \) by

\[
X_m(i_1, i_2, \ldots) := x(i_1, \ldots, i_m).
\]

Then we have, by definition of the \( T_n \) and by construction of the \( x(i_1, \ldots, i_n) \), that

\[
\mathbb{E}[T_n(X_m)] = T_n(a) \quad \text{if} \quad m \geq n. \quad \text{(\(\ast\ast\))}
\]

Furthermore, by \( (\ast) \), for every \( \omega = (i_1, i_2, \ldots) \in \Omega \) and every fixed \( n \) the sequence

\[
T_n[X_1(\omega)], T_n[X_2(\omega)], \ldots
\]

is constant from \( m \geq n \) on. Besides that we have

\[
T_n[X_m(\omega)] \in \mathcal{E}xt(K_n) \quad \text{for all} \quad m \geq n. \quad \text{(***)}
\]

3
Altogether this means that for all \( \omega \in \Omega \) the sequence \( X_1(\omega), X_2(\omega), \ldots \) converges to an element \( X(\omega) \) in \( K \). This limit \( X \) represents, of course, a stochastic variable defined on \( (\Omega, \mathcal{F}, \mathbb{P}) \). For this variable \( X \) we have, by \((***)\), that

\[
T_n[X(\omega)] \in \operatorname{Ext}(K_n) \quad \text{for all } n \text{ and all } \omega \in \Omega.
\]

It is easily deduced from this that

\[
X(\omega) \in \operatorname{Ext}(K) \quad \text{for all } \omega \in \Omega.
\]

The set \( \operatorname{Ext}(K) \) being Borel, the above implies that the probability distribution \( \mathbb{P}_X \) of \( X \) is concentrated on \( \operatorname{Ext}(K) \). Furthermore, by \((**)\), we have for every fixed \( n \) that

\[
T_n[\mathbb{E}(X)] = \mathbb{E}[T_n(X)] = \lim_{m \to \infty} \mathbb{E}[T_n(X_m)] = T_n(a).
\]

Hence

\[
\mathbb{P}_X = \mathbb{E}(X) = a.
\]

It thus appears that \( \mathbb{P}_X \) is a measure that makes it all true. □

Choquet’s theorem has been generalised in several directions. For example, in [5] a non-compact version is proved for sets \( K \) which are convex, closed, bounded subsets with the Radon-Nikodym property in a separable Banach space. The Radon-Nikodym property turns out to be crucial in this (see [10]). As already noticed by G. Choquet himself, the theory finds its natural framework in the context of convex cones rather than convex compact or convex bounded sets. Recent results in this setting, for so-called ‘conuclear cones’, are to be found in [11].

References


