The Generalized Stochastic Preference Choice Model

October 18, 2018

Gerardo Berbeglia*

Abstract. We propose a new discrete choice model that generalizes the random utility model (RUM). We show that this model, called the *Generalized Stochastic Preference* (GSP) model can explain several choice phenomena that can't be represented by a RUM. In particular, the model can easily (and also exactly) replicate some well known examples that are not RUM, as well as controlled choice experiments carried out since 1980's that possess strong regularity violations. One of such regularity violation is the *decoy effect* in which the probability of choosing a product increases when a similar, but inferior product is added to the choice set. An appealing feature of the GSP is that it is non-parametric and therefore it has very high flexibility. The model has also a simple description and interpretation: it builds upon the well known representation of RUM as a stochastic preference, by allowing some additional consumer types to be non-rational.

1. Introduction

Discrete choice models, which have been studied for more than 50 years, are essential to understand and make predictions about the choices made by individuals in different settings. For example, choice models are used to estimate customer purchases in several markets such as retail, air travel, and accommodation. These sales estimates obtained with choice models are a key component of *revenue management*, where the availability of products as well as their prices are optimized to maximize expected profits.

The first choice model used in revenue management was the independent demand (ID) model. Under ID, the probability of purchasing a product is considered to be independent on what other products are being offered. This property makes the model easy to analyze. However, the main disadvantage is that it cannot even explain very simple substitution behaviour such as the additional demand for a product might have for the absence of another product (spilled demand). More sophisticated discrete choice models that can incorporate substitution behaviour have been studied by the revenue management community. Some

---

*Melbourne Business School, University of Melbourne. Email: `g.berbeglia@mbs.edu`

of the most prominent ones are the multinomial logit model (MNL) [19], the nested logit, the latent class MNL, and the Markov chain model [4]. A common feature from these choice models is that they imposed a specific structure to the consumer preferences. For example, the MNL imposes that the ratio between the likelihood of choosing two products is independent on what other products are being offered. This condition, known as the independence of irrelevant alternatives, typically doesn't hold in reality.

All choice models mentioned above belong to a large family known as random utility models (RUM), originally proposed introduced by Thurstone [25]. In the last decade, there has been an increasing interest in estimating a general RUM model without imposing any a-priori specific structure in the solution (such as those imposed by the MNL, nested logit, or Markov chain). One of the first methods proposed to estimate efficiently a general RUM was carried out by Farias et al. [11]. The authors used the constraint sampling method to find a distribution over strict preference rankings (which is equivalent to a RUM, see Block and Marschak [5]) that produces the worst-case revenue compatible with the available data. Another method to estimate a general RUM was proposed by van Ryzin and Vulcano [26]. In this case, the authors developed a column-generation procedure to compute the maximum likelihood estimates. While the problem of estimating a non-parametric model such as the RUM faces multiple challenges such as computational time complexity, over-fitting issues and the non-identifiability problem (a single RUM may have multiple ways to describe it), there is a considerable benefit. Namely, by not imposing any particular structure, one can potentially capture complex substitution behaviour seen in the data that falls outside any special case of the RUM. This idea is sometimes referred as "let data speak". Indeed, [11] conducted an empirical study with sales data from 14 products types (vehicles) along 16 months from a major US automaker. The authors calibrated a standard MNL, a mixed MNL and a general RUM using a training data-set and then compared the three models using out-of-sample data. The results showed that their non-parametric model improved prediction accuracy of the other two methods by around 20 percent. In addition, they authors showed that this improvement can be translated into a 10 percent increase in revenues. In a study for a large US fashion retailer to improve sales purchases, Farias et al. [12] also applied a non-parametric approach to estimate a general RUM. Their proposed solution, which included an assortment optimization procedure based on the RUM, has increased revenues by around 7 percent.

Despite the generality of the RUM and the efficient methods to perform a non-parametric estimation of it, there are still important challenges ahead. It is easy to show that every RUM model must satisfy a property known as regularity: the probability of choosing an alternative cannot increase if the offer set is enlarged. Although this property seems natural, economists have carried out several controlled experiments (e.g. Huber et al. [16]; Simonson and Tversky [23]) from which it becomes clear that individuals (in certain contexts) violate it. Clearly, in contexts where regularity doesn't hold, it is impossible to fit a RUM regardless on how sophisticated the non-parametric estimation procedure is. Although strong regularity violations have been observed in controlled experiments, one may raise the question on whether these violations also appear in more business relevant contexts such as in aggregate retail sales. Very recently, Jagabathula and Rusmevichientong [17] have developed an efficient procedure to quantify the *loss of rationality* (LoR) which is the cost of approximating a dataset of aggregate sales using a RUM. Based on the analysis consisting of real transactions of consumer packaged goods, they have shown that for some categories of products such as coffee and yogurt, the LoR is high. This means that *regardless* of the RUM choice model used, it will not be able to approximate well the observed data.

As mentioned in Jagabathula and Rusmevichientong [17], there is no *de facto* non-RUM model class to which to go to in the case the RUM class performs poorly. Almost all choice models outside the RUM class do not subsume it, and are parametric, imposing considerable structure onto the consumer preferences. Some examples of choice models outside RUM (but do not subsume the RUM class) are the GAM model [14] [1], the general Luce model [8], the perception adjusted model (PALM) [9] and the comparison-based choice model [18]. As it was pointed out in Jagabathula and Rusmevichientong [17], the problem of using any of these models is that they can potentially be worse than a RUM model (in fitting and making choice predictions).

1.1. **Contributions.** In this paper we propose a new discrete choice model, which we called *Generalized Stochastic Preference*(GSP) model. The GSP can explain several choice phenomena that can't be represented by a RUM such as the *attraction effect* and the *compromise effect*, both of which are examples of regularity violations [22]. We believe this model is appealing for the following reasons.

---

[1]The GAM can only fall outside RUM when the shadow attractions are larger than the natural attractions.

First, the generalized stochastic preference (GSP) model is a non-parametric model; and it subsumes the RUM class, thus, it has a high flexibility to explain (fit) complex choice behaviour (data). Second, the GSP is amenable to sparse solutions. For example, sparse GSP's can easily (and also exactly), replicate several well known non-RUM examples (e.g. McFadden and Richter [20]), as well as well known choice experiments that possess strong regularity violations (e.g. Simonson and Tversky [23]). Third, the GSP is built upon well known concepts and has a simple interpretation. Indeed, the model is a natural extension of the stochastic preference model (another representation of RUMs) in which additional "non-rational consumer types" as those proposed in Kleinberg et al. [18] are incorporated. The similarity (in terms of the construction) between the GSP and the stochastic preference model (i.e. ranked list, or RUM) allows to use the same non-parametric estimation framework as those proposed for the RUM (see, e.g. Farias et al. [11], van Ryzin and Vulcano [26]).

The remaining of this paper is organized as follows. The generalized stochastic preference model is introduced in Section 2. Section 2.1 provides a small example from a well known social experiment to illustrate the model and the concepts involved. In Section 3 we show an important property about the GSP class and compare the GSP with a broad class of choice models recently introduced in Cattaneo et al. [7]. Section 4 describes the framework to estimate the model used for our five examples: it follows similar ideas as those used for the estimation of the RUM class (see [11] and [26]). The assortment problem under the GSP is discussed in Section 5. Finally, some conclusions and future research directions are discussed in Section 6.

2. The model

Let $\mathcal{C} = \{1, 2, \ldots, N\}$ denote a set of product types or alternatives. We consider the standard setting in which consumers are faced some choice set $S \subseteq \mathcal{C}$ and they have to select at most one of the elements in S (if the individual selects nothing we interpret this as if they select a default element labeled as 0). Based on this, a *discrete choice model* is a function $\mathcal{P} : 2^{\mathcal{C}} \times \mathcal{C} \cup \{0\} \to [0, 1]$ such that $\mathcal{P}(i, S)$ is the probability that a consumer will select alternative i when she is offered the choice set of alternatives $S \subseteq \mathcal{C}$. Thus, a function $\mathcal{P} : 2^{\mathcal{C}} \times \mathcal{C} \cup \{0\} \to \mathbb{R}$ is a discrete choice model if and only if it satisfies the following three axioms [3]:

(1) $\mathcal{P}(x, S) \geq 0$ for every $x \in \mathcal{C} \cup \{0\}$ and $S \subseteq \mathcal{C}$;

(2) $\mathcal{P}(x, S) = 0$ for every $x \in \mathcal{C}$ and $S \subseteq \mathcal{C} \setminus \{x\}$;

(3) $\sum_{x \in S} \mathcal{P}(x, S) \leq 1$ for every $S \subseteq \mathcal{C}$.

Depending on how the function $\mathcal{P}$ is defined (while satisfying those three axioms), we obtain different discrete choice models such as for example the MNL, the Markov chain [4], the Nested logit, or the broader random utility model (RUM).

The *generalized stochastic preference* choice model (GSP), is a choice model based on a probability distribution among a set of consumer types (or consumer preferences) we call Ω. Let Γ, be the set of all sequences of elements in $\mathcal{C}$ *without* repetition. Formally, let $\Gamma = \{(y_1, \ldots, y_k) : 1 \leq k \leq |N|, y_i \in \mathcal{C}, y_i \neq y_j \text{ for all } i = 1, \ldots, k, j = 1 \ldots, k, j \neq i\}$. The set of *consumer types* Ω is defined as $\Omega = \{(\ell, i) \in (\Gamma, [N]) : 0 \leq i \leq |\ell|\}$. We will now define the choice function associated to each consumer type which tell us which alternative is chosen by j depending on what is being offered. Whenever a consumer type (ℓ, i) is offered a choice set of alternatives S, she will first construct a subsequence $s(\ell, S)$ from ℓ by removing *all* alternatives that are *not* in S, and then she chose the alternative at the $i^t h$ position in s if it exists [2]. If $|s| < i$, or $i = 0$, the consumer will select the no-choice option, labeled by the number zero. The choice function is described more formally below.

Definition 1. *Faced with a choice set $S \subseteq \mathcal{C}$, the binary choice function $C_j(x, S)$ which states whether consumer type $j = (\ell, i)$ chooses alternative $x \in \mathcal{S} \cup \{0\}$ is defined as follows:*

$$C_j(x, S) = \mathbb{1}(i, \ell, x, S) := \begin{cases} 1 & \text{if } x \in S \text{ and } x \text{ is in the } i^{th} \text{ position in } s(\ell, S) \\ 0 & \text{otherwise} \end{cases}$$

and letting $C_j(0, S) = \mathbb{1}(i, \ell, 0, S) := 1 - \sum_{x \in S} C_j(x, S)$.

Definition 2. *A discrete choice model $\mathcal{P}$ belongs to the class of generalized stochastic preference models (GSP) if there exists a probability distribution P over consumer types set Ω (as defined above) such that*

$$\mathcal{P}(x, S) = \sum_{j \in \Omega} \mathsf{P}(j) C_j(x, S) \tag{1}$$

for all $S \subseteq \mathcal{C}$ and $x \in S \cup \{0\}$.

Observe that a consumer type (ℓ, i) with $i = 1$ always chooses the most preferred alternative among those which are offered. Thus, we call those type of consumers *rational*. The

---

[2]Positions in a sequence start with position 1 (and not zero).

consumer types $(\ell, 0)$, who always choose the no-choice options are also called *rational*. On the other hand, all the remaining consumer types, namely those types (ℓ, i) with $i > 1$ are called *irrational*.

We highlight that the GSP is also a generalization of the recently proposed mixed comparison-based choice model [18]. The key difference is that in that model, all consumer types share the same preference list. While the comparison-based choice model can also explain choice phenomena outside RUM (such as the attraction effect and the compromise effect), by fixing all consumer types to have the same preference list restricts considerably the choice behaviours that can capture. In particular, it can't capture the RUM class.

Observation 1. *The class of random utility models is contained by the class of generalized stochastic preference models.*

Proof. Block and Marschak [5] proved that every RUM can be represented by a probability distribution P over rankings among the alternatives in $\mathcal{C} \cup \{0\}$ where consumers choose the first alternative available. This coincides with the special case of GSP models in which all consumer types with positive probability are rational. $\square$

We finish this section with a small example coming from a well known social experiment to illustrate the model and the concepts involved. We also fitted (exactly) a GSP model into another four well known social choice experiments reported in Ariely [2], Simonson and Tversky [23], and Herne [15]; as well as a classical non-RUM example that satisfies regularity by McFadden and Richter [20]. None of those examples can explained by a RUM. The fitting method used for all these examples is described in Section 4. The interested reader can find all these examples in the appendix.

2.1. **An example.** Simonson and Tversky [23] performed a choice experiment with three alternatives: (1) Minolta X-370 camera; (2) Minolta MAXXUM 3000i camera; and (3) Minolta MAXXUM 7000i camera. Participants were split into two conditions. Those under the first condition (condition 1) were asked to choose between alternatives $\{1, 2\}$ whereas those under the second condition (condition 2) had to choose a camera among all three. The experiment results were:

TABLE 1. A generalize stochastic preference that explains Experiment 1 by Simonson and Tversky [23].

Consumer type			Probability
label	ℓ	i	
1	(1,3,2)	1	0.22
2	(2,3,1)	1	0.29
3	(3,2,1)	1	0.21
4	(3,2,1)	2	0.28

Model	Price (USD)	Condition 1 : $\{1,2\}$	Condition 2 $\{1,2,3\}$
1: X-370	169.99	50%	22%
2: MAXXUM 3000i	239.99	50%	57%
3: MAXXUM 7000i	469.99	not available	21%

Next, we show that these results can be explained *exactly* with a generalized stochastic preference model that has only four consumer types as shown in Table 2.1.

The first three consumer types of this GSP are rational, whereas the fourth type (which has a probability of 0.28), is irrational. In particular, this type has the preference list (3,2,1) but will select the *second* alternative (if it exists) after removing from the sequence (3,2,1) the alternatives not present in the choice set. As one can see, if the expensive MAXXUM 7000i camera is on the choice set, this consumer type will select the MAXXUM 3000i (if available). In the case the choice set is composed of only the X-370 (alternative 1) and the MAXXUM 3000i (alternative 2), the consumer will select the X-370.

For illustration, we calculate $\mathcal{P}(2, \{1,2\})$ and $\mathcal{P}(2, \{1,2,3\})$ using the proposed GSP model to show that it coincides with the experimental results. Following the same logic, one can recover the remaining market shares from the experiment under both conditions.

$$
\begin{aligned}
\mathcal{P}(2, \{1,2\}) &= 0.22 \cdot C_1(2, \{1,2\}) + 0.29 \cdot C_2(2, \{1,2\}) + \\
&\quad 0.21 \cdot C_3(2, \{1,2\}) + 0.28 \cdot C_4(2, \{1,2\}) \\
&= 0 + 0.29 + 0.21 + 0.0 \\
&= 0.50
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{P}(2, \{1,2,3\}) &= 0.22 \cdot C_1(2, \{1,2,3\}) + 0.29 \cdot C_2(2, \{1,2,3\}) + \\
&\quad\ 0.21 \cdot C_3(2, \{1,2,3\}) + 0.28 \cdot C_4(2, \{1,2,3\}) \\
&= 0 + 0.29 + 0 + 0.28 \\
&= 0.57
\end{aligned}
$$

Observe that by adding alternative 3 to the choice set, the probability of choosing alternative 2 increased by 7 percent. This is the result of gaining consumer type 4 (28 percent) and losing consumer type 3 (21 percent) who now prefer to choose alternative 3.

3. Properties

In this section we show a basic property satisfied by all GSP, and also prove that the GSP falls outside the broad class recently proposed by Cattaneo et al. [7] known as the *random attention model.*

The example from Section 2.1 showed the class of GSP admits violations to regularity. That is, it is not always true that

$$
\mathcal{P}(x, S) \geq \mathcal{P}(x, S \cup \{y\}) \text{for every} x \in S \cup \{0\}, S \subseteq \mathcal{C} \tag{2}
$$

The following lemma shows, however, that regularity as it is stated in equation (2), is satisfied at least for the no-choice option $x = 0$. That is, every GSP model satisfies that the probability of choosing nothing cannot increase if the offer set is enlarged. Thus, this shows that despite its flexibility to model complex behaviour, the class of GSP model possesses some structure.

Lemma 1. *Let $\mathcal{P}$ denote be a generalized stochastic preference model. Then, $\sum_{i \in S} \mathcal{P}(i, S) \leq \sum_{i \in S'} \mathcal{P}(i, S')$ for every $S \subseteq S' \subseteq \mathcal{C}$.*

Proof. It is enough to show that for every consumer type $j = (\ell, i) \in \Omega$, whenever $C_j(0, S') = 1$, it holds that $C_j(0, S) = 1$. If $i = 0$ this is clearly true, as this consumer type will always choose the no-choice option regardless of what is the offer set. Suppose now that $i \neq 0$ and that $C_j(0, S') = 1$. This means that the subsequence $s(\ell, S')$ of ℓ which is constructed by removing all alternatives that are not in S' satisfies that $|s(\ell, S')| < i$. Since $S \subseteq S'$, we also have that $|s(\ell, S)| < i$, and therefore $C_j(0, S) = 1$. $\qquad\square$

One may wonder whether lemma 1 provides a characterization of the choice models that are GSP. Namely, is it true that every choice model $\mathcal{P}$ that satisfies that $\sum_{i \in S} \mathcal{P}(i, S) \leq \sum_{i \in S'} \mathcal{P}(i, S')$ for every $S \subseteq S' \subseteq \mathcal{C}$ belongs to the GSP class? The answer is no, as seen by the following theorem. The proof is deferred to the appendix.

Theorem 1. *There are discrete choice models that satisfy that $\sum_{i \in S} \mathcal{P}(i, S) \leq \sum_{i \in S'} \mathcal{P}(i, S')$ for every $S \subseteq S' \subseteq \mathcal{C}$ but don't belong to the GSP class.*

Very recently, Cattaneo et al. [7] introduced a new choice model based on consideration sets which they called the *random attention model* (RAM). A RAM consists of a ranking (i.e. a permutation) π of the elements in $\mathcal{C}$, and a function $\mu(T, S)$ (with $T \subseteq S$) which is interpreted as the probability that an individual would consider the set T when it is shown the offer set S. In a RAM, this function must satisfy that $\mu(T, S) \geq \mu(T, S')$ for all $T \subseteq S \subset S'$. Intuitively, the inequality says that the probability of considering the set T cannot increase if instead of offering the set S ($T \subseteq S$) we offer a superset of it. Equipped with μ and π, the probability of choosing alternative i when offered assortment S is given by

$$\mathcal{P}(i, S) = \sum_{T \subseteq S} \mu(T, S)\mathbb{1}(i \text{ is ranked first (under } \pi) \text{ among the alternatives in } T) \qquad (3)$$

The probability of selecting the no-choice option is given by $\mathcal{P}(0, S) := 1 - \sum_{i \in S} \mathcal{P}(i, S) = \mu(\emptyset, S)$.

The RAM model is very general, it contains the RUM class as well as other parametric choice model that fall outside the RUM class. For example, it contains the additive perturbed utility (APU) model [13], as well as the menu-dependent stochastic feasibility model [6]. The next theorem shows however, that not every GSP model belongs to the RAM class.

Theorem 2. *The GSP class is not contained in the RAM class.*

Proof. We consider a GSP with four alternatives $\mathcal{C} = \{1, 2, 3, 4\}$ and only six consumer types. Table 3 shows the six consumer together with their probabilities.

Based on this GSP, we can compute the following probabilities:

TABLE 2. A generalized stochastic preference that is not a random attention model

Consumer type			Probability
label	ℓ	i	
1	(2,3,1,4)	1	0.41
2	(2,4,1,3)	1	0.09
4	(2,1,3,4)	2	0.1
3	(3,1,2,4)	2	0.01
5	(1,3,2,4)	2	0.09
6 (no-choice)	$(1,2,3,4)$	0	0.30

$$
\begin{aligned}
\mathcal{P}(1,\{1,2,3,4\}) &= \mathsf{P}(((2,1,3,4),2)) + \mathsf{P}(((3,1,2,4),2)) \\
&= 0.1 + 0.01 = 0.11 \\
\mathcal{P}(1,\{1,3,4\}) &= \mathsf{P}(((3,1,2,4),2)) = 0.01 \\
\mathcal{P}(2,\{2,3,4\}) &= \mathsf{P}(((2,3,1,4),1)) + \mathsf{P}(((2,4,1,3),1)) + \mathsf{P}(((3,1,2,4),2)) + \mathsf{P}(((1,3,2,4),2)) \\
&= 0.41 + 0.09 + 0.01 + 0.09 \\
&= 0.6 \\
\mathcal{P}(2,\{2,4\}) &= \mathsf{P}(((2,3,1,4),1)) + \mathsf{P}(((2,4,1,3),1)) \\
&= 0.41 + 0.09 = 0.5 \\
\mathcal{P}(3,\{1,3\}) &= \mathsf{P}(((2,3,1,4),1)) + \mathsf{P}(((2,4,1,3),1)) + \mathsf{P}(((2,1,3,4),2)) \\
&= 0.41 + 0.09 + 0.1 = 0.6 \\
\mathcal{P}(3,\{3\}) &= \mathsf{P}(((2,3,1,4),1)) + \mathsf{P}(((2,4,1,3),1)) \\
&= 0.41 + 0.09 = 0.5
\end{aligned}
$$

Cattaneo et al. [7] consider the following binary relation $\prec \in 2^{\mathcal{C}^2}$ among the elements in $\mathcal{C}$: $(x,y) \in \prec$ if and only if there exists $S \subseteq \mathcal{C}$ such that $x, y \in S$ and $\mathcal{P}(x, S - \{y\}) < \mathcal{P}(x, S)$. They showed that a discrete choice model can be represented by RAM if and only if that relationship doesn't have a cycle. Observe now that the probabilities computed above show that the pairs $(1,2), (2,3), (3,1)$ all belong to $\prec$. Therefore, there is a cycle in $\prec$. It follows that this choice model doesn't belong to the RAM class. □

One may wonder whether GSP subsumes the class of random attention models. The following theorem shows this is not the case.

Theorem 3. *The RAM class is not contained in the GSP class.*

Proof. Consider the choice model used in the proof of Theorem 1 (see Table 7). In Theorem 1, we proved that the model doesn't belong to the GSP class. However, based on the characterization by Cattaneo et al. [7] (Theorem 2), one can deduce that this model does belong to the RAM class as there are no cycles in the relationship $\prec$ among the alternatives. In particular, observe that the only alternative in the model from table 7 that violates regularity is alternative 1. □

4. MODEL ESTIMATION

The task of estimating a generalized stochastic preference is similar to that of estimating a standard stochastic preference model (i.e. a general RUM) (see, e.g. [26], [11]). In what follows, we revisit the mathematical formulation of the estimation problem following [11], [17], [12], adapting it to the GSP, and briefly comment on some methods to solve it. In particular, all the examples provided in this paper (see appendix) have been estimated by solving the optimization problem (4) written at the end of this section.

Let $\mathcal{M} \subseteq 2^{\mathcal{C}}$ denote a family of assortments that the firm has offered to consumers. For each $S \in \mathcal{M}$, let $f_{i,S}$ denote the fraction of consumers who chose product $i \in S$ when offered assortment S. Let $\mathbf{f}_{\mathcal{M}} = (f_{i,S} : i \in S, S \in \mathcal{M})$. Informally, our problem consists of finding a generalized stochastic preference model such that $\mathcal{P}(i, S)$ is as close as possible to $f_{i,S}$ for every $i \in S$, $S \in \mathcal{M}$.

Let P be probability distribution over consumer types Ω, that is, P represents a generalized stochastic preference model. Let $\mathbf{x}_{\mathcal{M}}^{\mathsf{P}} = (\mathcal{P}_{i,S} : i \in S, S \in \mathcal{M})$. Where $\mathcal{P}(i, S) = \sum_{(\ell,k)\in\Omega} \mathsf{P}((\ell, k))\mathbb{1}(k, \ell, i, S)$ is the probability of choosing alternative i when offered assortment S under the generalized stochastic model P.

The estimation problem consist then in finding P^* such that

$$\mathsf{P}^* = \arg \min_{\mathsf{P}} \{\mathbf{loss}(\mathbf{x}_{\mathcal{M}}^{\mathsf{P}}, \mathbf{f}_{\mathcal{M}}) : \mathsf{P} : \Omega \to [0, 1], \sum_{j\in\Omega} \mathsf{P}(j) = 1\}$$

Where **loss** is a non-negative function such that $\mathbf{loss}(\mathbf{x}_{\mathcal{M}}^{\mathsf{P}}, \mathbf{f}_{\mathcal{M}}) = 0$ if and only if $\mathbf{x}_{\mathcal{M}}^{\mathsf{P}} = \mathbf{f}_{\mathcal{M}}$. Typical examples of the loss function are (1) the additive inverse of the maximum

likelihood and (2): the norm function. van Ryzin and Vulcano [26] proposed a method to solve this problem (defining the loss as the maximum likelihood) using a column generation technique. Although their method was originally designed to solve the standard stochastic preference model (i.e. a ranked-list), it can be adapted to solve the generalized stochastic preference model. To this end, one would need to add additional terms into the likelihood function to account for the irrational consumer types.

Following Farias et al. [12], the GSP estimation problem can be written using matrices. Let $\lambda \in 1 \times \mathbb{R}^{|\Omega|}$ denote the vector in which we index each of the consumer types in Ω. Let also $\mathbf{y} = (f_{i,S})_{i \in S, S \in \mathcal{M}}$ be the vector containing the market shares $f_{i,S}$ indexed by i, S. If we let A denote the $L \times |\Omega|$ 0-1 matrix where $L = \sum_{S \in \mathcal{M}} |S|$ such that $A_{(i,S),(\ell,k)} = \mathbb{1}(k, \ell, i, S)$, then we can write the norm minimization problem as:

$$\min_{\lambda \geq 0, \|\lambda\|=1} \|y - A\lambda\| + c \cdot \|\lambda\|_0$$

where $\|\lambda\|_0$ counts the number of non-zero elements in λ and $c > 0$ acts as a penalty factor for the model complexity which is measured by $\|\lambda\|_0$.

Farias et al. [12] solved the problem using for the special case of standard stochastic preference model using optimization methods over a polytope appropriately defined. Extending their method to solve the generalized stochastic preference model is an interesting research direction.

We finish this section with some comments with regards the type of solution λ that is desired. For the standard RUM, one seeks for a solution that is sparse, i.e. it the number of consumer types (permutations) that have non-zero probability is small.

In the generalized stochastic preference model, besides finding a sparse model it may be beneficial to find a model where the weight associated to the irrational consumer types is low. Indeed, if the weight of the irrational consumer types is low enough, one may discard them to get a standard RUM. This will allow the analyst to use standard methods to solve pricing and assortment optimization that work well under RUM. To that end, suppose we partition the vector λ into λ_1 and λ_2 where λ_1 (λ_2) contains all rational (irrational) consumer types. Then, one may wish to solve the following problem:

$$\min_{\lambda \geq 0, \|\lambda\|=1} \|y - A\lambda\| + c_2 \cdot \|\lambda_2\|_0 + c \cdot \|\lambda\|_0 \tag{4}$$

where $c_2 > c > 0$ in order to penalize solutions which contain more a higher weight of irrational consumer types and at the same type keep penalizing solutions with high number of consumer types with non-zero probability.

5. Assortment optimization

The assortment optimization problem in revenue management consists of choosing an assortment (i.e. offer set of products) in order to maximize the expected revenue. Formally, given a discrete choice model $\mathcal{P}$ and a revenue function $r : \mathcal{C} \to \mathbb{R}_{>0}$, the assortment problem consists of

$$\max_{S \subseteq \mathcal{C}} \sum_{x \in S} \mathcal{P}(i, S) r(i) \tag{5}$$

The strongest negative result about the assortment problem is due to Aouad et al. [1]. For the RUM class, they proved that it is NP-hard to approximate to within a factor of $\Omega(1/n^{1-\epsilon})$ and to within a factor of $\Omega(1/\log^{1-\epsilon}(r_k/r_1))$, for every $\epsilon > 0$. The problem is hard even for specific families of choice models inside the RUM class. For example, it is NP-hard to solve the assortment problem for the latent class MNL model even when there are only two-classes [21].

The heuristic that provides the best revenue guarantees (for a general RUM or a broader class) is *revenue ordered assortments* proposed by [24], which we briefly describe below. Consider $\{r_1, r_2, \ldots, r_k\}$ to be the different values taken by the revenue function r, sorted in increasing order (i.e. $0 < r_1 < r_2 < \cdots < r_k$). Let $S_i \subseteq \mathcal{C}$ be the set consisting of all products of revenue *at least* r_i (with $i = 1, \ldots, k$). The revenue ordered assortment simply compares the revenue obtained by each of the k sets $S_1, \ldots, S_k$, and chooses one with maximum revenue. Berbeglia and Joret [3] proved that, for a choice model satisfying regularity (this includes the RUM class), the revenue ordered assortment provides a revenue of at least a $\max\{\frac{1}{k}, \frac{1}{1+\ln(r_k/r_1)}\}$ fraction of the optimal revenue, and this bound is exactly tight [3]. Thus, revenue ordered assortment achieves the best possible approximation guarantees among all computationally efficient strategies. Unfortunately, the worst-case performance guarantees of the revenue ordered assortments strategy deteriorates considerably under the GSP class.

---

[3]The bound no longer holds even if the factor is multiplied by $(1+\epsilon)$ for any $\epsilon > 0$. Restricted to the RUM class, Aouad et al. [1] obtained, independently, essentially the same revenue guarantees (up to a constant factor).

Theorem 4. *Under the GSP, revenue-ordered assortments approximate the optimum revenue to within a factor of $\frac{r_1}{r_k}$ and this factor is tight.*

Proof. First, observe that the revenue obtained using revenue ordered assortments is never less than what can be achieved by showing all alternatives and this is $\sum_{i \in \mathcal{C}} \mathcal{P}(i, \mathcal{C}) r(i)$. It is then straightforward to obtain the revenue guarantee as follows:

$$
\begin{aligned}
\text{OPT} \quad &= \quad \sum_{i \in S^*} \mathcal{P}(i, S^*) r(i) \\
&\leq \quad \sum_{i \in S^*} \mathcal{P}(i, S^*) r_k \\
&= \quad r_k \sum_{i \in S^*} \mathcal{P}(i, S^*) \\
&\leq \quad r_k \sum_{i \in \mathcal{C}} \mathcal{P}(i, \mathcal{C}) \\
&= \quad \frac{r_k}{r_1} \sum_{i \in \mathcal{C}} \mathcal{P}(i, \mathcal{C}) r_1 \\
&\leq \quad \frac{r_k}{r_1} \sum_{i \in \mathcal{C}} \mathcal{P}(i, \mathcal{C}) r(i)
\end{aligned}
$$

where the second inequality is due to lemma 1.

Next, we show that this factor is tight. Consider a GSP choice model with $N = 3$ alternatives and a single consumer type $j = ((1, 2, 3), 2) \in \Omega$. Let the revenue function be defined as follows: $r_1 = r(1) = r(2) < r(3) = r_2$. Observe than an optimal solution for the assortment problem is $S^* = \{1, 3\}$, which provides a revenue of $r(3) = r_2$. Revenue ordered assortments on the other hand would provide a revenue of r_1. $\qquad\square$

It is an open question whether there are other efficient algorithms that can provide higher worst-case revenue guarantees for a general GSP.

6. Conclusions and future research

The RUM is the most studied discrete choice model studied in Economics, Marketing, Psychology and Operations. Its success can be explained by the fact that (1) the RUM contains as special cases important choice models used in those fields (e.g. MNL, mixed MNL, nested logit etc) and (2) it is a non-parametric model meaning that is has a huge flexibility. Despite its generality, it's been known for decades, thanks to controlled choice

experiments, that sometimes some human choices cannot be explained by any RUM. There is now evidence that aggregate retail purchases under certain categories such as coffee and yogurt cannot be explained by a RUM. Although there are several choice models that fall outside the RUM (e.g. general luce model, perception adjusted luce model, etc), most of them are parametric, and impose a significant structure into the choice probabilities.

In this paper, we introduced the generalized stochastic preference (GSP), a choice model that generalizes the RUM and can explain violations to regularity such as the attraction effect and the compromise effect. One key feature of the GSP is that it is non-parametric, and that it can explain (exactly) multiple choice experiments with strong violations to regularity. The GSP has a simple interpretation as well as description: it is based upon a generalization of the stochastic preference model (which is equivalent to the RUM).

There are multiple research directions for further study of the generalized stochastic preference model. One of them is about the GSP characterization. Given a choice model $\mathcal{P}$ as the complete system of probabilities $\mathcal{P}(i, S)$ for each $i \in S \cup \{0\}, S \subseteq \mathcal{C}$, how can we know whether $\mathcal{P}$ belongs to the GSP class? In 1978, Falmagne [10] proposed a family of inequalities and proved that a choice model $\mathcal{P}$ given as a system of choice probabilities is indeed a RUM if and only if it satisfies every one of them. Although each GSP must satisfy the inequality $\sum_{i \in S} \mathcal{P}(i, S) \leq \sum_{i \in S'} \mathcal{P}(i, S')$ for all $S \subseteq S' \subseteq \mathcal{C}$ (Lemma 1), we know that this is not enough (Theorem 1). Obtaining a characterization of the GSP class as that provided by Falmagne [10] for the RUM family we believe is an interesting open question.

Another important research direction is to develop efficient methodologies to estimate a GSP from data. In Section 4 we showed that this problem can be seen as a loss minimization problem subject to linear constraints over a huge solution space. The formulation is very similar to the one of estimating a RUM model and we believe the methods developed for RUM (e.g. Farias et al. [11], van Ryzin and Vulcano [26]) could be a good starting point to build upon.

Finally, it would be interesting to study special cases of the GSP that still allow regularity violations. For example, consider the standard MNL model, but instead of all consumers selecting the alternative with the highest utility (among those offered), a certain fraction of consumers buy the highest alternative, another fraction buys the second highest alternative, etc. An appealing feature of this model is that it has only $2N$ parameters, subsumes the MNL, and still allows violations to regularity. This idea, in fact, can be extended to any choice model based on random utility.

References

[1] Ali Aouad, Vivek F Farias, Retsef Levi, and Danny Segev. The approximability of assortment optimization under ranking preferences. *Available at SSRN 2612947*, 2015.

[2] D. Ariely. *Predictably irrational: The hidden forces that shape our decisions, Year = 2009*. Harper-Collins, New York.

[3] Gerardo Berbeglia and Gwenaël Joret. Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments. *Available at ArXiv*, 2016.

[4] Jose Blanchet, Guillermo Gallego, and Vineet Goyal. A markov chain approximation to choice modeling. In *14th ACM Conference on Electronic Commerce*, pages 103–104, 2013.

[5] Henry David Block and Jacob Marschak. Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 2:97–132, 1960.

[6] Richard L Brady and John Rehbeck. Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223, 2016.

[7] Matias D Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. A random attention model. *arXiv preprint arXiv:1712.03448*, 2017.

[8] Federico Echenique and Kota Saito. General luce model. 2015.

[9] Federico Echenique, Kota Saito, and Gerelt Tserenjigmid. The perception-adjusted luce model. *Mathematical Social Sciences*, 2018.

[10] Jean-Claude Falmagne. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72, 1978.

[11] Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.

[12] Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. Building optimized and hyperlocal product assortments: A nonparametric choice approach. 2017.

[13] Drew Fudenberg, Ryota Iijima, and Tomasz Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

[14] Guillermo Gallego, Richard Ratliff, and Sergey Shebalov. A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research*, 63(1):212–232, 2014.

[15] Kaisa Herne. Decoy alternatives in policy choices: Asymmetric domination and compromise effects. *European Journal of Political Economy*, 13(3):575–589, 1997.

[16] Joel Huber, John W Payne, and Christopher Puto. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1):90–98, 1982.

[17] Srikanth Jagabathula and Paat Rusmevichientong. The limit of rationality in choice modeling: Formulation, computation, and implications. *Management Science*, 2018.

[18] Jon Kleinberg, Sendhil Mullainathan, and Johan Ugander. Comparison-based choices. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 127–144, 2017.

[19] D. Luce. *Individual Choice Behavior*. John Wiley and Sons, 1965.

[20] Daniel McFadden and Marcel K Richter. Stochastic rationality and revealed stochastic preference. *Preferences, uncertainty, and optimality, essays in honor of Leo Hurwicz*, pages 161–186, 1990.

[21] Paat Rusmevichientong, David Shmoys, Chaoxu Tong, and Huseyin Topaloglu. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, 23(11):2023–2039, 2014.

[22] Itamar Simonson. Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research*, 16(2):158–174, 1989.

[23] Itamar Simonson and Amos Tversky. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29(3):281, 1992.

[24] Kalyan Talluri and Garrett Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

[25] Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.

[26] Garrett van Ryzin and Gustavo Vulcano. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300, 2014.

Appendix

The appendix contains additional examples as well as the proofs omitted from the main text.

Example 2. *[Ariely [2]] Let's consider an experiment reported in Ariely [2] where students were asked to choose between some subscription options for The Economist magazine. The subscription options were: (1) Online version only, (2) Print version only and (3) Print*

and online bundle. Half of the students were shown options $\{1,3\}$ and the other half were asked to choose between $\{1,2,3\}$. These are the results:

S	$\mathcal{P}(1,S)$	$\mathcal{P}(2,S)$	$\mathcal{P}(3,S)$
$\{1,3\}$	0.68	–	0.32
$\{1,2,3\}$	0.16	0	0.84

Observe that by increasing the choice set from $\{1,3\}$ to $\{1,2,3\}$, the probability of choosing alternative 3 increased considerably. This violates regularity, which implies that there is no choice model based on random utility that can be fitted to that data.

In what follows, we show that there exists a simple generalized stochastic preference model which yields the same outcome. Consider a GSP P where only three consumer types have a non-zero probability of being chosen. $\mathsf{P}((3,1,2),1) = 0.16$; $\mathsf{P}((3,2,1),1) = 0.16$; $\mathsf{P}((2,3,1),2) = 0.68$. One can easily check that this GSP model reproduces the same table.

Example 3. *[Simonson and Tversky [23]] In another experiment reported in Simonson and Tversky [23], participants were split into two groups ($N = 60$ and $N = 61$). First, all subjects were given pictures and descriptions of five microwave ovens. Second, participants were asked to choose among some of those products. Those in the first group had to choose among microwaves Emerson and Panasonic I (condition 1), whereas those in the second group had to choose among 3 microwaves: Emerson, Panasonic I and Panasonic II (condition 2). Below we display the results of this experiment.*

Model	Price (USD)	Condition 1 : $\{1,2\}$	Condition 2$\{1,2,3\}$
1: Emerson	109.99	57%	27%
2: Panasonic I	179.99	43%	60%
3: Panasonic II	199.99	not available	13%

These results cannot be explained using a RUM, as there is a violation of regularity. Specifically, the proportion of participants choosing alternative 2 (Panasonic I) increase considerably when the alternative 3 (Panasonic II) is added to the choice set. However, it is possible to perfectly fit a generalized stochastic preference model with only four consumer types to the results of this experiment. Table 3 shows the four consumer types with their associated probabilities.

TABLE 3. A generalized stochastic preference that fits the experiment from Example 3

Consumer type			Probability
label	ℓ	i	
1	(1,2,3)	1	0.27
2	(2,1,3)	1	0.43
3	(3,1,2)	1	0.13
4	(3,2,1)	2	0.17

Note that again in this example, the first three consumer types are rational: they prefer the first alternative from their ranking. The fourth type (which has a probability of 0.17), has the preference list (3,2,1) and will select the alternative in the second position (if it exists) after removing from the sequence (3,2,1) the alternatives not present in the choice set. If the more expensive Panasonic II microwave is on the choice set, this consumer type will select the cheaper Panasonic model (if available). But if the choice set is composed of only the Emerson (alternative 1) and the Panasonic (alternative 2), the consumer will select the Emerson.

We now calculate $\mathcal{P}(2, \{1,2\})$ and $\mathcal{P}(2, \{1,2,3\})$ and show that it matches with the experiment results. Again, one can calculate the remaining probabilities and recover the different market shares from the experiment under both conditions.

$$
\begin{aligned}
\mathcal{P}(2, \{1,2\}) &= 0.27 \cdot C_1(2, \{1,2\}) + 0.43 \cdot C_2(2, \{1,2\}) + \\
&\quad\ 0.13 \cdot C_3(2, \{1,2\}) + 0.17 \cdot C_4(2, \{1,2\}) \\
&= 0 + 0.43 + 0 + 0 = 0.43
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{P}(2, \{1,2,3\}) &= 0.27 \cdot C_1(2, \{1,2,3\}) + 0.43 \cdot C_2(2, \{1,2,3\}) + \\
&\quad\ 0.13 \cdot C_3(2, \{1,2,3\}) + 0.17 \cdot C_4(2, \{1,2,3\}) \\
&= 0 + 0.43 + 0 + 0.17 = 0.60
\end{aligned}
$$

S	$\mathcal{P}(0,S)$	$\mathcal{P}(1,S)$	$\mathcal{P}(2,S)$	$\mathcal{P}(3,S)$
$\{1\}$	.5	.5	-	-
$\{2\}$	.5	-	.5	-
$\{3\}$	.5	-	-	.5
$\{1,2\}$	.4	.3	.3	-
$\{1,3\}$	.4	.3	-	.3
$\{2,3\}$	.4	-	.3	.3
$\{1,2,3\}$	.25	.25	.25	.25

TABLE 4. Example of a choice model that is not a RUM but satisfies regularity.

Example 4. *[McFadden and Richter [20]] Consider the choice model displayed in Table 4 which was originally proposed by McFadden and Richter [20] and recently adapted by Berbeglia and Joret [3] to incorporate the no-choice option.*

Unlike the previous examples, this choice model satisfies the regularity condition. However it is not difficult to prove that it is not a RUM (see McFadden and Richter [20], Berbeglia and Joret [3]).

In what follows, we show that this choice model belongs to the class of generalized stochastic choice models. Indeed, the model can be represented by a generalized stochastic choice with 11 consumer types. Among the 11 consumer types, 3 are irrational and they account for only 20 percent of the market share.

As in the other examples, it is a simple (but tedious) exercise to check that the proposed generalized stochastic preference model provided in Table 5 fits exactly the choice probabilities given in Table 4.

Example 5. *[Herne [15]] The decoy effect has also been observed when people are asked to make political choices. Herne [15] performed a choice experiment in which participants need to pick one economic union for an imaginary country (initially, this country doesn't belong to any economic union).*

Again, participants were split into two conditions. Those under the condition 1 were asked to choose between the unions $\{T,C\}$ whereas those under the second condition (condition 2) had to choose among all three $\{T,C,D\}$. The main features displayed about the unions and the experiment results are summarized below:

TABLE 5. A generalized stochastic preference that fits the choice model proposed by McFadden and Richter [20]

Consumer type			Probability
label	ℓ	i	
1	(1)	0	0.25
2	(1)	1	0.05
3	(3,2)	1	0.05
4	(1,2,3)	1	0.1
5	(1,3,2)	1	0.05
6	(2,1,3)	1	0.05
7	(2,3,1)	1	0.15
8	(3,1,2)	1	0.1
9	(1,2)	2	0.05
10	(2,3)	2	0.1
11	(3,1)	2	0.05

Economic Union	Inflation(%)	Economic growth(%)	Condition 1	Condition 2
T	1	2.1	53%	63%
C	2.2	3.9	47%	37%
D	1.5	2	not available	0%

Observe that the union D acted as a decoy to attract participants to union T which dominates D in both dimensions: inflation and economic growth. We now show that only 3 consumer types are needed to explain (exactly) the results of this experiment. The three consumer types [4], together with their probability are given in Table 6.

The type 1 is rational type that has a weight of 53 percent and prefers T over C and C over the decoy D. Type 2 is also a rational type with a weight of 37 percent with the strict preference (C,T,D). Finally, type 3 is an irrational type, who prefers the alternative placed in the second position and has the ranking (D,T,C). Thus, the results of this experiment can be explained with a GSP whose irrational types have a weight of only 10 percent. Those

---

[4]In this context 'participant type' rather than a 'consumer type' would be more appropriate

TABLE 6. A generalized stochastic preference that explains the experiment
reported in Herne [15].

Consumer type			Probability
label	ℓ	i	
1	(T,C,D)	1	0.53
2	(C,T,D)	1	0.37
3	(D,T,C)	2	0.10

S	$\mathcal{P}(0,S)$	$\mathcal{P}(1,S)$	$\mathcal{P}(2,S)$	$\mathcal{P}(3,S)$
$\{1\}$	0	1	-	-
$\{2\}$	0	-	1	-
$\{3\}$	0	-	-	1
$\{1,2\}$	0	1	0	-
$\{1,3\}$	0	0	-	1
$\{2,3\}$	0	-	1	0
$\{1,2,3\}$	0	1	0	0

TABLE 7. Example of a choice model that is not in the GSP class.

*participants are the ones that make the alternative T increase the market share from 53 to
63 percent whenever the union D is added to the choice set.*

Theorem B1 (Theorem 1). *There are discrete choice models that satisfy $\sum_{i \in S} \mathcal{P}(i,S) \leq \sum_{i \in S'} \mathcal{P}(i,S')$ for every $S \subseteq S' \subseteq \mathcal{C}$ but don't belong to the GSP class.*

Proof. Let $\mathcal{C} = \{1,2,3\}$ and consider a choice model defined in Table 7.

For the purpose of contradiction, suppose there exists a generalized stochastic prefer-
ence model (i.e. a probability distribution P over consumer types $(\ell,i) \in (\Omega,[N])$) that
characterizes the choice model defined above.

Given that $\mathcal{P}(1,\{1,2\}) = 1$, we have that

$$\mathsf{P}((1,2,3),1) + \mathsf{P}((1,3,2),1) + \mathsf{P}((3,1,2),1) + \mathsf{P}((2,1,3),2) + \mathsf{P}((2,3,1),2) + \mathsf{P}((3,2,1),2) = 1 \tag{6}$$

Above, the first three terms account for all the consumer types (ℓ,i) who consider the
alternative in the first position $(i = 1)$ and who rank 1 above 2. The remaining three terms

are from all the consumer types who rank alternative 2 higher than alternative 1 and choose the alternative in the second position $(i = 2)$.

Similarly, using the fact that $\mathcal{P}(2, \{2, 3\}) = 1$, and $\mathcal{P}(3, \{1, 3\}) = 1$, we can derive the following two equations.

$$\mathsf{P}((2,3,1),1) + \mathsf{P}((2,1,3),1) + \mathsf{P}((1,2,3),1) + \mathsf{P}((3,2,1),2) + \mathsf{P}((3,1,2),2) + \mathsf{P}((1,3,2),2) = 1 \quad (7)$$

$$\mathsf{P}((3,1,2),1) + \mathsf{P}((3,2,1),1) + \mathsf{P}((2,3,1),1) + \mathsf{P}((1,3,2),2) + \mathsf{P}((1,2,3),2) + \mathsf{P}((2,1,3),2) = 1 \quad (8)$$

Based on equations (6), (7), and (8), the existence of a generalized stochastic preference implies that the following matrix equation has a solution where each variable is non-negative.

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
\mathsf{P}((1,2,3),1) \\
\mathsf{P}((1,3,2),1) \\
\mathsf{P}((3,1,2),1) \\
\mathsf{P}((2,3,1),1) \\
\mathsf{P}((2,1,3),1) \\
\mathsf{P}((3,2,1),1) \\
\mathsf{P}((2,1,3),2) \\
\mathsf{P}((2,3,1),2) \\
\mathsf{P}((3,2,1),2) \\
\mathsf{P}((3,1,2),2) \\
\mathsf{P}((1,3,2),2) \\
\mathsf{P}((1,2,3),2)
\end{bmatrix}
=
\begin{bmatrix}
1 \\ 1 \\ 1 \\ 1
\end{bmatrix}
;
$$

where all variables must be non-negative [5]

If there exists a feasible solution, by Farkas Lemma, there is no vector $(y_1, y_2, y_3, y_4) \in \mathbb{R}^4$ such that the following two conditions are respected.

---

[5]The equation induced by the first row of the matrix must hold because P is a probability distribution together with the fact that the sum of a subset of the variables is equal to 1.

$$
\begin{bmatrix}
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 \\
1 & 0 & 0 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{bmatrix}
\geq
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix} ;
$$

and that

$$
\begin{bmatrix}
1 & 1 & 1 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{bmatrix}
< \begin{bmatrix} 0 \end{bmatrix} ;
$$

However, the following vector $(y_1, y_2, y_3, y_4) = (-1, -1, -1, 2)$ satisfies the previous two inequalities. This is a contradiction, and therefore we proved that there is no generalized stochastic preference that can characterize the proposed system of choice probabilities.

$\square$