Stochastic Expected Utility for Binary Choice

A 'Modular' Axiomatic Foundation

Matthew Ryan

Auckland University of Technology

26 July 2019

Probabilistic choice

- It is routine to observe experimental subjects making **different** choices in successive presentations of the **same** binary choice problem (Mosteller and Nogee, 1951)
  - Indifference? Deliberate randomisation? (Machina, 1985; Agranov and Ortoleva, 2016; Cerreia-Vioglio *et al*., 2019)
- Mathematical psychologists have long accepted that choice is a fundamentally probabilistic process
  - An assertion about preference is not "the record of a particular observation [...] but is a theoretical assertion inferred from data and subject to errors of inference" (Suppes *et al*., 1989, p.300)
  - Probabilistic choice may be characterised as inconstancy of "preference" (*multiple utility* models) or random imperfections in "utility maximisation" (*single utility* models)
  - **We focus on single utility models here**

Axiomatic foundations

- Various classes of single utility models have been proposed to describe probabilistic choice
- The best known are those in the *Fechnerian* family in which (binary) choice probabilities depend on *utility differences*
  - There is a utility function, u, such that probability of choosing option α over option β is a non-decreasing function of $u(\alpha) - u(\beta)$
  - The familiar (binary) logit model – also known as the *Luce* or *strict utility* model – is in the Fechnerian class
- Some single utility models have yielded simple and intuitive axiomatisations; others have proved more resistant

Axiomatic foundations

- When axiomatising *deterministic* models of choice, it is well known that *mixture set* domains may simplify the task: convex structure and independence conditions are powerful tools
- Similar benefits are known to accrue for multiple utility models of probabilistic choice (Gul and Pesendorfer, 2006; Wu, 2018): intuitive independence conditions replace complex Block-Marschak inequalities
- Analogous benefits exist for some Fechnerian models (e.g., Blavatskyy, 2008; Dagsvik, 2008; Ryan, 2018a) but the picture is less tidy:
  - Very different axiomatisations of very similar models; and
  - Imperfect axiomatic separation of "linearity" of u (i.e., expected utility preferences) and "linearity" of Fechnerian noise
- **We provide a unified axiomatisation that separates the two types of linearity** (with ancillary benefits for experimental testing of the "EU maximisation with Fechnerian noise" hypothesis)

Binary choice probabilities for general domains

- Let A be a set of alternatives
- A given decision maker (DM) is characterised by a set of choice probabilities (stochastic choice function)
  - We focus on *binary* choices; choices from binary subsets of A
- For any $\{a, b\} \subseteq A$ with $a \neq b$, let $P(a, b)$ denote the probability that alternative a is chosen from the binary choice set $\{a, b\}$
- There is no "outside option" so P must satisfy the *completeness* (or *balance*) condition: for any $a, b \in A$,

$$P(a, b) + P(b, a) = 1 \tag{C}$$

- Note that (C) implies $P(a, a) = \frac{1}{2}$ for any $a \in A$, but no behavioural meaning attaches to these "diagonal" terms

Binary choice probabilities for general domains

Definition

A **binary choice probability (BCP)** is a mapping $P : A \times A \longrightarrow [0, 1]$ that satisfies (C)

- If P is a BCP, its associated "base relation" $\succsim^P \subseteq A \times A$ is defined as follows:

$$a \succsim^P b \quad \Leftrightarrow \quad P(a, b) \geq \frac{1}{2} \quad \overset{(C)}{\Leftrightarrow} \quad P(a, b) \geq P(b, a)$$

- A function $u : A \longrightarrow \mathbb{R}$ is a *weak utility* for P iff the following holds for any $a, b \in A$:

$$P(a, b) \geq \frac{1}{2} \quad \Leftrightarrow \quad u(a) \geq u(b)$$

Thus, u is a weak utility for P iff it represents the base relation, $\succsim^P$

Single utility models

- We next describe a range of single utility models for BCPs.
- Each model will characterise choice as the noisy maximisation of some weak utility function
- In other words, choice is the noisy expression of "preferences" described by the base relation
- For each model, we describe what is known about its axiomatisation: necessary and sufficient conditions on P

Model S1: Strict scalability

Definition (Ryan, 2018b)

Binary choice probability P is *strictly scalable* if there exists a pair (u, F) such that

$$P(a, b) = F(u(a), u(b))$$

for any $a, b \in A$, where $u : A \to \mathbb{R}$ is a weak utility for P and $F : \mathbb{R}^2 \to [0, 1]$ is non-decreasing (respectively, non-increasing) in its first (respectively, second) argument and satisfies

$$F(x, y) + F(y, x) = 1 \tag{C^*}$$

for all $x, y \in \mathbb{R}$. We say that P is *strictly scalable* through (u, F).

- Strict scalability means that:
  1. alternatives are evaluated on a unidimensional, context-independent "scale"; and
  2. choice may be characterised as noisy weak utility maximisation

Strict scalability and stochastic transitivity

Strong Stochastic Transitivity (SST): *For any $a, b, c \in A$,*

$$\min\left\{P\left(a,b\right),P\left(b,c\right)\right\}\geq\frac{1}{2}\quad\Rightarrow\quad P\left(a,c\right)\geq\max\left\{P\left(a,b\right),P\left(b,c\right)\right\}$$

- Note that SST implies the transitivity of $\succsim^{P}$ (WST); the base relation is also complete by construction.

Theorem (Ryan, 2018b)

A binary choice probability is strictly scalable iff it satisfies SST and a weak utility exists.

- If A is countable, SST is necessary and sufficient
- There is also a multinomial version of this theorem, involving suitably generalised notions of strict scalability and SST to choice sets of any finite size (Ryan, 2018b)

Model S2: Simple scalability

Definition (Tversky and Russo, 1969)

Binary choice probability P is *simply scalable* iff there exists a pair (u, F) such that

$$P(a, b) = F(u(a), u(b))$$

for any $a, b \in A$, where $F : \mathbb{R}^2 \to [0, 1]$ is strictly increasing (respectively, strictly decreasing) in its first (respectively, second) argument and satisfies (C^*) for all $x, y \in \mathbb{R}$. We say that P is *simply scalable* through (u, F).

- If P is simply scalable by (u, F) then u is a weak utility for P:

$$\begin{aligned} P(a, b) \geq P(b, a) \quad &\Leftrightarrow \quad F(u(a), u(b)) \geq F(u(b), u(a)) \\ &\Leftrightarrow \quad u(a) \geq u(b) \end{aligned}$$

  Hence, if P is simply scalable then it is also strictly scalable
- Simple scalability restricts the $P = 0$ and $P = 1$ contours

Simple scalability and stochastic transitivity

"Stronger" Stochastic Transitivity (SSST): *For any* $a, b, c \in A$*,*

$$\min\left\{P(a,b), P(b,c)\right\} \geq [>] \frac{1}{2} \;\Rightarrow\; P(a,c) \geq [>] \max\left\{P(a,b), P(b,c)\right\}$$

Theorem (Tversky and Russo, 1969)

A binary choice probability is simply scalable iff it satisfies SSST.

- SSST ensures the existence of a weak utility: fixing some $e \in A$, SSST implies (albeit not obviously) that, for any $a, b \in A$,

$$a \succsim^P b \quad \Leftrightarrow \quad P(a,e) \geq P(b,e)$$

  Hence $u(\cdot) \equiv P(\cdot, e)$ is a weak utility function for P
- Once again, there is a multinomial version: Tversky (1972)

From scalability to Fechnerian structure: Models F1 and F2

Definition

If P is strictly (respectively, simply) scalable by (u, F) and F depends only on utility differences (i.e., $F(x, y) = F(x', y')$ whenever $x - y = x' - y'$) then we have a **strict** (respectively, **strong**) **Fechner model** for P. In this case:

$$P(a, b) = G(u(a) - u(b))$$

for any $a, b \in A$, where u is a weak utility for P and G is non-decreasing (respectively, strictly increasing) on its domain $\Gamma = u(A) - u(A)$ and satisfies $G(x) + G(-x) = 1$ for any $x \in \Gamma$.

- Of course, a strong Fechner model (F2) is a strict Fechner model (F1)
- We return to the axiomatisation of F1 and F2 after a brief detour...

Adding richness and continuity

- In "rich" domains (such as the domain of lotteries in the context of risk), it is sometimes useful to focus on models with some additional structure:
  1. weak utilities that have interval range, and
  2. transformations of weak utilities into choice probabilities that are continuous (i.e., continuity of G or F)
- Let us refer to these as models S1*, S2*, F1* and F2*

Axiomatic foundations for Fechner models

- Strict Fechner models on general domains have yet to be characterised
- Various characterisations of strong Fechner models (F2 or F2*) exist, but only for specific domain restrictions
  - The case of F2 with finite domain (i.e., finite A) is especially problematic (Scott, 1964)
  - Debreu (1958) considered BCP's that satisfy **solvability**: for any $a, b, c \in A$ and any $\rho \in (0, 1)$

$$P(a, b) \geq \rho \geq P(a, c) \quad \Rightarrow \quad P(a, d) = \rho \text{ for some } d \in A$$

    - Solvability implies a "rich" domain (excluding trivial cases)
    - Debreu showed, using results from topology, that solvability and the *quadruple condition (QC)* suffice for the existence of a model within the F2* class...

Axiomatic foundations for Fechner models

Quadruple Condition (QC): *For any $a, b, a', b' \in A$,*

$$P(a, b) \geq P(a', b') \quad \Leftrightarrow \quad P(a, a') \geq P(b, b')$$

- QC implies SSST but not conversely
- Some refinement of Debreu's result was obtained by Doignon and Falmagne (1974), who showed that QC can be replaced by conditions intermediate in strength between QC and SSST
- Further refinements were shown by Köbberling (2006), who proved, in particular, that the Solvability (richness) restriction can be somewhat relaxed without sacrificing F2 structure

Solvability and scalability

- Axiomatic foundations for models S1* and S2* have not, to my knowledge, been studied previously
- However, it is not hard to show that:

Theorem

A binary choice probability has a model of type S2* iff it satisfies SSST plus solvability

The state of play (general domains)...

	1	**2**	$\mathbf{1}^*$	$\mathbf{2}^*$
S	SST $+$ weak utility	SSST	**?**	SSST $+$ solvability
F	**?**	Scott ('64) Köbberling ('06) etc.	**?**	Debreu ('58) Doignon/Falmagne ('74) etc.

Mixture set domains

- Let A be a mixture set
  - If $a, b \in A$ and $\lambda \in [0, 1]$ then $a\lambda b \in A$ is the λ-mixture of a and b
  - In particular, for any $a, b \in A$ and any $\lambda, \mu \in [0, 1]$:
    - $a1b = a$
    - $a\lambda b = b\,(1-\lambda)\,a$
    - $a\lambda\,(a\mu b) = a\,(\mu\lambda)\,b$
  - Examples include spaces of lotteries, Anscombe-Aumann acts, etc.
- A function $u : A \longrightarrow \mathbb{R}$ is *mixture-linear* if, for any $a, b \in A$ and any $\lambda \in [0, 1]$,

$$u\,(a\lambda b) = \lambda u\,(a) + (1-\lambda)\,u\,(b)$$

  - It is easy to show that $u\,(A)$ is an interval when u is mixture-linear

Mixture set domains

- When A is a mixture set it is useful to characterise models with mixture-linear weak utilities (EU form)
  - Hence, we define models MS1, MS1*, MF1, MF1* and so forth
    - For the "$*$" models, only continuity of G or F is salient since $u(A)$ is necessarily an interval
- There has been a limited amount of work on the axiomatic foundations of these models
  - Blavatskyy (2008) characterises MF1
  - Dagsvik (2008) characterises MF2*
    - Each author takes A to be the set of lotteries over a fixed, finite set of prizes (i.e., the unit simplex in $\mathbb{R}^n$)

Blavatskyy (2008)

Continuity (CT): *For any* $a, b, c \in A$ *the following two sets are closed:*

$$\left\{ \lambda \in [0,1] \;\middle|\; P\left(a\lambda b, c\right) \geq \frac{1}{2} \right\}$$

$$\left\{ \lambda \in [0,1] \;\middle|\; P\left(a\lambda b, c\right) \leq \frac{1}{2} \right\}$$

Common Consequence Independence (CCI): *For any* $a, b, c, d \in A$ *and any* $\lambda \in [0,1]$,

$$P\left(a\lambda c, b\lambda c\right) = P\left(a\lambda d, b\lambda d\right)$$

Theorem (Blavatskyy, 2008 [as modified by Ryan, 2015])

A binary choice probability has a model of type MF1 iff it satisfies SST, CT and CCI.

Dagsvik (2008)

Archimedean property: *For any* $a, b, c \in A$*:*

$$P(a,b) > \frac{1}{2} > P(c,b) \quad \Rightarrow \quad P(a\lambda c, b) > \frac{1}{2} > P(a\mu c, b)$$

for some $\lambda, \mu \in (0,1)$

Strong Independence (SI): *For any* $a, b, a', b', c \in A$ *and any* $\lambda \in [0,1]$*:*

$$P(a,b) \geq P\left(a', b'\right) \quad \Rightarrow \quad P(a\lambda c, b\lambda c) \geq P\left(a'\lambda c, b'\lambda c\right)$$

Theorem (Dagsvik, 2008)

A binary choice probability has a model of type MF2 iff it satisfies QC, solvability, the Archimedean property and SI.*

The state of play (mixture set domains)...

	1	**2**	**1^***	**2^***
MS	**?**	**?**	**?**	**?**
MF	SST + CT + CCI	**?**	**?**	QC + solvability + Archimedean + SI

The state of play (mixture set domains)...

- Blavatskyy's proof relies only on the linear algebra familiar from standard EU representation theorems
- Dagsvik, by contrast, appeals to the topological results used by Debreu, plus results on functional equations – a familiar part of the toolkit in mathematical psychology
- Another notable feature of Blavatskyy's result is the fact that CCI ensures *both* mixture-linearity of u *and* the Fechnerian structure of "noise"
  - Continuity and SST together guarantee neither the existence of an MS1 model nor the existence of a model of type F1 (Example 3.1)
- By separating these two aspects of CCI we can obtain intuitive axiomatisations of all eight models

"Decomposing" CCI

CCI: *For any* $a, b, c, d \in A$ *and any* $\lambda \in [0,1]$,

$$P(a\lambda c, b\lambda c) = P(a\lambda d, b\lambda d).$$

CCI.1: *For any* $a, b, c \in A$,

$$P\left(a\frac{1}{2}c, b\frac{1}{2}c\right) > \frac{1}{2} \quad \Rightarrow \quad \min\left\{P\left(a, a\frac{1}{2}b\right), \ P\left(a\frac{1}{2}b, b\right)\right\} > \frac{1}{2}$$

- CCI.1 says that replacing the common consequence c with common consequence a or common consequence b does not affect the base-relation ranking; equivalently:

$$a\frac{1}{2}c \succ^P b\frac{1}{2}c \quad \Rightarrow \quad \left[a\frac{1}{2}a \succ^P b\frac{1}{2}a \text{ and } a\frac{1}{2}b \succ^P b\frac{1}{2}b\right]$$

  - This has the flavour of a "betweenness" restriction (Chew, 1983; Dekel, 1986) on the base relation

"Decomposing" CCI

CCI: *For any* $a, b, c, d \in A$ *and any* $\lambda \in [0, 1]$,

$$P(a\lambda c, b\lambda c) = P(a\lambda d, b\lambda d).$$

CCI.2: *For any* $a, b \in A$ *and any* $\lambda \in [0, 1]$,

$$P(a, a\lambda b) = P(b\lambda a, b)$$

- CCI.2 is equivalent to CCI with the additional requirement that $\{c, d\} \subseteq \{a, b\}$
  - CCI.2 is a mixture-symmetry of "preference intensity" condition in the spirit of the "symmetry" axiom in SSB utility theory (Fishburn, 1984)
- CCI.1 and CCI.2 are jointly *strictly* weaker than CCI (Example 3.2)

New results (mixture set domains)...

	1	**2**	**1^***	**2^***
MS	SST + CT & CCI.1	SSST + CT & CCI.1	SST + CT & CCI.1 + solvability	SSST + CT & CCI.1 + solvability
MF	SST + CT & CCI.1 + CCI.2	SSST + CT & CCI.1 + CCI.2	SST + CT & CCI.1 + CCI.2 + solvability	SSST + CT & CCI.1 + CCI.2 + solvability

New results (mixture set domains)...

	1	**2**	$\mathbf{1}^*$	$\mathbf{2}^*$
MS	**SST + CT & CCI.1**	SSST + CT & CCI.1	SST + CT & CCI.1 + solvability	SSST + CT & CCI.1 + solvability
MF	SST + CT & CCI.1 + CCI.2	SSST + CT & CCI.1 + CCI.2	SST + CT & CCI.1 + CCI.2 + solvability	SSST + CT & CCI.1 + CCI.2 + solvability

New results (mixture set domains)...

	1	**2**	**1***	**2***
MS	SST + CT & CCI.1	SSST + CT & CCI.1	SST + CT & CCI.1 + solvability	SSST + CT & CCI.1 + solvability
MF	SST + CT & CCI.1 + **CCI.2**	SSST + CT & CCI.1 + **CCI.2**	SST + CT & CCI.1 + **CCI.2** + solvability	SSST + CT & CCI.1 + **CCI.2** + solvability

New results (mixture set domains)...

	1	**2**	**1***	**2***
MS	SST + CT & CCI.1	**SSST** + CT & CCI.1	SST + CT & CCI.1 + solvability	**SSST** + CT & CCI.1 + solvability
MF	SST + CT & CCI.1 + CCI.2	**SSST** + CT & CCI.1 + CCI.2	SST + CT & CCI.1 + CCI.2 + solvability	**SSST** + CT & CCI.1 + CCI.2 + solvability

New results (mixture set domains)...

	1	**2**	**1***	**2***
MS	SST + CT & CCI.1	SSST + CT & CCI.1	SST + CT & CCI.1 + **solvability**	SSST + CT & CCI.1 + **solvability**
MF	SST + CT & CCI.1 + CCI.2	SSST + CT & CCI.1 + CCI.2	SST + CT & CCI.1 + CCI.2 + **solvability**	SSST + CT & CCI.1 + CCI.2 + **solvability**

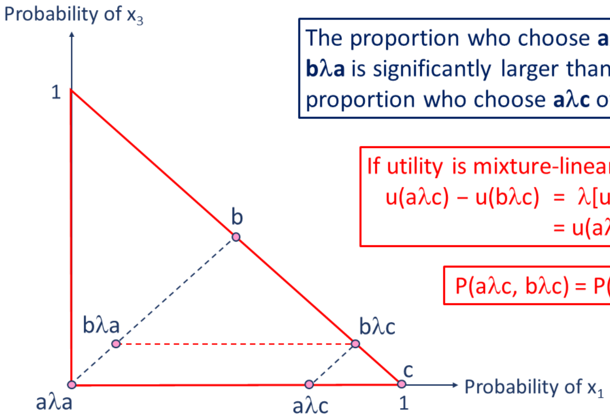
Revisiting Blavatskyy (2008) and Dagsvik (2008)

	1	**2**	**1***	**2***
MS	SST + CT & CCI.1	SSST + CT & CCI.1	SST + CT & CCI.1 + solvability	SSST + CT & CCI.1 + solvability
MF	**SST + CT & CCI.1 + CCI.2**	SSST + CT & CCI.1 + CCI.2	SST + CT & CCI.1 + CCI.2 + solvability	**SSST + CT & CCI.1 + CCI.2 + solvability**

Testing "EU maximisation with Fechnerian noise"

- The "modular" structure of these axiomatisations may also be useful for empirical testing
- For example, CCI is known to be empirically vulnerable...

Testing "EU maximisation with Fechnerian noise"



The proportion who choose **aλa** over **bλa** is significantly larger than the proportion who choose **aλc** over **bλc**

If utility is mixture-linear:

u(aλc) – u(bλc) = λ[u(a) – u(b)]

= u(aλa) – u(bλa)

P(aλc, bλc) = P(aλa, bλa)

CCI and the Common Consequence Effect

Testing "EU maximisation with Fechnerian noise"

- Since tests of CCI are tests of a joint hypothesis, it is hard to interpret this failure: should we blame von Neumann and Morgenstern, or Fechner?
  - If CCI fails then model MF1 is invalid – is the problem with "M" or with "F"?
- If the data were to pass the CCI.2 test, this would point the finger of blame firmly at von Neumann and Morgenstern (under a maintained hypothesis that BCPs are strictly scalable: S1)
  - If CCI fails but CCI.2 holds then CCI.1 cannot hold, which excludes model MS1
- Of course, we could test CCI.1 directly but such tests are empirically challenging
  - CCI and CCI.2 can be tested on data aggregated over heterogeneous individuals – if they hold for all individual choice probabilities then they must hold for average choice proportions – but this is not true of CCI.1 (Ballinger and Wilcox, 1997)