Elect. Comm. in Probab. **4** (1999) 51–59

ELECTRONIC COMMUNICATIONS in PROBABILITY

ON STRASSEN'S THEOREM ON STOCHASTIC DOMINATION

TORGNY LINDVALL
Mathematical statistics, Chalmers and GU, 41296 Göteborg, Sweden.
email: lindvall@math.chalmers.se

submitted August 1, 1998; *final version accepted* June 1, 1999

AMS subject classification: 60B05 (60E15, 60J10)
Strassen's theorem; coupling; pre-ordering; maximal diagonal probability

Abstract

The purpose of this note is to make available a reasonably complete and straightforward proof of Strassen's theorem on stochastic domination, and to draw attention to the original paper. We also point out that the maximal possible value of $\mathbb{P}(Z = Z')$ is actually not reduced by the requirement $Z \preceq Z'$. Here, Z, Z' are stochastic elements that Strassen's theorem states exist under a stochastic domination condition. The consequence of that observation to stochastically monotone Markov chains is pointed out. Usually the theorem is formulated with the assumption that $\preceq$ is a partial ordering; the proof reveals that a pre-ordering suffices.

1 Introduction, preliminaries

Strassen's theorem on stochastic domination is now established as a canonical result in probability: it is a crucial tool in the theory of interacting particle systems, cf. Liggett (1985), and has also found many interesting applications in other areas, cf. Preston (1976), Kamae, Krengel and O'Brien (1977), and Lindvall (1992). But there seems to be no proof of it easily available. The one in Liggett (1985) and Lindvall (1992) treats only the compact space case, and in Strassen (1965) the result is rather implicit, only briefly mentioned as a special case of a general theorem. This makes a straightforward proof worth our while, considering the importance of the theorem.

Strassen (1965) works in a Polish space, i.e., it is (homeomorphic to) a complete and separable metric space. This condition has later been relaxed in several steps, cf. Skala (1993), but the main arguments of the original proof have not been altered. This gives a reason to draw attention to the natural and beautiful strand of ideas that leads to the domination result. Polish spaces suffice for most purposes in probability theory, and the proof is not more complicated than that in Liggett (1985) for the compact space case. Due to the lack of highlighting of the domination result in Strassen (1965), there is a risk that too few have cared to consult the original paper. Indeed: a sample among some competent and honest colleagues revealed that the word "risk" is not really the right one...

This section continues with the preliminaries and proof of the theorem; the proof of a crucial supporting result is, however, deferred to Section 4.

The new observation offered by this note, presented in Section 2, is that the maximal possible value of $\mathbb{P}(Z = Z')$ is actually not reduced by the requirement $Z \preceq Z'$. Here, Z, Z' are stochastic elements that Strassen's theorem states exist under a stochastic domination condition. The consequence of that observation to stochastically monotone Markov chains is pointed out in Section 3.

Let E be a Polish space endowed with a partial ordering $\preceq$, i.e., a relation satisfying

$$\begin{array}{ll} \text{(i)} & x \preceq x \quad \text{for all} \quad x \in E, \\ \text{(ii)} & x \preceq y,\, y \preceq z \;\Rightarrow\; x \preceq z, \text{ and} \\ \text{(iii)} & x \preceq y,\, y \preceq x \;\Rightarrow\; x = y. \end{array} \tag{1}$$

When (iii) is dropped, the relation $\preceq$ is called a pre-ordering; from now on, we assume $\preceq$ to be such an ordering.

A real-valued function f on E is said to be increasing if

$$x \preceq y \Rightarrow f(x) \leq f(y).$$

We let $\mathcal{E}$ denote the Borel σ-field on E as well as the class of real-valued functions measurable w.r.t. that σ-field; $ib\mathcal{E}$ is the class of functions $\in \mathcal{E}$ which are increasing and bounded.

For probability measures P and P' on $(E, \mathcal{E})$, P is stochastically dominated by P' if

$$\int f dP \leq \int f dP' \quad \text{for all} \quad f \in ib\mathcal{E}.$$

We write that $P \overset{\mathcal{D}}{\preceq} P'$. The meaning of $Y \overset{\mathcal{D}}{\preceq} Y'$ for random elements in $(E, \mathcal{E})$ is that the distributions P, P' of Y, Y' satisfy $P \overset{\mathcal{D}}{\preceq} P'$.

We assume that the partial ordering $\preceq$ is closed, i.e., the set

$$M = \{(x, x');\, x \preceq x'\}$$

is closed in the product topology on E^2; we abbreviate $E \times E, \mathcal{E} \times \mathcal{E}$ to E^2 and $\mathcal{E}^2$, as is standard. Strassen's theorem on stochastic domination states that

$$\begin{array}{l}\text{if } P \text{ and } P' \text{ satisfy } P \overset{\mathcal{D}}{\preceq} P', \text{ then there exists a} \\ \text{probability measure } \hat{P} \text{ on } (E^2, \mathcal{E}^2) \text{ with marginals} \\ P \text{ and } P' \text{ such that } \hat{P}(M) = 1.\end{array} \tag{2}$$

The random element version is the following: under the assumption of (2),

$$\begin{array}{l}\text{there exist random elements } Z, Z' \text{ in } (E, \mathcal{E}) \\ \text{with distribution } P, P' \text{ respectively, on some} \\ \text{probability space, such that } Z \preceq Z' \text{ a.s.}\end{array} \tag{3}$$

The result (2) is often referred to as "Strassen's Theorem", which is formally misleading: in Strassen (1965) it is only briefly mentioned as one possible application of Theorem 11 in that paper, and the condition $P \overset{\mathcal{D}}{\preceq} P'$ does not appear explicitly. The crucial result of Strassen (1965), that leads to the mentioned Theorem 11 among other things, is his Theorem 7. It states that if P and P' are probability measures on $(E, \mathcal{E})$, Π is the class of such measures on $(E^2, \mathcal{E}^2)$, and Λ a convex subset of Π closed with respect to weak convergence, then there exists a $\hat{P} \in \Lambda$ with marginals P, P' if and only if

$$\int f dP + \int g dP' \leq \sup_{Q \in \Lambda} \{ \int (f(x) + g(x')) dQ(x, x') \} \tag{4}$$

for all continuous functions f, g on E satisfying $0 \leq f,\ g \leq 1$.

To prove this theorem, which certainly has a strong appeal to intuition, it is inevitable to rely on some functional analysis. The proof, a version of Strassen's original one, is deferred to Section 4.

Let

$$\Lambda = \{Q \in \Pi; Q(M) = 1\}.$$

That Λ is indeed convex and weakly closed in Π (throughout, we use the term "weak" in the standard way of probability theory). Now assume $P \preceq P'$. To verify (4), we pick the relevant ideas from the proof of the mentioned Theorem 11 in Strassen (1965). Let f and g be continuous functions on E, with $0 \leq f,\ g \leq 1$.

We have

$$\int g dP' \leq \int g^* dP' \leq \int g^* dP$$

where $g^*(x) = \sup\{g(y); x \preceq y\}$; this is due the facts that $g \leq g^*$ and that g^* is decreasing.

We get

$$\int f dP + \int g dP' \leq \int (f + g^*) dP$$
$$\leq \sup_x (f(x) + g^*(x)) \leq \sup\{f(x) + g(x');\ x \preceq x'\}$$
$$= \sup\{\int (f(x) + g(x')) dQ(x, x');\ Q \in \Lambda\}.$$

Hence (4) holds.

2 The maximal diagonal probability

For probability measures Q and Q' on $(E, \mathcal{E})$, we know that

$$\|Q - Q'\| \leq 2 \cdot \mathbb{P}(Y \neq Y') \tag{5}$$

for any random elements Y and Y' such that $Y \stackrel{\mathcal{D}}{=} Q$ and $Y' \stackrel{\mathcal{D}}{=} Q'$, defined on some space with probability measure $\mathbb{P}$; here $\|\cdot\|$ denotes the total variation norm for signed bounded measures on $(E, \mathcal{E})$. For the difference of two probability measures, we have $\|Q - Q'\| = 2 \cdot sup_{A \in \mathcal{E}}(Q(A) - Q'(A))$. The inequality (5) is the well-known basic coupling inequality, cf., e.g., Lindvall (1992), §I.2. Hence we always have

$$\mathbb{P}(Y = Y') \leq 1 - \|Q - Q'\|/2. \tag{6}$$

But equality may be achieved in (5)-(6) using the so-called γ-coupling; cf. Lindvall (1992), §I.5. The idea is the following. Let Q^0 have density $g \wedge g'$ with respect to a reference measure λ relative to which both Q and Q' are absolutely continuous, with densities g and g' respectively (we may take $\lambda = Q + Q'$). Lift the sub-probability Q^0 to the diagonal $\Delta = \{(x, x) \in E^2\}$ using the mapping $\psi : E \to E^2$ defined by

$$\psi(x) = (x, x),\ x \in E.$$

With $Q^* = Q^0 \psi^{-1}$ we have

$$Q^*(\Delta) = Q^0(E) = \int g \wedge g' d\lambda. \tag{7}$$

Now

$$\|Q - Q'\| = 2 \cdot (1 - \int g \wedge g' d\lambda\ ) \tag{8}$$

since we indeed have $sup_{A\in\mathcal{E}}(Q(A)-Q'(A)) = sup_{A\in\mathcal{E}}(\int_A gd\lambda - \int_A g'd\lambda)$ $= 1-\int g\wedge g'd\lambda$; cf. Lindvall (1992), p. 19. So (6), (7) and (8) yield that if (Y,Y') takes values on Δ according to Q^*, we have maximized $\mathbb{P}((Y,Y')\in\Delta)$. Denote the maximal possible value $(=\int g\wedge g'd\lambda)$ by γ.

Now the question is: if we know that $Q \overset{\mathcal{D}}{\preceq} Q'$, is it possible to sharpen (2) to the effect that the coupling has mass γ on Δ? To that end, assume $\gamma<1$ (the case $\gamma=1$ is trivial), and let

$$Q_0 = Q - Q^0 \quad \text{and} \quad Q_0' = Q' - Q^0.$$

Then

$$Q_1 = Q_0/(1-\gamma) \quad \text{and} \quad Q_1' = Q_0'/(1-\gamma)$$

are probability measures.

The crucial observation is that

$$\text{if} \quad Q \overset{\mathcal{D}}{\preceq} Q' \quad \text{then} \quad Q_1 \overset{\mathcal{D}}{\preceq} Q_1'.$$

To prove this, let $f\in ib\mathcal{E}$. We get

$$\int fdQ_0' - \int fdQ_0 = \int fdQ' - \int fdQ \geq 0,$$

and division by $(1-\gamma)$ renders

$$\int fdQ_1 \leq \int fdQ_1'.$$

Now let $\hat{Q}_1$ be a coupling of Q_1 and Q_1' according to (2), and

$$\tilde{Q} = Q^* + (1-\gamma)\hat{Q}_1. \tag{9}$$

Then $\tilde{Q}$ is indeed a coupling of Q and Q', $\tilde{Q}(M)=1$, and we have maximized the probability of Δ.

There is a theorem to formulate. We return to the setting of Section 1; the meaning of γ and $\tilde{P}$ is the obvious one.

Theorem. *For a $\hat{P}$ satisfying* (2)*we have $\hat{P}(\Delta)\leq\gamma$. By choosing $\hat{P}=\tilde{P}$ we achieve $\tilde{P}(\Delta)=\gamma$.* □

What we have achieved so far may be summed up in terms of random elements as follows. If $Z \overset{\mathcal{D}}{=} P$ and $Z' \overset{\mathcal{D}}{=} P'$, we always have that $\mathcal{P}(Z=Z')\leq\gamma$, and equality is obtained using γ-coupling. That equality may be retained for a sharpening of (3): if we also have $P \overset{\mathcal{D}}{\preceq} P'$, (Z,Z') may actually be constructed such that

$$Z \preceq Z' \text{ a.s., and } \mathbb{P}(Z=Z')=\gamma. \tag{10}$$

3 An application to Markov chains

An application of Theorem 2 in Kamae et. al. (1977) to time-homogeneous Markov chains is the following. Let the assumptions established above be at force, with E endowed with a closed partial ordering $\preceq$. Let us assume that the Markov kernel $P(x, A), x \in E, A \in \mathcal{E}$, on $(E, \mathcal{E})$ is monotone in the sense that

$$P(x, \cdot) \overset{\mathcal{D}}{\preceq} P(x', \cdot) \quad \text{whenever} \quad x \preceq x'. \tag{11}$$

Then, for initial distribution λ and λ' satisfying $\lambda \overset{\mathcal{D}}{\preceq} \lambda'$, we may produce Markov chains $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$, both governed by $P(\cdot, \cdot)$, such that (X_n, X'_n), $n = 0, \ldots$ is a Markov chain and

$$X_n \preceq X'_n \quad \text{for all} \quad n \geq 0 \tag{12}$$

and $X_0 \overset{\mathcal{D}}{=} \lambda, X'_0 \overset{\mathcal{D}}{=} \lambda'$. The method is the natural one: use Strassen's theorem first to get X_0 and X'_0 satisfying $X_0 \preceq X'_0$ and with distributions λ, λ' respectively, then repeatedly for $n = 0, 1, \ldots$ That is, after n steps, $n = 0, \ldots,$ we have obtained

$$X_0 \preceq X'_0, \ldots, X_n \preceq X'_n\,.$$

Another application of Strassen's theorem using (11) renders $X_{n+1} \preceq X'_{n+1}$, and (12) is set by induction.

Using the theorem of Section 2, we find that X and X' may be constructed, in addition to having the property (12), such that

$$\mathbb{P}(X_{n+1} = X'_{n+1} | X_n = x, X'_n = x') = 1 - \|P(x, \cdot) - P(x', \cdot)\|/2 \tag{13}$$

for all $x \preceq x'$ and for all $n = 0, \ldots,$ and that probability can never be strictly larger, under the condition that $(X_n, X'_n), n = 0, \ldots$ is a Markov chain.

Also: we noticed in Section 2 that (12)-(13) hold just as well in a pre-ordered space.

We content ourselves with this sketchy account. At the side of Kamae et. al. (1977), see also Lindvall (1992), §IV.2, for Markov chain domination.

4 The proof of Strassen's Theorem 7

For a Banach space B with norm $\| \cdot \|$, the dual space B^* is the set of continuous linear functionals on B. The weak* topology on B^* is defined as the weakest one such that for all $\mathsf{z} \in \mathsf{B}$ the mapping $L \to L\mathsf{z}$ is continuous. The weak*-topology renders the linear space B^* locally convex.

For our purposes, let

$$\mathsf{B} = \{(f, g); f, g \text{ are bounded continuous functions on } E\}.$$

We may use $\|(f,g)\| = |f| + |g|$ as norm on B, where $|\cdot|$ is the supremum norm for functions on E.

In (4), the restrictions $0 \leq f, g \leq 1$ were introduced to facilitate the estimates that followed; they are now dropped without effect on the problem. So from now on, f and g denote any bounded continuous functions.

Let

$$H : \text{all pairs } [\mu, \mu'] \text{ of probability measures } \mu, \mu' \text{ on } (E, \mathcal{E}).$$

Notice that such a pair defines a linear functional $\in \mathsf{B}^*$ through the definition

$$[\mu, \mu'](f, g) = \int f d\mu + \int g d\mu' \quad \text{for} \quad (f, g) \in \mathsf{B}.$$

Further, let

$$H_\Lambda = \{[\mu, \mu'] \in H;\ \exists \lambda \in \Lambda \text{ with } \mu \text{ and } \mu' \text{ as marginals}\}.$$

Due to the convexity assumption on Λ, we have that H_Λ is convex in B^*. The reader is now deferred to Rudin (1973), Theorems 3.4(b) and 3.10, and §3.14, for proofs of the following:

(i) all weak*-continuous functionals on B^* are of the form $L \to L(f, g)$ for some $(f, g) \in \mathsf{B}$,

(ii) if B is a closed and convex set $\subset \mathsf{B}^*$ and $L_0 \notin B$, then there exists a weak*-continuous linear functional $\mathcal{T}$ such that $\mathcal{T}(L_0) > \sup_{L \in B} \mathcal{T}(L)$.

Now suppose $[P, P'] \notin \bar{H}_\Lambda$. Find a functional $\mathcal{T}_0$ according to (ii) such that

$$\mathcal{T}_0[P, P'] > \sup_{L \in \bar{H}_\Lambda} \mathcal{T}_0(L).$$

Then find (f_0, g_0) representing $\mathcal{T}_0$ according to (i). This means

$$[P, P'](f_0, g_0) > \sup_{L \in \bar{H}_\Lambda} L(f_0, g_0).$$

But this is just another formulation of

$$\int f_0 dP + \int g_0 dP' > \sup_{Q \in \Lambda} \{\int (f(x) + g(x')) dQ(x, x')\},$$

violating (4). Hence (4) implies that $[P, P'] \in \bar{H}_\Lambda$.

It remains to prove that H_Λ is closed in B^*. To that end, we first note that the relativized weak* topology on H is metrizable. Indeed, let $\mathbf{P}$ denote the probability measures on $(E^2, \mathcal{E}^2)$. Then the mapping $\phi : H \to \mathbf{P}$ defined by

$$\phi([\mu, \mu']) = \mu \times \mu'$$

is a homeomorphism from H onto a closed subset of **P** endowed with the weak topology. But that topology is metrizable by the Prohorov metric, cf. Billingsley (1968), p. 238.

This means that in order to prove that H_Λ is weak* closed, it suffices to show that the limit of a convergent sequence of elements in H_Λ still is in H_Λ. For that, let $[\mu_n, \mu'_n], n \geq 1$, be a sequence in H_Λ which is weak* convergent to $[P, P']$. We adopt the $\Rightarrow$ notation for weak convergence of sequences of probability measures: cf. Billingsley (1968), where all the standard theory needed is easily found.

We have $\mu_n \Rightarrow P$ and $\mu'_n \Rightarrow P'$. This implies that any sequence $(\lambda_n)_1^\infty$ such that λ_n has marginals μ_n, μ'_n for $n \geq 1$ is tight. Indeed, take any $\epsilon > 0$ and let K_ϵ, K'_ϵ be compact sets such that

$$\inf_n \mu_n(K_\epsilon) > 1 - \epsilon/2 \quad \textit{and} \quad \inf_n \mu_n(K'_\epsilon) > 1 - \epsilon/2.$$

It is immediate that $\inf_n \lambda_n(K_\epsilon \times K'_\epsilon) > 1 - \epsilon$, and $K_\epsilon \times K'_\epsilon$ is compact in E^2.

Now apply Prohorov's theorem to find a cluster point λ of $(\lambda_n)_1^\infty$. Since Λ is closed, we have $\lambda \in \Lambda$; that λ has marginals P and P' follows from the continuous mapping theorem, applied to the mappings $(x, x') \to x$ and $(x, x') \to x'$.

It seems as Strassen's domination theorem has an easy proof only when E is either finite or is a subset of the real line.

Acknowledgement

Ali Falahati, a graduate student at my department, showed me a proof of the theorem in Section 2 for the case $E = \{0, 1, \ldots\}$. His construction, suitable for simulations, was one inspiration for this note.

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] Kamae, T., Krengel, U. and O'Brien, G.L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5**, 899-912.

[3] Liggett, T. (1985). *Interacting Particle Systems.* Springer, New York.

[4] Lindvall, T. (1992). *Lectures on the Coupling Method.* Wiley, New York.

[5] Preston, C. (1976). *Random Fields.* Lecture Notes in Mathematics **534**. Springer, Berlin.

[6] Rudin, W. (1973). *Functional Analysis.* McGraw-Hill, New Dehli.

[7] Skala, H.J. (1993). The existence of probability measures with given marginals. *Ann. Probab.* **21**, 136-142.

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