DEPARTMENT OF MATHEMATICS

TECHNICAL REPORT

A SHORT NOTE ON STRASSEN'S THEOREMS

DR. MOTOYA MACHIDA

OCTOBER 2004

No. 2004-6

TENNESSEE TECHNOLOGICAL UNIVERSITY
Cookeville, TN 38505

A Short note on Strassen's theorems

Motoya Machida

Tennessee Technological University

Abstract

This short note will discuss Theorems 1, 3 and 4 of Strassen's paper [6] from the viewpoint of completely modern treatment of conditional distributions.

1 Preliminary

1.1 Hahn-Banach and separation theorem.

A real-valued function h on a real linear space E is called a *Minkowski functional* if it satisfies for $x, y \in E$ and $a \geq 0$,

$$h(x+y) \leq h(x) + h(y) \quad \text{and} \quad h(ax) = ah(x).$$

Suppose that f_0 is a linear functional on a subspace E_0 of E such that $f_0 \leq h$ on E_0. Then f_0 can be extended to a linear functional f on E satisfying $f \leq h$ on E (*Hahn-Banach theorem*; see 5.2 of [3]).

Now let E be a locally convex linear topological space, and let A and B be non-empty convex subset of E. Then there exists a non-trivial continuous linear functional f on E such that $\sup f(A) \leq \inf f(B)$ if and only if $\mathrm{int}(A) \cap B = \emptyset$, where $\sup f(A) := \sup_{x \in A} f(x)$ and $\inf f(B) := \inf_{x \in B} f(x)$, and by $\mathrm{int}(A)$ we denote the interior of A (*Separation theorem*; see Theorem 14.2 of [5]). Furthermore, there exists a continuous linear functional f on E such that $\sup f(A) < \inf f(B)$ if and only if the closure $\overline{(B - A)}$ does not contain 0 (*Strong separation theorem*; see 14.3 of [5]).

1.2 Weak* topology and continuous Minkowski functionals.

Let X be a Banach space and let X^* be its dual (i.e., the linear space of "norm"-continuous real linear functions on X). Here we consider the weak* topology on X^*,

AMS 2000 *subject classifications.* Primary 60B05; secondary 60E05

Key words and phrases. Hahn-Banach theorem, Minkowski functional, conditional distribution, capacity

that is, the smallest topology such that the linear functional $x^*(x)$ on X^* is continuous for every $x \in X$ (see, e.g., page 194 of [2]). Then a linear functional f on X^* is weak* continuous if and only if we have $f(x^*) = x^*(x)$ for some $x \in X$ (17.6 of [5]). Also note that the weak* topology is always Hausdorff and locally convex (see the paragraph above 17.6 of [5]).

Theorem 1.1. *Let K be a non-empty, convex and weak* compact subset of X^*, and let*

$$h(x) := \sup_{x^* \in K} x^*(x) \quad \text{for all } x \in X. \tag{1}$$

Then (a) *h is a continuous Minkowski functional on X, and* (b) $K = \{x^* \in X^* : x^* \leq h\}$.

Proof. In (a) it is easy to see that h is a Minkowski functional. To verify the continuity of h, observe that a Minkowski functional h is norm-continuous if and only if

$$\|h\| := \sup_{\|x\| \leq 1} |h(x)| < \infty.$$

(The proof of the above statement is analogous to that of Theorem 6.1.2 of [2].) Let $U_i = \{x^* \in X^* : \|x^*\| < i\}$ be an open set in X^* for each $i = 1, 2, \ldots$. Since K is compact, there is an integer N such that $K \subset \cup_{i=1}^{N} U_i$. Hence we have $\|h\| \leq \sup_{x^* \in K} \|x^*\| \leq K$; thus, h is continuous.

Clearly we have $K \subseteq \{x^* \in X^* : x^* \leq h\}$ in (b). Suppose $z^* \notin K$ such that $z^* \in h$. Then by the strong separation theorem there exists a continuous linear functional f on X^* satisfying $\sup_{x^* \in K} f(x^*) < f(z^*)$. Moreover, there is some $x \in X$ satisfying $f(x^*) = x^*(x)$ for all $x^* \in X^*$; thus, implying $h(x) < z^*(x)$. However, this is a contradiction. ☐

2 Strassen's theorem for kernel functions

In what follows we assume that X is a separable Banach space, and that $(\Omega, \mathcal{B}, \mu)$ is a μ-complete measure space with σ-algebra $\mathcal{B}$ and probability measure μ. Let $\{h_\omega\}_{\omega \in \Omega}$ be a collection of continuous Minkowski functionals on X. Then we call h_ω a *kernel* if

(1) the map $\omega \to h_\omega(x)$ is $\mathcal{B}$-measurable for each $x \in X$;

$$\int \|h_\omega\| \, d\mu(\omega) < \infty. \tag{2}$$

The kernel h_ω is said to be linear if h_ω is linear for each $\omega \in \Omega$. Note that $\|h_\omega\| < \infty$ for each $\omega \in \Omega$, and that $\|h_\omega\|$ is $\mathcal{B}$-measurable by the separability of X. It is easily observed that

$$h(x) := \int h_\omega(x) \, d\mu(\omega) \tag{3}$$

is a continuous Minkowski functional if h_ω is a kernel. Furthermore, if h_ω is linear, so is h. Then we can state Theorem 1 of Strassen [6].

Theorem 2.1. *Let $x^* \in X^*$, and let h_ω be a kernel. Then the following statements are equivalent:*

(i) $x^* \leq h$ *with the Minkowski functional* (3)*;*

(ii) *there is a linear kernel x^*_ω such that*

$$x^*_\omega \leq h_\omega \quad \text{for all } \omega \in \Omega; \tag{4}$$

$$x^*(x) = \int x^*_\omega(x)\, d\mu(\omega) \quad \text{for all } x \in X. \tag{5}$$

In Theorem 2.1 (ii) is clearly sufficient for (i). We first present the proof of its necessity when Ω is discrete, following the remark at page 426 of Strassen [6]. Let K be the subset of X^* in which each x^* is expressed in the form of (5) with some linear kernel x^*_ω satisfying (4). Then it is not difficult to show that K is non-empty, convex, and weak* compact subset of X^*; the weak* compactness follows from the fact that $\|x^*\| \leq \|h\|$ for all $x^* \in K$. Furthermore, we claim that (1) with the above K coincides with (3). Let $x \in X$ be fixed. For each $\omega \in \Omega$, we can apply the separation theorem with $A = \{(x, h_\omega(x))\}$ and $B = \{(z, r) : h_\omega(x) \leq r\}$ on $X \times \mathbb{R}$ to obtain $-x^*_\omega(x) + a h_\omega(x) \leq -x^*_\omega(z) + ar$ for $(z, r) \in B$, where $x^*_\omega \in X^*$ (but $x^*_\omega \not\equiv 0$) and $a \in \mathbb{R}$. Since we can find some $(z, r) \in B$ satisfying $x^*_\omega(z) - x^*_\omega(x) > 0$ and $r - h_\omega(x) > 0$, we can assume $a = 1$ without loss of generality. By observing that $-x^*_\omega(x) + h_\omega(x) \leq 0$ at $(0, 0) \in B$, we have $x^*_\omega \leq h_\omega$ with $x^*_\omega(x) = h_\omega(x)$; thus, implying $\sup_{x^* \in K} x^*(x) = h(x)$. Therefore, (ii) is necessary for (i) in view of Theorem 1.1.

The difficulty arises in this proof when a continuous case is considered: The application of the separation theorem does not guarantee the $\mathcal{B}$-measurability of the map $\omega \to x^*_\omega(x)$. Taking this into account we now introduce Strassen's proof.

Proof of Theorem 2.1. Assuming (i), we prove the existence of linear kernel x^*_ω desired in (ii).

I. *Construction of a $\mathcal{B}$-measurable function $q_\omega(x)$.* Let $\mathcal{L}$ be the linear space of X-valued simple measurable functions on Ω, and let $\mathcal{L}_0$ be the subspace of constant functions in $\mathcal{L}$. Define for any element $\xi(\omega) = \sum_{i=1}^n x_i I_{A_i}(\omega)$ of $\mathcal{L}$,

$$H(\xi) := \sum_{i=1}^n \int_{A_i} h_\omega(x_i)\, d\mu(\omega),$$

and define for any constant function $\xi(\omega) = x I_\Omega(\omega)$,

$$Q_0(\xi) := x^*(x).$$

Then we can easily check that H is a Minkowski functional on $\mathcal{L}$, and that Q_0 is linear on $\mathcal{L}_0$ satisfying $Q_0 \leq H$ on $\mathcal{L}_0$. Thus, we can apply the Hahn-Banach theorem to extend a linear functional Q on $\mathcal{L}$ satisfying (a) $Q(\xi) = Q_0(\xi)$ for all $\xi \in \mathcal{L}_0$ and (b)

$Q \le H$ on $\mathcal{L}$. Let $x \in X$ be fixed. We can define a signed measure Q_x on $(\Omega, \mathcal{B})$ by setting $Q_x(A) := Q(xI_A)$. Observing that

$$-H(-xI_A) \le Q_x(A) \le H(xI_A) \quad \text{for every } A \in \mathcal{B}, \tag{6}$$

we obtain $Q_x \ll \mu$. Thus, there is a Radon-Nikodym derivative $q_\omega(x)$ such that $Q_x(A) = \displaystyle\int_A q_\omega(x)\, d\mu(\omega)$.

II. *μ-almost everywhere properties of $q_\omega(x)$ and construction of $\tilde{x}^*_\omega(x)$.* Given $a, b \in \mathbb{R}$ and $x, y \in X$, the linearity and Property (b) of the extension Q imply respectively that (a) $q_\omega(ax + by) = a\, q_\omega(x) + b\, q_\omega(y)$ and (b) $q_\omega(x) \le h_\omega(x)$ for μ-a.e. $\omega \in \Omega$. By introducing a countable dense subset X_0 of X satisfying $ax + by \in X_0$ whenever $x, y \in X_0$ and $a, b \in \mathbb{Q}$, we can find a subset N_0 of measure zero for which both (a) and (b) in the previous sentence hold for all $\omega \in \Omega \setminus N_0$ and for all $x, y \in X_0$ and $a, b \in \mathbb{Q}$. Then, for each $\omega \in \Omega \setminus N_0$, there is a unique extension to X, denoted by $\tilde{x}^*_\omega$, of the restriction q_ω on X_0; furthermore, it satisfies $\|\tilde{x}^*_\omega\| = \sup_{x \in X_0} q_\omega(x) \le \|h_\omega\|$, and

$$\tilde{x}^*_\omega(x) = \lim_{x_n \to x} q_\omega(x_n) \quad \text{for every sequence } \{x_n\} \text{ of } X_0 \text{ converging in } X. \tag{7}$$

It is easy to see that the map $\omega \to \tilde{x}^*_\omega(x)$ is $\mathcal{B}$-measurable for each $x \in X$, and that $\tilde{x}^*_\omega \in X^*$ satisfying $q_\omega \le h_\omega$ for each $\omega \in \Omega \setminus N_0$.

III. *μ-almost everywhere equivalence with $q_\omega(x)$ and construction of $x^*_\omega(x)$.* By using (6) and the continuity of h_ω with integrable upper bound $\|h_\omega\|$, we can see that

$$\lim_{x_n \to x} \sup_{A \in \mathcal{B}} |Q_{x_n}(A) - Q_x(A)| = 0,$$

which in turn indicates that $\lim_{x_n \to x} \int |q_\omega(x_n) - q_\omega(x)|\, d\mu(\omega) = 0$. Therefore, together with (7), we can show for every $x \in X$ that $\tilde{x}^*_\omega(x)$ and $q_\omega(x)$ must coincide almost everywhere. Thus, we obtain

$$\int \tilde{x}^*_\omega(x)\, d\mu(\omega) = \int q_\omega(x)\, d\mu(\omega) = Q(xI_\Omega) = x^*(x).$$

Finally set $x^*_\omega = \tilde{x}^*_\omega$ for $\omega \in \Omega \setminus N_0$. For each $\omega \in N_0$ we can find $x^*_\omega \in X^*$ satisfying $x^*_\omega \le h_\omega$ by the Hahn-Banach theorem. Note that the map $\omega \to x^*_\omega(x)$ is $\mathcal{B}$-measurable for each $x \in X$ by the μ-completeness of $\mathcal{B}$. Therefore, x^*_ω is a linear kernel as desired in (ii). $\square$

3 Strassen's theorem for conditional distributions.

3.1 Conditional distributions.

Let $(R, \mathcal{F})$ be another measurable space, and let P be a real-valued function defined on $\mathcal{F} \times \Omega$. Then P is called a *conditional distribution* if (a) $P(\cdot, \omega)$ is a probability measure on $\mathcal{F}$ for μ-a.e. $\omega \in \Omega$, and (b) the map $\omega \to P(E, \omega)$ is $\mathcal{B}$-measurable for every $E \in \mathcal{F}$. A conditional distribution P defines the probability measure $(P \times \mu)$ on $(R \times \Omega, \mathcal{F} \otimes \mathcal{B})$ via

$$\int g \, d(P \times \mu) = \int \int g(r, \omega) \, P(dr, \omega) \, d\mu(\omega) \tag{1}$$

for any $(P \times \mu)$-integrable function g (see, e.g, Theorem 10.2.1 of [2]). It should be noted that, given the measures μ and $(P \times \mu)$, the existence of conditional distribution P cannot be guaranteed unless R is a Polish space (i.e., a complete separable metric space) with Borel σ-algebra $\mathcal{F}$ (see Theorem 10.2.2 and page 351 of [2]). When it exists, the map $\omega \to P(E, \omega)$ coincides with a Radon-Nikodym derivative of $(P \times \mu)(E \times \cdot)$ with respect to μ for every $E \in \mathcal{F}$ (see Theorem 10.2.5 of [2]).

3.2 Strassen's theorem.

Let R be a Polish space with Borel σ-algebra $\mathcal{F}$, and let $C(R)$ be the Banach space of bounded continuous real-valued functions on R. Let $\{h_\omega\}_{\omega \in \Omega}$ be a collection of continuous Minkowski functionals on $C(R)$ such that (a) the map $\omega \to h_\omega(x)$ is $\mathcal{B}$-measurable for every $x \in C(R)$, and (b) $h_\omega(x) \le \sup x(R)$. Then h_ω is continuous on $C(R)$, and therefore, it is a kernel.

Theorem 3.1. *Let ν be a probability measure on $(R, \mathcal{F})$. Then the following statements are equivalent:*

(i) $\int x(r) \, d\nu(r) \le \int h_\omega(x) \, d\mu(\omega)$ *for all $x \in C(R)$;*

(ii) *there is a conditional distribution P on $\mathcal{F} \times \Omega$ such that*

$$\int x(r) \, P(dr, \omega) \le h_\omega(x) \quad \text{for every } x \in C(R) \text{ and for all } \omega \in \Omega; \tag{2}$$

$$\nu(E) = \int P(E, \omega) \, d\mu(\omega) \quad \text{for every } E \in \mathcal{F}. \tag{3}$$

It clearly follows from (1) that (ii) is sufficient for (i). Here we explore the subtle difference between Theorem 2.1 and 3.1 in connection with the necessity of (ii). Define the continuous linear functional

$$x^*(x) := \int x(r) \, d\nu(r), \quad x \in C(R), \tag{4}$$

and similarly define the linear kernel x^*_ω via $x^*_\omega(x) := \int x(r) \, P(dr, \omega)$. Then (3) implies (5). In fact, the converse is true: Given a linear kernel x^*_ω satisfying Theorem 2.1(ii), we can construct a conditional distribution P desired for Theorem 3.1(ii) in the spirit of Riesz representation theorem. (R is not a compact space as in the usual setting for the representation theorem, but a variation of this theorem can be found for locally convex space R in Section 56 of [4].) Therefore, we can rewrite Theorem 3.1 in the form of Theorem 2.1. However, the Banach space $C(R)$ in the place of X is not separable in general. The following proof of Theorem 3.1 introduces an additional technique to get around.

Proof of Theorem 3.1. Assuming (i), we claim that there exists a desired conditional distribution P in (ii).

I. *Introduction of separable Banach space X.* Let $\mathcal{V}$ be the countable algebra generated by countable open base. For each $B \in \mathcal{V}$ choose a sequence $\{B_n\}$ of increasing compact subsets such that $B = \lim_{n\to\infty} B_n$. Then we can construct a countable algebra $\mathcal{U}$ which includes the algebra $\mathcal{V}$ and all the sequences $\{B_n\}$'s for all $B \in \mathcal{V}$ (cf. the proof of Theorem 10.2.2 in [2]). Now for each $A \in \mathcal{U}$ and $\varepsilon > 0$ choose a continuous function $x_{A,\varepsilon}$ on R such that

$$x_{A,\varepsilon}(r) = \begin{cases} 1 & \text{if } r \in A; \\ 0 & \text{if } d(r, A) := \inf\{d(r, s) : s \in A\} > \varepsilon, \end{cases}$$

and $0 \le x_{A,\varepsilon}(r) \le 1$ for all $r \in R$. Thus, we can construct a separable subspace X of $C(R)$ which contains all $x_{A,\varepsilon}$'s for all $A \in \mathcal{U}$ and all $\varepsilon > 0$.

II. *Construction of probability measure $P(\cdot,\omega)$.* When restricted on X, the kernel h_ω and the linear functional x^* in (4) satisfies Theorem 2.1(i), and therefore, there exists a linear kernel x^*_ω satisfying Theorem 2.1(ii). Observe for $A \in \mathcal{U} \setminus \{R, \emptyset\}$ that

$$0 = -\sup(-x_{A,\varepsilon}(R)) \le -h_\omega(-x_{A,\varepsilon}) \le x^*_\omega(x_{A,\varepsilon}) \le h_\omega(x_{A,\varepsilon}) \le \sup(x_{A,\varepsilon}(R)) = 1.$$

For each $\omega \in \Omega$ we can define a finitely additive nonnegative measure $P(\cdot,\omega)$ on the algebra $\mathcal{U}$ via

$$P(A,\omega) = \lim_{k\to\infty} x^*_\omega(x_{A,1/k}), \quad A \in \mathcal{U}.$$

Then the map $\omega \to P(A,\omega)$ is $\mathcal{B}$-measurable, and satisfies (3) for every $A \in \mathcal{U}$.

Let $\omega \in \Omega$ be fixed. Then $P(\cdot,\omega)$ satisfies for each $B \in \mathcal{V}$,

$$P(B,\omega) = \sup\{P(K,\omega) : K \in \mathcal{U} \text{ and } K \text{ is a compact subset of } B\},$$

and is called *regular on $\mathcal{V}$ for $\mathcal{U}$*. According to Theorem 10.2.4 of [2], the regular finitely additive $P(\cdot,\omega)$ is countably additive on $\mathcal{V}$. Thus, we can extend it uniquely to a measure $P(\cdot,\omega)$ on the Borel σ-algebra $\mathcal{F}$ (see, e.g., Theorem 3.1.4 and 3.1.10 of [2]) satisfying (2). We can also show that $P(\cdot,\omega)$ is a probability measure for μ-a.e. $\omega \in \Omega$. Since R is separable and ν is tight (cf. Theorem 7.1.4 of [2]), for any $\delta > 0$ we can find a sequence $\{K_n\}$ of compact subsets such that $R = \lim_{n\to\infty} K_n$ and $\nu(K_n) > 1 - \delta 2^{-2n}$. We can immediately see that $P(R,\omega) = \lim_{n\to\infty} P(K_n,\omega) \le 1$ for every $\omega \in \Omega$. Let $a_n := \mu(\{\omega : P(K_n,\omega) > 1 - 2^{-n}\})$ for each $n = 1, 2, \ldots$. Then we have $a_n > 1 - \delta 2^{-n}$ since

$$1 - \delta 2^{-2n} < \nu(K_n) = \int P(K_n,\omega)\, d\mu(\omega) \le (1 - 2^{-n})(1 - a_n) + a_n.$$

Therefore, we obtain

$$\mu(\{\omega : P(K_n,\omega) > 1-2^{-n} \text{ for all } n\ \}) > 1-\sum_{n=1}^{\infty}(1-a_n) \geq 1-\delta,$$

which implies that $P(R,\omega)=1$ for μ-a.e. $\omega \in \Omega$.

III. *Monotone class argument for the existence of P.* Let $\mathcal{E}$ be the collection of $E \in \mathcal{F}$ such that the map $\omega \to P(E,\omega)$ is $\mathcal{B}$-measurable, satisfying (3). It is easy to check that $\mathcal{E}$ is a monotone class; thus, $\mathcal{E}=\mathcal{F}$ by the monotone class theorem (see, e.g., 4.4.2 of [2]). Therefore, P is a conditional distribution as desired in (ii). □

4 Capacity and Strassen's Theorem 4.

Let $\mathcal{G}$ be the family of open subsets in R. A real-valued function f is called a *normalized capacity alternating of order 2* if (a) $f(\emptyset)=0$ and $f(R)=1$, (b) $f(U) \leq f(V)$ whenever $U \subset V$, (c) $f(U)=\lim_{n\to\infty} f(U_n)$ whenever $U_n \uparrow U$, and (d) $f(U\cup V)+f(U\cap V) \leq f(U)+f(V)$. The normalized capacity f alternating of order 2 defines the continuous Minkowski functional $h(x)$ on $C(R)$ via

$$h(x) = \inf x(R) + \int_{\inf x(R)}^{\infty} f(\{r : x(r) > t\})\,dt, \quad x \in C(R). \tag{1}$$

Once h is found to be a Minkowski functional (see [1] for the proof), h is clearly continuous since $h(x) \leq \sup x(R)$.

Let F be a real-valued function on $\mathcal{G}\times\Omega$. Then F is said to be a *kernel alternating of order 2* if (a) $F(\cdot,\omega)$ is a normalized capacity alternating of order 2 for every $\omega \in \Omega$, and (b) the map $\omega \to F(U,\omega)$ is $\mathcal{B}$-measurable for every $U \in \mathcal{G}$. If F is a kernel alternating of order 2, then

$$f(U) = \int F(U,\omega)\,d\mu(\omega), \quad U \in \mathcal{G} \tag{2}$$

defines a normalized capacity alternating of order 2.

Theorem 4.1. *Let ν be a probability measure on $(R,\mathcal{F})$, and let F be a kernel alternating of order 2. Then the following statements are equivalent:*

(i) *$\nu(U) \leq f(U)$ in (2) for all $U \in \mathcal{G}$;*

(ii) *there is a conditional distribution P on $\mathcal{F}\times\Omega$ such that*

$$P(U,\omega) \leq F(U,\omega) \quad \textit{for every } U \in \mathcal{G} \textit{ and for all } \omega \in \Omega;$$
$$\nu(E) = \int P(E,\omega)\,d\mu(\omega) \quad \textit{for every } E \in \mathcal{F}.$$

Proof. Again (ii) is sufficient for (i). Here we outline the proof for its necessity. We can define a kernel h_ω via

$$h_\omega(x) = \inf x(R) + \int_{\inf x(R)}^{\infty} F(\{r : x(r) > t\}, \omega)\, dt, \quad x \in C(R).$$

Clearly we have $h_\omega(x) \leq \sup x(R)$ as in the setting of Theorem 3.1. Furthermore, we can observe in (1) that $h(x) = \int h_\omega(x)\, d\mu(\omega)$ for all $x \in C(R)$. Then it is not difficult to show that Theorem 4.1(i) implies Theorem 3.1(i), and that the conditional distribution P in Theorem 3.1(ii) is the one desired in Theorem 4.1(ii). $\square$

References

[1] Choquet, G. (1953). Theory of capacities. *Ann. Inst. Fourier.* **5** 131–292.

[2] Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge University Press, Cambridge, UK.

[3] Folland, G. B. (1984). *Real Analysis. Modern techniques and their applications.* John Wiley & Sons, New York.

[4] Halmos, P. R. (1950). *Measure Theory.* Springer-Verlag, New York.

[5] Kelley, J. L. and Namioka, I. (1963). *Linear Topological Spaces.* Van Nostrand, Princeton.

[6] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.