Inferential Choice Theory*

Narayanaswamy Balakrishnan† Efe A. Ok‡ Pietro Ortoleva§

February 17, 2021

Abstract

Despite being the fundamental primitive of the study of decision-making in economics, choice correspondences are not observable: even for a single menu of options, we observe at most one choice of an individual at a given point in time, as opposed to the *set* of all choices she deems most desirable in that menu. However, it may be possible to observe a person choose from a feasible menu at various times, repeatedly. We propose a method of inferring the choice correspondence of an individual from this sort of choice data. First, we derive our method axiomatically, assuming an ideal dataset. Next, we develop statistical techniques to implement this method for real-world situations where the sample at hand is often fairly small. As an application, we use the data of two famed choice experiments from the literature to infer the choice correspondences of the participating subjects.

JEL Classification: C81, D11, D12, D81.
Keywords: Choice Correspondences, Estimation, Stochastic Choice Functions, Transitivity of Preferences

*We thank José Luis Montiel Olea and Gerelt Tserenjigmid for their valuable observations and comments on an earlier draft of this paper. We are also grateful to Mauricio Almeida Couri Riberio for his superb research assistance.

†Department of Mathematics and Statistics, McMaster University. E-mail: bala@mcmaster.ca

‡Department of Economics and the Courant Institute of Applied Mathematics, New York University. E-mail: efe.ok@nyu.edu

§Department of Economics and School of Public and International Affairs, Princeton University. E-mail: pietro.ortoleva@princeton.edu

1 Introduction

At the core of revealed preference theory is the idea that, while preferences and utilities are unobservable constructs, choices of an individual are observable. An advantage is thus given to theories derived from choice. The main primitive of this approach is the notion of *choice correspondence* which is a function that maps each feasible menu of options to the set of choices from that menu. Most textbooks use choice correspondences as the starting point of microeconomic theory from which preferences, and then utility functions, are derived. Over the last century, choice correspondences proved to be exceptionally useful for the development of the theory of rational decision-making as well as its boundedly rational alternatives.

On closer scrutiny, however, one has to concede that choice correspondences – in fact, their values – are unobservable as well. After all, such a correspondence assigns to any given menu a *set* of choices, but it is practically impossible to observe such a set in its entirety. For, each time an individual makes a choice from a menu, we only observe a single option being chosen from that menu. For example, when a rational individual is indifferent between two options x and y, the value of her choice correspondence at $\{x, y\}$ must be $\{x, y\}$, and yet we can only observe her choose x or y, but not both, at any one choice trial.[1] Indeed, while the use of choice correspondences is often motivated in first-year Ph.D. courses as based on observability, this often leads to embarrassing questions—in our experience invariably asked by some alert student—about how such set-valued functions can really be observed, and how their properties can actually be tested. Many prominent textbooks, as well as lectures on the web, either avoid the discussion of this issue (thereby choosing to treat choice correspondences as purely theoretical constructs), or simply assume it away by working only with single-valued choice correspondences.[2] A few of them suggest informal ways of thinking about choice correspondences as observable entities, but this never goes beyond offering a few passing sentences to this effect.[3]

All in all, despite the original aspirations of classical revealed preference theory that go back to the seminal works of Samuelson (1938, 1950), Houthakker (1950) and Arrow (1959), choice correspondences remain as analytic constructs. This does not diminish their value for microeconomic theory and beyond, to be sure. But their fundamental unobservability

[1] There are a few papers that have proposed experimental methods that try to circumvent this issue—we discuss them below. Whether one thinks these methods are effective or not, it is plain that these techniques only apply to specially constructed experiments. We should also mention that the empirical approach to revealed preference theory that builds on the famous Afriat Theorem never attempts to elicit one's choice correspondence; it is rather primed to obtain a finite selection from this correspondence.

[2] Simplifying as it is, this latter approach not only fails solving the unobservability problem, it serves rather poorly as a foundation for even the most basic economic models of decision making. After all, if a choice correspondence is single-valued, it can never be rationalized by a preference relation that allows for indifference, thereby ruling out the standard model of consumer choice, as well as all non-degenerate cases of expected utility theory under risk, among others. In addition, it is known that single-valuedness hypothesis cannot be satisfied by *continuous* choice correspondences, unless one makes severe assumptions on the grand space of alternatives (cf. Nishimura and Ok (2014)).

[3] For instance, Mas-Colell, Whinston and Green (1995) says that the set of choices of an individual from a menu B "can be thought of as containing those alternatives that we would actually see chosen if the decision maker were repeatedly to face the problem of choosing an alternative from set B." In many ways, our work can be thought of as trying to extract formal content from this intuitive statement in a manner that allows one infer choice correspondences in practice.

nonetheless leads to some practical difficulties. This stems mainly from the fact that behavioral traits such as rationality, or particular types of bounded rationality, are modeled in the form of properties of a choice correspondence. For instance, quite a bit of attention in the literature has been devoted to testing if a given individual abides by the Weak Axiom of Revealed Preference. Clearly, answering this requires observing one's choice correspondence in full to identify potential indifferences. Or, if we want to test whether or not a person is always decisive, we need to distinguish between when she is indecisive and when indifferent. Similarly, to test for a particular boundedly rational choice theory (say, one that uses consideration sets or reference-dependent decision-making procedures), we need to be able to "compute" a person's choice correspondence in full. Besides, as thoroughly explored by Bernheim and Rangel (2007, 2009), performing unambiguous individual welfare comparisons for an economic agent requires *all* values of her choice correspondence across a variety of menus, thereby pointing to the importance of being able to carry out this sort of a computation. Simply put, eliciting choice correspondences is essential for the testing of any sort of a choice theory as well as for behavioral welfare economics at large.

The principal objective of the present paper is to introduce a general method of inferring choice correspondences from data obtained through repeated observations of choices made by individuals. More precisely, we aim to "compute" a choice correspondence from the number of times each option is chosen by a person from a given menu when asked repeatedly. As such, our starting point is that what is observable about one's choice behavior comes in the form of a vector of relative frequencies (of choices), that is, in the form of stochastic choice data.[4]

We approach the problem in two stages. These are distinct from each other both procedurally and conceptually, and are primed to capture different aspects of the problem. In the first stage, we discuss methods of constructing a choice correspondence if the analyst had access to an 'ideal' data set that provides the actual relative choice frequencies of the subject with perfect accuracy. In the second stage, we address the issue that real data includes only a finite (and often small) sample of observations, and suggest a statistical procedure to take this into account to construct our (empirical) choice correspondence. We summarize what we actually do in these stages next.

Choice Imputation with Ideal Data. At the outset, we look at the problem at hand theoretically, assuming that the analyst has access to an 'ideal' data set that somehow provides the probability $\mathbb{P}(x, S)$ with which each option x is chosen by a subject from a given menu S. In other words, we study the functions that map any given stochastic choice function $\mathbb{P}$ to a choice correspondence. We refer to any such function as a *choice imputation* (provided that it never includes an option with zero probability of being chosen in a choice set).

There are many interesting types of choice imputations. For example, we may declare as a "choice" in a menu S any option that has a positive probability of being chosen in S. But this may well be too permissive. If the probability of x being chosen from $\{x, y\}$ is negligibly small, it may be reasonable to think of it as a "mistake," instead of a bona fide "choice." We may also go to the opposite extreme, and consider only the options with maximum likelihood

[4]The term stochastic choice is used both to indicate the relative frequency of the choices of an individual in repeated trials, and across individuals when each person is observed only once. In this paper we exclusively focus on multiple choices made by the same individual.

of being chosen in a menu. But, obviously, this may well be too restrictive; for instance, it would not consider x as a choice from $\{x, y\}$ even if the probability of x being chosen is as high as .49. And, of course, there are many intermediate imputations that possess a less extreme makeup. No imputation is likely to be suitable in all contexts; there does not appear to be a reason *a priori* to work with any one specific imputation.

To address this issue systematically, we adopt an axiomatic approach, and consider some basic properties that characterize an interesting one-parameter family of choice imputations. An element of this family either maps to the choice correspondence that chooses in a menu all elements that have a positive probability of being chosen, or there exists a constant $\lambda \in (0, 1]$ such that it maps any $\mathbb{P}$ to the choice correspondence that declares

$$\left\{x \in S : \mathbb{P}(x, S) \geq \lambda \max_{y \in S} \mathbb{P}(y, S)\right\}$$

as the set of "choices" from any menu S. We refer to any such choice imputation as a *Fishburn imputation*.[5] This family includes the two examples above, but it allows for many intermediate cases: For any $\lambda \in (0, 1)$, the associated imputation declares x as a choice in a menu S when its probability of being selected is higher than a factor (namely, λ) of the choice probability of any other option in S. It is worth noting that, for menus with at least three options, this is not the same as focusing on alternatives chosen with probability higher than a certain threshold.[6]

The value of λ here determines how selective the associated Fishburn imputation really is. As such, the decision of which λ to adopt belongs to the analyst and should best be tailored to the problem at hand. For example, in highly noisy environments—e.g., when choices are made under time pressure, under disturbance, or with low incentives—the analyst may adopt a higher λ, discarding options that have a low probability of being chosen. In other contexts, a lower λ may be more reasonable to adopt.

Imputation with Real Data. In reality, of course, we do not observe $\mathbb{P}(\cdot, S)$ directly, but rather get information about it through finitely many observations. This brings us to the second stage of our construction: Even if we have decided to use a Fishburn imputation with a particular λ, our ultimate elicitation problem requires us decide whether or not to include an option x in the set of "choices" from S, given the empirical distribution of observed choices. This leads to a hypothesis testing problem with the null hypothesis

$$H_{0,x} : \mathbb{P}(x, S) \geq \lambda \max_{y \in S} \mathbb{P}(y, S)$$

which is to be tested by using empirical choice frequencies. At this junction we adopt the assumption that choice trials are independent and the probability of choice of any alternative in S is the same in each trial. In addition to λ, the associated test procedure depends, of course, on the sample size n (i.e., the total number of times we see the agent choose from S),

[5]This map was first introduced in Fishburn (1978), although with a different goal; see below.

[6]Put precisely, for any $\theta \in (0, 1)$, consider the function that maps any stochastic choice function $\mathbb{P}$ to the choice correspondence that declares $\{x \in S : \mathbb{P}(x, S) \geq \theta\}$ as the set of "choices" from any menu S. Easy examples show that this is distinct from any Fishburn imputation (unless the choice domain consists only of pairwise problems).

the number of times x is chosen in S, and the level of significance α of the test (to be chosen by the analyst). Given these parameters, we develop a statistical procedure to compute

$$\{x \in S : H_{0,x} \text{ is not rejected at the significance level } \alpha\}$$

to be designated as the set of all potential choices of the individual from the menu S. Put a bit more precisely, we propose a method of determining a threshold level $w_{\lambda,n}$ (that also depends on α) such that the set of choices from S is inferred as

$$\{x \in S : \text{the number of times } x \text{ is chosen in } S > w_{\lambda,n}\}.$$

As such, this procedure assigns to every data set pertaining to repeated choice trials a particular choice correspondence that depends on the observed data, the number of repetitions, as well as two parameters chosen by the analyst, namely, the significance level α and the threshold that would be used if $\mathbb{P}$ were observable, that is, λ.

Applications. To demonstrate the applicability of the elicitation method we develop in this paper, we present two empirical applications. In the first one, we pool the experimental pairwise choice data of Tversky (1969) and Regenwetter et al. (2011) to obtain a repeated (20 times) choice data set for 26 subjects, and use this to estimate a deterministic choice correspondence for each subject. There are many papers in the literature on boundedly rational choice that cites Tversky (1969) as providing evidence for cyclic *deterministic* choice behavior. But as Tversky (1969) only reports the relative choice frequencies in his study, and hence a stochastic choice function, this is not really warranted. After all, Tversky (1969) only tests for violations of Weak Stochastic Transitivity, not for cyclic preference relations. Our brief application here fills this gap, by inferring deterministic choice correspondences which can readily be tested for violations of the Weak Axiom of Revealed Preference (WARP). For certain parameter choices, we in fact found here that the latter violations are, while still significant, less pronounced than what Tversky (1969) found.

The statistical method we implement to estimate empirical choice correspondences becomes more involved when the menus at hand consist of more than two alternatives. To illustrate, in our second application we revisit the experimental pairwise choice data of Tversky (1972). This data is repeated for 8 subjects 30 times for menus of three alternatives, and 20 times for menus of two alternatives. We again used our method to estimate deterministic choice correspondence for each subject. While this is only a side issue for the present study, we also tested for violations of WARP in terms of these correspondences, and found that about 40% of them violate this axiom (under certain parameter choices).

Related Literature. While the issues of observability of choice correspondences are well-known, only a few papers have attempted to elicit them from data, and to our knowledge all have done so introducing novel experimental procedures instead of using standard (repeated) choice data. For instance, Bouacida (2019) asks subjects to choose from a set of alternatives, but allows them to choose multiple options and give them an additional (small) payment if they do so; in that case, the agent receives one of her choices randomly. Other papers use unincentivized additional questions after the choice to elicit the strength of preferences, and to identify indifferences. Some of these papers, notably, Costa-Gomes et al. (2019), use this information to construct choice correspondences. Whether one believes these procedures to be effective, or that they in fact introduce additional confounds, it is plain that they are tailored for particular experiments, and hence cannot be applied to typical choice data. (In particular, none of these procedures is applicable to data collected in past experiments with repeated choice trials, such as the ones we use in our empirical applications below.)

We should also note that some authors have used different experimental methods to elicit multiple choices, such as allowing subjects to use randomization devices[7] or choice deferrals.[8] As is well-known, however, these methods generate information that is markedly different from that needed for deriving a choice correspondence.[9]

Finally, we emphasize that the literature on revealed preference theory based on choice from budget sets, which started with the seminal work of Afriat (1967), does take into account the issue of limited observability of choices and allows for the possibility of (unobserved) choice correspondences. The papers that belong to this strand treat each observation as a selection from the demand set of the agent at a given price configuration, and do not attempt to construct the entire demand set at the associated budget. (See Chambers and Echenique (2016) for a review.) As such, they implicitly treat each observation as equally informative of one's demand correspondence, which relates it to the special case of Fishburn imputations with $\lambda = 0$, where each element chosen, however infrequently, is considered to be a choice.

2 Imputation of Choice Correspondences from Ideal Data

Throughout the paper, X stands for an arbitrarily fixed nonempty finite set with $|X| \geq 3$. We denote by $\mathfrak{X}$ the collection of all nonempty subsets of X and, for any positive integer k, by $\mathfrak{X}_k$ the collection of all subsets that contain k elements.

2.1 Choice Imputations

Choice Correspondences. By a **choice correspondence** on $\mathfrak{X}$, we mean a set-valued map $C : \mathfrak{X} \to \mathfrak{X}$ such that $C(S) \subseteq S$. The standard (if a bit ambiguous) interpretation is that $C(S)$ includes all feasible alternatives that the individual deems worth choosing. (Giving empirical content to this interpretation is in fact one of the main objectives of the present paper.) We denote the collection of all choice correspondences on $\mathfrak{X}$ by $\textsf{\textbf{cc}}(X)$.

[7] See Cohen, et al. (1985, 1987), Rubinstein (2002), Kircher, et al. (2013), Agranov and Ortoleva (2017, 2021), Dwenger et al. (2018), Miao and Zhong (2018), Cettolin and Riedl (2019), and Feldman and Rehbeck (2020).

[8] See Danan and Ziegelmeyer (2006), Sautua (2017), and Costa-Gomes et al. (2019).

[9] For example, Agranov and Ortoleva (2017, 2020) observe that many subjects are willing to randomize between two alternatives x and y. This is conceptually distinct from saying that both alternatives belong to these subjects' choice sets from the menu $\{x, y\}$; instead, it suggests merely that some agents may prefer a particular randomization over the given options (as it would be the case, for instance, if they possessed quasiconcave utility functions over lotteries). See Cerreia-Vioglio, et al. (2019) for more on this.

Stochastic Choice Functions. By a **stochastic choice function** on $\mathfrak{X}$, we mean a function $\mathbb{P} : X \times \mathfrak{X} \to [0,1]$ such that

$$\sum_{x\in S} \mathbb{P}(x,S) = 1 \quad \text{and} \quad \mathbb{P}(y,S) = 0$$

for every $S \in \mathfrak{X}$ and $y \in X \backslash S$. The collection of all such functions is denoted by $\mathsf{scf}(X)$.

For any $\mathbb{P} \in \mathsf{scf}(X)$, the map $x \mapsto \mathbb{P}(x,S)$ defines a probability distribution on S. From an individualistic perspective, we interpret this distribution by imagining that a decision maker has been observed making choices from the feasible menu S multiple times, and the relative frequency of the times x is chosen from S is $\mathbb{P}(x,S)$ in the limit, as the number of observations tends to infinity. Thus, $\mathbb{P}(x,S)$ is not an observable quantity, just like the probability of getting heads in a particular coin toss is not observable. Instead, from the viewpoint of an outside observer, it is a random entity. Put informally, it is *approximately* observable in the sense that any choice experiment that tracks the choices of the agent from S repeatedly provides a sample wherein the empirical value of the relative frequency of the times x is chosen in S is a strongly consistent estimator of $\mathbb{P}(x,S)$.[10] As we discussed in the introduction, we will first consider below how to derive choices from $\mathbb{P}$ as if this function is known, and only later account for the unobservability of $\mathbb{P}$.

Lastly, one extra bit of notation: Given a stochastic choice function $\mathbb{P}$ and $S \in \mathfrak{X}$, we put

$$M_{\mathbb{P}}(S) := \max_{z\in S} \mathbb{P}(z,S) \quad \text{and} \quad m_{\mathbb{P}}(S) := \min_{z\in S} \mathbb{P}(z,S)$$

That is, $M_{\mathbb{P}}(S)$ and $m_{\mathbb{P}}(S)$ are the choice probabilities of items in S with the maximum and the minimum likelihood, respectively.

Choice Imputations. At the center of our analysis is a map that assigns a choice correspondence to any stochastic choice function $\mathbb{P}$, that is, a map of the form

$$\Psi : \mathsf{scf}(X) \to \mathsf{cc}(X).$$

The only condition that we impose on this map at the outset is that

$$\mathbb{P}(x,S) = 0 \quad \text{implies} \quad x \notin \Psi(\mathbb{P})(S) \tag{1}$$

for any $S \in \mathfrak{X}$ and $\mathbb{P} \in \mathsf{scf}(X)$. This condition forbids designating an item that is never chosen in S as a choice from S. We refer to any Ψ that satisfies this property as a **choice imputation**. In words, a choice imputation Ψ is a method of transforming the behavior of an individual represented by a stochastic choice function into a deterministic choice correspondence. Loosely speaking, we wish this method to associate a choice correspondence $C_{\mathbb{P}}$ to $\mathbb{P}$ in such a way that, for any menu S, the set $C_{\mathbb{P}}(S)$ consists of all items in S that have a "significant" probability of being chosen in S, eliminating, for instance, items that are chosen by mistake, in a rush, etc..

[10] This statement is readily formalized by means of the Glivenko-Cantelli Theorem.

Example 1. An interesting choice imputation is one that includes anything chosen with positive probability in the associated choice set. Formally, this imputation, which we denote by Ψ_0, maps any $\mathbb{P} \in \textsf{scf}(X)$ to the choice correspondence $C_{\mathbb{P},0}$ on $\mathfrak{X}$ defined by

$$C_{\mathbb{P},0}(S) := \{x \in S : \mathbb{P}(x, S) > 0\}.$$

The choice imputation of Example 1 provides a natural starting point, and indeed, it is implicitly suggested by Mas-Colell, Whinston and Green (1995); see footnote 3. Nevertheless, it appears too inclusive. It is arguable that if x is chosen from $\{x, y\}$ with probability 0.001, it should not be included in the choice set of the agent at the menu $\{x, y\}$.

Example 2. For any menu S, one may wish to include in the set of all choices only those options in S with the maximum probability of being chosen. This method is captured by the choice imputation Ψ_1 which maps any $\mathbb{P} \in \textsf{scf}(X)$ to the choice correspondence $C_{\mathbb{P},1}$ on $\mathfrak{X}$ defined by

$$C_{\mathbb{P},1}(S) := \{x \in S : \mathbb{P}(x, S) \geq \mathbb{P}(y, S) \text{ for all } y \in S\}.$$

The choice imputations Ψ_0 and Ψ_1 are extreme members of an interesting one-parameter family.

Example 3. For any $\mathbb{P} \in \textsf{scf}(X)$ and $\lambda \in (0, 1]$, define the map $C_{\mathbb{P},\lambda} : \mathfrak{X} \to \mathfrak{X}$ by

$$C_{\mathbb{P},\lambda}(S) := \{x \in S : \mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S)\}.$$

In words, $C_{\mathbb{P},\lambda}(S)$ contains a feasible alternative $x \in S$ iff there are no alternatives in S that are chosen at least $\frac{1}{\lambda}$ times more frequently than x. For λ close to 1, it seems unexceptionable that we qualify the members of $C_{\mathbb{P},\lambda}(S)$ as potential "choices" of the agent from S; conversely, for λ close to 0, it makes sense to think of the members of $S \backslash C_{\mathbb{P},\lambda}(S)$ as objects that are chosen due to occasional mistakes. For any given $\lambda \in [0, 1]$, the map Ψ_λ defined on $\textsf{scf}(X)$ by $\Psi_\lambda(\mathbb{P}) := C_{\mathbb{P},\lambda}$ is a choice imputation (where $C_{\mathbb{P},0}$ is defined in Example 1). We refer to this as a **Fishburn imputation** with factor λ.[11]

When a set includes only two items, say x and y, it is easy to see that $x \in C_{\mathbb{P},\lambda}\{x, y\}$ iff $\mathbb{P}(x, \{x, y\}) \geq \frac{\lambda}{1+\lambda}$. That is, an item is deemed as a "choice" in a doubleton menu iff it is chosen with a probability above a fixed threshold. For larger sets, however, Fishburn imputations are more complex. Whether or not x is included in $C_{\mathbb{P},\lambda}(S)$ depends not only on the probability $\mathbb{P}(x, S)$, but also on the highest choice probability in S, namely, $M_{\mathbb{P}}(S)$. For a given probability of choosing x in S, that option will be included in the set of choices from S only when the maximum probability of choice is not too high. Thus, the criterion is more selective for sets in which some option is chosen with a very high chance, less selective for sets in which all options are chosen with low probability.[12]

[11] The correspondences $C_{\mathbb{P},\lambda}$ were first considered by Fishburn (1978) who sought the characterization of $\mathbb{P}$ such that for every $\lambda \in [0, 1]$, there is a (utility) function $u_\lambda : X \to \mathbb{R}$ with $C_{\mathbb{P},\lambda}(S) = \arg\max u_\lambda(S)$ for every $S \in \mathfrak{X}$. In turn, they were recently used by Ok and Tserenjigmid (2019, 2020) to produce rationality criteria for stochastic choice rules.

[12] It is thus easy to see that this criterion, like most models of stochastic choice, is not immune from adding duplicates into a menu. For example, consider x, y, y' where y and y' are duplicates, and suppose

The extent of selectivity of a Fishburn imputation depends on the value of λ, which is to be chosen by the analyst for the problem at hand. It may be reasonable to pick higher values of λ—that is, a more selective criterion—for environments in which there is noise, or more generally, when mistakes are expected. In those cases one may wish to disregard options that are not chosen with sufficiently high probability. On the other hand, lower values of λ may be appropriate when there is reason to consider objects chosen with low probability as genuine selections as well.

Remark 1. It may be of theoretical interest to compute the Fishburn imputations of some well-known models of stochastic choice. To illustrate, take any two real injective maps on X and any map $\theta : \mathfrak{X} \to (0, 1)$. The stochastic choice function $\mathbb{P}$ on X where

$$\mathbb{P}(x, S) := \theta(S)\mathbf{1}_{\arg\max u(S)}(x) + (1 - \theta(S))\mathbf{1}_{\arg\max v(S)}(x)$$

for any $x \in S$ and $S \in \mathfrak{X}$, is said to be a **dual random utility model**; this model has recently been characterized by Manzini and Mariotti (2018). For any $S \in \mathfrak{X}$ and any injection $f : X \to \mathbb{R}$, let $x(f, S)$ stand for the (unique) maximizer of f in S. Then, the Fishburn imputation of $\mathbb{P}$ (with factor λ) is readily computed. If $\theta(S) \geq \frac{1}{2}$, then $\Psi_\lambda(\mathbb{P})(S) = \{x(u, S)\}$ when $\frac{1-\theta(S)}{\theta(S)} < \lambda$ and $\Psi_\lambda(\mathbb{P})(S) = \{x(u, S), x(v, S)\}$ otherwise. And if $\theta(S) < \frac{1}{2}$, we have $\Psi_\lambda(\mathbb{P})(S) = \{x(v, S)\}$ when $\frac{1}{\lambda} < \frac{1-\theta(S)}{\theta(S)}$ and $\Psi_\lambda(\mathbb{P})(S) = \{x(u, S), x(v, S)\}$ otherwise. $\square$

In this paper we mostly work with Fishburn imputations, but there are several other types of choice imputations that may be useful in empirical work. For good measure, we next present a selection of such alternatives.

Example 4. Consider the map $\Psi : \mathbf{scf}(X) \to \mathbf{cc}(X)$ with

$$\Psi(\mathbb{P})(S) := \{x \in S : \mathbb{P}(x, S) \geq \min\{\theta, M_{\mathbb{P}}(S)\}\}, \quad S \in \mathfrak{X}$$

for some $\theta \in (0, 1)$. In any menu S, the map Ψ includes an item as a "choice" if either that item is chosen with a probability above a threshold θ, or if it is the item with the maximum likelihood of being chosen in S. Ψ is then a choice imputation, but it is not a Fishburn imputation.

Example 5. Define $c_{\mathbb{P}}(S) := C_{\mathbb{P},1}(S) \cup C_{\mathbb{P},1}(S \backslash C_{\mathbb{P},1}(S))$ (with the convention that $C_{\mathbb{P},1}(\varnothing) := \varnothing$) for any $S \in \mathfrak{X}$ and $\mathbb{P} \in \mathbf{scf}(X)$. (Thus, $c_{\mathbb{P}}(S)$ contains all alternatives in S that are the most probably chosen in S as well as those that are second most probably chosen.) The map Ψ defined on $\mathbf{scf}(X)$ by $\Psi(\mathbb{P}) := c_{\mathbb{P}}$ is a choice imputation (but again, it is not a Fishburn imputation).

Example 6. Let $\lambda : \mathbb{N} \to (0, 1]$ be any non-constant decreasing function, and consider the map $\Psi : \mathbf{scf}(X) \to \mathbf{cc}(X)$ with

$$\Psi(\mathbb{P})(S) := \{x \in S : \mathbb{P}(x, S) \geq \lambda(|S|)M_{\mathbb{P}}(S)\}, \quad S \in \mathfrak{X}.$$

$\mathbb{P}(x, \{x, y\}) = 0.2$, $\mathbb{P}(y, \{x, y\}) = 0.8$, while $\mathbb{P}(x, \{x, y, y'\}) = 0.2$, $\mathbb{P}(y, \{x, y, y'\}) = 0.4$, and $\mathbb{P}(y', \{x, y, y'\}) = 0.4$ (which may be natural since y and y' are duplicates). Then for certain values of λ we have $x \in C_{\mathbb{P},\lambda}(\{x, y\})$ but $x \notin C_{\mathbb{P},\lambda}(\{x, y, y'\})$, that is, adding duplicates affects the selection. Depending on one's view of mistakes, this may be a feature or a concern. In either case, this is shared by most models on stochastic choice data, and simply calls for the careful identification of choice items.

This is a choice imputation that acts over menus with the same size just like a Fishburn imputation, but it may use different factors over menus with different cardinalities. It may be useful if one subscribes to the view that it gets harder to achieve a given fraction of the maximum likelihood in larger menus.[13]

2.2 Foundations

Recall that in the first step of our analysis we presume that the actual stochastic choice function $\mathbb{P}$ of a given individual is known. (Or, equivalently, in this first step we suppose that the data is perfectly informative about the true stochastic choice function of the decision maker.) As such, our problem is to decide which sort of choice imputation to use in order to transform $\mathbb{P}$ into a deterministic choice correspondence.

It is plain that every choice imputation has its advantages and disadvantages. We thus start our analysis by looking at axiomatic ways of evaluating such procedures in the abstract. The postulates below are imposed on an arbitrarily given choice imputation Ψ; for ease of notation, we denote the value of Ψ at $\mathbb{P}$ by $C_{\mathbb{P}}$.

Anonymity. We begin by positing that if the choice probability distributions of two individuals with stochastic choice functions $\mathbb{P}$ and $\mathbb{Q}$ over a menu S are identical, then the value of the choice correspondences we attribute to them should be equal to each other at S. This seems like a reasonable property to impose on a choice imputation.

A. Anonymity. For every $\mathbb{P}, \mathbb{Q} \in \mathbf{scf}(X)$ and $S \in \mathfrak{X}$,

$$\mathbb{P}(\cdot, S) = \mathbb{Q}(\cdot, S) \quad \text{implies} \quad C_{\mathbb{P}}(S) = C_{\mathbb{Q}}(S).$$

A choice imputation that satisfies this property uses only the information about the choice behavior of a person in a menu S to infer her set of choices from that menu. While it constitutes a natural starting point, it would not be suitable for a method of imputation that looks at the choice behavior of an individual across all menus to decide how low a low a choice probability in S should be to count as a "mistake."

Imputations for Pairwise Choices. Suppose that, for a stochastic choice function $\mathbb{P}$ on X, we have somehow deemed the choice probability $\mathbb{P}(x, \{x, y\})$ large enough to include x in the choice set $C_{\mathbb{P}}\{x, y\}$. Let $\{z, w\}$ be another menu, and suppose z is primed to be chosen from $\{z, w\}$ even more frequently than x is from $\{x, y\}$, that is, $\mathbb{P}(z, \{z, w\}) \geq \mathbb{P}(x, \{x, y\})$. Since we have declared $x \in C_{\mathbb{P}}\{x, y\}$, consistency demands that we also declare $z \in C_{\mathbb{P}}\{z, w\}$. That is:

[13] Suppose we set $\lambda = \frac{1}{2}$, and consider the two menus $S := \{x, y\}$ and $T := \{x, z_1, ..., z_6, y\}$. If $\mathbb{P}(x, S) = \frac{1}{3}$ while $\mathbb{P}(x, T) = \mathbb{P}(z_1, T) = \cdots = \mathbb{P}(z_6, T) = 0.1$, we have $\mathbb{P}(x, S) \geq \frac{1}{2}\mathbb{P}(y, S)$ and $\mathbb{P}(x, T) < \frac{1}{2}\mathbb{P}(y, T)$, so the Fishburn imputation $\Psi_{1/2}$ deems x as a choice from S, but not from T, while one may argue that the latter conclusion is not acceptable, and choose to use a factor less than $\frac{1}{3}$ to work with in the context of menus that contain more than two alternatives.

B. Monotonicity across Pairwise Choice Data. For every $\mathbb{P} \in \mathbf{scf}(X)$ and $S, T \in \mathfrak{X}_2$,

$$x \in C_{\mathbb{P}}(S) \text{ and } \mathbb{P}(z, T) \geq \mathbb{P}(x, S) \quad \text{imply} \quad z \in C_{\mathbb{P}}(T).$$

We next consider a fairly weak form of continuity.

C. Continuity on Pairwise Menus. For every $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, ... \in \mathbf{scf}(X)$ and $S \in \mathfrak{X}_2$ with $S = C_{\mathbb{P}_k}(S)$ for each $k = 1, 2, ...,$

$$m_{\mathbb{P}_k}(S) \to m_{\mathbb{P}}(S) > 0 \quad \text{implies} \quad S = C_{\mathbb{P}}(S).$$

In words, given a doubleton menu S, if the choice sets imputed from each term of a sequence of stochastic choice functions include both elements of S, and if the associated smallest choice probabilities converge to a strictly positive value, then both elements of S should be included in the set of choices from S in the limit as well.[14]

Independence of Irrelevant Alternatives. The three properties we have considered above discipline the behavior of a choice imputation only with respect to pairwise choice situations. By contrast, the following property controls that behavior on arbitrary menus, and it does this by forcing the imputation be consistent with that used for pairwise choice problems.

To illustrate, suppose a person chooses x from a menu S with probability 0.2, while the item she chooses from S with the maximum likelihood is y, and that with probability 0.6. Suppose we also know that this individual chooses x against y in pairwise comparisons 25 percent of the time, that is, $\mathbb{P}(x, \{x, y\}) = .25$ and $\mathbb{P}(y, \{x, y\}) = .75$. How should then $C_{\mathbb{P}}(S)$ relate to $C_{\mathbb{P}}\{x, y\}$? Observe that we have chosen our numbers so that the relative probability of choosing x against y in the set S is the same as that in $\{x, y\}$ (for $\frac{.2}{.6} = \frac{.25}{.75}$). Thus, if we wish to abide by the principle of *Independence of Irrelevant Alternatives* (as formulated by, say, Luce (1959)), it would be natural to include x in $C_{\mathbb{P}}(S)$ if, and only if, $x \in C_{\mathbb{P}}\{x, y\}$. This principle maintains that an analyst may or may not find $\mathbb{P}(x, S)$ too small to include x in $C_{\mathbb{P}}(S)$, but whatever is her decision in this regard, it should be the same in the context of $C_{\mathbb{P}}\{x, y\}$.

D*. Independence of Irrelevant Alternatives*. For every $\mathbb{P} \in \mathbf{scf}(X)$, $S \in \mathfrak{X}$, and $x, y \in S$ such that $\mathbb{P}(y, S) = M_{\mathbb{P}}(S)$ and

$$\frac{\mathbb{P}(x, \{x, y\})}{\mathbb{P}(y, \{x, y\})} = \frac{\mathbb{P}(x, S)}{\mathbb{P}(y, S)},$$

we have $x \in C_{\mathbb{P}}(S)$ iff $x \in C_{\mathbb{P}}\{x, y\}$.

[14]The condition $m_{\mathbb{P}}(S) > 0$ is essential in the formulation of this property. For instance, if (x_k) and (y_k) are two sequences in X such that $\mathbb{P}_k(x, \{x, y\}) = \frac{1}{k}$ for each k, and $\mathbb{P}(x, \{x, y\}) = 0$, we do not wish $C_{\mathbb{P}}\{x, y\}$ to include x (otherwise $\mathbb{P} \mapsto C_{\mathbb{P}}$ would not define a choice imputation; recall (1)).

Independence of Irrelevant Alternatives type axioms (such as Luce's Choice Axiom) are much discussed in the literature on stochastic choice. They typically lead to ratio-scale representations and easy-to-use formulae in making probabilistic computations. They rule out various ways in which choices may be menu-dependent and are well-known to be sensitive to the presence of perfectly substitutable alternatives. Having said that, it seems desirable to explore the consequences of Property D before entertaining its relaxations that allow for menu dependent methods of choice imputation.

Remark 2. The four postulates we have considered above are logically independent. In particular, any of the choice imputations presented in Examples 4, 5 and 6 satisfies properties A, B and C, but not D.

Characterization Theorem. We are now ready to state our characterization result.

> ***Theorem 1.*** A choice imputation $\Psi : \mathsf{scf}(X) \to \mathsf{cc}(X)$ satisfies the properties A, B, C, and D if, and only if, it is a Fishburn imputation, that is, there exists a (unique) $\lambda \in [0, 1]$ such that $\Psi(\mathbb{P}) = C_{\mathbb{P},\lambda}$ for every $\mathbb{P} \in \mathsf{scf}(X)$.

Our postulates thus characterize the family of Fishburn imputations completely. For any choice imputation Ψ that satisfies these properties, there exists a unique $\lambda \in [0, 1]$ such that an item x is included in $C_{\mathbb{P}}(S)$ iff it is chosen with a probability at least as high as λ times the choice probability of an option in S with the maximum likelihood of being chosen, and this, for any menu S. As discussed above, different values of λ make the imputation more or less inclusive, and should be chosen by the analysis according to the problem at hand.

Remark 3. Independence of Irrelevant Alternatives (D) is not needed to identify the behavior of $\Psi(\mathbb{P})$ on *pairwise* choice situations. Put more precisely, if a choice imputation $\Psi : \mathsf{scf}(X) \to \mathsf{cc}(X)$ satisfies the properties A, B and C, it must act on pairwise choice problems as a Fishburn imputation would, that is, there exists a (unique) $\lambda \in [0, 1]$ such that $\Psi(\mathbb{P})(S) = C_{\mathbb{P},\lambda}(S)$ for every $S \in \mathfrak{X}_2$ and $\mathbb{P} \in \mathsf{scf}(X)$. (See Lemma A.3 in the Appendix.) This observation shows that the appeal of Fishburn imputations on pairwise choice situations is particularly pronounced. This is important because a vast majority of the experimental work that produce within-subject repeated choice data do so in the context of pairwise choice problems.

3 Imputation of Choice from Sample Data

3.1 The General Hypothesis Testing Problem

We have so far acted as if $\mathbb{P}$ is known, and looked at methods of deducing a choice correspondence from $\mathbb{P}$. In reality, however, all we have at best is a data set that reports a person's choice frequencies per menu for a finite number of observations. To explain the situation formally, let S be a menu in $\mathfrak{X}$, and suppose we have observed an individual make a choice from this menu n times. Then, the data at hand is in the form of a *realization* of the random variable

$$L_n(x, S) := \text{the number of times } x \text{ is chosen in } S \text{ in } n \text{ observations}$$

where $x \in S$. (Here n may of course depend on S.) Our problem is to impute a choice correspondence from every possible realization of this random variable as x varies over S.

Let $\ell_n(\cdot, S) := \frac{1}{n} L_n(\cdot, S)$, and note that every realization of $\ell_n(\cdot, S)$ is a probability distribution over the contents of S, but this distribution may be a poor representative of $\mathbb{P}(\cdot, S)$, especially when the sample size n is small. For example, if $n = 3$, observing x chosen once in a doubleton menu is hardly proof that it would be chosen about $\frac{1}{3}$ of the time from that menu if we had a larger sample. Thus, even if we have decided which choice imputation Ψ to employ if $\mathbb{P}$ were known, we cannot approach the problem of eliciting the choices of the individual from the menu S simply by applying Ψ at the given realization of $\ell_n(\cdot, S)$. This matter is not about the choice of an ideal Ψ; it is instead related to the finite-sample nature of real-world data.

In the abstract, once a certain choice imputation Ψ is agreed upon, one is confronted with a general hypothesis testing problem of the following form:

$$\begin{cases} H_{0,x}: & x \in C_{\mathbb{P}}(S) \\ H_{\mathrm{A},x}: & x \notin C_{\mathbb{P}}(S) \end{cases}$$

where, of course, $C_{\mathbb{P}}(S) = \Psi(\mathbb{P})(S)$. As usual, to solve this problem we need to agree on the probability of rejecting $H_{0,x}$ when this hypothesis is true, that is, on the *level of significance* of the test α. Once this is done, we have all the ingredients we need, namely, a realization of $L_n(\cdot, S)$ (the choice data from the menu S), Ψ (the method of imputation) and α (the significance level of the test), to impute the choice correspondence. Put precisely, given these ingredients, the set of choices of the person from the menu S is determined simply as the set of all x in S for which $H_{0,x}$ is not rejected at the significance level α given the choice data at hand.

In view of the analysis presented in the previous section, we study this problem in more concrete terms by choosing Ψ to be a Fishburn imputation. Thus, for a fixed $\lambda \in [0, 1]$ to be chosen by the analyst, our hypothesis testing problem (in the context of a specific menu S) is

$$\begin{cases} H_{0,x}: & \mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S) \\ H_{\mathrm{A},x}: & \mathbb{P}(x, S) < \lambda M_{\mathbb{P}}(S). \end{cases} \tag{2}$$

This leads us directly to the value of the target choice correspondence at S as

$$\begin{aligned} c_{\lambda,\alpha}(S) \;=\; \{x \in S : H_{0,x} & \text{ is not rejected at the significance level } \alpha \\ & \text{given the realization of } L_n(\cdot, S)\}. \end{aligned}$$

The next two subsections are devoted to the analysis of the test (2). But before we move to this analysis, it may be useful to take stock. The present approach decomposes the elicitation of the choice correspondence (from data) into two stages. First, a decision is made as to which choice imputation one would use if $\mathbb{P}$ were observable. In particular, if one is set on using a Fishburn imputation, a value for the factor λ is chosen (but of course one may choose to carry out the procedure in terms of several choices for λ). Second, one deals with issues that arise due to the finiteness of data sets by means of a statistical procedure that accounts for sampling errors. These two stages, and hence the choices for λ and α are kept separate as they pertain to different domains and are conceptually distinct. This being said, our procedure taken as a whole naturally generates a choice imputation that maps the observed choice frequencies (i.e., the realization of $L_n(\cdot, S)$) to the choice set $c_{\lambda,\alpha}(S)$, where n is the total number of choice trials from the menu S. (Observe that $c_{\lambda,\alpha}(S)$ is really a random set; for each realization of $L_n(\cdot, S)$, we obtain a particular choice set.)

An alternative approach would be to use the realization of ℓ_n itself as a stochastic choice function to compute $\Psi(\ell_n)(S)$ directly for a suitable Ψ. This seems less straightforward, however. While our approach allows the analyst to separately think of the "right" λ and α for the problem at hand, choosing an imputation method to be applied directly to the choice data requires one identify a rule that accounts for these aspects jointly. In particular, it does not seem obvious how the inclusivity of the rule should consistently evolve with the sample size, or how one could then account for sampling errors appropriately.

3.2 The Case of Pairwise Choice Data

We begin by studying the hypothesis testing problem (2) in the special, but important, case of pairwise choice situations, that is, when $|S| = 2$. In this case, it is possible to obtain precise, closed-form solutions for an arbitrary sample size n.

Fix an $S \in \mathfrak{X}_2$, and suppose λ and α are chosen. Let x be an element of S. Our problem is to decide whether or not we should declare x as a choice from S given the choice data, that is, a realization of $L_n(\cdot, S)$ that pertains to a particular decision maker. For ease of notation, let us put $p := \mathbb{P}(x, S)$, and rewrite (2) as:

$$\begin{cases} H_0: & p \geq \lambda M_{\mathbb{P}}(S) \\ H_{\mathrm{A}}: & p < \lambda M_{\mathbb{P}}(S). \end{cases}$$

Given that S contains two alternatives, a routine manipulation shows that $p \geq \lambda M_{\mathbb{P}}(S)$ iff $p \geq \frac{\lambda}{1+\lambda}$. Consequently, our hypothesis testing problem can be stated as:

$$\begin{cases} H_0: & p \geq \frac{\lambda}{1+\lambda} \\ H_A: & p < \frac{\lambda}{1+\lambda}. \end{cases}$$

Clearly, we need to make a sampling assumption to deal with this. We adopt the following standard hypothesis in this regard, which is maintained in the remainder of the paper.

> ***Assumption.*** The choice trials are independent and the probability of choice of any alternative in S is the same in each trial.

Now, we wish to obtain a threshold test of the form: Reject H_0 if the realization of $L_n(x, S)$ is smaller than some suitably chosen nonnegative integer $w_{\lambda,n}$. More precisely, given the desired level of significance α, we would like to find the largest integer $w_{\lambda,n}$ that satisfies

$$\mathrm{Prob}(L_n(x, S) \leq w_{\lambda,n}) \leq \alpha \quad \text{whenever } p \geq \tfrac{\lambda}{1+\lambda}. \tag{3}$$

(We do not make the dependence of $w_{\lambda,n}$ on α explicit in our notation only for expositional purposes.)

Let us recall the following elementary, and well-known, property of the binomial distribution.

Lemma 2. Let n be a positive integer, $\theta \in \{0, ..., n\}$, and $p, q \in [0, 1]$. For any two binomially distributed random variables u and v with parameters n and p, and n and q, respectively, we have $\text{Prob}(v \leq \theta) < \text{Prob}(u \leq \theta)$ whenever $p < q$.

As $L_n(x, S)$ is binomially distributed with parameters n and p, Lemma 2 entails that $\text{Prob}(L_n(x, S) \leq w_{\lambda,n})$ *when* $p > \frac{\lambda}{1+\lambda}$ is strictly smaller than $\text{Prob}(L_n(x, S) \leq w_{\lambda,n})$ *when* $p = \frac{\lambda}{1+\lambda}$. It follows that (3) holds iff

$$\text{Prob}(L_n(x, S) \leq w_{\lambda,n}) \leq \alpha \quad \text{when } p = \tfrac{\lambda}{1+\lambda}. \tag{4}$$

Consequently, the threshold we are after is

$$w_{\lambda,n} := \max \{\theta \in \{0, ..., n\} : \text{Prob}(u \leq \theta) \leq \alpha\} \tag{5}$$

where u is any binomially distributed random variable with parameters n and $\frac{\lambda}{1+\lambda}$. Thence, our test procedure rejects H_0 when the realization of $L_n(x, S)$ is less than or equal to $w_{\lambda,n}$. In other words, the procedure says:

Do not include x in $c_{\lambda,\alpha}(S)$ if the realization of $L_n(x, S)$ is less than or equal to $w_{\lambda,n}$,

that is,

$$c_{\lambda,\alpha}(S) := \{x \in S : \text{ the realization of } L_n(x, S) > w_{\lambda,n}\} \tag{6}$$

where $w_{\lambda,n}$ is determined by (5).

We note that $w_{\lambda,n}$, and thus the choice correspondence given in (6), is straightforward to compute in practice. At any rate, the supplementary material includes a python code to compute $w_{\lambda,n}$ for an combination of parameter values.

Some features of $w_{\lambda,n}$ are worth emphasizing. It is easy to compute that, when $n = 6$, we have $\frac{w_{\lambda,n}}{n} = 0$ for all λs considered, which means that any option chosen *at least once* is included in the choice correspondence. Put another way, any time we have at most 6 repetitions, and the level of significance is set at .05, the choice of λ is irrelevant and we include any item chosen at least once out 6 times in the (doubleton) choice set. This illustrates how our criterion tends to be inclusive in the face of (relatively) small sample. This is to be expected, because we include an item as a choice unless we can reject with sufficient confidence that it should be excluded. If we observe only 6 repetitions, we cannot reject that an option chosen only once out of six times, may in fact have a probability of choice as high as 50 percent (so this item is included in the choice set even for $\lambda = 1$.) Thus, in typical experimental data where less than six repetitions are not uncommon, the small sample could be a significant concern that overcomes the relevance of analysts' choices for λ.

The construction becomes more selective when, say, $n = 20$. The values of $\frac{w_{\lambda,20}}{20}$ are 0%, 5%, 10%, 20%, 25%, 25% for λ equal to .1, .3, .5, .7, .9, and 1, respectively. Then, an item chosen 6 times out 20 (so that its relative choice frequency is 30%) is always included in the choice set for any λ. But an item chosen 3 times out of 20 (so that its relative choice frequency is 15%) is included for $\lambda \leq 0.5$, but excluded for $\lambda \geq 0.7$. For $n = 50$, on the other hand, the corresponding values of $\frac{w_{\lambda,n}}{n}$ become 0%, 12%, 20%, 28%, 34%, 36%. In particular, an item chosen 19 out of 50 times is always included in the choice set (from a doubleton menu), even for $\lambda = 1$. But if an item is chosen 18 times, then it is excluded from the choice correspondence when $\lambda = 1$. When $n = 50$ and $\lambda \leq 0.1$, every item chosen at least once is included in the choice set.

These examples suggest a few properties of the threshold $w_{\lambda,n}$ as a function of λ and n. First, and this is a consequence of the fact that the the cdf of the binomial distribution with parameters n and p is decreasing in n, $w_{\lambda,n}$ is increasing in n. By contrast, $\frac{w_{\lambda,n}}{n}$ is decreasing in n only on average. Moreover, $\frac{w_{\lambda,n}}{n}$ converges (from below) to $\frac{\lambda}{1+\lambda}$ as n tends to infinity, which is the target threshold. When n is small, it is harder to confidently reject the hypothesis that the true underlying frequency is above a threshold, making our criterion more inclusive. As n grows, however, the small-sample nature of the data becomes less problematic, and the sample threshold converges to the one we would have with ideal data.

On the other hand, it is plain from Lemma 2 and the definition of $w_{\lambda,n}$ that this function is increasing in λ, that is, we have $w_{\lambda,n} \leq w_{\lambda',n}$ whenever $0 \leq \lambda \leq \lambda' \leq 1$ (for any n). In particular, $w_{\lambda,n} \leq w_{1,n}$ for all $\lambda \in [0,1]$ and $n \in \mathbb{N}$. This is, of course, in the nature of things. As λ increases, it becomes harder to be admitted as a choice, and the estimated choice correspondence gets "thinner." The hardest test obtains at $\lambda = 1$, which yields the most exclusive (empirical) choice correspondence.

We note that the present test always yields a nonempty-valued choice correspondence.

> ***Proposition 3.*** For any $n \in \mathbb{N}$, $S \in \mathfrak{X}_2$ and $\lambda \in [0,1]$, we have $c_{\lambda,\alpha}(S) \neq \varnothing$ (at every realization of $L_n(\cdot, S)$).

Proof. Fix any $n \in \mathbb{N}$ and $\lambda \in [0,1]$. Note that if u is a binomially distributed random variable with parameters n and $\frac{1}{2}$, we have

$$\text{Prob}(u \leq \lfloor \tfrac{n}{2} \rfloor) = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n!}{i!(n-i)!} 2^{-i} 2^{-(n-i)} \approx \tfrac{1}{2} > \alpha,$$

so $w_{\lambda,n} \leq w_{1,n} < \lfloor \frac{n}{2} \rfloor \leq \frac{n}{2}$. If there were realizations of $L_n(x,S)$ and $L_n(y,S)$, say, L_x and L_y, respectively, for which $c_{\lambda,\alpha}(S) = \varnothing$, then both L_x and L_y would be less than $w_{\lambda,n}$, whence $n = L_x + L_y \leq 2w_{\lambda,n} < n$, a contradiction. ■

We show next that, with high probability, our procedure infers a choice correspondence that contains the "true" choice correspondence of the subject:

> ***Proposition 4.*** For any $n \in \mathbb{N}$, $S \in \mathfrak{X}_2$ and $\lambda \in [0,1]$, we have $C_{\lambda,\mathbb{P}}(S) \subseteq c_{\lambda,\alpha}(S)$ with probability at least $1 - 2\alpha$.

Proof. Fix any $n \in \mathbb{N}$ and $\lambda \in [0,1]$. Let $p := \mathbb{P}(x,S)$ so that the choice probability of y in S is $1-p$. Without loss of generality, assume that y is more likely to be chosen in S than x, that is, $p \leq 1-p$. Let A and B denote the events that "$x \notin c_{\lambda,\alpha}(S)$" and "$y \notin c_{\lambda,\alpha}(S)$," respectively. By Proposition 3, these events are disjoint, so $\text{Prob}(A \cup B) = \text{Prob}(A)+ \text{Prob}(B)$. Now, if $\lambda(1-p) \leq p$, we have $C_{\lambda,\mathbb{P}}(S) = \{x,y\}$, so our claim fails iff $A \cup B$ occurs. But given that $p \geq \frac{\lambda}{1+\lambda}$, the definition of $w_{\lambda,n}$ entails $\text{Prob}(A) = \text{Prob}(L_n(x,S) \leq w_{\lambda,n}) \leq \alpha$, and similarly, $\text{Prob}(B) \leq \alpha$, whence $\text{Prob}(A \cup B) \leq 2\alpha$. If, on the other hand, $\lambda(1-p) > p$, we have $C_{\lambda,\mathbb{P}}(S) = \{y\}$, so our claim fails iff B occurs. As $\text{Prob}(B) \leq \alpha$, our proof is complete. ■

Remark 4. Describing the exact nature of the function $w_{\lambda,n}$ is not a trivial task, but it can be shown that this function is increasing in n, and it satisfies $\frac{w_{\lambda,n}}{n} \leq \frac{\lambda}{1+\lambda}$ for all $\lambda \in [0,1]$ and $\alpha \in [0, \frac{1}{4}]$ for sufficiently large n. (This is not an asymptotic result; it holds as soon as $n > \frac{1+\lambda}{\lambda}$.) Finally, and perhaps most important, we have $\lim \frac{w_{\lambda,n}}{n} = \frac{\lambda}{1+\lambda}$, that is, our test statistic $\frac{w_{\lambda,n}}{n}$ is a consistent estimator of the true threshold $\frac{\lambda}{1+\lambda}$ regardless of the choice of λ and α. (Proofs are given in the Appendix.)

Remark 5. Experimental datasets are almost always quite small, so one cannot do better than applying the above small-sample procedure as stated. Nevertheless, we should note that if n is large and λ is not small, one can approximately determine the threshold $w_{\lambda,n}$ using the normal distribution. A long-standing convention in statistics is that one can safely approximate a binomial distribution with parameters n and p with the normal distribution provided that $np > 5$. Thus, in the present setup, so long as $\frac{n\lambda}{1+\lambda} > 5$ holds, we have

$$\alpha = \text{Prob}(L_n(x,S) \leq w_{\lambda,n}) \approx \text{Prob}\left(\ell_n(x,S) \leq \tfrac{w_{\lambda,n}+0.5}{n}\right) \approx \Phi\left(\frac{\frac{w_{\lambda,n}+0.5}{n} - \frac{\lambda}{1+\lambda}}{\sqrt{\frac{\lambda}{n(1+\lambda)^2}}}\right)$$

by continuity correction. (Here Φ stands for the standard normal (cumulative) distribution function.) From this approximation, we readily get a normal-based approximation for the critical value we are after as

$$w_{\lambda,n} = n\tfrac{\lambda}{1+\lambda}\left(1 + \frac{\Phi^{-1}(\alpha)}{\sqrt{n\lambda}}\right) - 0.5.$$

3.3 The Case of Choice Data with Menus of Arbitrary Size

We now turn to consider larger choice sets. Let k be any integer with $k \geq 3$, take any $S \in \mathfrak{X}_k$, and suppose again that λ and α are chosen. For concreteness, we enumerate the menu S as

$$S = \{x_1, ..., x_k\},$$

and simplify the notation by setting

$$p_i := \mathbb{P}(x_i, S) \quad \text{and} \quad \mathcal{L}_i := L_n(x_i, S), \quad i = 1, ..., k.$$

Thus, p_i is the (unknown) probability of the agent choosing x_i from the menu S, while $\mathcal{L}_i$ is the number of times the individual has been observed to choose the item x_i from S in a choice experiment that is repeated n times. Obviously, $(p_1, ..., p_k)$ belongs to the $(k-1)$-dimensional unit simplex, while $\mathcal{L}_1 + \cdots + \mathcal{L}_k = n$. Moreover, under our Assumption, we have

$$\text{Prob}(\mathcal{L}_1 = a_1, ..., \mathcal{L}_k = a_k) = \frac{n!}{a_1! \cdots a_k!} p_1^{a_1} \cdots p_k^{a_k}$$

where $(a_1, ..., a_k) \in \{0, ..., n\}^k$ and $a_1 + \cdots + a_k = n$. Thus, the distribution of $(\mathcal{L}_1, ..., \mathcal{L}_k)$ is multinomial with parameters n and $p_1, ..., p_k$ (but note that this distribution is singular in the sense that the covariance matrix of $(\mathcal{L}_1, ..., \mathcal{L}_k)$ is of rank $k-1$, due to the restriction $\mathcal{L}_1 + \cdots + \mathcal{L}_k = n$). Consequently, as proved in Chapter 35 of Johnson, Kotz and Balakrishnan (1997), we have:

Lemma 5. (***a***) $\mathcal{L}_1$ is binomially distributed with parameters n and p_1;

(***b***) If $p_1 = 1$, then $\mathcal{L}_1 = n$ and $\mathcal{L}_2 = \cdots = \mathcal{L}_k = 0$ almost surely. If $p_1 < 1$, then for any $(a_1, ..., a_k) \in \{0, ..., n\}^k$ with $a_1 + \cdots + a_k = n$,

$$\text{Prob}(\mathcal{L}_2 = a_2, ..., \mathcal{L}_k = a_k \mid \mathcal{L}_1 = a_1) = \frac{(n-a_1)!}{a_2! \cdots a_k!} \prod_{i=2}^{k} \left(\frac{p_i}{1-p_1} \right)^{a_i};$$

that is, conditional on $\mathcal{L}_1 = a_1$, $(\mathcal{L}_2, ..., \mathcal{L}_k)$ is multinomially distributed with parameters $n - a_1$ and $\frac{p_2}{1-p_1}, ..., \frac{p_k}{1-p_1}$.

Without loss of generality, we may state the null hypothesis of our problem in terms of the item x_1 as:

$$H_0 : p_1 \geq \lambda \max\{p_1, ..., p_k\},$$

or equivalently, as

$$H_0 : np_1 \geq \lambda \max\{np_2, ..., np_k\},$$

for any positive integer n. To obtain a threshold test for this hypothesis, we consider the random variable

$$V_{a_1} := \mathcal{L}_2 \vee \cdots \vee \mathcal{L}_k$$

conditional on $\mathcal{L}_1 = a_1$, where $a_1 \in \{0, ..., n\}$. In words, V_{a_1} is the maximum value of the random variables $\mathcal{L}_2, ..., \mathcal{L}_k$ given that x_1 is chosen from S exactly a_1 many times. By Lemma 5,

$$\begin{aligned} \text{Prob}(V_{a_1} < v) &= \text{Prob}(\mathcal{L}_2 < v, ..., \mathcal{L}_k < v \mid \mathcal{L}_1 = a_1) \\ &= \sum_{(a_2,...,a_k) \in S(v;a_1)} \frac{(n-a_1)!}{a_2! \cdots a_k!} \prod_{i=2}^{k} \left(\frac{p_i}{1-p_1} \right)^{a_i} \end{aligned}$$

where $S(v; a_1)$ is the set of all $(k-1)$-vectors $(a_2, ..., a_k)$ of nonnegative integers such that $a_i < v$ for each $i = 2, ..., k$, and $a_2 + \cdots + a_k = n - a_1$.

We now define our test statistic (that depends on the chosen λ) as

$$W_\lambda := \mathcal{L}_1 - \lambda \left(\mathcal{L}_2 \vee \cdots \vee \mathcal{L}_k \right), \tag{7}$$

and note that this is a simple random variable whose (finite) range we denote by $\text{rng}(W_\lambda)$.[15] Our test will reject the null hypothesis H_0 when the observed value of W_λ is smaller than or

[15] To be precise, let $J_0 := \{0\}$, and for any positive integer N, set $J_N := \{N, N-1, ..., \lfloor \frac{N}{k} \rfloor\} \backslash \{0\}$. Then, $\text{rng}(W_\lambda)$ is contained within the finite set $\bigcup_{i=0}^{n} (i - \lambda J_{n-i})$.

equal to a certain critical value. To derive this critical value, note that

$$\begin{aligned}
\text{Prob}(W_\lambda \leq w) &= 1 - \text{Prob}(W_\lambda > w) \\
&= 1 - \text{Prob}(\mathcal{L}_2 \vee \cdots \vee \mathcal{L}_k < \tfrac{1}{\lambda}(\mathcal{L}_1 - w)) \\
&= 1 - \sum_{a_1=0}^{n} \text{Prob}(V_{a_1} < \tfrac{1}{\lambda}(\mathcal{L}_1 - w) \mid \mathcal{L}_1 = a_1)\text{Prob}(\mathcal{L}_1 = a_1) \\
&= 1 - \sum_{a_1=0}^{n} \left(\sum_{(a_2,\ldots,a_k)\in S\left(\frac{a_1-w}{\lambda};a_1\right)} \frac{(n-a_1)!}{a_2!\cdots a_k!} \prod_{i=2}^{k} \left(\frac{p_i}{1-p_1}\right)^{a_i} \right) \frac{n! p_1^{a_1}(1-p_1)^{n-a_1}}{a_1!(n-a_1)!} \\
&= 1 - \sum_{a_1=0}^{n} \sum_{(a_2,\ldots,a_k)\in S\left(\frac{a_1-w}{\lambda};a_1\right)} \frac{n!}{a_1! a_2! \cdots a_k!} \prod_{i=1}^{k} p_i^{a_i}
\end{aligned}$$

for any real number w. This completely characterizes the distribution of W_λ for any probability vector $(p_1, ..., p_k)$.

We now use small values of W_λ to form a critical region of the form $\{W_\lambda \leq w\}$ where we choose w as the largest real number with $\text{Prob}(W_\lambda \leq w) \leq \alpha$ when H_0 is true, or more precisely, the largest $w \in \mathbb{R}$ such that

$$\sup_{\substack{(p_1,\ldots,p_k)\in\Delta^{k-1} \\ p_1 \geq \lambda \max\{p_2,\ldots,p_k\}}} \text{Prob}(W_\lambda \leq w) \leq \alpha$$

with α being the desired level of significance. Again, this is the same thing as choosing w as the largest number in the range of W_λ with

$$\text{Prob}(W_\lambda \leq w) \leq \alpha \quad \text{when} \quad p_1 = \lambda \max\{p_2, ..., p_k\}.$$

We may thus define our test procedure in precise terms as follows:

STEP 1. Estimate p_i by $\hat{p}_i := \ell_n(x_i, S)$, $i = 2, ..., k$;

STEP 2. Set $\hat{p}_1 := \lambda \max\{\hat{p}_2, ..., \hat{p}_k\}$;

STEP 3. Put $\eta := \hat{p}_1 + \cdots + \hat{p}_k$. If $\eta = 0$, do not reject the H_0, and thus include x_1 in $c_{\lambda,\alpha}(S)$.[16] If $\eta > 0$, normalize the estimated probabilities as $\hat{q}_i := \frac{1}{\eta}\hat{p}_i$, $i = 1, ..., k$;[17]

[16] This is an exceptional case in which the agent is observed to choose x_1 from S in every repetition.

[17] A clarification is in order here. Given the setup of our problem, "assuming that the null hypothesis H_0 is true" is a nontrivial statement, as it pertains to the (unknown) probabilities $p_1, ..., p_k$. Our procedure uses the sample relative frequencies as proxies for $p_2, ..., p_k$, and to account for the null hypothesis being true, sets p_1 as λ times the largest of these frequencies. As the resulting numbers, $\hat{p}_1, ..., \hat{p}_k$, need not add up to 1, we then normalize these numbers to obtain a probability vector $(\hat{q}_1, ..., \hat{q}_k)$ with respect to which the null hypothesis holds, that is, $\hat{q}_1 := \lambda \max\{\hat{q}_2, ..., \hat{q}_k\}$. The procedure uses the probability vector $(\hat{q}_1, ..., \hat{q}_k)$ to compute the distribution of the sample test statistic $\hat{W}_\lambda$.

STEP 4. Compute

$$w_{\lambda,n} := \max\left\{w \in \text{rng}(W_\lambda) : \text{Prob}(\hat{W}_\lambda \leq w) \leq \alpha\right\}$$

where $\hat{W}_\lambda$ is the random variable whose cumulative distribution is given as

$$\text{Prob}(\hat{W}_\lambda \leq w) = 1 - \sum_{a_1=0}^{n} \sum_{(a_2,\ldots,a_k)\in S\left(\frac{a_1-w}{\lambda};a_1\right)} \frac{n!}{a_1!a_2!\cdots a_k!} \prod_{i=1}^{k} \hat{q}_i^{a_i};$$

STEP 5. Reject H_0 if $\ell_n(x_1, S) - \lambda \max\{\hat{p}_2, ..., \hat{p}_k\} \leq \frac{1}{n} w_{\lambda,n}$, or equivalently, do not include x_1 in $c_{\lambda,\alpha}(S)$ if the realization of $\mathcal{L}_n(x_1, S) - \lambda\mathcal{L}_n(x_i, S)$ is less than or equal to $w_{\lambda,n}$ for each $i = 2, ..., k$.

This procedure generalizes the one we developed in Section 3.2, and it allows one infer a choice correspondence over menus of any size, provided that the analyst has decided on which values to use for λ and α.

4 Applications

To illustrate the elicitation method developed above, we now apply it to infer empirical choice correspondences from existing data. We consider two famous within-subject repeated choice experiments. For binary choices, we use Experiment 1 of Tversky (1969) which was originally designed for testing the transitivity of preferences. For ternary choices, we use the main experiment of Tversky (1972) which was designed to test for choice by elimination by aspects.

4.1 Estimation for Pairwise Choice Situations

In this section we consider Experiment 1 of Tversky (1969). The original experiment had only 8 subjects, but the same experiment was later replicated by Regenwetter et al. (2011) with 18 additional subjects.[18] We pool these two data sets to obtain a choice data set of 26 individuals, and implement our method to infer a deterministic choice correspondence for each of them.

The Original Experiment. The goal of Tversky's experiment was to evaluate the rationality of individuals by focusing on the transitivity of their choices from doubleton choice problems. It recorded the choice from all pairs of 5 gambles, named, $a, b, ..., e$, where each trial was repeated 20 times (with additional 'decoy' choices in between), thereby generating a sample stochastic choice data. Each gamble was characterized by a probability of winning and an amount won. The gambles with names adjacent in the alphabet had similar probabilities of winning, while the difference of their payoffs were more pronounced. Only across gambles

[18]The only differences in the replication are the use of computers and updated payoffs (as decades have passed), and the implementation of the experiment in only one session (as opposed to Tversky's five sessions).

with names further apart in the alphabet the probability of winning differed substantially.[19] Tversky conjectured that this would lead to failure of rationality by generating violations of transitivity of choices. To investigate this, he tested whether the empirical stochastic choice function of the subjects satisfied Weak Stochastic Transitivity (WST). He found that this property failed significantly for 5 out of the 8 subjects (62%).[20]

Part of the fame of Tversky's finding is that it provides evidence of cyclical choice behavior. Indeed, it is routinely cited as evidence of cyclical choice in the recent literature on boundedly rational *deterministic* choice (cf. Manzini and Mariotti (2007), Masatlioglu, Nakajima, and Ozbay (2012), and Tserenjigmid (2015)). On closer scrutiny, this is not really warranted because Tversky's experiment provides stochastic choice data, not data on one's choice correspondences. To see if it indeed provides evidence of widespread violation of WARP, one needs to infer subjects' choice correspondences from Tversky's data. We take on precisely this exercise here by using the method prescribed in Sections 2 amd 3.2 above.

Applying the procedure. We wish to apply our procedure to infer a choice correspondence for each subject, and then test for rationality by checking if the elicited correspondences satisfy WARP. Maintaining $\alpha = .05$, we consider values of $\lambda \in \{0, .3, .5, .7, 1\}$. Panel (a) of Table 1 includes, for each λs, the fraction of subjects whose computed choice correspondence abides by WARP, as well as the value of the threshold $w_{\lambda,20}$ used in our procedure. The former goes from 57.7%, with $\lambda = 0$, to 30.8%, when $\lambda = 1$.[21]

λ	satisfy WARP	$w_{\lambda,20}$
0	57.7%	0
.3	53.8%	1
.5	46.2%	2
.7	38.5%	4
1	30.8%	5

Table 1.a

λ	violates both WARP and WST	violates only WST	violates only WARP	no violation
0	15.4%	38.5%	26.9%	19.2%
.3	23.1%	30.8%	23.1%	23.1%
.5	34.6%	19.2%	19.2%	26.9%
.7	46.2%	7.7%	15.4%	30.8%
1	53.8%	0%	15.4%	30.8%

Table 1.b

Weak Stochastic Transitivity is violated by about 54% of the subjects in this data set. Importantly, however, focusing on WARP and Weak Stochastic Transitivity capture different types of violations, which are neither nested nor perfectly correlated. To see this, Panel (b)

[19] If we identify gambles by (p, x), for a probability p of winning an amount x, the gambles were: $a = (\frac{7}{24}, 5)$, $b = (\frac{8}{24}, 4.75)$, $c = (\frac{9}{24}, 4.50)$, $d = (\frac{10}{24}, 4.25)$, $e = (\frac{11}{24}, 4)$. When comparing, for example, a and b, the difference between payoffs (5 vs. 4.75) seems more relevant than the difference between probabilities ($\frac{7}{24}$ vs. $\frac{8}{24}$).

[20] Recall that a stochastic choice function satisfies *Weak Stochastic Transitivity* if $P(x, \{x, y\}) \geq .5$ and $P(y, \{y, z\}) \geq .5$ imply $P(x, \{x, z\}) \geq .5$. Subsequent literature questioned the appropriateness of the statistical tests used in Tversky (1969), and argued that more apt tests fail to find significant violations of Weak Stochastic Transitivity; cf. Iverson and Falmagne (1985) and Regenwetter et al. (2011). As mentioned in the latter, these misgivings are rarely discussed, which is particularly striking given the prominence of Tversky's original paper.

[21] The fact that the fraction of rational subjects decreases with λ is a feature of this data but not a prediction of the model. Indeed, as λ increases, fewer items are included in the choice correspondences, which may eliminate previous violations of WARP but may well add new ones. For example, in the data, subject #17 satisfies WARP with $\lambda = .3$ but violates it with $\lambda = .5$, while subject #5 violates WARP with $\lambda = .3$ but satisfies it with $\lambda = .5$.

of Table 1 reports the fraction of subjects that violate each of the conditions, separately and jointly. Clearly, there are sizable fractions of subjects who violate one condition but not the other.

Weak Stochastic Transitivity vs. WARP. There is a formal, albeit superficial, relation between tests of Weak Stochastic Transitivity and WARP in the case of pairwise choice situations. Suppose we use our method to infer choice correspondences, but (i) we take $\lambda = 1$; and (ii) we treat the observed data as if it were $\mathbb{P}$, that is, we disregard sampling errors. Then, an alternative x belongs to the computed choice correspondence if, and only if, it is chosen at least half of the times. Thus, testing transitivity of the choice correspondence becomes identical to testing Weak Stochastic Transitivity.

However, the choice of $\lambda = 1$ seems quite extreme. In the case of pairwise choice problems that are repeated 20 times, it seems quite unreasonable to rule out an option that is chosen, say, 30 percent of the time as a choice. (Even a standard utility-maximizer must break his indifference between two options in some way, and this may well yield her choose one option only 30% of the time.) Moreover, it seems desirable to account for sampling errors in data directly. Doing this yields multi-valued choice correspondences, and draws a markedly different picture that allows us derive further information from the data.[22] For example, applying our method with $\lambda = .3$ yields choice correspondences about 54% of which satisfy WARP. By contrast, only 46% of the (sample) stochastic choice correspondences satisfy Weak Stochastic Transitivity.[23] It seems to us that analyzing choice data by employing both approaches (that is, inferring a deterministic choice correspondence as well as deducing a stochastic choice function) yields better insight into the nature of data.

4.2 Estimation with Larger Choice Sets

We now turn to an illustration of our procedure in the context of larger choice sets, and consider the main experiment of Tversky (1972). In this experiment eight subjects were asked to make a choice from menus with three options, each repeated 30 times over the course of 12 sessions, and from menus with options, each repeated 20 times. Choices were made in three distinct domains. The first involved random dot patterns, for which subjects were instructed to choose the pattern with the most dots. The second involved college applicants, for which subjects were instructed to choose the most promising. Finally, the third one consisted of monetary lotteries, for which subjects were asked to choose the gamble they preferred. Choices in this third domain were incentivized. We refer to Tversky (1972) for details.

For each domain, Tversky (1972) reports the fraction of choices for the set of three elements, and for two out of the three subsets of two elements. To illustrate the typical application of the statistical test devised in Section 3.3, let us consider the estimation of the choice

[22] Again, we are not the first to take issue with how sampling errors are handled by Tversky in his original paper, and to provide a different take on his own findings. See, among others, Iverson and Falmagne (1985), and Regenwetter et al. (2011).

[23] Note that, in the abstract, there is no reason for either approach to favor transitivity. Expanding the values of a choice correspondence may increase or decrease the violations of WARP. Indeed, Table 1 shows that there are non-trivial fractions of subjects who violate WARP but not Weak Stochastic Transitivity, and vice versa.

correspondence of subject 1 over monetary lotteries (domain 3). In this illustration, we choose $\lambda = 0.6$, and set the significance level of the test at $\alpha := 0.05$ as usual. This individual has chosen x from $\{x, z\}$ 7 out of 20 times, and y from $\{y, z\}$ 10 out of 20 times. Applying the pairwise choice test of Section 3.2 readily shows that we cannot reject the hypothesis that the probability of choice of any one alternative in these problems is at least 0.375 $(= \frac{.6}{1+.6})$, so, where we write c for $c_{.6,.05}$ for simplicity, we have $c\{x, z\} = \{x, z\}$ and $c\{y, z\} = \{y, z\}$.

On the other hand, from the menu $S = \{x, y, z\}$, this person has chosen x 2 times, y 13 times and z 15 times. Thus, Steps 1 and 2 of our test procedure (for testing if x should be included in $c(S)$ or not) yields $\hat{p}_1 = 0.6 \max\{\frac{13}{30}, \frac{15}{30}\} = \frac{9}{30}$, $\hat{p}_2 = \frac{13}{30}$ and $\hat{p}_3 = \frac{15}{30}$. Consequently, in Step 3, we find $\eta = \frac{37}{30} = 1.23$, and hence, dividing $\hat{p}_i$s by 1.23, we arrive at the estimated probabilities $\hat{q}_1 \approx 0.24$, $\hat{q}_2 \approx 0.35$ and $\hat{q}_3 \approx 0.40$. The bulk of the calculation then takes place in Step 4 with these probabilities. In this step we find $\text{Prob}(\hat{W}_{0.6} \leq -7) \approx 0.024$ and $\text{Prob}(\hat{W}_{0.6} \leq -6) \approx 0.052$; the critical value $w_{.6,30}$ is thus determined as -7. Finally, in Step 5, we observe that the realization of $\mathcal{L}_{30}(x, S) - 0.6 \max\{\mathcal{L}_{30}(y, S), \mathcal{L}_{30}(z, S)\}$ is less than or equal to $w_{.6,30}$ – in this case, coincidentally, these numbers are the same – and hence reject the null hypothesis, that is, do not include x in $c(S)$. Repeating this analysis for y and z (playing the role of x), shows that we cannot reject the null hypothesis in these cases, so we conclude: $c\{x, y, z\} = \{y, z\}$. In particular, the inferred choice correspondence violates WARP.

By repeating this procedure for all subjects, we have estimated the choice correspondences of the involved individuals across all three domains, and with a variety of choices for λ (but taking the size of the test as 0.05 throughout). Our findings are fairly consistent across the three domains. For instance, with $\lambda = 0.6$, 2 out of 8 subjects violate WARP in the case of dots and college applicants; only one (namely, the subject we considered in the previous paragraph) in the case of gambles. Moreover, we find that different subjects violate rationality across the three domains: overall, 37.5% of the subjects violate WARP at least once.

The present exercise is meant only to serve as an illustration of how our statistical imputation method could be applied in practice. But it also points to what sort of decision-theoretic inquiries can be addressed with this method. Once choice correspondences are estimated in different domains, one can test whether rationality of an individual is stronger in some environments vs. others: for example, whether violations increase or decrease in settings in which options are more complex, when more details are given, or when alternatives are less familiar to the subject. One could then also test whether violations are committed by the same subject across all domains or whether different participants exhibit different degrees of compliance depending on the domain, e.g., some exhibiting many violations with lotteries but few with, say, food items. Finally, it would also allow us test for the "source" of one's violations of WARP, such as indecisiveness, presence of consideration sets, reference dependence, aspiration effects, changing preferences, etc., and this, again, across choice domains of different nature.

Remark 6. It is plain that the "inclusiveness" of our test procedure crucially depends on the choice of λ; lower values of λ yield more inclusive tests. We should emphasize that this also depends on the sample size at hand. To illustrate, take the case of subject 1 we considered above; this person chose x only 2 times out of thirty from the menu $S = \{x, y, z\}$, while choosing y 13 times and z 15 times. If we then take $(\frac{2}{30}, \frac{13}{30}, \frac{15}{30})$ as the actual choice probabilities of these items, then any Fishburn imputation with $\lambda \geq 0.2$ would not include x

in the set of choices from S, while that with $\lambda = 0.1$ would. The statistical test we propose, however, is more inclusive due to its account of sampling errors.[24] Indeed, as we have seen above, our test does not include x in the set of choices from S when $\lambda = 0.6$, but it would certainly include it in $c_{\lambda,.05}(S)$ for smaller values of λ. This may be a reason for working with larger values of λ when the data at hand come from a small number of repetitions. For data sets with large number of repetitions, the sampling errors are likely to become less influential, and the "inclusiveness" of the test would rest more squarely on the choice of λ.

5 Conclusion

Despite being perhaps the most fundamental primitive of microeconomics, the values of choice correspondences are not observable. Barring some specially designed experiments, all we can observe in general is a single choice made by an individual at a given time, and not the *set* of all her potential choices. In this paper we propose a method to "compute" a choice correspondence using data that come in the form of repeated observations of choices made by a decision maker.

Our approach constructs one's choice correspondence in two stages. First, the analyst needs to decide how to impute the choice correspondence if she had access to an 'ideal' dataset that provides the true choice probability of each option. There is no unexceptionable method of doing this, but we have underscored here a one-parameter family of choice imputations. These have the advantage of being mathematically simple, and as we have shown in Section 2, are erected on an appealing axiomatic foothold (especially in the context of pairwise choice problems). Any one member of this family either includes everything that is chosen with positive probability in the choice set from a menu S, or keeps in that set only the options whose probability of choice is higher than λ times the choice probability of any other alternative in S. The parameter λ determines how inclusive the criterion is, and would be chosen by the analyst according to the problem at hand. It seems reasonable to choose more selective criteria (higher λ) for more noisy environments, more permissive ones (lower λ) for cases when there is reason to consider even low-probability selections as one's potential 'choices.'

The second stage pertains to applying such a rule in the real world, where the analyst does not have access to an ideal data set, but is instead confronted with a finite number of observations. This brings a set of issues concerning sampling errors to the fore. To address these, we develop statistical methods to estimate a choice correspondence by means of hypothesis testing.

When combined, these two stages require the analyst to select two parameters—λ, indicating the inclusivity with ideal data, and α, indicating the level of significance for hypothesis testing—and provides practical formulae to infer choice correspondences. To illustrate the use of our overall method of elicitation, we considered here the repeated (within-subject) choice experiments of Tversky (1969) and Tversky (1972) and estimated the deterministic choice correspondences of the subjects of those experiments.

[24] This is inescapable with small data sets. As an extreme example, consider the case where there is only one observation at hand in the case of a menu. Any reasonable test would then *not* eliminate any of the choice items from contention, as the data at hand is obviously insufficient to rule out the observed choice being but a result of sampling error.

Let us conclude by noting that, instead of trying to construct a choice correspondence, one may instead choose to altogether abandon choice correspondences and focus instead directly on stochastic choice as a primitive. Indeed, a growing literature in economics has been developing models of decision-making based on stochastic choice, making use of the richer nature of this data. However, shifting to a broader stochastic perspective to individual decision-making implies a radical departure from the standard approach. With only few exceptions, the entirety of economic analysis is built on deterministic choices—an approach that has arguably proven very fruitful.[25] Embracing stochasticity at the individual level would move economic analysis away from this and bring it closer to psychology and neuroscience, where mistakes, misperceptions, and variations in preferences are incorporated in the fundamental models of choice. Whether such shift is advisable or not is far beyond our scope. Our goal was simply to suggest a practical method of deriving the fundamental primitive of choice correspondences from real-world data.

[25] This is different than adopting models of stochastic choice to study the choices of heterogenous agents, where the stochasticity in the data derives from (unobserved) heterogeneity in preferences. The nature of the latter type of models is in fact more in line with deterministic choice theory. In those models, each agent makes a deterministic choice, but the analyst sees the choice data only in the aggregate, and hence evaluates it "as if" it is stochastic.

APPENDIX: PROOFS

Proof of Theorem 1

We begin with two lemmata.

> ***Lemma A.1.***[26] Let $\Psi : \textsf{scf}(X) \to \textsf{cc}(X)$ be a choice imputation that satisfies the properties A and B. Then, for every $\mathbb{P}, \mathbb{Q} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}_2$,

$$x \in \Psi(\mathbb{P})(S) \text{ and } \mathbb{Q}(z,S) \geq \mathbb{P}(x,S) \quad \text{imply} \quad z \in \Psi(\mathbb{Q})(S).$$

Proof. As usual, we put $C_{\mathbb{P}} := \Psi(\mathbb{P})$ for any $\mathbb{P} \in \textsf{scf}(X)$. Now take any $\mathbb{P}, \mathbb{Q} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}_2$ and pick any $x, z \in S$ such that $x \in C_{\mathbb{P}}(S)$ and $\mathbb{Q}(z,S) \geq \mathbb{P}(x,S)$. We wish to show that $z \in C_{\mathbb{Q}}(S)$. To this end, we take any $y \in X \backslash \{x, z\}$, put $A := \{x, y\}$, and pick a $\mathbb{P}_1 \in \textsf{scf}(X)$ such that

$$\mathbb{P}_1(x,S) = \mathbb{P}(x,S) \quad \text{and} \quad \mathbb{P}_1(y,A) = \mathbb{P}(x,S).$$

As $\mathbb{P}_1(\cdot,S) = \mathbb{P}(\cdot,S)$, we have $C_{\mathbb{P}_1}(S) = C_{\mathbb{P}}(S)$ by property A, so $x \in C_{\mathbb{P}_1}(S)$. But then, since $\mathbb{P}_1(y,A) = \mathbb{P}(x,S) = \mathbb{P}_1(x,S)$, property B entails that $y \in C_{\mathbb{P}_1}(A)$.

Let us now pick any $\mathbb{P}_2 \in \textsf{scf}(X)$ such that

$$\mathbb{P}_2(y,A) = \mathbb{P}_1(y,A) \quad \text{and} \quad \mathbb{P}_2(z,S) = \mathbb{Q}(z,S).$$

As $\mathbb{P}_2(\cdot,A) = \mathbb{P}_1(\cdot,A)$, we have $C_{\mathbb{P}_2}(A) = C_{\mathbb{P}_1}(A)$ by property A, so by what we have just found, $y \in C_{\mathbb{P}_2}(A)$. Moreover,

$$\mathbb{P}_2(z,S) = \mathbb{Q}(z,S) \geq \mathbb{P}(x,S) = \mathbb{P}_1(y,A) = \mathbb{P}_2(y,A),$$

so by property B, $z \in C_{\mathbb{P}_2}(S)$. But as $\mathbb{P}_2(\cdot,S) = \mathbb{Q}(\cdot,S)$, property A says that $C_{\mathbb{P}_2}(S) = C_{\mathbb{Q}}(S)$, so we conclude that $z \in C_{\mathbb{Q}}(S)$. ■

> ***Lemma A.2.*** Let $\Psi : \textsf{scf}(X) \to \textsf{cc}(X)$ be a choice imputation that satisfies the properties A and B. Then, for every $\mathbb{P}, \mathbb{Q} \in \textsf{scf}(X)$ and $S, T \in \mathfrak{X}_2$,

$$x \in \Psi(\mathbb{P})(S) \text{ and } \mathbb{Q}(z,T) \geq \mathbb{P}(x,S) \quad \text{imply} \quad z \in \Psi(\mathbb{Q})(T).$$

Proof. As usual, we put $C_{\mathbb{P}} := \Psi(\mathbb{P})$ for any $\mathbb{P} \in \textsf{scf}(X)$. Now take any $\mathbb{P}, \mathbb{Q} \in \textsf{scf}(X)$ and $S, T \in \mathfrak{X}_2$ and pick any $(x,z) \in T \times S$ such that $x \in C_{\mathbb{P}}(S)$ and $\mathbb{Q}(z,T) \geq \mathbb{P}(x,S)$. If $S = T$, we are done by Lemma A.1. Suppose, then, S and T are distinct. Let $\mathbb{P}_0$ be any element of $\textsf{scf}(X)$ with $\mathbb{P}_0(\cdot,S) = \mathbb{P}(\cdot,S)$ and $\mathbb{P}_0(\cdot,T) = \mathbb{Q}(\cdot,T)$. By the property A, we then have $C_{\mathbb{P}_0}(S) = C_{\mathbb{P}}(S)$ and $C_{\mathbb{P}_0}(T) = C_{\mathbb{Q}}(T)$. It follows that $x \in C_{\mathbb{P}_0}(S)$ while $\mathbb{P}_0(z,T) \geq \mathbb{P}_0(x,S)$, whence property B entails $z \in C_{\mathbb{P}_0}(T)$, so we again find $z \in C_{\mathbb{Q}}(T)$, as desired. ■

> ***Lemma A.3.*** Let $\Psi : \textsf{scf}(X) \to \textsf{cc}(X)$ be a choice imputation that satisfies the properties A, B and C. Then, there exists a (unique) $\lambda \in [0,1]$ such that $\Psi(\mathbb{P})(S) = C_{\mathbb{P},\lambda}(S)$ for every $S \in \mathfrak{X}_2$ and $\mathbb{P} \in \textsf{scf}(X)$.

[26] Gerelt Tserenjigmid has suggested this lemma to us, which simplifies the subsequent argument.

Proof. For any $\mathbb{P} \in \textsf{scf}(X)$, we put $C_{\mathbb{P}} := \Psi(\mathbb{P})$, and define

$$\lambda_{\mathbb{P}} := \min\left\{\frac{m_{\mathbb{P}}(S)}{M_{\mathbb{P}}(S)} : C_{\mathbb{P}}(S) = S \in \mathfrak{X}_2\right\}.$$

Clearly, $0 < \lambda_{\mathbb{P}} \leq 1$. (If $\lambda_{\mathbb{P}} = 0$ were the case, then there would be an $S \in \mathfrak{X}_2$ with $S = C_{\mathbb{P}}(S)$ and $m_{\mathbb{P}}(S) = 0$, but this would contradict (1).) Now take any $\mathbb{P} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}_2$. Denote the elements of S by x and y so that $\mathbb{P}(x, S) \leq \mathbb{P}(y, S)$. If $x \in C_{\mathbb{P}}(S)$, then $\frac{\mathbb{P}(x,S)}{\mathbb{P}(y,S)} \geq \lambda_{\mathbb{P}}$ by definition of $\lambda_{\mathbb{P}}$. Conversely, suppose $\mathbb{P}(x, S) \geq \lambda_{\mathbb{P}}\mathbb{P}(y, S)$. Since $\mathfrak{X}_2$ is finite, there is an $T \in \mathfrak{X}_2$ with $C_{\mathbb{P}}(T) = T$ and $\frac{m_{\mathbb{P}}(T)}{M_{\mathbb{P}}(T)} = \lambda_{\mathbb{P}}$. Then, $\frac{\mathbb{P}(x,S)}{1-\mathbb{P}(x,S)} \geq \frac{m_{\mathbb{P}}(T)}{1-m_{\mathbb{P}}(T)}$, and it follows that $\mathbb{P}(x, S) \geq m_{\mathbb{P}}(T)$. But then, by property B, we obtain $x \in C_{\mathbb{P}}(S)$. Thus, $x \in C_{\mathbb{P}}(S)$ iff $\mathbb{P}(x, S) \geq \lambda_{\mathbb{P}}\mathbb{P}(y, S)$. As this property (as well as Lemma A.1) implies that $y \in C_{\mathbb{P}}(S)$, and $\mathbb{P}(y, S) \geq \lambda_{\mathbb{P}} M_{\mathbb{P}}(S)$ holds trivially, and because S was arbitrarily chosen above, we conclude that

$$C_{\mathbb{P}}(S) = \{x \in S : \mathbb{P}(x, S) \geq \lambda_{\mathbb{P}} M_{\mathbb{P}}(S)\}$$

for every $\mathbb{P} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}_2$.

To complete our proof, we define

$$\lambda := \inf_{\mathbb{P}\in\textsf{scf}(X)} \lambda_{\mathbb{P}}.$$

Let us again fix arbitrary $\mathbb{P} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}_2$, and denote the elements of S by x and y so that $\mathbb{P}(x, S) \leq \mathbb{P}(y, S)$. Besides, let us pick a sequence $(\mathbb{Q}_k)$ in $\textsf{scf}(X)$ such that $\lambda_{\mathbb{Q}_k} \downarrow \lambda$. As $\mathfrak{X}_2$ is finite, for every positive integer k, there is a $T_k \in \mathfrak{X}_2$ with $C_{\mathbb{Q}_k}(T_k) = T_k$ and $\lambda_{\mathbb{Q}_k} = \frac{m_{\mathbb{Q}_k}(T_k)}{M_{\mathbb{Q}_k}(T_k)}$. Since $\mathfrak{X}_2$ is finite, there must exist a constant subsequence of (T_k), so it is without loss of generality to assume that $T_1 = T_2 = \cdots = T$ for some $T \in \mathfrak{X}_2$. Then, $\lambda_{\mathbb{Q}_k} = \frac{m_{\mathbb{Q}_k}(T)}{M_{\mathbb{Q}_k}(T)}$ and $C_{\mathbb{Q}_k}(T) = T$ for each k. Next, for each $k \in \mathbb{N}$ we take any $\mathbb{P}_k \in \textsf{scf}(X)$ with $\mathbb{P}_k(x, S) = m_{\mathbb{Q}_k}(T)$. Then, by Lemma A.2, we have $C_{\mathbb{P}_k}(S) = S$ for each k. Thus, $\lambda_{\mathbb{Q}_k} = \frac{m_{\mathbb{P}_k}(S)}{M_{\mathbb{P}_k}(S)}$ and $C_{\mathbb{P}_k}(S) = S$ for each k.

Consider first the case in which $\lambda = 0$. In this case, $\frac{m_{\mathbb{P}_k}(S)}{M_{\mathbb{P}_k}(S)} \downarrow 0$, whence $m_{\mathbb{P}_k}(S) \downarrow 0$. If $\mathbb{P}(x, S) > 0$, therefore, we have $\mathbb{P}(x, S) > m_{\mathbb{P}_k}(S)$ for large enough k. But then, Lemma A.2 entails that $x \in C_{\mathbb{P}}(S)$. Thus, $\text{supp}(\mathbb{P}(\cdot, S)) \subseteq C_{\mathbb{P}}(S)$. As the converse inequality is ensured by (1), we conclude that $C_{\mathbb{P}}(S) = C_{\mathbb{P},0}(S)$, as desired.

We now assume that $\lambda > 0$. If $x \in C_{\mathbb{P}}(S)$, then $\mathbb{P}(x, S) \geq \lambda_{\mathbb{P}} M_{\mathbb{P}}(S) \geq \lambda M_{\mathbb{P}}(S)$, so $C_{\mathbb{P}}(S) \subseteq C_{\mathbb{P},\lambda}(S)$. Conversely, assume $\mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S)$, and note that this implies $\mathbb{P}(x, S) > 0$. If $\mathbb{P}(x, S) > \lambda M_{\mathbb{P}}(S)$, then, by definition of λ, there is a k large enough that $\frac{\mathbb{P}(x,S)}{M_{\mathbb{P}}(S)} > \lambda_{\mathbb{Q}_k} \geq \lambda$, whence $\frac{\mathbb{P}(x,S)}{1-\mathbb{P}(x,S)} > \frac{m_{\mathbb{P}_k}(S)}{1-m_{\mathbb{P}_k}(S)}$. Thus, $\mathbb{P}(x, S) > m_{\mathbb{P}_k}(T)$, whence, by Lemma A.2, we find $x \in C_{\mathbb{P}}(S)$, as desired. Finally, suppose that $\mathbb{P}(x, S) = \lambda M_{\mathbb{P}}(S)$. In this case, we have $\frac{m_{\mathbb{P}_k}(S)}{1-m_{\mathbb{P}_k}(S)} \downarrow \frac{\mathbb{P}(x,S)}{1-\mathbb{P}(x,S)}$, whence $m_{\mathbb{P}_k}(S) \downarrow \mathbb{P}(x, S)$. By property C, we thus again find $x \in C_{\mathbb{P}}(S)$. As Lemma A.1 implies $y \in C_{\mathbb{P}}(S)$, and $\mathbb{P}(y, S) \geq \lambda M_{\mathbb{P}}(S)$ holds trivially, we conclude that $C_{\mathbb{P}}(S) = \{x \in S : \mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S)\} = C_{\mathbb{P},\lambda}(S)$, as we sought. ■

We are now prepared to prove Theorem 1. The proof of the "if" part of the claim is straightforward, so we focus on its "only if" part. Assume that $\Phi : \textsf{scf}(X) \to \textsf{cc}(X)$ satisfies the properties A, B, C and D. By Lemma A.3, there exists a (unique) $\lambda \in [0, 1]$ such that $C_{\mathbb{P}}(S) = C_{\mathbb{P},\lambda}(S)$ for every $S \in \mathfrak{X}_2$ and $\mathbb{P} \in \textsf{scf}(X)$, where, as usual, we write $C_{\mathbb{P}}$ for $\Phi(\mathbb{P})$. Now take any $\mathbb{P} \in \textsf{scf}(X)$ and $S \in \mathfrak{X}$ with $|S| \geq 3$. Let y be an element of S with $\mathbb{P}(y, S) = M_{\mathbb{P}}(S)$, and note that $y \in C_{\mathbb{P}}(S)$ by the property D (applied for $x = y$). Next, take any x in S, put $A := \{x, y\}$, and define $\mathbb{Q} \in \textsf{scf}(X)$ as $\mathbb{Q}(\cdot, T) := \mathbb{P}(\cdot, T)$ for every $T \in \mathfrak{X}\backslash\{A\}$, and $\mathbb{Q}(x, A) := \frac{\mathbb{P}(x,S)}{\mathbb{P}(x,S)+\mathbb{P}(y,S)}$ and $\mathbb{Q}(y, A) := 1 - \mathbb{Q}(x, A)$.

Now suppose $x \in C_{\mathbb{P}}(S)$. As $A \neq S$ (because S does not belong to $\mathfrak{X}_2$), we have $\mathbb{P}(\cdot, S) = \mathbb{Q}(\cdot, S)$, and the property A thus implies that $x \in C_{\mathbb{Q}}(S)$. Besides,

$$\frac{\mathbb{Q}(x, A)}{\mathbb{Q}(y, A)} = \frac{\mathbb{P}(x, S)}{\mathbb{P}(y, S)} = \frac{\mathbb{Q}(x, S)}{\mathbb{Q}(y, S)},$$

so property D entails $x \in C_{\mathbb{Q}}(A)$. Since $A \in \mathfrak{X}_2$, this means that $\mathbb{Q}(x, A) \geq \lambda\mathbb{Q}(y, A)$, whence $\mathbb{P}(x, S) \geq \lambda\mathbb{P}(y, S) = \lambda M_{\mathbb{P}}(S)$. Conversely, suppose $\mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S)$ holds. Then,

$$\mathbb{Q}(x, A) = \left(\frac{\mathbb{Q}(x, A)}{\mathbb{Q}(y, A)}\right)\mathbb{Q}(y, A) = \left(\frac{\mathbb{P}(x, S)}{\mathbb{P}(y, S)}\right)M_{\mathbb{Q}}(A) \geq \lambda M_{\mathbb{Q}}(A)$$

so $x \in C_{\mathbb{Q}}(A)$. It then follows from the property D that $x \in C_{\mathbb{Q}}(S)$. Since $\mathbb{P}(\cdot, S) = \mathbb{Q}(\cdot, S)$, the property A thus implies $x \in C_{\mathbb{P}}(S)$. In view of the arbitrary choice of x in S, we thus conclude that

$$C_{\mathbb{P}}(S) = \{x \in S : \mathbb{P}(x, S) \geq \lambda M_{\mathbb{P}}(S)\},$$

that is, $C_{\mathbb{P}}(S) = C_{\mathbb{P},\lambda}(S)$. In view of the arbitrary choice of S and $\mathbb{P}$, we are done.

Proof of Lemma 2

For any $n \in \mathbb{N}$ and $c \in \{0, ..., n\}$, define the self-map $\varphi_{n,c}$ on $[0, 1]$ by

$$\varphi_{n,c}(p) := \sum_{i=0}^{c} \binom{n}{i} p^i(1-p)^{n-i}$$

and note that

$$\begin{aligned}
\varphi'_{n,c}(p) &= \sum_{i=0}^{c} \binom{n}{i} ip^{i-1}(1-p)^{n-i} - \sum_{i=0}^{c} \binom{n}{i}(n-i)p^i(1-p)^{n-i-1} \\
&= \sum_{i=1}^{c} \frac{n!}{(i-1)!(n-i)!}p^{i-1}(1-p)^{n-i} - \sum_{i=0}^{c} \frac{n!}{i!(n-i-1)!}p^i(1-p)^{n-i-1} \\
&= n\sum_{i=1}^{c} \binom{n-1}{i-1}p^{i-1}(1-p)^{n-i} - n\varphi_{n-1,c}(p) \\
&= n\left(\sum_{i=0}^{c-1} \binom{n-1}{i}p^i(1-p)^{n-i-1} - \varphi_{n-1,c}(p)\right) \\
&= n\left(\varphi_{n-1,c-1}(p) - \varphi_{n-1,c}(p)\right).
\end{aligned}$$

It follows that $\varphi'_{n,c} < 0$, and the lemma follows.[27] ■

Proofs of the Claims of Remark 4

[27] This argument proves a bit more than what is needed for Lemma 2. It shows that $\frac{1}{n}\varphi'_{n,c}(p)$ equals Prob($w \leq c-1$)− Prob($w \leq c$) where w is a random variable with $w \sim$ Binomial($n-1, p$).

For any $p \in [0,1]$, let (x_m) be a sequence of Bernoulli random variables on a given probability space that are i.i.d. with parameter p. For any positive integer n, put $u_n := x_1 + \cdots + x_n$; u_n is binomially distributed with parameters n and p. Let F_n stand for the cumulative distribution function of u_n. Notice that $F_n \geq F_{n+1}|_{(-\infty,n]}$ implies $w_{\lambda,n} \leq w_{\lambda,n+1}$, so it is enough to prove the former inequality to conclude that $w_{\lambda,n}$ is increasing in n. To this end, fix an $n \in \mathbb{N}$, take any $\theta \in \{0,...,n\}$, and note that

$$\begin{aligned}
F_{n+1}(\theta) &= \text{Prob}(u_{n+1} \leq \theta) \\
&= \text{Prob}(u_n < \theta) + \text{Prob}(u_{n+1} \leq \theta \mid u_n = \theta)\text{Prob}(u_n = \theta) \\
&= F_n(\theta - 1) + (1-p)\binom{n}{\theta} p^\theta (1-p)^{n-\theta} \\
&= \left(F_n(\theta - 1) + \binom{n}{\theta} p^\theta (1-p)^{n-\theta} \right) - p\binom{n}{\theta} p^\theta (1-p)^{n-\theta} \\
&= F_n(\theta) - p\binom{n}{\theta} p^\theta (1-p)^{n-\theta} \\
&\leq F_n(\theta).
\end{aligned}$$

This proves our first assertion in Remark 4.

Next, fix any $n \in \mathbb{N}$ and $p \in [0,1]$ such that $p > \frac{1}{n}$. By Theorem 1 of Greenberg and Mohri (2014), we have $\text{Prob}(u \leq np) \geq \frac{1}{4}$ for any binomially distributed random variable u with parameters n and p. Setting $p = \frac{\lambda}{1+\lambda}$ and picking any $\alpha \in [0, \frac{1}{4})$, it then follows from (5) that $w_{\lambda,n} \leq \frac{n\lambda}{1+\lambda}$ whenever $n > \frac{1+\lambda}{\lambda}$, as we claimed in Remark 4.

To prove the consistency assertion made in Remark 4, take any $\lambda \in (0,1]$ and $n \in \mathbb{N}$, and let u_n be a binomially distributed random variable with parameters n and $p := \frac{\lambda}{1+\lambda}$. (Our claim in the case where $\lambda = 0$ is trivially true.) Note that $\text{Prob}(np < u_n \leq \lceil np \rceil)$ is either 0 (which happens when np is an integer) or it equals $\text{Prob}(u_n = \lceil np \rceil)$. Since $\binom{n}{i} p^{i-1}(1-p)^{n-i} \to 0$ as $n \uparrow \infty$ for any positive integer i, therefore, we have

$$\text{Prob}(np < u_n \leq \lceil np \rceil) \to 0 \quad \text{as } n \uparrow \infty.$$

Since every median of the binomial distribution with parameters n and p lies within $\lfloor np \rfloor$ and $\lceil np \rceil$ – see, for instance, Kaas and Buhrman (1980) – it follows that for every $\varepsilon > 0$ there is a positive integer N large enough that $\text{Prob}(u_n \leq np) \geq \frac{1}{2} - \varepsilon$ for each $n \geq N$. In particular, for any α within, say, $(0, \frac{1}{4})$, we have

$$\text{Prob}(u_n \leq np) > \alpha \quad \text{for each } n \geq N,$$

which, by (5), means $w_{\lambda,n} < np$ for each $n \geq N$. We conclude that $\limsup \frac{w_{\lambda,n}}{n} \leq p$.

On the other hand, by Hoeffding's Inequality,

$$\text{Prob}(u_n \leq (p-\varepsilon)n) \leq e^{-2\varepsilon^2 n} \quad \text{for every } n \in \mathbb{N} \text{ and } \varepsilon > 0.$$

Consequently, for any $\varepsilon > 0$, $\text{Prob}(\frac{u_n}{n} \leq p - \varepsilon) \to 0$ as $n \uparrow \infty$. But by (5), the definition of $w_{\lambda,n}$,

$$\text{Prob}(u_n \leq w_{\lambda,n}) \leq \alpha < \text{Prob}(u_n \leq w_{\lambda,n} + 1) = \text{Prob}\left(\frac{u_n}{n} \leq \frac{w_{\lambda,n}}{n} + \tfrac{1}{n}\right)$$

for every $n \in \mathbb{N}$. It follows that, for any $\varepsilon > 0$, there is a positive integer N large enough that

$$\text{Prob}\left(\frac{u_n}{n} \leq p - \varepsilon\right) < \text{Prob}\left(\frac{u_n}{n} \leq \frac{w_{\lambda,n}}{n} + \tfrac{1}{n}\right) \quad \text{for each } n \geq N,$$

whence

$$p - \varepsilon < \frac{w_{\lambda,n}}{n} + \tfrac{1}{n} \quad \text{for each } n \geq N.$$

We conclude that $p - \varepsilon \leq \liminf \frac{w_{\lambda,n}}{n}$ for any $\varepsilon > 0$, which means $p \leq \liminf \frac{w_{\lambda,n}}{n}$. Combining this with what we have found in the previous paragraph yields $\lim \frac{w_{\lambda,n}}{n} = p = \frac{\lambda}{1+\lambda}$, as we sought.

References

Afriat, S., The construction of utility functions from expenditure data, *Int. Econ. Rev.* 8 (1967), 67-77.

Agranov, M., and P. Ortoleva, Stochastic Choice and Preferences for Randomization, *J. Pol. Econ.* 125 (2017), 40-68.

Agranov, M., and P. Ortoleva, Ranges of Randomization, *mimeo*, Princeton Univ., 2021.

Arrow, K.J., Rational choice functions and orderings, *Economica,* 102 (1959), 121-127.

Bernheim, D., and A. Rangel, Toward choice-theoretic foundations for behavioral welfare economics, *Amer. Econ. Rev.* 97 (2007), 464-70.

Bernheim, D., and A. Rangel, Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics, *Quart. J. Econ.* 124 (2009), 51-104.

Bouacida, E., Eliciting choice correspondences: A general method and an experimental implementation, *mimeo*, Paris School of Economics, 2018.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella, Deliberately stochastic, *Amer. Econ. Rev.* 109 (2019), 2425-45.

Cettolin, E. and A. Riedl, Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization, *J. Econ. Theory* 181 (2019), 547–85.

Chambers, C., and F. Echenique, *Revealed Preference Theory*, Cambridge, Cambridge University Press, 2016.

Cohen, M., J-Y. Jaffray, and T. Said, Individual behavior under risk and uncertainty: An experimental study, *Theory Decis.* 18 (1985), 203-228.

Cohen, M., J-Y. Jaffray, and T. Said, Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses, *Organ. Behav. Hum. Decis. Process.* 39 (1987), 1-22.

Costa-Gomes, M., C. Cueva, G. Gerasimou, and M. Tejiščák, Choice, deferral and consistency, *mimeo*, Univ. St. Andrews,, 2019.

Danan, E. and A. Ziegelmeyer, Are preferences complete?: An experimental measurement of indecisiveness under risk, *mimeo*, Max Planck Inst. of Economics, 2006.

Dwenger, N., D. Kübler, and G. Weizsäcker, Flipping a coin: Evidence from laboratory and field, *J. Pub. Econ.* 167 (2018), 240-250.

Feldman, P. and J. Rehbeck, Revealing a preference for mixing: An experimental study of risk, *mimeo*, UC-San Diego, 2020.

Fishburn, P., Choice probabilities and choice functions, *J. Math. Psych.* 18 (1978), 201-219.

Houthakker, H.S., Revealed preference and the utility function, *Economica*, 17 (1950), 159-174.

Greenberg, S. and M. Mohri, Tight lower bound on the probability of a binomial exceeding its expectation, *Stat. Prob. Lett.*, 86 (2014), 91-98.

Iverson, G., and J.-C. Falmagne, Statistical issues in measurement, *Math. Soc. Sci.* 10 (1985), 131-153.

Johnson, N., S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distributions*, New York, Wiley, 1997.

Kaas, R. and J. Buhrman, Mean, median and mode in binomial distributions, *Stat. Neerlandica,* 34 (1980) 13-18.

Kircher, P., S. Ludwig, and A. Sandroni, On the difference between social and private goods, *BE J. Theoret. Econ.* 13 (2013), 1-27.

Luce, D., *Individual Choice Behavior: A Theoretical Analysis*, New York, Wiley, 1959.

Manzini, P. and M. Mariotti, Sequentially rationalizable choice, *Amer. Econ. Rev.* 97 (2007), 1824-1830.

Manzini, P. and M. Mariotti, Dual random utility maximisation, *J. Econ. Theory* 177 (2018), 162-182.

Masatlioglu, Y., D. Nakajima, and E. Ozbay, Revealed attention, *Amer. Econ. Rev.* 102 (2012), 2183-2205.

Mas-Colell, A., M. Whinston, and J. Green, *Microeconomic Theory*, New York: Oxford University Press, 1995.

Miao, B. and S. Zhong, Probabilistic social preference: How Machina's Mom randomizes her choice, *Econ Theory* 65 (2018), 1-24.

Nishimura, H., and E. A. Ok, Non-Existence of Continuous Choice Functions, *J. Econ. Theory* 153 (2014), 376-391.

Ok, E. A., and G. Tserenjigmid, Indifference, indecisiveness, experimentation and stochastic choice, *Theoretical Econ.*, forthcoming.

Ok, E. A., and G. Tserenjigmid, Comparative rationality of stochastic choice rules, *mimeo*, New York Univ., 2020.

Regenwetter, M., J. Dana, and C. Davis-Stober, Transitivity of preferences, *Psych. Rev.* 118 (2011), 42-56.

Rubinstein, A., Irrational diversification in multiple decision problems, *European Econ. Rev.* 46 (2002), 1369–78.

Samuelson, P.A., A note on the pure theory of consumer's behaviour, *Economica* 5 (1938), 61-71.

Samuelson, P.A., Consumption theory in terms of revealed preference, *Economica* 60 (1950), 243-253.

Sautua, S., Does uncertainty cause intertia in decision making? An experimental study of the role of regret aversion and indecisiveness, *J. Econ. Behav. Org.* 136 (2017), 1-14.

Tserenjigmid, G., Theory of decisions by intra-dimensional comparisons, *J. Econ. Theory* 159 (2015), 326-338.

Tversky, A., Intransitivity of preferences, *Psych. Rev.* 76 (1969), 31-48.

Tversky, A., Elimination by aspects: A theory of choice, *Psych. Rev.* 79 (1972), 281-299.