Notes on Savages Foundations of Statistics

Conor Mayo-Wilson

April 21, 2014

1 Purpose

The purpose of this document is to provide quick and detailed proofs of theorems that are omitted from [Savage, 1972]. In a few cases, I have adopted notational conventions that are not adopted by Savage. This is done to make the statements of Savage's axioms more perspicuous and to make proofs more mechanical.

2 Definitions

Let S be a set called **states** and C a set called **consequences**. A set of states E is called an **event**. Given an event E, let $\neg E$ denote its complement.

An **act** is a function $f : S \to C$. Let $\preceq$ denote a binary relation on A. The relation $\preceq$ is intended to represent preference. That is, the interpretation of $f \preceq g$ is that f is not preferred to g. Several axioms restricting the interpretation of $\preceq$ are introduced in the next section. Say one is **indifferent** between two acts f and g if and only if $f \preceq g$ and $g \preceq f$. In this case, write $f \approx g$. Say f is **strictly preferred to** g if and only if $f \preceq g$ and $g \not\preceq f$, and in this case write[1] $f \prec g$.

For any consequence $c \in C$, let $\widetilde{c}$ denote the function $\widetilde{c}(s) = c$ for all $s \in S$. The function $\widetilde{c}$ will be called a **constant act**. The ordering $\preceq$ on actions, therefore, induces an ordering $\trianglelefteq$ on C as follows. Let $c, d \in C$ be consequences. Then write $c \trianglelefteq d$ if and only if $\widetilde{c} \preceq \widetilde{d}$.

Given an event E and two actions f and g, say that f **agrees with** g **on** E if the restriction $f \upharpoonright E$ of f to E is equal to $g \upharpoonright E$. In this case, write $f =_E g$.

[1] Note [Savage, 1972] defines $f \prec g$ to hold precisely if and only if $g \not\preceq f$. His definition is equivalent to the one in these notes under assumption P1 as, if $g \not\preceq f$, then $f \preceq g$ by totality of $\preceq$.

In Chapter 2.7, Savage defines the notion of f not being preferred to g on condition that event E obtains. His definition occurs mostly in prose, and so for the sake of clarity, I introduce a bit of notation to make the proofs below easier. For any two acts f and g and any event E, let f_E^g denote the action such that for all $s \in S$:

$$f_E^g(s) := \begin{cases} f(s) \text{ if } s \in E \\ g(s) \text{ if } s \notin E \end{cases}$$

Say the action f **is not preferred to** g **given** E if and only if $f_E^g \preceq g$. In this case, write $f \preceq_E g$. Define $f \prec_E g$ to hold if $f \preceq_E g$ and $g \not\preceq_E f$. Write $f \approx_E g$ if $f \preceq_E g$ and $g \preceq_E f$.

An event E is called **null** if and only if $f \approx_E g$ if for all actions f and g. Null events are intended to represent those to which one assigns essentially no likelihood of occurring, and so, one is completely indifferent among all available actions on the condition that a null event occurs.

3 Savage's Axioms

The following axioms are employed throughout Savage's work. P1 is stated on page 18; P2 on page 23; P3 on page 26; P4 and P5 on page 31, and P6' on page 38.

P1: $\preceq$ is a **simple ordering** on A, or in other words:

- **Transitivity:** For all $f, g, h \in A$:

$$f \preceq g \text{ and } g \preceq h \Rightarrow f \preceq h$$

- **Totality:** For all $f, g \in A$, either $f \preceq g$ or $g \preceq f$ or both.

Note that $\preceq$ is also reflexive (i.e., that $f \preceq f$ for all $f \in A$), as the totality of the relation $\preceq$ entail that either $f \preceq f$ or $f \preceq f$.

P2: For any event E and any four acts f, f', g and g':

- $f =_E f'$ and $g =_E g'$
- $f =_{\neg E} g$ and $f' =_{\neg E} g'$
- $f \preceq g$

together entail that

$$f' \preceq g'$$

P2 is also called the "Sure-Thing Principle."

P3: For any two consequences c and d and any non null event E:

$$\widetilde{c} \preceq_E \widetilde{d} \Leftrightarrow \widetilde{c} \preceq \widetilde{d}$$

P4: For all consequences $a, b, x, y \in C$ and all events E, F, if

1. $a \prec b$ and $x \prec y$,
2. $\widetilde{b}_E^{\widetilde{a}} \preceq \widetilde{b}_F^{\widetilde{a}}$

Then $\widetilde{y}_E^{\widetilde{x}} \preceq \widetilde{y}_F^{\widetilde{x}}$.

P5: There exist at least one pair of consequences c and d such that $c \lhd d$.

4 Theorems

Theorem 1

1. *$\emptyset$ is null.*
2. *E is null if and only if $f \preceq_E g$ for all actions f and g.*
3. *If E is null and $F \subseteq E$, then F is null.*
4. *If $\neg E$ is null, then*

   $$f \preceq_E g \text{ if and only if } f \preceq g$$

5. *$f \preceq_S g$ if and only if $f \preceq g$, and*
6. *If S is null, then $f \approx g$ if for all actions f and g.*

Proof:

1. For all actions g, one has that $g \preceq g$ by reflexivity of $\preceq$ (i.e., Condition 1 of P1). By definition $f_\emptyset^g = g$, and so it follows that $f_\emptyset^g \preceq g$ for all actions f and g. Thus, $f \preceq_\emptyset g$ by the definition of $\preceq_E$, and hence, $\emptyset$ is null by the definition of "null."

2. Follows immediately from definitions.

3. Suppose $F \subseteq E$. Let f and g be arbitrary acts and E be any event. I claim that $(f_F^g)_E^g = f_F^g$. Why? If $s \in E$, then it immediately follows that $(f_F^g)_E^g(s) = f_F^g(s)$. If $s \notin E$, then $(f_F^g)_E^g = g(s)$ by definition. Moreover, as $F \subseteq E$, it follows that $s \notin F$. Hence, $f_F^g(s) = g(s)$ by definition. So $(f_F^g)_E^g(s) = f_F^g(s)$ as desired.

   As E is null, by the second part of this theorem, it follows that $(f_F^g) \preceq_E g$. In other words:
   $$(f_F^g)_E^g \preceq g.$$
   Since $(f_F^g)_E^g = f_F^g$, it follows that
   $$f_F^g \preceq g.$$
   By definition of $\preceq_F$, this entails that
   $$f \preceq_F g$$
   As f, g and E were arbitrarily chosen, we have shown that, if E is null and $F \subseteq E$, then $f \preceq_F g$ for all actions f and g. Again by the second part of the theorem, it follows that F is null.

4. Assume $\neg E$ is null. We first show that if $f \preceq_E g$, then $f \preceq g$. To do so, note that

   - $f_{\neg E}^g =_{\neg E} f$ and $g =_{\neg E} f_E^g$,
   - $f_{\neg E}^g =_{\neg\neg E} g$ and $f =_{\neg\neg E} f_E^g$, and
   - $f_{\neg E}^g \preceq g$

   where the third assertion follows from the fact that $\neg E$ is null, and so $f \preceq_{\neg E} g$. Applying P2 yileds the conclusion that $f \preceq f_E^g$. By transitivity of $\preceq$ and the fact that $f_E^g \preceq g$, the conclusion follows.

   In the reversion direction, suppose that $f \preceq g$.

   - $f_{\neg E}^g =_E g$ and $f =_E f_E^g$,
   - $f_{\neg E}^g =_{\neg E} f$ and $g =_{\neg E} f_E^g$, and
   - $f_{\neg E}^g \preceq f$

   where the third assertion follows from the fact that $\neg E$ is null. Applying P2 yields that $f_E^g \preceq g$, or in other words, that $f \preceq_E g$.

5. Follows from Parts 1 and 4, as $\emptyset = \neg S$ is null.

6. If S is null, then $f \approx_S g$ for all f and g. By Part 5, it follows that $f \approx g$ for all f and g.

□

The following lemma is the inductive step to Theorem 2 on page 24. It is the formal result that motivates calling P2 the "Sure Thing" principle. The result entails, for example, that if I prefer to reading to jogging when it's cloudy outside, and if I prefer reading to jogging when it's not cloudy outside, then I prefer to reading to jogging without qualification.

Lemma 1 *Suppose P1 and P2 and let E be any event. If $f \preceq_E g$ and $f \preceq_{\neg E} g$, then $f \preceq g$. If in addition, $f \prec_E g$, then $f \prec g$.*

Proof: By definition of $f \preceq_E g$, we know that $f^g_E \preceq g$. Similarly, $f^g_{\neg E} \preceq g$. It suffices to show that $f \preceq f^g_E$ because, by P1, the relation $\preceq$ is transitive (and hence $f \preceq f^g_E$ and $f^g_E \preceq g$ together entail that $f \preceq g$).

Notice that

- $f^g_{\neg E} =_E g$ and $f =_E f^g_E$,
- $f^g_{\neg E} =_{\neg E} f$ and $g =_{\neg E} f^g_E$, and
- $f^g_{\neg E} \preceq g$

So by P2, it follows that $f \preceq f^g_B$ as desired.

For the second part of the theorem, note that $f \preceq f^g_E$ and $f^g_E \prec g$ immediately entail that $f \prec g$.

□

5 Qualitative Personal Probability

Given events E and F, write $E \leq F$ if and only if for all consequences $c, d \in C$:

$$c \lhd d \Rightarrow \widetilde{\widetilde{d^c_E}} \preceq \widetilde{\widetilde{d^c_F}}$$

In this case, say E is **not more probable** than F. Write $E < F$ if and only if $E \leq F$ and $F \not\leq E$.

A binary relation $\sqsubset$ between events is called a **qualitative probability** if and only if for all events E, F, and G, the following three conditions hold.

- $\sqsubseteq$ is a simple ordering.
- If $E \cap G = F \cap G = \emptyset$, then

$$E \sqsubseteq F \Leftrightarrow E \cup G \sqsubseteq F \cup G.$$

- $\emptyset \sqsubseteq E$ and $E \sqsubset S$.

Theorem 2 *P1-P5 together entail that the relation $\leq$ is a qualitative probability.*

Proof:

1. To show that $\leq$ is a neg ple ordering, we must show that it is reflexive, transitive, and total. By P1, $\preceq$ is reflexive, transitive, and total.

   Reflexivity: By definition, $E \leq E$ if and only if $\widetilde{d_E^c} \preceq \widetilde{d_E^c}$ for all consequences $c, d \in C$ such that $c \lhd d$. The latter is true because $\preceq$ is reflexive by P1.

   Transitivity: Suppose $E \leq F$ and $F \leq G$. We want to show that $E \leq G$. So let $c, d \in C$ be consequences such that $c \lhd d$. As $E \leq F$, it follows that $\widetilde{d_E^c} \preceq \widetilde{d_F^c}$, and as $F \leq G$, it follows that $\widetilde{d_F^c} \preceq \widetilde{d_G^c}$. By the transitivity of $\preceq$, it follows that

   $$\widetilde{d_E^c} \preceq \widetilde{d_G^c}$$

   and so $E \leq G$ as desired.

   Totality: Finally, we want to show that $\leq$ is total. Let E and F be given. By P5, there are two consequences c, d such that $c \prec d$. By the totality of $\preceq$ on actions, it follows that either:

   $$\widetilde{d_E^c} \preceq \widetilde{d_F^c} \text{ or } \widetilde{d_F^c} \preceq \widetilde{d_E^c}$$

   Without loss of generality, assume that $\widetilde{d_E^c} \preceq \widetilde{d_F^c}$. Now let $x, y \in C$ be any constants such that $x \lhd y$. As (i) $c \lhd d$, (ii) $\widetilde{d_E^c} \preceq \widetilde{d_F^c}$, and (iii) $x \lhd y$, it follows from P4 that $\widetilde{x_E^y} \preceq \widetilde{x_F^y}$. As x and y were chosen arbitrarily, it follows from the definition of $\leq$ that $E \leq F$.

2. Next, we must show that if $E \cap G = F \cap G = \emptyset$, then

   $$E \leq F \Leftrightarrow E \cup G \leq F \cup G.$$

In the left to right direction, assume that $E \leq F$ and that $E \cap G = F \cap G = \emptyset$. The proof employs P2, i.e., the Sure-Thing principle.

First, note that (i) $\widetilde{d^c_{E\cup G}}$ agrees with $\widetilde{d^c_F}$ over $\neg(E \cup F)$. Why? If $s \notin E \cup F$, then there are two cases to consider. If $s \notin G$, then s is not an element of either $E \cup G$ or $F \cup G$ and hence,

$$\widetilde{d^c_{E\cup G}}(s) = c \text{ and } \widetilde{d^c_{F\cup G}}(s) = c.$$

On the other hand, if $s \in G$, then

$$\widetilde{d^c_{E\cup G}}(s) = d \text{ and } \widetilde{d^c_{F\cup G}}(s) = d.$$

In either case, $\widetilde{d^c_{E\cup G}}(s) = \widetilde{d^c_{F\cup G}}(s)$ as desired.

Next, note that (ii) $\widetilde{d^c_E}$ agrees with $\widetilde{d^c_F}$ over $\neg(E \cup F)$, as both are identically c on $\neg(E \cup F)$.

Third, note that, (iii) $\widetilde{d^c_{E\cup G}}$ agrees with $\widetilde{d^c_E}$ over $E \cup F$. Why? If $s \in E \cup F$, either $s \in E$ or $s \in F \setminus E$. In the former case, both $\widetilde{d^c_{E\cup G}}(s) = d$ and $\widetilde{d^c_E} = d$. In the latter case, note that $F \cap G = \emptyset$. Hence, it follows that if $s \in F \setminus E$, then s is neither an element of E nor G. From this it follows that $\widetilde{d^c_{E\cup G}}(s) = c$ and $\widetilde{d^c_E} = c$.

By analogous reasoning, it follows that (iv) $\widetilde{d^c_{F\cup G}}$ agrees with $\widetilde{d^c_F}$ over $E \cup F$.

Finally, note that because $E \leq F$, we have that $\widetilde{d^c_E} \preceq \widetilde{d^c_F}$ (by definition of $\leq$). Putting (i)-(v) together, we have shown that

- $\widetilde{d^c_E} =_{E\cup F} \widetilde{d^c_{E\cup G}}$ and $\widetilde{d^c_F} =_{E\cup F} \widetilde{d^c_{F\cup G}}$,
- $\widetilde{d^c_{E\cup G}} =_{\neg(E\cup F)} \widetilde{d^c_{F\cup G}}$ and $\widetilde{d^c_E} =_{\neg(E\cup F)} \widetilde{d^c_F}$, and
- $\widetilde{d^c_E} \preceq \widetilde{d^c_F}$

By P2, we obtain that $\widetilde{d^c_{E\cup G}} \preceq \widetilde{d^c_{F\cup G}}$. By definition of $\leq$, this entails that $E \cup G \leq F \cup G$ as desired.

In the reverse direction, suppose that $E \cup G \leq F \cup G$ and $E \cap G = F \cap G = \emptyset$. We want to show that $E \leq F$. This follows from the exact same reasoning as in the left to right direction, except one uses the fact that $E \cup G \leq F \cup G$ to instantiate the third premise of P2.

3. Next, we must show that $\emptyset \leq E$ and $\emptyset < S$ for all events E. In the former case, this amounts to showing that if $c \lhd d$, then $\widetilde{d^c_\emptyset} \preceq \widetilde{d^c_E}$. Now, note that $\widetilde{d^c_\emptyset} = \widetilde{c}$. So we must show that $\widetilde{c} \preceq \widetilde{d^c_E}$, or in other

words, that $\widetilde{c} \preceq_E \widetilde{d}$. If E is null, then $\widetilde{c} \preceq_E \widetilde{d}$ by definition of null. If E is not null, then because $\widetilde{c} \preceq \widetilde{d}$, by P3 it follows that $\widetilde{c} \preceq_E \widetilde{d}$, as desired.

Finally, we must show that $\emptyset < S$. So we must show that $S \not\leq \emptyset$. Suppose for the sake of contradiction that $S \leq \emptyset$. By P5, there are consequences c, d such that $c \lhd d$. As $S \leq \emptyset$, it follows that $\widetilde{d}_S^{\widetilde{c}} \preceq \widetilde{d}_\emptyset^{\widetilde{c}}$. Now $\widetilde{d}_S^{\widetilde{c}} = \widetilde{d}$ and $\widetilde{d}_\emptyset^{\widetilde{c}} = \widetilde{c}$. So it follows that $\widetilde{d} \preceq \widetilde{c}$. By definition of preference among consequences, it follows that $d \unlhd c$, contradicting the assumption that $c \lhd d$.

□

References

L. J. Savage. *The foundation of statistics*. Dover publications, 1972.