Invariant Modeling for Joint Distributions[*]

Christopher P. Chambers[†]   Yusufcan Masatlioglu[‡]   Ruodu Wang[§]

October 28, 2025

Abstract

A common theme underlying many problems in statistics and economics involves the determination of a systematic method of selecting a joint distribution consistent with a specified list of categorical marginals, some of which have an ordinal structure. We propose guidance in narrowing down the set of possible methods by introducing Invariant Aggregation (IA), a natural property that requires merging adjacent categories in one marginal not to alter the joint distribution over unaffected values. We prove that a model satisfies IA if and only if it is a copula model. This characterization ensures i) robustness against data manipulation and survey design, and ii) allows seamless incorporation of new variables. Our results provide both theoretical clarity and practical safeguards for inference under marginal constraints.

JEL: D01, D91

Keywords: Copula | Invariance | Joint and Marginal Distribution |

1   Introduction

Recovering a joint distribution from its marginal distributions is a fundamental problem in statistics, economics, machine learning, and related fields Sklar (1959), Genest & Rivest (1993), Li (1999), Beare (2010), Joe (2014), McNeil et al. (2015), Fan & Henry (2023), Chambers et al. (2025). However, marginal distributions alone are generally insufficient to uniquely determine the joint distribution—infinitely many joint distributions may be consistent with the same set of marginals. This ambiguity grows with dimensionality, as the number of free parameters in the joint distribution increases exponentially with the number of random variables in discrete statistical models.

To resolve this indeterminacy, additional structure must be imposed—such as assumptions about dependence patterns, joint moments, or other inter-variable relationships. A fundamental approach is to introduce a functional mapping, denoted by f, that takes a set of discrete marginal distributions and returns a unique joint distribution consistent with them. We refer to such a mapping f as a model. Each model imposes certain structural assumptions, such as positive dependence, negative dependence, and independence.

In this paper, we propose and analyze an intuitive structural property that a model f may satisfy: aggregating two adjacent values in the support of one marginal should not affect the joint distribution over the remaining, unaltered values. The determination of adjacency requires us to specify an order on each marginal. Our approach is motivated by examples in economics, but the ideas here may be of interest in other fields such as political science, statistics, or psychology.

To illustrate, consider electricity and water consumption over a fixed period of time in a neighborhood. Each consumer in the neighborhood consumes some amount of water, and some amount of electricity (in the jargon of economists, we speak of them *choosing* water consumption and electricity consumption). Water and electricity consumption are each naturally modeled as real-valued, but we suppose that each of these real-valued consumption levels is partitioned into bins: low, medium, and high: (L, M, and H). Each bin represents an interval of possible real-valued consumption levels. Importantly, a consumer whose water choice lies in bin L consumes less water than one whose choice lies in bin M; we write $L < M < H$. The economist observes the empirical distribution of choices from

bins L, M and H for both water consumption across all consumers, and for electricity consumption. We refer to these as *marginals*.

Even though each consumer chooses both water and electricity consumption, we assume that these joint choices are unobserved. That is, we imagine data comes from different sources, leaving the joint distribution of water and electricity consumption unknown. We seek to study methods of selecting joint distributions. As a modeler or policymaker, we may hypothesize a positive association—for instance, we may conjecture that higher electricity usage may tend to coincide with higher water usage. Given the marginals, our model produces a corresponding joint distribution. But many other hypotheses are possible.

Now, imagine that due to a different survey design, the medium and high electricity categories are merged (written MH), while the water usage categories remain unchanged. The marginal distribution for electricity becomes coarser, and the model generates a new joint distribution. Crucially, since the low category is unaffected, we expect the joint probabilities involving this category recommended by our selection to remain unchanged. Our proposed property, called Invariant Aggregation (IA), ensures this invariance, enhancing robustness and interpretability. IA captures the idea that predictions of a model would not contradict themselves in similar situations. In other words, the prediction of a model is invariant to the small changes in response categories, which is a desirable property in categorical data analysis Agresti (2010), Johnson & Albert (2006).

Violating IA would render social policies sensitive to arbitrary data-reporting choices—especially problematic when reporting methods vary across contexts. Worse, it could allow unscrupulous data collectors to manipulate outcomes by altering data presentation. Many, but not all, models satisfy IA. A trivial example of a model satisfying IA is the independence model, which assumes statistical independence across variables, constructing the joint distribution via multiplication of marginals. However, this is not the only such model. Our main result provides a complete characterization of all IA-compliant models: they are precisely those defined by copulas, a well-studied class of statistical objects. Indeed, for statisticians, to get from marginal distributions on the real line to a joint distribution, copulas come to mind first Nelsen (2006). Sklar's theorem states that a joint distribution on an Euclidean space can be described by its marginal distributions and a copula. Our task is, however, different: in our setting, both the marginal distributions

and the joint distribution vary (and they are categorical), while keeping the invariance property.

Finally, IA has an important implication for introducing new variables. If the i-th dimension initially contains no data, increasing its support cardinality is akin to introducing new data. Critically, any IA model accommodates such extensions without altering the joint distribution over existing dimensions.

Other sections of the paper are organized as follows. Section 2 introduces the notation and definitions of IA. Section 3 introduces and proves the main result of the paper. Section 4 investigated an additional result characterizing a family of independence copulas. Finally, Section 5 concludes.

2 The Model

We have in mind the following problem. There are several categorical random variables. To represent these random variables, fix $n \in \mathbb{N}$, and assume given an ordered list of n finite sets $\{A_1, \ldots, A_n\}$, each of which is endowed with an ordinal ranking $<_i$ over A_i. Though the variables are categorical in nature, the underlying implicit structure we envision consists of a real-valued random variable that has been partitioned into a finite number of bins, each of which represents an interval. For example, political positions in the "spatial model" of politics are often modeled as a unidimensional spectrum (*e.g.* Liberal, Moderate, Conservative) Downs (1957), and the importance of global warming could be categorized into five bins (extremely, very, somewhat, not too, not at all important) Chryst et al. (2018).

For each set A_i, a probability distribution p_i is given over A_i. Our goal is to find a systematic method of constructing a joint probability distribution p on $\prod_{i=1}^{N} A_i$ whose marginals coincide with each p_i; that is marg $p|_{A_i} = p_i$.

Our framework is ordinal, so that the actual names of the elements of each A_i are not relevant. Instead, only the cardinality and order structure of each A_i matters. Therefore, the variable n represents the number of sets A_i for which we anticipate receiving data. For example, consider the example discussed in the introduction, where a distribution of water consumption is specified, and a distribution of electriy consumption is specified. In

4

this example, we can imagine that A_1 represents possible water consumption levels, and A_2 represents possible electricity consumption levels.

As is standard, for any countable set A, let $\Delta(A)$ be the set of all probability distributions on $(A, 2^A)$. For a probability distribution q on $\mathbb{R}^n$, we denote by $[q]_i$ its ith marginal distribution for $i \in N$.[1] Distributions are identified with their probability mass functions; that is, we write $q(a) = q(\{a\})$ for $a \in A$.

For any $k \in \mathbb{N}$, let $[k] = \{1, \ldots, k\}$. Then, for $z = (z_1, \ldots, z_n) \in \mathbb{N}^n$, denote by $[z] = \prod_{i=1}^n [z_i]$. For each $i \in N$, each $[z_i]$ reflects the finite number of bins (z_i bins) from which individuals may choose, or the cardinality of A_i. Return to our water and electricity example. If A_1 reflects water consumption, then z_i is simply a label of the highest bin (our introductory example featured labels of the form $L < M < H$, here the integers $1 < 2 < 3$ are placeholders for these labels. In this example then, $z_i = 3$). Thus, in the economics example, an element of $[z_i]$ constitutes a choice of the i-th good. In this case, a member of s of $[z]$ reflects a hypothetical or possible *joint* consumer choice: a possible consumer who consumed s_1 units of water and s_2 units of electricity.

Though the class of models we will ultimately characterize is quite large, our specification involves an implicit independence assumption. We think our specification is useful: it keeps our analysis both tractable and applicable in a broad variety of situations. A finite list of ordered bins in our model is always written as $\{1, \ldots, z_i\}$, for some $z_i \in \mathbb{N}$, without reference to the names of the alternatives in the bins. Returning to the $L < M < H$ example from the introduction, observe that $\{L, M, H\}$ is a three element set, so we would set $z_i = 3$ in this case. In merging L and M, we obtain $LM < H$, where $\{LM, H\}$ is a two element set, so $z_i = 2$. Similarly, merging M and H results in a two element set $L < MH$, with $z_i = 2$. Our framework therefore does not include a language for distinguishing the ordered set $\{L, M\} < H$ from $L < \{M, H\}$. We discuss this further after the statement of the main result.

Next, define

$$\Delta^n(z) = \prod_{i=1}^n \Delta([z_i]) = \{(p_1, \ldots, p_n) : p_i \in \Delta([z_i]), \ i \in N\}.$$

[1]Similarly for a distribution on any subset of $\mathbb{R}^n$.

For a given z, any member of $\Delta^n(z)$ constitutes the input of the model. In practical settings, the input could consist of the empirical distributions of variables of interest. It is a list of probability distributions, consisting of a probability distribution over $[z_i]$ for each $i \in N$.

We need a method for describing how data can change as z changes—this was our primary motivation. The bins into which choices can be put can be refined further. To this end, let

$$\Delta_{\mathrm{ma}}^n = \bigcup_{z \in \mathbb{N}^n} \Delta^n(z) \ \text{ and } \ \Delta_{\mathrm{jo}}^n = \bigcup_{z \in \mathbb{N}^n} \Delta\left([z]\right).$$

The set Δ_{ma}^n lists all possible datasets we may ever possibly encounter as z varies. On the other hand, members of Δ_{jo}^n reflect our possible *models* of data. For any $[z]$, a member of $\Delta\left([z]\right)$ is a joint distribution over $\prod_{i \in N}[z_i]$; Δ_{jo}^n collects all of these as z varies.

Consider the following running example. There are two random variables (A_1 and A_2 are water and electricity consumptions). Hence $n = 2$. There are three categories in A_1 and four categories in A_2, hence $(z_1, z_2) = (3, 4)$. Then data comes in the form of a pair of distributions: a distribution p_1 over $\{1, 2, 3\}$ and a distribution p_2 over $\{1, 2, 3, 4\}$. Our aim is to construct a joint distribution on $\{1, 2, 3\} \times \{1, 2, 3, 4\}$ whose marginals are p_1 and p_2 respectively. We seek a method for doing this no matter what our choice of (z_1, z_2) is and no matter what the marginals may be. Table 1 illustrates $(z_1, z_2) = (3, 4)$. The top row summarizes z_2, and the leftmost column represents z_1. This idea motivates the following definition, which is our main object of study.

For vectors $z, s \in \mathbb{N}^n$ and $p \in \Delta_{\mathrm{ma}}^n$, as well as similar quantities, we tacitly denote their components by $(z_1, \ldots, z_n)$, $(s_1, \ldots, s_n)$, and $(p_1, \ldots, p_n)$, respectively. For notational convenience, in what follows, when $p_1, \ldots, p_n$ appear, their domains are $[z_1], \ldots, [z_n]$, unless otherwise specified.

Let f be a mapping from Δ_{ma}^n to Δ_{jo}^n satisfying $[f(p)]_i = p_i$ for $p \in \Delta_{\mathrm{ma}}^n$ and $i \in N$. Such a mapping f is called a *model*. Perhaps the simplest example of a specific model is the *independence model*, given by

$$f(p)(s) = \prod_{i \in N} p_i(s_i), \quad s \in [z]$$

6

Return to the preceding example, whereby $(z_1, z_2) = (3, 4)$. We will show how to apply the independence model to data. Suppose that $p_1 = (1/2, 1/4, 1/4)$ and $p_2 = (1/4, 1/4, 1/4, 1/4)$. If f is the independence model, then for any pair $(s_1, s_2) \in [z]$, $f(p)(s_1, s_2) = p_1(s_1)p_2(s_2)$. So, for example, $f(p)(1, 2) = 1/8$. Observe that $f(p)$ is a probability distribution over $[z]$ and further that the marginal of $f(p)$ in each dimension coincides with the original data. Table 1 illustrates the application of the independence model. The bottom row summarizes the marginal probabilities associated with p_1, and the rightmost column the marginal probabilities associated with p_2. So, we can find the joint probability of $s = (2, 3)$, for example, by multiplying these values: it is $1/16$.

	1	2	3	4	p_1
1	1/8	1/8	1/8	1/8	1/2
2	1/16	1/16	1/16	1/16	1/4
3	1/16	1/16	1/16	1/16	1/4
p_2	1/4	1/4	1/4	1/4	

Table 1: Marginal Distributions p_1 and p_2 and the independent joint distribution of (p_1, p_2)

Our notion of a model needs to specify what happens for more refined data. Let us now imagine $(z_1', z_2) = (4, 4)$. For example, A_1 (the water consumption) is provided with finer categories. Consider $p_1' \in \Delta([z_1'])$ given by $(1/2, 1/4, 1/8, 1/8)$. This data is consistent with the idea that the last bin in $[z_1]$ splits into two bins while the other two bins do not change: the probability of the initial two bins remains the same while $p_1'(3) + p_1'(4) = p_1(3)$. Thus, p_1' is consistent with the idea that the data described by p_1 has become more refined. Now, let us apply the independence model to (p_1', p_2). For example, $f(p_1', p_2)(3, 2) = p_1'(3)p_2(2) = 1/32$ and $f(p_1', p_2)(4, 2) = p_1'(4)p_2(2) = 1/32$.

	1	2	3	4	p_1'
1	1/8	1/8	1/8	1/8	1/2
2	1/16	1/16	1/16	1/16	1/4
3	1/32	1/32	1/32	1/32	1/8
4	1/32	1/32	1/32	1/32	1/8
p_2	1/4	1/4	1/4	1/4	

Observe that $f(p_1', p_2)(1,2) = 1/8 = f(p_1, p_2)(1,2)$. Indeed, the entries in the first two rows did not change. That is, the recommendations made by the independence model for the coarse and the fine datasets are consistent.

To formalize the idea behind the preceding example, we need to introduce some new notation. Specifically, we need to describe a notation that allows us to merge two bins into one.

Take any $w \in \{2, 3, \dots\}$ and $q \in \Delta([w])$. For $j \in [w-1]$, let $q^j \in \Delta([w-1])$ be given by, for all $t \in [w-1]$, $q^j(t) = q(t)\mathbb{1}_{\{t \leqslant j\}} + q(t+1)\mathbb{1}_{\{t \geqslant j\}}$.

Then, given q, the measure q^j is the measure that results from "collapsing" categories j and $j+1$.

When q^j appears for a probability $q \in \Delta([w])$, we always assume $j \in [w-1]$. For $i \in N$, e^i is defined as the unit vector $(0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$ where 1 is at the ith position. For any distribution q, we write F_q the cumulative distribution function (cdf) of q.[2] Let T_i^j be the mapping that maps p to $(p_1, \dots, p_{i-1}, p_i^j, p_{i+1} \dots, p_n)$, for all $i \in [n]$ and $j \in [z_i - 1]$.

Thus, T_i^j is simply a profile of marginal choices when, for set i, alternatives j and $j+1$ are collapsed into one.

Our primary axiom can now be expressed as follows. It requires that when merging two adjacent alternatives into one, the model should respond consistently.

Invariant Aggregation(IA): For all $p \in \Delta_{\mathrm{ma}}^n$, all $i \in N$, all j and all s,

$$f\left(T_i^j(p)\right)(s) = f(p)(s)\mathbb{1}_{\{s_i \leqslant j\}} + f(p)(s + e^i)\mathbb{1}_{\{s_i \geqslant j\}} \tag{1}$$

---

[2]For multidimensional distributions, this is defined as $F_q(s) = q(\{t : t_i \leqslant s_i \text{ for all } i \in N\})$.

Importantly, this axiom is not vacuously satisfied. Many models violate it. We illustrate below with an example of a *maximal coupling*, an object in statistics which is intended to maximize the agreement between variables. For example, for $n = 2$, *maximal coupling* maximizes the probability of $\{(s, s) : s \leqslant \min\{z_1, z_2\}\}$.

Returning to our example, we apply a maximal coupling as the underlying model. First, consider (p_1, p_2), then the maximal coupling model results in

	1	2	3	4	p_1
1	1/4	0	0	1/4	1/2
2	0	1/4	0	0	1/4
3	0	0	1/4	0	1/4
p_2	1/4	1/4	1/4	1/4	

Notice that the probability $f(p)(s, s)$ is maximized. Consider now the case $(z_1, z_2) = (4, 4)$ and (p'_1, p_2). Then we have

	1	2	3	4	p'_1
1	1/4	0	1/8	1/8	1/2
2	0	1/4	0	0	1/4
3	0	0	1/8	0	1/8
4	0	0	0	1/8	1/8
p_2	1/4	1/4	1/4	1/4	

Observe now that IA is obviously violated. Were it satisfied, the probability of $(1, 4)$ would need to remain the same across the two environments, but it does not.

As a final ingredient, we need the notion of copulas. A *copula* is a function C from $[0, 1]^n$ to $[0, 1]$ that is a cumulative distribution function for a distribution with standard uniform (i.e., uniform on $[0, 1]$) marginals. So, we may conclude that C is both nondecreasing and continuous, and satisfies other properties, such as $C(1, \ldots, 1, x, 1, \ldots, 1) = x$, for example.

There are many copulas; a good introduction to the theory is Nelsen (2006).[3] Two classical ones are as follows:

- The *independence copula* is given by $C(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i$.

- The *Fréchet-Hoeffding upper-bound* is defined as $C(x_1, \ldots, x_n) = \min_i x_i$.

These two copulas happen to be described by symmetric functions (permuting the coordinates does not change the value of the copula). This need not be the case. For a matrix $\Sigma \in \mathbb{R}^{n \times n}$ corresponding to the correlation coefficients of a multivariate normal, the *Gaussian copula* is given by $C(x_1, \ldots, x_n) = \Phi_\Sigma(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_n))$, where Φ is the cdf of a standard normal and Φ_Σ is the joint cdf of a mean-zero normal with correlation matrix Σ. The independence copula is the special case where Σ is the identity matrix.

A *copula model* is a model f for which there is a copula $C : [0,1]^n \to [0,1]$ such that

$$F_{f(p)}(s) = C(F_{p_1}(s_1), \ldots, F_{p_n}(s_n)) \tag{2}$$

for all $p \in \Delta_{\mathrm{ma}}^n$ and $s \in \mathbb{R}^n$. The left-hand side of (2) reports the cumulative probability $f(p)$ assigned to s, that is, the probability of $\{s' : s'_i \leqslant s_i \text{ for all } i\}$. As in the standard single-dimensional case, this object uniquely pins down $f(p)$, through an inclusion-exclusion principle. The right-hand side applies the copula C to the cumulative probabilities of the marginals p_i.

If C is the independence copula, the corresponding copula model is the *independence model*. We now demonstrate what happens when the Fréchet-Hoeffding upper bound is applied. First, consider (p_1, p_2), then the Fréchet-Hoeffding upper bound model results in

	1	2	3	4	p_1
1	1/4	1/4	0	0	1/2
2	0	0	1/4	0	1/4
3	0	0	0	1/4	1/4
p_2	1/4	1/4	1/4	1/4	

---

[3]Copulas have also been applied in other areas of economic theory, such as bargaining Bastianello & LiCalzi (2019) and behavioral game theory Frick et al. (2022).

To find $f(p)(2, 3)$ for the Fréchet upper bound, we observe that the cumulative probability of $(2, 3)$ is $3/4$; then similarly subtract the cumulative probabilities of $(2, 2)$ $(1/2)$ and $(1, 3)$ $(1/2)$, and add the cumulative probability of $(1, 2)$ $(1/2)$, resulting in $f(p)(2, 3) = 1/4$. This model actually always recommends a joint distribution whose support is weakly ordered: in this case, the support is concentrated on the elements $(1, 1) \leqslant (1, 2) \leqslant (2, 3) \leqslant (3, 4)$; all other elements get probability zero.

Consider now the case $(z_1, z_2) = (4, 4)$ and (p'_1, p_2). Then we have

	1	2	3	4	p'_1
1	1/4	1/4	0	0	1/2
2	0	0	1/4	0	1/4
3	0	0	0	1/8	1/8
4	0	0	0	1/8	1/8
p_2	1/4	1/4	1/4	1/4	

Note that this model also satisfies IA.

3 Main Result

We now present our main result: (i) every copula-based model satisfies the IA property, and (ii) no other models beyond these satisfy IA.

Theorem 1 *A model satisfies the IA property if and only if it is a copula model.*

Proof: We first prove the "only if" statement. For $t \in [0, 1]$, let b_t be the binary distribution on $\{1, 2\}$ given by $b_t(1) = t = 1 - b_t(2)$. For $x \in [0, 1]^n$, write $b_x = (b_{x_1}, \ldots, b_{x_n})$. Let $C : [0, 1]^n \to [0, 1]$ be given by $C(x) = F_{f(b_x)}(1, \ldots, 1)$ for $x \in [0, 1]^n$. Repeatedly applying IA, we get, for any $p \in \Delta^n_{\text{ma}}$ and $s \in [z]$,

$$
\begin{aligned}
F_{f(p)}(s) &= f(b_{F_{p_1}(s_1)}, \ldots, b_{F_{p_n}(s_n)})(1, \ldots, 1) \\
&= C(F_{p_1}(s_1), \ldots, F_{p_n}(s_n)).
\end{aligned}
\tag{3}
$$

Therefore, f is a copula model if we could prove that C is a copula. From $C(x) = F_{f(b_x)}(1, \ldots, 1)$, we know that $C(1, \ldots, 1, t, 1, \ldots, 1) = t$ for any $t \in [0, 1]$; this justifies that C has standard uniform marginals. It remains to show that C is a distribution function.

11

Let us verify that C is increasing. Fix $x \in [0,1]^n$ and $\delta \in (0, 1 - x_1]$. Take $p \in \Delta_{\text{ma}}^n$, $s \in [z]$ and $\varepsilon > 0$ such that such that $(F_{p_1}(s_1), \ldots, F_{p_n}(s_n)) = x$ and $F_{p_1}(s_1 + \varepsilon) = x_1 + \delta$. This implies $(F_{p_1}(s_1 + \varepsilon), F_{p_2}(s_2), \ldots, F_{p_n}(s_n)) = x + (\delta, 0, \ldots, 0)$. By (3), we have $C(x + (\delta, 0, \ldots, 0)) \geqslant C(x)$ because the distribution function $F_{f(p)}$ is increasing. Repeating this argument for the other components, we get that C is increasing.

To show that C is a distribution function, take $k \in \mathbb{N}$ and let $G_k = \{j/2^k : j \in [2^k]\}$, that is, the $(1/2^k)$-grid. Let F^k be the distribution function on $(G_k)^n$ given by $F^k(s/2^k) = F_{f(p)}(s)$ where $p_1, \ldots, p_n$ are uniform distributions on $[2^k]$; that is, we scale the distribution $F_{f(p)}$ by $1/2^k$. Using (3), we get, for $s \in [2^k]^n$,

$$C\left(\frac{s_1}{2^k}, \ldots, \frac{s_n}{2^k}\right) = F_{f(p)}(s) = F^k\left(\frac{s_1}{2^k}, \ldots, \frac{s_n}{2^k}\right).$$

Therefore, C coincides with a distribution function F^k on $(G_k)^n$. For any $x \in [0,1]^n$, let $a^k, b^k \in (G_k)^n$ be the componentwise floor and ceiling x with respect to the grid $(G_k)^n$, respectively. We have, by the monotonicity of C and F^k, as $k \to \infty$,

$$|F^k(x) - C(x)| \leqslant F^k(b^k) - C(a^k)$$

$$= F^k(b^k) - F^k(a^k) \leqslant \|b^k - a^k\| \leqslant n2^{-k} \to 0,$$

where $\|\cdot\|$ is the 1-norm and the penultimate inequality follows from that marginals of F^k are uniform on G_k. Hence, $F^k \to C$ weakly as $k \to \infty$. Because the support of F^k is contained in $[0,1]^n$, the set $\{F^k : k \in \mathbb{N}\}$ is compact, by e.g., Prokhorov's theorem. Therefore, its weak limit C is a distribution function, and thus C is a copula.

Next we prove the "if" statement. Suppose that f is a copula model for copula C. Let P_C be the probability measure with distribution function C. For $p \in \Delta_{\text{ma}}^n$ and $s \in [z]$,

$$f(p)(s) = P_C\left(\underset{k \in N}{\times} [F_{p_k}(s_k - 1), F_{p_k}(s_k)]\right).$$

For any $i \in [n]$ and j, we have

$$f(T_i^j(p))(s) = P_C\left(\underset{k \in N}{\times} \left[\hat{F}_{p_k}(s_k - 1), \hat{F}_{p_k}(s_k)\right]\right),$$

where $\hat{F}_{p_k} = F_{p_k}$ for $k \neq i$ and $\hat{F}_{p_k} = F_{p_i^j}$ if $k = i$. Note that $[F_{p_i^j}(s_i - 1), F_{p_i^j}(s_i)]$ is equal to $[F_{p_i}(s_i - 1), F_{p_i}(s_i)]$ if $s_i < j$, to $[F_{p_i}(s_i - 1), F_{p_i}(s_i + 1)]$ if $s_i = j$, and to $[F_{p_i}(s_i), F_{p_i}(s_i + 1)]$ if $s_i > j$. This shows (1), and hence IA holds.

Theorem 1 asserts that the only models satisfying IA are the copula models. Three technical remarks are in order.

- Our notion of a model allows any $z \in \mathbb{N}^N$. However, we could restrict the domain of z to any $\times_i [1, \hat{z}_i]$, where $\hat{z}_i \in \mathbb{N} \cup \{+\infty\}$. So long as $\hat{z}_i \geqslant 3$, the characterization still holds (a different, more tedious proof is necessary, however).

- We assume that f depends on $\mathbf{z}$, but is otherwise independent of the names of the ordered objects in $[z_i]$. This means that our approach is implicitly ordinal, but also means our approach is highly flexible. From a practical standpoint, the underlying map carrying "real world data" to the members of $[z_i]$ can be ignored in applications. However, we can envision a more general approach, whereby f is permitted to depend on the names of the elements of $[z_i]$. For example, consider the "low, medium, high" example described in the introduction. $L < M < H$ constitutes a three alternative ordered set, and is modeled by $z_i = 3$. If low and medium are merged, we get $LM < H$, a two element set modeled by $z_i = 2$; similarly when medium and high are merged $L < MH$. Our framework applies the same function f to the two distinct merged environments. Instead, we could allow f to depend explicitly on the names of the alternatives, rather than on z_i. The most natural way to do this would be to assume ambient totally ordered sets X_i for each $i \in [n]$ ($X_i = \mathbb{R}$ seems a natural example), and for each i, a finite ordered partition of X_i, say π_i; the set of π_i could be restricted to some class. In our running example, $X_i = \{L, M, H\}$ and when L and M are merged, we would consider $\pi_i = \{\{L, M\}, \{H\}\}$. In our current framework, $z_i = |\pi_i|$, the cardinality of the ordered partition, but we could allow f to depend on $\times_i \pi_i$ instead of only on the cardinalities of each π_i. An IA property is easy to define for this environment. While we leave the study of this property to future research, we conjecture that it relates to a parameteric generalization of the notion of copula.

- Models satisfying IA allow us to consider the introduction of a new dimension. If n is very large, then at any given time we may consider only some subset $M \subseteq N$; the remaining probability putting a point mass on the members of $N \backslash M$. However, even if we do not wish to bound n, the introduction of a new dimension, say, $n + 1$, poses no difficulty. Copulas can always be extended to include another dimension. Even

more, we are free to work with $[0, 1]^{\mathbb{N}}$, endowed with the Borel product σ-algebra. We could define a copula in this case to be any measure μ with uniform marginals, and consider at any point in time only finite subsets of $\mathbb{N}$ with the induced marginal cdf on those elements. Such probability measures μ obviously exist, owing to the Kolmogorov extension theorem, for example a joint distribution of an independent infinite sequence of uniform random variables on $[0, 1]$. Symmetric ones are easy to describe; for example analogues of the product copula and Frechet lower-bound are natural candidates.

4 Unordered Categorical Variables

The preceding results dealt with the case in which choices came from some ordered set. In this section, we investigate categorical data, where observations need not be ordered. A familiar example of such choices might be transportation choices: observations consist of a choice between bus and train. Now, suppose we can refine choices further, so that we can observe whether a bus is red or blue. There is no natural order across red bus, blue bus, and train. Because there is no natural notion of ordering here, there is no natural notion of what it means for two choices to be adjacent.

We can discuss this issue in our setup. Up to now, IA allows us to combine two adjacent bins (represented by integers j and $j + 1$) in a marginal distribution. If two non-adjacent bins j and k are combined, it is not immediately clear whether we should place this new bin at the location of j or k. Indeed, to allow for combining arbitrary bins for a probability $q \in \Delta([w])$, the interpretation of location among the points in $[w]$ is lost. Moreover, without an ordinal structure, adjacency cannot be properly discussed.

Inspired by this observation, we propose the following symmetry property that removes the location from points in $[w]$. Let Σ_w be the set of all permutations on $[w]$ for $w \in \mathbb{N}$. For $q \in \Delta([w])$ and $\sigma \in \Sigma_w$, $q^{\sigma}(s) = q(\sigma(s))$ for all $s \in [w]$. Let T_i^{σ} be the mapping that maps p to $(p_1, \ldots, p_{i-1}, p_i^{\sigma}, p_{i+1} \ldots, p_n)$, for all $i \in [n]$ and $\sigma \in \Sigma_{z_i}$. When T_i^{σ} is applied to a vector s, it means $T_i^{\sigma}(s) = (s_1, \ldots, s_{i-1}, \sigma(s_i), s_{i+1}, \ldots, s_n)$.

Let $M = [m]$ be a subset of N. In the following definitions, M should be understood as the set of categorical choices, whereas the remaining members of N are to be understood

as ordinal. So the following definitions allow for a combination of categorical as well as ordinal data.

M-neutrality: For $p \in \Delta_{\mathrm{ma}}^n$, $i \in M$, $\sigma \in \Sigma_{z_i}$ and $s \in [z]$, $f(T_i^\sigma(p))(s) = f(p)(T_i^\sigma(s))$.

The interpretation of M-neutrality is that switching the labels of two bins (or reshuffling them among several bins) within M does not change the joint distribution. Intuitively, the names of the bins for variables in M do not matter (the standard neutrality notion of social choice).

Theorem 2 *A model satisfies IA and M-neutrality if and only if it is a copula model with $C(x) = (\prod_{i \in M} x_i)\tilde{C}((x_j)_{j \in N \setminus M})$, where $\tilde{C}$ is a copula on $\mathbb{R}^{n-m}$.*

Proof: Let P_C be the probability measure with distribution function C. If f is the copula model with copula C, then IA follows from Theorem 1, and for $i \in M$ we have

$$f(T_i^\sigma(p))(s) = p_i(\sigma(s_i))P_C(s_{-i}) = f(p)(T_i^\sigma(s)),$$

where $P_C(s_{-i})$ is the probability of the set $\{t \in [z] : t_j = s_j,\ j \neq i\}$ under P_C. Hence, M-neutrality holds. Next, we show that if a copula model with copula C satisfies M-neutrality, then C must be the copula in the theorem. Let $p_1, \ldots, p_n$ be uniform distributions on $[k]$ for $k \in \mathbb{N}$. For $s \in [k]^n$, $f(p)(s) = P_C(\{(s_1/k, \ldots, s_n/k)\})$. Note that $T_i^\sigma(p) = p$ for each $i \in N$. Hence, for $i \in M$, we have

$$P_C(\{(s_1/k, \ldots, s_n/k)\}) = P_C(\{T_i^\sigma(s_1/k, \ldots, s_n/k)\})$$

for all $\sigma \in \Sigma_k$. Repeatedly applying this equality, we get that, for each given $(s_i)_{i \in N \setminus M}$, P_C assigns equal probability to each point in the set $\{(s_1/k, \ldots, s_n/k) : (s_i)_{i \in M} \in [k]^m\}$. Since k is arbitrary, we know that the M-marginal distribution of P_C is uniform on $[0, 1]^m$, and under P_C, the components in M are independent of all remaining components in $N \setminus M$. Hence, the copula has the form $C(x) = (\prod_{i \in M} x_i)\tilde{C}((x_j)_{j \in N \setminus M})$.

If all variables are categorical, and we insist on the neutrality property, only the independence model can be used. On the other hand, we have some freedom in the choice of our copula for non-categorical variables; for categorical variables, the model behaves as the independence model, but otherwise is completely flexible.

5 Conclusion

In this paper, we proposed and characterized the property of Invariant Aggregation (IA)—a condition that requires merging adjacent categories in marginal distributions not to distort the joint distribution over unaffected categories. We proved that IA holds if and only if the model is based on a copula, thereby offering a precise, axiomatic foundation for when and why copula-based modeling is appropriate in discrete choice settings. This result highlights the powerful structure of copulas in maintaining consistency under data coarsening or refinement, which is common in empirical contexts.

Our investigation was motivated by problems in stochastic choice theory, especially in contexts where joint choice data is incomplete or unobserved. In economics, the rationalization of data is usually motivated by an underlying behavioral model, and in some such environments, there may be either no rationalization or multiple rationalizations. The canonical example of such a framework is the *random utility model* (RUM), which we will not describe here (it is an example of a categorical model; no exogenous order on X is specified) Block & Marschak (1960), Falmagne (1978), McFadden (1978), Kono et al. (2023), Yildiz (2016), Dardanoni et al. (2022).

Recent contributions focus on ordered choices. A major contribution is due to Apesteguia et al. (2017), Filiz-Ozbay & Masatlioglu (2023), the so-called *progressive random choice* (PRC) model (see also Yildiz (2022), Apesteguia & Ballester (2023), Petri (2023), Masatlioglu & Vu (2024)). In this model, the set X is endowed with a linear order and each $A_i \subseteq X$ inherits the restriction of this order. Chambers et al. (2025) introduces copula-based models for stochastic choice and shows that the progressive random choice model is a special case. Indeed, the progressive random choice model coincides with the Fréchet-Hoeffding upper bound model. In economics problems, we sometimes imagine new data arriving in the form of a new subset $B \subseteq X$ and a distribution over B. We realized that the application of the PRC model when including B in the data did not distort the marginal probabilities assigned to each member of $\times_i A_i$. This naturally led to the investigation of the IA property here.

For future research, it is worth considering the more general notion of a model, which allows for empty or multi-valued sets of rationalizations. It can be proved that RUM

satisfies an analogue of IA tailored to such an environment, but we do not know the general class. The class of models satisfying IA in a more general cardinally specified framework is also of interest.

References

Agresti, A. (2010), *Analysis of ordinal categorical data*, John Wiley & Sons. 3

Apesteguia, J. & Ballester, M. A. (2023), 'Random utility models with ordered types and domains', *Journal of Economic Theory* **211**, 105674. 16

Apesteguia, J., Ballester, M. A. & Lu, J. (2017), 'Single-crossing random utility models', *Econometrica* **85**(2), 661–674. 16

Bastianello, L. & LiCalzi, M. (2019), 'The probability to reach an agreement as a foundation for axiomatic bargaining', *Econometrica* **87**(3), 837–865. 10

Beare, B. K. (2010), 'Copulas and temporal dependence', *Econometrica* **78**(1), 395–410. 2

Block, H. & Marschak, J. (1960), *Contributions to Probability and Statistics*, Stanford University Press, Stanford, CA. 16

Chambers, C. P., Masatlioglu, Y. & Yildiz, K. (2025), 'Ordered probabilistic choice', *arXiv preprint arXiv:2504.01227* . 2, 16

Chryst, B., Marlon, J., van der Linden, S., Leiserowitz, A., Maibach, E. & Roser-Renouf, C. (2018), 'Global warming's "six americas short survey": Audience segmentation of climate change views using a four question instrument', *Environmental Communication* **12**(8), 1109–1122. 4

Dardanoni, V., Manzini, P., Mariotti, M., Petri, H. & Tyson, C. (2022), 'Mixture choice data: Revealing preferences and cognition', *Journal of Political Economy* . 16

Downs, A. (1957), *An Economic Theory of Democracy*, Harper. 4

Falmagne, J.-C. (1978), 'A representation theorem for finite random scale systems', *Journal of Mathematical Psychology* **18**(1), 52–72. 16

Fan, Y. & Henry, M. (2023), 'Vector copulas', *Journal of Econometrics* **234**(1), 128–150. 2

Filiz-Ozbay, E. & Masatlioglu, Y. (2023), 'Progressive random choice', *Journal of Political Economy* **131**(3), 716–750. 16

Frick, M., Iijima, R. & Ishii, Y. (2022), 'Dispersed behavior and perceptions in assortative societies', *American Economic Review* **112**(9), 3063–3105. 10

Genest, C. & Rivest, L.-P. (1993), 'Statistical inference procedures for bivariate archimedean copulas', *Journal of the American statistical Association* **88**(423), 1034–1043. 2

Joe, H. (2014), *Dependence modeling with copulas*, CRC press. 2

Johnson, V. E. & Albert, J. H. (2006), *Ordinal data modeling*, Springer Science & Business Media. 3

Kono, H., Saito, K. & Sandroni, A. (2023), 'Axiomatization of random utility model with unobservable alternatives', *arXiv preprint arXiv:2302.03913* . 16

Li, D. X. (1999), 'On default correlation: A copula function approach', *Available at SSRN 187289* . 2

Masatlioglu, Y. & Vu, T. P. (2024), 'Growing attention', *Journal of Economic Theory* **222**, 105929. 16

McFadden, D. (1978), 'Modeling the choice of residential location', *Transportation Research Record* (673). 16

McNeil, A. J., Frey, R. & Embrechts, P. (2015), *Quantitative risk management: concepts, techniques and tools-revised edition*, Princeton university press. 2

Nelsen, R. B. (2006), *An introduction to copulas*, Springer. 3, 10

Petri, H. (2023), 'Binary single-crossing random utility models', *Games and Economic Behavior* **138**, 311–320. 16

Sklar, M. (1959), Fonctions de répartition à n dimensions et leurs marges, *in* 'Annales de l'ISUP', Vol. 8, pp. 229–231. 2

Yildiz, K. (2016), 'List-rationalizable choice', *Theoretical Economics* **11**(2), 587–599. 16

Yildiz, K. (2022), Self-progressive choice models. 16