Munich Center for
Mathematical Philosophy
27 June 2014

STRASSEN CAPACITIES AS CONSTRAINTS ON PROBABILITY REVISION

Carl Wagner
University of Tennessee

1. Jeffrey Conditioning (JC)

(Θ, **A**, p) = a probability space

E = a countable family of nonempty, pairwise disjoint events, with p(E) > 0 for all E ϵ **E**

{ u_E : E ϵ **E** } = a family of positive real numbers with ∑ u_E = 1

New evidence prompts you to revise your prior p to a posterior q .

● Based on the *total* evidence, old as well as new, you decide to set

(1) q(E) = u_E for all E ϵ **E** .

Unless each E ϵ **E** is a singleton, there are infinitely many q's satisfying (1).

But you also judge that nothing in the evidence alters the relevance of any E ϵ **E** to any A ϵ **A**, and so you set

(2) $q(A \mid E) = p(A \mid E)$

for all A ϵ **A** and all E ϵ **E** (*Rigidity*).

- (1) and (2) are equivalent to

(JC) $q(A) = \sum_{E \epsilon \mathbf{E}} u_E \, p(A \mid E)$

for all A ϵ **A**.

EXAMPLE (Jeffrey) : *The Mudrunner*

Θ = { wm, lm, wd, ld }
W = { wm, wd }, M = { wm, lm } , etc.
$p(W \mid M) = 0.8$, $p(W \mid D) = 0.1$
$p(M) = 0.3$, and $p(D) = 0.7$, so

$p(W) = (0.3)(0.8) + (0.7)(0.1) = 0.31$

A fresh weather forecast yields revised probabilities $q(M) = 0.6$ and $q(D) = 0.4$.

It seems clear that $p(W|M)$ and $p(W|D)$ should remain unchanged in passing from p to q. So by JC,

$$q(W) = (0.6)(0.8)+ (0.4)(0.1) = 0.52$$

Remark. $q(M)$ and $q(D)$ are actually *inferred* from a probability u on $\Omega = \{ \text{rain, clear} \}$, with u(rain) = 0.6, etc., along with the assumptions that

rain => a muddy track, and

clear weather => a dry track.

This is a special case of a basic result in probability theory:

2. Point-valued Mappings of a Probability Space

If

(Ω, **A**, u) is a probability space,

(Θ, $\mathbf{A}^*$) is a measurable space,

$t : \Omega \rightarrow \Theta$, with $t^{-1}(A) \in \mathbf{A}$ for all $A \in \mathbf{A}^*$,

and

w: $\mathbf{A}^* \rightarrow [0,1]$ by $w(A) = u(t^{-1}(A))$,

then w is a probability measure (henceforth, "pm") on $\mathbf{A}^*$.

What if t is replaced by a set-valued map $T: \Omega \rightarrow \mathbf{A}^*$, with $T(\omega) =$ the set of Θ-states compatible with ω ?

Consider a simple example…

3. Marginal Bounds for an Incomplete Contingency Table

	Black	Red	Green	
Sphere	0			0.1
Cylinder	0			0.2
Cube			0	0.5
Cone	0	0	(0.2)	0.2
	b	r	g	1.0

$0 \leq b \leq 0.5 \qquad 0 \leq r \leq 0.8 \qquad 0.2 \leq g \leq 0.5$

$0.5 \leq b+r \leq 0.8 \qquad 0.2 \leq b+g \leq 1.0$

$0.5 \leq r+g \leq 1.0$

The lower and upper bounds above, regarded as set functions on the power set of C = {black, red, green} are simple examples of *Strassen capacities.*

4. Strassen Mappings and Capacities

V. Strassen, *Messfehler und Information*, ZW **2** (1964), 273-305.

(Ω, **A**, u) is a probability space.

Θ is *finite* ; T: $\Omega \rightarrow 2^{\Theta} \setminus \{\emptyset\}$.

- For all $E \subset \Theta$, let

$E_{\#} := T^{-1}(\{E\}) = \{ \omega \in \Omega : T(\omega) = E \}$,

and $m(E) := u(E_{\#})$.

- $\sum_{E \subset \Theta} m(E) = 1$, with $m(\emptyset) = 0$.

- Let **E** : = $\{ E \subset \Theta : m(E) > 0 \}$. Call members of the family **E** *focal events*.

● For all $A \subset \Theta$, let

$$\beta(A) := \sum_{E \subset A} m(E)$$

$$= u\{\omega \in \Omega : T(\omega) \subset A\}, \quad \text{and}$$

$$\alpha(A) := \sum_{E \cap A \neq \emptyset} m(E)$$

$$= u\{\omega \in \Omega : T(\omega) \cap A \neq \emptyset\}.$$

● Call α and β, respectively, the *Strassenian upper and lower probabilities induced by* u and T. Since

$$\alpha(A) = 1 - \beta(A^c) \quad (\textit{conjugacy}),$$

it suffices to restrict attention in what follows to β.

Some Properties of β:

- $0 \le \beta(A) \le 1$; $\beta(\varnothing) = 0$; $\beta(\Theta) = 1$.

- β is *r-monotone* for every $r \ge 2$:

$$\beta(A_1 \cup \ldots \cup A_r) \ge \sum \beta(A_i) - \sum \beta(A_i \cap A_j)$$

$$+ \ldots + (-1)^{r-1} \beta(A_1 \cap \ldots \cap A_r)$$, and so

- β is *superadditive* :

$A \cap B = \varnothing \Rightarrow \beta(A \cup B) \ge \beta(A) + \beta(B)$,

hence, *monotone*:

$$A \subset B \Rightarrow \beta(A) \le \beta(B).$$

- Monotone set functions satisfying certain continuity properties (which automatically hold for finite sets) are called *capacities* (Choquet 1953-54).

● One can recover m from β by the formula

$$m(E) = \sum_{H \subset E} (-1)^{|E \setminus H|} \beta(H).$$

This explains why m is often called the *Möbius transform of* β. Shafer(1976) calls m a *basic probability assignment*, and β a *belief function.*

Theorem 4.1. The Strassenian lower probability β induced on 2^{Θ} by u and T is a probability measure, and hence equal to α, if and only if every focal event is a *singleton* subset of Θ.

Generalizing JC to the case in which the possible revisions of a prior are bounded below by a Strassenian lower probability β :

5. A Lower Bound on the Revision of a Prior

● Θ and Ω are *both* finite sets of possible states of the world (a.k.a. *frames of discernment* – Shafer 1976).

● p = a prior pm on 2^{Θ}, with p(θ) > 0 for all θ ϵ Θ.

● New evidence yields a pm u on 2^{Ω}, with u(ω) > 0 for all ω ϵ Ω. The measure u is based on all the evidence, old as well as new.

● Ω-states are related to Θ-states by a Strassen mapping

$$T: \Omega \rightarrow 2^{\Theta} \setminus \{\emptyset\} ,$$

where T(ω) is the set of all Θ-states compatible with the Ω-state ω.

- $\beta(A) = u(\{\omega \in \Omega: T(\omega) \subset A\})$ is the Strassenian lower probability induced on 2^Θ by u and T.

- The possible revisions of p compatible with u and T are precisely those pm's q satisfying $q(A) \geq \beta(A)$ for all $A \subset \Theta$, since $\beta(A)$ is the sum of the probabilities of all those states ω which entail the event A.

Theorem 5.1. (Strassen). A pm q on 2^Θ dominates the Strassenian lower probability β induced by u and T iff q is a *smear* of the Möbius transform m of β, i.e., iff there exists a family $\{ w_E : E \in \mathbf{E} \}$ of *probability mass functions* on Θ such that

$$\theta \in E^c \text{ implies that } w_E(\theta) = 0 \text{ , and}$$

for all $\theta \in \Theta$,

$$q(\theta) = \sum_{E \in \mathbf{E}} w_E(\theta)\, m(E) \text{ .}$$

- Consider the set of probability measures Q on $2^{\Omega x \Theta}$ that are *compatible* with u and T, in the sense that

(1) $Q(\omega,\theta) = 0$ if $\theta \in (T(\omega))^c$

(2) $Q_\Omega = u$, where

Q_Ω is the *marginalization* of Q to Ω,
i.e., $Q_\Omega(E) = Q(E \times \Theta)$ for all $E \subset \Omega$.

Q's compatible with u and T :

	θ_1	…	θ_j	…	θ_n	
ω_1						$u(\omega_1)$
…						…
ω_i			$Q(\omega_i,\theta_j)$			$u(\omega_i)$
…						…
ω_m						$u(\omega_n)$
						1.0

(1) $Q(\omega_i,\theta_j) = 0$ if $\theta_j \in (T(\omega_i))^c$

(2) $\Sigma_j\, Q(\omega_i,\theta_j) = u(\omega_i)$, $i = 1,\ldots,m$

From Theorem 5.1, we get

Theorem 5.2. A pm q on 2^{Θ} satisfies $q \geq \beta$ iff there exists a pm Q on $2^{\Omega \times \Theta}$ such that

(1) Q is compatible with u and T, and

(2) $q = Q_{\Theta}$, the marginalization of Q to Θ

6. A Natural Generalization of JC

Let **Q :=** the set of all pm's Q on $2^{\Omega \times \Theta}$ that are

(1) compatible with u and T, and

(2) satisfy the additional "rigidity" condition

(GR) For all $A \subset \Theta$ and all $E \in \mathbf{E}$,

$$Q(\Omega \times A \mid E_{\#} \times \Theta) = p(A \mid E) .$$

To judge that (GR) should hold is to judge that the total impact of the occurrence of the Ω- event $E_{\#}$ is to preclude the occurrence of any $\theta \in E^c$, and that, within E, p remains operative in the assessment of relative uncertainties.

● In general , the set **Q** contains infinitely many probability measures (exceptions: the cases where $\omega_1 \neq \omega_2$ implies that $T(\omega_1) \neq T(\omega_2)$, or that $|T(\omega_1)| = |T(\omega_2)| = 1$, in which case **Q** consists of a single pm. BUT

Theorem 6.2. For every Q ϵ **Q**, the marginalization Q_Θ of Q to Θ is *identically equal* to q, defined for all $A \subset \Theta$ by

(GJC) $q(A) = \sum_{E \in \mathbf{E}} m(E)\, p(A \mid E)$

REMARK : Formula (GJC) has much wider applicability. Suppose $c: 2^\Theta \rightarrow [0,1]$ $c(\emptyset) = 0$, $c(\Theta) = 1$, and m_c is the Möbius transform of c, i.e.,

$$m_c(E) = \sum_{H \subset E} (-1)^{|E \setminus H|}\, c(H).$$

● Let **E** = $\{E \subset \Theta: m_c(E) \neq 0\}$. If p is any pm on 2^Θ such that $p(E) > 0$ for all $E \in$ **E**, let

$$q(A) := \sum_{E \in \mathbf{E}} m_c(E)\, p(A|E).$$

(note that q is a special kind of smear of m_c, where $p_E = p(\ \ |E)$ for a *single* p)

Theorem 6.3. If c is monotone, then q is a pm on 2^Θ. If c is 2-monotone, then also $q \geq c$. (See Sundberg & W, *J. Theoretical Prob.* **5** (1992), 159-167.)

Remark. Each $q \geq c$ is said to belong to the *core* of the *cooperative game* with *characteristic function* c. One member of the core of c, the *Shapley value*

$$q_s(\theta) = \sum_{\theta \in A \subset \Theta} |A-\theta|!\ |\Theta-A|!\ [c(A)-c(A-\theta)] \ /\ |\Theta|!$$

arises from the uniform pm p.

7. Recovering JC within GJC

The following theorems show how JC arises within GJC:

Theorem 7.1. If the set **E** of focal events consists of *pairwise disjoint* subsets of Θ, and q is a pm on 2^{Θ}, then $q(E) \geq \beta(E)$ for all $E \subset \Theta$ iff $q(E) = m(E)$ for all $E \in$ **E** .

Theorem 7.2. With **E** as above, stipulating of a pm q on 2^{Θ} that $q = Q_{\Theta}$ for some pm Q on $2^{\Omega x \Theta}$ compatible with u and T, and satisfying (GR) is equivalent to stipulating that

$q(E) = m(E)$ for all $E \in$ **E** , and that

$q(A|E) = p(A|E)$ for all $A \subset \Theta$ and all $E \in$ **E**.

8. Example: the amateur linguist

Is your fellow passenger

θ_1 = cn = a Catholic northerner ,
θ_2 = cs = a Catholic southerner ,
θ_3 = pn = a Protestant northerner , or
θ_4 = ps = a Protestant southerner ?

Your prior p : p(cn) = 0.2; p(cs) = 0.3; p(pn) = 0.4; p(ps) = 0.1

He utters indistinctly either

ω_1 = tc = a traditional Catholic piety,
ω_2 = ap = an anti-Protestant epithet
ω_3 = sr = a southern regionalism, or
ω_4 = us = a universal slang expression.

Your subjective probabilities of the above are: u(tc)=.4; u(ap)=.3; u(sr)=.2; u(us) = .1

Contingency tables for Q's compatible with u and (the obvious) T:

	cn	cs	pn	ps	
tc	x	.4-x	0	0	.4
ap	y	.3-y	0	0	.3
sr	0	z	0	0.2-z	.2
us	t	u	v	w	.1
					1.0

Condition (GR):

1. The class of utterances {tc, ap}, taken as a whole, provides no information (differing from that incorporated in p) regarding a catholic's geographical origins.

2. {sr} is similarly uninformative regarding a southerner's religion.

3. {ue} is similarly uninformative regarding any nonempty, proper subset of Θ = {cn, cs, pn, ps}.

You judge that (GR) is the case. Contingency tables of Q's compatible with u and T and satisfying (GR):

	cn	cs	pn	ps	
tc	x	.4-x	0	0	.4
ap	.28-x	x- .02	0	0	.3
sr	0	.15	0	.05	.2
us	.02	.03	.04	.01	.1
	.3	.6	.04	.06	1.0

So, e.g., q(N) = .34, where N ={cn,pn}.

Derivation of q(N) using (GJC):
S = {cs,ps} = southerner
C = {cn, cs} =catholic
P = {pn, ps} = protestant

with **E** = { C, S, Θ } . Then

q(N) = m(C)p(N|C) + m(S)p(N|S)
+ m(Θ) p(N|Θ) =
(0.7)(0.4) + (0.2)(0) + (0.1)(0.6) = 0.34 .

APPENDIX:

9. Mechanical updating

THEOREM. Let p be a positive discrete probability measure on 2^{Θ} , with **E** a countable family of pairwise disjoint events. Given a family $\{ u_E : E \in \mathbf{E} \}$ with each $u_E > 0$, and $\sum_{E \in \mathbf{E}} u_E = 1$, the unique $q \in \{ q \text{ on } 2^{\Theta} : q(E) = u_E \text{ for all } E \in \mathbf{E} \}$ that minimizes the Kullback-Leibler divergence,

$$I(q,p) := \sum_{\theta \in \Theta} q(\theta) \log (q(\theta) / p(\theta)) ,$$

is given by JC.

(P.M. Williams, *British J. Phil. Sci.* **31** (1980), 131-144)

Call updating p to q by minimizing I(q,p) *minimum relative entropy* (MRE) *updating.*

Comment:

(1) Since MRE leads to

$$q(A) = \sum_{E \in \mathbf{E}} u_E \, p(A \mid E)$$

and q(A|E) = p(A|E) for all for all $A \subset \Theta$ and $E \in \mathbf{E}$ (Rigidity), the "mechanical" user of MRE is *implicitly assuming rigidity*. If that assumption is questionable, so is the result of MRE.

(2) But if a potential user of MRE makes the explicit and considered judgment that rigidity is reasonable, then she can use JC directly, and so MRE is superfluous.

Can GJC be derived, even mechanically, by MRE ?

- If you attempt to use MRE to update p to a posterior q that dominates β, and minimizes I(q,p), then, if p itself dominates β, MRE will spit back p, even if you have made the considered judgment that generalized rigidity is reasonably assumed, and can thus apply GJC directly.

10. A Family of Generalizations

A template for creating numerous (formal) generalizations of JC(and GJC):

1. Let b and β be two infinitely monotone capacities on 2^{Θ}, with m = the Möbius transform of β, and **E**={E⊂Θ: m(E) > 0}.

2. Choose *any* "conditional-like" function b(A,E) of b, A⊂ Θ, and E ϵ **E** such that b(A,E) = p(A|E) = p(A∩E)/p(E) when b = p, a pm on 2^{Θ}.

3. For all A ⊂ Θ, set

$$\beta \square b(A) = \Sigma_{E \in \mathbf{E}} m(E)\, b(A,E).$$

When b = a pm p, this yields GJC, and when members of **E** are pairwise disjoint as well, it yields JC.

Some candidates for b(A,E) proposed by Ichihashi and Tanaka (*Int'l J. Approx. Reasoning* **3** (1991),143-156.)

(1) $b_1(A,E) = b(A \cap E)/b(E)$ (= $b_{SZ}(A|E)$)

(2) $b_2(A,E) = [b(A)- b(A \cap E^c)] / [1-b(E^c)]$

(Note: it is possible that $b_2(E,E) < 1$!)

(3) $b_3(A,E) =[b(A \cup E^c)- b(E^c)] / [1-b(E^c)]$

(= $b_{LD}(A|E)$)

● Shafer (*Phil. Sci.* **48** (1981), 337-362) shows that when members of **E**, the family of focal events of β, are *pairwise disjoint*, then

$$\Sigma_{E\in \mathbf{E}}\, m(E)\, b_{LD}(A|E) = \beta \,\Delta\, b\,(A) ,$$

where Δ combines β and b by Dempster's rule.

Another obvious possibility:

$b_B(A|E) = \inf \{p(A|E): p \geq b, \text{ with } p(E)>0\}$

$= b(A \cap E) / [b(A \cap E) + 1 - b(A \cup E^c)]$

All of the above, inserted in the template

$\beta \,\square\, b\,(A) = \Sigma_{E \in \mathbf{E}}\, m(E)\, b(A,E)$

provide *merely formal generalizations* of JC, with no scenarios delineated in which they might arise, and no criteria offered for when they are applicable.