

Hefti, Andreas

Working Paper

A note on symmetric random vectors with an application to discrete choice

Working Paper, No. 419

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Hefti, Andreas (2022) : A note on symmetric random vectors with an application to discrete choice, Working Paper, No. 419, University of Zurich, Department of Economics, Zurich,
<https://doi.org/10.5167/uzh-221673>

This Version is available at:

<https://hdl.handle.net/10419/266242>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 419

A Note on Symmetric Random Vectors with an Application to Discrete Choice

Andreas Hefti

October 2022

A Note on Symmetric Random Vectors with an Application to Discrete Choice

Andreas Hefti*

August 2022

Abstract

This paper studies random vectors X featuring symmetric distributions in that i) the order of the random variables in X does not affect its distribution, or ii) the distribution of X is symmetric at zero. We derive a number of characterization results for such random vectors, thereby connecting the distributional symmetry to various notions of how (Euclidean) functions have been regarded as symmetric. In addition, we present results about the marginals and conditionals of symmetrically distributed random vectors, and apply some of our results to various transformations of random vectors, e.g., to sums or products of random variables, or in context of a choice probability system known from economic models of discrete choice.

Keywords: Symmetric Distributions, Symmetric Random Vectors, Symmetric Random Variables, Symmetric Functions, Choice Probability System

JEL classification: X

*Department of Economics, University of Zurich, and School of Management and Law, Zurich University of Applied Science.
Email: ahefti@heftynomics.com.

1 Introduction

The symmetry of certain shapes or functions have pervaded the mathematical profession and related fields, such as geometry, crystallography, quantum mechanics, statistics or economics. In probability theory, the notion of symmetry mostly pertains to the distribution of some random variable (RV). The Normal distribution, Student's distribution, the Uniform or the Logistic distribution all are examples of distributions that are symmetric around some $x \in \mathbb{R}$. Among such RV's, those that are zero symmetric embody a strong notion of *unbiasedness* of the distribution: They have a zero mean (if it exists), and always display equal tail probabilities $P(X \leq x) = P(X \geq -x)$, which is a helpful property, e.g., to statistics or econometrics for doing inference.

Regarding random vectors $X = (X_1, \dots, X_n)$, there are multiple ways how the respective distributions could be called symmetric. First, one can generalize the notion of zero symmetry to hold for some or all component RV's of X . However, as these RV's can be correlated – an important general aspect of the distributions of random vectors – it is not obvious in general how the zero symmetry of the various RV's in X matter for the overall zero symmetry of X itself. Second, another intuitive notion of symmetry concerns the order of the random variables (RV's) constituting X , asking whether one can permute these RV's without altering the distribution of X . This is a reasonable property if one thinks of the RV's in X as representing different observations from a random sample. Symmetry in this sense then states that the sampling order itself should not have any predictive power. This is evidently the case if one is willing to assume that the RV's in X are iid, as is a standard supposition in applied statistics and econometrics. But what if the various independent RV's fail to be identically distributed, or if at least some RV's are correlated?

While the symmetry of random variables or vectors matters in fields as diverse as algebraic theory, crystallography, quantum mechanics, statistics or economics, the definite connections between the various ways how functions have been called “symmetric” and the symmetry of related distributions has not been comprehensively unfolded yet, to our knowledge. With this note, we aim at making the corresponding picture more complete. In particular, we provide a number of characterization results for the above two notions of symmetric random vectors, e.g., in terms of CDF's, Characteristic Functions or PDF's, thereby relating the distributional symmetry to the symmetry of certain functions. Further, we are interested in transformations that preserve the relevant symmetry of the random vectors, and correspondingly consider marginal and conditional distributions, respectively. In addition, we identify a simple class of Euclidean functions – those featuring center symmetric transformations – that preserves the zero symmetry of any random vector. This result is powerful enough to conclude that various important transformations, such as (weighted) sums or certain products of RV's from a zero symmetric random vector will retain the zero symmetry as its key property.

To provide a more elaborated example for the latter, applied economics frequently depicts choice as selecting one of $n \in \mathbb{N}$ different products, where tastes (i.e., individual preferences) are dispersed over a consumer population of, say, measure one. Let $n = 2$ for the sake of illustration, and suppose that consumer valuation for

products $j = 1, 2$ is $U_j = q_j - p_j + X_j$, where $p_j > 0$ is the price of the product, $q_j \geq 0$ is product quality, and X_j is an RV expressing how well product j matches consumers' tastes. Thus, consumer tastes are distributed over the population according to the random vector $X = (X_1, X_2)$, defined on some probability space (Ω, \mathcal{F}, P) . It follows that utilities themselves are dispersed according to a random vector (U_1, U_2) .¹ Ignoring the possibility of binding budget constraints, a rational consumer chooses option j if $U_j = \max\{U_1, U_2\}$. Thus, the choice probability for, say, option $j = 1$ (or the fraction of consumers who choose $j = 1$) is $P_1 = P(U_2 - U_1 \leq 0)$. Plugging in the definition of U_j reveals that P_1 depends decisively on the distribution of *valuation differences* $Y_1 \equiv (X_2 - X_1)$ for given prices and qualities, as $P_1 = P(Y_1 \leq w_1)$, $w_1 \equiv (a_1 - a_2) + (p_2 - p_1)$. In applications, the literature frequently assumes that tastes X_1, X_2 are independent and zero symmetric, which is a strong form of assuring that tastes are *unbiased* within the consumer population.² It is an exercise to see that if tastes X_1, X_2 are independent and zero symmetric, then the difference distribution Y_1 must also be zero symmetric for $j = 1, 2$, meaning that rational choices must also reflect the unbiasedness of tastes. For example, for equal prices and qualities, each firm gets exactly half of the consumer population. Likewise, for given prices and quality levels, a permutation of the firm index just permutes the respective choice probabilities, reflecting that no firm has an unilateral advantage in terms of consumer-side willingness-to-pay. But what if X_1 and X_2 are correlated? Would this not generally imply that the difference distribution Y_1 , and hence the choice probability system, could feature some form of bias towards a certain product, which reflects the correlation structure? The answer to this question is a direct corollary to one of the transformation theorems in this paper. In particular, we show that even if $n \geq 2$ and we do not restrict the correlation of the RV's in $X = (X_1, \dots, X_n)$ other than possibly by the requirement of zero symmetry, the difference distributions Y_1, \dots, Y_n must always be zero symmetric as well. Thus zero symmetry of the taste distribution X is all that is needed to assure that the correlation among the X_j 's washes out.³

The paper is structured as follows. Section 2 introduces the relevant concepts from probability theory, in particular the two symmetry notions for random vectors outlined above (Section 2.1), and the symmetry notions for functions we require (Section 2.2). The main results are in Section 3. We present various characterization results for symmetric random vectors (Section 3.1), and derive analogous results for the marginal and conditional distributions in Section 3.2. Finally, Section 3.3 shows that zero symmetry is preserved under arbitrary center symmetric transformations and discusses various implications, including our motivational example of unbiased

¹Such Random Utility Models generally allow for two possible interpretations (Anderson et al., 1992). In the first, one thinks of modeling the choice of a single consumer, about whose precise tastes the observer (e.g., the econometrician) has limited information (see Manski, 1977 for a discussion). This is frequently the relevant interpretation when estimating discrete choice models. In the second, one thinks of a continuum of consumers with tastes distributed according to X over the consumer population. Then, X can reflect some form of "locational" allocation of consumers and products across a metric space (or a preference space) as well as imperfect cognitive abilities of consumers.

²This does not mean that the two products must be symmetric. It can well be that, *ceteris paribus*, one product is more preferred than the other as it features a higher exogenously given quality. In fact, assuming iid zero symmetric RV's X_1, X_2 with exogenously given quality levels (q_1, q_2) is equivalent to assuming iid RV's which are symmetric at (q_1, q_2) .

³Allowing that the RV's in X can be correlated is vital in that if valuations U_j are micro-founded by some form of locational model – such as a Salop circle – then the corresponding random utility model naturally features correlated tastes; see, e.g., Hefti et al. (2022).

choice probability systems. Longer proofs are in Appendix A.1.

2 Basic Definitions and Concepts

Consider the measurable space $(\mathbb{R}^k, \mathcal{B}^k)$, where \mathcal{B}^k denotes the Borel field on \mathbb{R}^k . We denote the elements of \mathbb{R}^k as k -vectors $x = (x_1, \dots, x_k)$. A measurable measurable function g from $(\mathbb{R}^n, \mathcal{B}^n)$ to $(\mathbb{R}^m, \mathcal{B}^m)$ is a Borel function. Whenever the domain of g is $S \subset \mathbb{R}^n$, we assume that $S \in \mathcal{B}^n$ and also call g a Borel function. With respect to probability, we need the following standard notions. Let Ω be a non-empty set, and (Ω, \mathcal{F}, P) is a probability space. A measurable function $X : \Omega \rightarrow \mathbb{R}^n$ is a *random vector* with distribution dF_X on $(\mathbb{R}^n, \mathcal{B}^n)$; if $n = 1$ then X is a *random variable* (RV). Recall that the distribution dF_X of a random vector X is characterized by its Cumulative Distribution Function (CDF) $F_X : \mathbb{R}^n \rightarrow \mathbb{R}$, $F(x) \equiv dF_X((-\infty, x])$, by its Characteristic Function (ChF) $\varphi_X : \mathbb{R}^n \rightarrow \mathbb{C}$, $\varphi_X(t) \equiv E[e^{it \cdot X}]$, or by its Moment Generating Function (MGF) $m_X : \mathbb{R}^n \rightarrow \mathbb{R}$, $m_X(t) \equiv E[e^{t \cdot X}]$, provided that the latter expectation exists. Finally, two random vectors X, Y are equal in distribution, written as $X \stackrel{d}{=} Y$, if they have the same distribution ($dF_X(B) = dF_Y(B) \forall B \in \mathcal{B}^n$).

2.1 Symmetric Random Vectors: Exchangeability and Zero Symmetry

We next introduce the two notions of symmetric random vectors outlined in the introduction. Our first notion – *exchangeability* – states that the distribution dF_X of a random vector $X = (X_1, \dots, X_n)$ is invariant to the “order” of the random variables X_1, \dots, X_n .

Definition 1 (Exchangeability) *A random vector $X = (X_1, \dots, X_n)$ is ij -exchangeable if*

$$(X_1, \dots, X_i, \dots, X_j, \dots, X_n) \stackrel{d}{=} (X_1, \dots, X_j, \dots, X_i, \dots, X_n).$$

If $X \stackrel{d}{=} \sigma(X) \equiv (X_{\sigma(1)}, \dots, X_{\sigma(n)})$ for every permutation σ of the set $\{1, \dots, n\}$, then X is exchangeable.

If X is an exchangeable random vector, then every possible way how the RV's X_1, \dots, X_n can be written as a k -vector yields one and the same distribution. In this sense, an exchangeable random vector features an *irrelevance of order* of its constituting random variables.

Our second notion of symmetry – *zero symmetry* – captures the notion of *unbiasedness* of a random vector X suggested by the introduction.

Definition 2 (Zero Symmetry and Axial Symmetry) *A random vector X is zero symmetric if $X \stackrel{d}{=} -X$.*

Further, if X satisfies $(X_1, \dots, X_i, \dots, X_n) \stackrel{d}{=} (X_1, \dots, -X_i, \dots, X_n)$, then X is i -axially symmetrically distributed.

Finally, X is axially symmetrically distributed if X is i -axially symmetrically distributed for every $i = 1, \dots, n$.

A zero symmetric random vector is unbiased in the sense that its positive and negative values are equally likely. Likewise, if X is i -axially symmetrically distributed, then flipping the sign of the RV X_i does not affect the distribution of X . Put differently, if X is i -axially symmetrically distributed, then the RV X_i is zero symmetric. However, it may be of interest to observe that a zero symmetric random vector may fail to be axially symmetrically distributed, while axial symmetry always assures zero symmetry (see below). Further, our focus on zero symmetric random vectors is without loss in the followings sense. If X is zero symmetric, then $Y \equiv X + a$ is a -symmetric, for any given $a \in \mathbb{R}^n$, in that $(Y - a) \stackrel{d}{=} (a - Y)$. In reverse, any a -symmetric random vector Y can always be decomposed as $Y = X + a$, where X is zero symmetric.

2.2 Symmetric Functions

The above two symmetry definitions pertain to aspects of the distribution of a random vector. By contrast, the following symmetry notions pertain to functions. As we shall see, the various symmetry notions will be tightly related to each other. Let $S \subset \mathbb{R}^n$ denote a nonempty subset of \mathbb{R}^n , and $g : S \rightarrow \mathbb{R}^m$ be a function. We call a function g exchangeable if the value of this function is invariant to arbitrary permutations of its entries, e.g., if $g(x_1, x_2) = g(x_2, x_1)$ everywhere.

Definition 3 (Exchangeable Functions) *A function $g : S \rightarrow \mathbb{R}^m$ is ij -exchangeable if*

$g(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = g(x_1, \dots, x_j, \dots, x_i, \dots, x_n)$ on S . If

$$g(x_1, \dots, x_i, \dots, x_n) = g(x_{\sigma(1)}, \dots, x_{\sigma(i)}, \dots, x_{\sigma(n)})$$

holds on S for any permutation σ of $(1, \dots, n)$, then g is an exchangeable function.

In words, ij -exchangeability states that we can exchange the values of the i -th and j -th projection with each other without altering the value of g . Geometrically, this means that, for each coordinate function g_i of g and fixed values of $x_k \forall k \neq i, j$, the level sets of g_i are symmetric at the 45-degree line in the ij -plane (see Figure 1). Another well-known type of symmetry occurs if g is symmetric with respect to some coordinate axis.

Definition 4 (Axial Symmetry) *A function $g : S \rightarrow \mathbb{R}^m$ is i -axially symmetric if $g(x_1, \dots, x_i, \dots, x_n) = g(x_1, \dots, -x_i, \dots, x_n)$ on S . Further, g is axially symmetric if g is i -axially symmetric for all $i = 1, \dots, n$.*

If g is i -axially symmetric, this geometrically means that the level sets of g are rotated around the i -th coordinate axis (see Figure 1). If $m = 1$, a function g which is axially symmetric has sometimes been called an *even symmetric function*. Note that g is zero symmetric, i.e., $g(x) = g(-x)$ for all $x \in S$, whenever g is axially symmetric, while the converse does not hold. ⁴

⁴For instance, the function $g(x_1, x_2) = x_1 x_2$ verifies $g(-x) = g(x)$ on \mathbb{R}^2 , while it is not 1-axially symmetric.

As for our last notion of symmetry, a function may also be symmetric with respect to some point $a \in \mathbb{R}^n$, where the case $a = 0$ will be of relevance to us.

Definition 5 (Center Symmetry) *A function $g : S \rightarrow \mathbb{R}^m$ is center symmetric if $g(-x) = -g(x)$ on S .*

A function that is center symmetric is reflection symmetric at the origin. If $m = 1$, a center symmetric function sometimes has been called an *odd symmetric* function, or a *parity inversion* in quantum mechanics. For $m > 1$, observe that g is center symmetric if and only if each of its coordinate functions g_i , $i = 1, \dots, m$ is center symmetric.

It is easy to gauge that exchangeability, axial symmetry and center symmetry are different properties. Figure 1 depicts three different functions defined on $S = [-1, 1] \times [-1, 1]$. The function in Panel A is exchangeable and zero symmetric (i.e. $g(x) = g(-x)$), but it is not i -axially symmetric for $i = 1, 2$ and also not center symmetric. The function in Panel B is exchangeable, axially symmetric and zero symmetric but not center symmetric. Finally, the function in Panel C is not exchangeable, 1-axially symmetric but not 2-axially symmetric, and center symmetric.

While the notions of center symmetry and axial symmetry are different, they are related in the following sense. If g is a center symmetric C^1 -function, then the points x and $-x$ must have the same derivative, i.e., $Dg(x) = Dg(-x)$. Put differently, the derivative function Dg of a center symmetric function itself must be axially symmetric. This also implies that g has one and the same directional derivative at the points x and $-x$, i.e., $D_v g(x) = D_v g(-x)$. If g is real-valued, this geometrically implies that the level sets at the point $-x$ can be obtained from the ones at x by rotating the latter around the origin, where however the gradients at x and $-x$, respectively, are *not* rotated but rather point in the same direction (see Figure 1). We use this occasion to establish the converse of the above derivative condition.

Proposition 1 *Let M be open and convex. A C^1 -function $h : M \rightarrow \mathbb{R}^m$ is a translated version of a center symmetric function $g : M \rightarrow \mathbb{R}^m$, i.e., $h(x) = g(x) + c$ for some $c \in \mathbb{R}^m$, if and only if the derivative of h is zero symmetric, i.e., $Dh(-x) = Dh(x)$ on M .*

3 Main Results

3.1 Symmetric Random Vectors: Characterization Results

We first present various characterizations for our two notions of symmetric random vectors

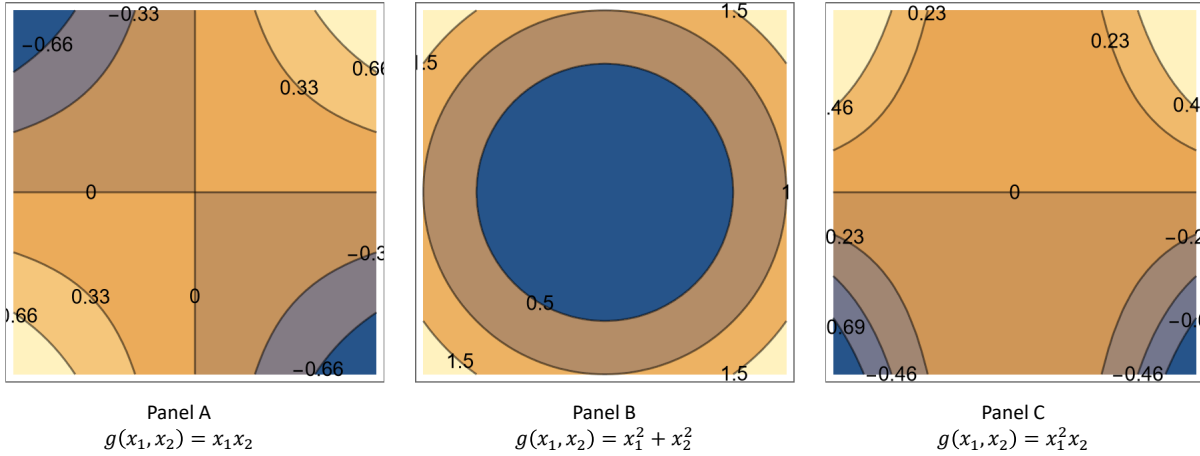


Figure 1: Symmetric Functions

3.1.1 Exchangeable Random Vectors

Let $X = (X_1, \dots, X_n)$ be a random vector. It is easy to see that if all RV's constituting X are iid, then X must be exchangeable.⁵ If one thinks of such RV's as iid observations from a random n -sample, exchangeability simply paraphrases the standard premise of applied statistics that the sampling order of the observations should be irrelevant for probabilistic considerations. One of the conclusions in this section is that the converse of this statement is false, as the "irrelevance of sampling order" can hold even if there is correlation among the RV's in X . We begin by establishing equivalence between exchangeability of a random vector and exchangeability of important related functions.

Proposition 2 *Let X be a random vector. The following statements are equivalent: 1) X is ij -exchangeable, 2) $F_X(x)$ is ij -exchangeable, 3) $\varphi_X(t)$ is ij -exchangeable, 4) $m_X(t)$ is ij -exchangeable (if $m_X(t)$ exists).*

If X allows for a density function, then a similar result holds:

Corollary 1 *Suppose that $X = (X_1, \dots, X_n)$ is either a discrete random vector or absolutely continuous with respect to Lebesgue measure. Then X has a density function $f_X(x)$, and X is ij -exchangeable if and only if f_X is ij -exchangeable almost everywhere. In addition, if the density $f_X(x)$ is ij -exchangeable almost everywhere, then $f_X(x)$ also is a density for the random vector where X_i and X_j have been permuted.*

The next result shows that for a random vector X to be exchangeable it is necessary and sufficient that one can permute any two arbitrary RV's in X without changing dF_X . It formally rests on the following Lemma, stating that any permutation of a finite set can be decomposed into at most $(n - 1)$ binary permutations.

⁵If the RV's in X are only independent but not identically distributed, X will not be exchangeable in general.

Lemma 1 Let σ_{ij} denote the permutation of $(1, \dots, n)$ that only exchanges i with j , and $\sigma(1, \dots, n) \equiv (\sigma(1), \dots, \sigma(n))$ be an arbitrary permutation of $(1, \dots, n)$. Then $\sigma(1, \dots, n) = \sigma_{(n-1)\sigma(n-1)} \circ \dots \circ \sigma_{2\sigma(2)} \circ \sigma_{1\sigma(1)}(1, \dots, n)$

Corollary 2 A random vector X is exchangeable if and only if X is ij -exchangeable for every $i, j \in \{1, \dots, n\}$.

By Corollary 2, we can directly use the characterizations in Proposition 2 and Corollary 1 to check whether X is exchangeable by verifying, e.g., whether F_X, φ_X etc. are ij -exchangeable for any index pair (i, j) . Moreover, exchangeability by no means requires that the RV's in X must be iid. To illustrate, suppose that $X = (X_1, X_2, X_3)$ has a (Lebesgue-) density $f_X(x_1, x_2, x_3) = \frac{2}{3}(x_1 + x_2 + x_3)$ with support $\text{supp}(X) = [0, 1] \times [0, 1] \times [0, 1]$. While the RV's in X are not iid, X still is exchangeable as the density f_X is ij -exchangeable for every pair (i, j) .

We end this section with a remark on the expectation and covariance, respectively, of an ij -exchangeable random vector.

Proposition 3 Let X be ij -exchangeable and Y be the correspondingly permuted random vector. Then, $E[X_i] = E[X_j]$ and $\text{Cov}(X_i, X_k) = \text{Cov}(X_j, X_k)$ for every $k = 1, \dots, n$, whenever the respective expressions exist.

Proof: The first claim follows as $X \stackrel{d}{=} Y \Rightarrow E[X] = E[Y]$ (provided that the expectation exists). Recall that $\text{Cov}(X_h, X_k) = E[(X_h - E[X_h])(X_k - E[X_k])]$. Using the Borel function $f_i(x_1, \dots, x_n) \equiv (x_i - E[X_i])(x_k - E[X_k])$ for some $k = 1, \dots, n$, and the fact that $f_i(X) \stackrel{d}{=} f_i(Y)$, yields $\text{Cov}(X_i, X_k) = E[f_i(X)] = E[f_i(Y)] = \text{Cov}(X_j, X_k)$. ■

It readily follows from Proposition 3 that $E[X] = (E[X_1], E[X_1], \dots, E[X_1])$ for an exchangeable random vector. As $\text{Var}(X)$ is the symmetric $n \times n$ matrix with the covariances $\text{Cov}(X_h, X_k)$ as its entries, Proposition 3 further shows that if X is exchangeable, then $\text{Var}(X)$ must be a perfectly symmetric matrix in that all entries of its diagonal (the variances) are the same, and all entries of its off-diagonal are also the same (but not necessarily equal to zero).

3.1.2 Zero Symmetric Random Vectors

We next derive characterization results for axially symmetrically distributed and zero symmetric random vectors.

Proposition 4 Let X be a random vector. The following statements are equivalent: 1) X is i -axially symmetrically distributed, 2) $\varphi_X(t)$ is i -axially symmetric, 3) $m_X(t)$ is i -axially symmetric (if $m_X(t)$ exists).

Similar to ij -exchangeability, the notion of a random vector with an i -axially symmetrical distribution has a direct connection to i -axial symmetry of the ChF. Further, if X is an RV it is well known that X is zero symmetric

if and only if its ChF is real-valued. The following result summarizes the straightforward generalization to the case of zero symmetric random vectors.

Proposition 5 *A random vector X is zero symmetric if and only if $\varphi_X(t) = \varphi_X(-t)$ for every $t \in \mathbb{R}^n$, or equivalently if and only if $\varphi_X(t)$ is a real-valued function.*

Proof: For any $t \in \mathbb{R}^n$, we have $\varphi_{-X}(t) = \varphi_X(-t)$. Thus $\varphi_X(t) = \varphi_{-X}(t)$, i.e., $X \stackrel{d}{=} -X$, iff $\varphi_X(t) = \varphi_X(-t) \forall t$. The remaining equivalence follows from the fact that $\varphi_{-X}(t) = \overline{\varphi_X(t)}$ (complex conjugate of $\varphi_X(t)$) for any $t \in \mathbb{R}^n$. Thus $\varphi_X(t)$ is real-valued, i.e., $\varphi_X(t) = \overline{\varphi_X(t)}$, iff $\varphi_X(t) = \varphi_X(-t) \forall t$. ■

We next establish that every axially symmetric random vector must also be zero symmetric.

Proposition 6 *Let X be axially symmetrically distributed. Then X is zero symmetric, and also $X \stackrel{d}{=} Y$, where Y coincides with X except for the fact that one or more random variables of X enter Y with negative sign.*

Note that zero symmetry of X is, in general, weaker than X being i -axially symmetrically distributed for all i ; an example is presented after Corollary 4 below.

There is no general analogue between axial or zero symmetry of a random vector and its CDF, in contrast to what we found for exchangeability. However, for the special case where the RV's in X are independent, the following result applies.

Proposition 7 *Let $X = (X_1, \dots, X_n)$ consist of n independent RV's. Then X is zero symmetric if and only if*

$$\prod_{i=1}^n F_i(x_i) = \prod_{i=1}^n \left(1 - \lim_{s_i \uparrow (-x_i)} F_i(s_i) \right) \quad (1)$$

If, in addition, X is a continuous random vector, then X is zero symmetric if and only if $\prod_{i=1}^n F_i(x_i) = \prod_{i=1}^n (1 - F_i(-x_i))$.

Proof: By independence, $F_X(x) = \prod_{i=1}^n F_i(x_i)$ and $F_{-X}(x) = P(X \geq -x) = \prod_{i=1}^n P(X_i \geq -x_i)$, where $P(X_i \geq -x_i) = 1 - P(X_i < -x_i)$ and $P(X_i < -x_i) = 1 - \lim_{s_i \uparrow (-x_i)} F_i(s_i)$ (left-side limit). Thus $F_X(x) = F_{-X}(x)$ iff condition (1) holds. The last claim follows from $P(X_i < -x_i) = P(X_i \leq -x_i)$ if X is continuous. ■

Proposition 7 includes $n = 1$ as a special case: If X is an RV, then X is zero symmetric if and only if its CDF satisfies $F_X(x) + (1 - \lim_{s \uparrow (-x)} F_X(s)) = 1$ for all x . Further, if X is continuous and zero symmetric *because* X is axially symmetrically distributed, the following result holds:

Proposition 8 *If $X = (X_1, \dots, X_n)$ is continuous and axially symmetrically distributed, then $F_X(0, \dots, 0) = \frac{1}{2^n}$.*

Proof Let Y be a random vector where one or several RV's in X enter with negative sign. Because X is axially symmetric, Proposition 6 shows that also $X \stackrel{d}{=} Y$. Let \mathcal{Y} be the set of all such Y , noting that $|\mathcal{Y}| = 2^n - 1$. As $P(X \leq 0) = P(Y \leq 0)$ for any $Y \in \mathcal{Y}$ and, because X is continuous, also $1 = P(X \leq 0) + \sum_{Y \in \mathcal{Y}} P(Y \leq 0)$, we obtain $1 = P(X \leq 0) + |\mathcal{Y}|P(X \leq 0)$, which shows that $F_X(0, \dots, 0) = P(X \leq 0) = \frac{1}{2^n}$. ■

Other than for the CDF, a nice characterization for zero symmetry (and i -axial symmetry) exists if X allows for a density function. This result builds on a following change-of-variable formula, which we derive next. Let B be a nonempty subset of \mathbb{R}^n , and denote by $-B \equiv \{-b : b \in B\}$ the negative of B .⁶

Lemma 2 *Let μ be a measure on $(\mathbb{R}^n, \mathcal{B}^n)$ with $\mu(B) = \mu(-B)$ on \mathcal{B}^n . For any measurable $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$*

$$\int_{-B} f(x) d\mu(x) = \int_B f(-x) d\mu(x). \quad (2)$$

Proposition 9 *Let $X = (X_1, \dots, X_n)$ be absolutely continuous with respect to some measure μ that satisfies $\mu(B) = \mu(-B)$ on \mathcal{B}^n . Then X is zero symmetric if and only if X has a density $f_X(x)$ satisfying $f_X(x) = f_X(-x)$ almost everywhere.*

The probably most important application of Proposition 9 is if X is discrete or has a density with respect to Lebesgue measure:

Corollary 3 *If $X = (X_1, \dots, X_n)$ is discrete or absolutely continuous with respect to Lebesgue measure, then $X \stackrel{d}{=} -X$ if and only if the respective density of X verifies $f_X(x) = f_X(-x)$ almost everywhere.*

Proof If X is discrete (continuous), then f_X is a density wrt Counting (Lebesgue) measure. The Counting measure γ obviously verifies $\gamma(B) = \gamma(-B)$. As for the Lebesgue measure λ , it is well known that for any linear transformation $\lambda(T(B)) = |\det(T)|\lambda(B)$. Consider the special linear transformation $T(b) = -Ib$, where $b \in \mathbb{R}^n$ and I is the $n \times n$ identity matrix. Then $T(B) = -B$ and thus $\lambda(-B) = \lambda(T(B)) = \lambda(B)$. The claim then follow from Proposition 9. ■

The statements of Proposition 9 and Corollary 3 apply, analogously, for the case of an i -axially symmetrically distributed random vector. We prove the analogue to Proposition 9 for i -axially symmetrical random vectors in Appendix A.1, and restrict attention here to the analogue of Corollary 3.

Corollary 4 *If $X = (X_1, \dots, X_n)$ is discrete or absolutely continuous with respect to Lebesgue measure, then X is i -axially symmetrically distributed if and only if the density of X is i -axially symmetric almost everywhere.*

⁶Note that $B \in \mathcal{B}^n \Leftrightarrow -B \in \mathcal{B}^n$.

We can use Corollary 4 to demonstrate that zero symmetry of X does not imply that X is axially symmetrically distributed. Consider $X = (X_1, X_2)$ with (Lebesgue-) density $f = \frac{3}{8}(x_1 + x_2)^2$ on $\text{supp}(X) = [-1, 1] \times [-1, 1]$. As $f(x) = f(-x)$ everywhere, X must be zero symmetric. By contrast, it is easy to verify that $f(x_1, x_2) \neq f(-x_1, x_2)$ on the positive measure set $B = (-1, 0) \times (-1, 0)$. By 4, this implies that (X_1, X_2) cannot be 1-axially symmetric. We end this section by remarks on the expectation and covariance, respectively, of an i -axially symmetrically distributed or a zero symmetric random vector.

Proposition 10 *Let X be an i -axially symmetrically distributed random vector. Then $E[X_i] = 0$, $\text{Var}(X_i) = E[X_i^2]$ and $\text{Cov}(X_i, X_k) = 0$ for any $k \neq i$, whenever the respective expressions exist.*

Proof The first two claims immediately follow from $E[X_i] = E[-X_i]$ and from $\text{Var}(X_i) = E[X_i^2] - E[X_i]^2$. For the last claim recall that $\text{Cov}(X_i, X_k) = E[X_i X_k] - E[X_i]E[X_k]$. Consider the continuous function with $f_k(x_1, \dots, x_n) = (x_1, x_k)$ for some fixed $k \neq i$. Then, as $f(X_1, \dots, X_i, \dots, X_n) \stackrel{d}{=} f(X_1, \dots, -X_i, \dots, X_n)$, we obtain $(X_i, X_k) \stackrel{d}{=} (-X_i, X_k)$. Therefore $E[X_i X_k] = E[-X_i X_k] = -E[X_i X_k]$ which implies $E[X_i X_k] = 0$, and hence $\text{Cov}(X_i, X_k) = 0$. ■

Thus, whenever X is an i -axially symmetrically distributed, then X_i and X_k , $k \neq i$, must necessarily be uncorrelated RV's. Moreover, if X is axially symmetrically distributed, then all RV's are uncorrelated, meaning that the variance-covariance matrix of X is a zero matrix except possibly for the diagonal, which consists of the variances of the various X_i 's in X . Further, $E[X] = 0$ also follows if X is zero symmetric (provided that the expectation exists). Note, however, that zero symmetry by itself does *not* imply that the RV's forming X must be uncorrelated. In fact, our previous example with $X = (X_1, X_2)$ and density $f = \frac{3}{8}(x_1 + x_2)^2$ is zero symmetric but has $\text{Cov}(X_1, X_2) = 1/3 > 0$.

3.2 Marginal and Conditional Distributions of Symmetric Random Vectors

We next present some results about the marginal and conditional distributions of symmetric random vectors.

Proposition 11 (Marginals) *If X is ij -exchangeable, then $X_i \stackrel{d}{=} X_j$. If X is i -axially symmetrically distributed, then X_i is a zero symmetric RV. Finally, if X is zero symmetric, then any sub-collection of RV's in X is zero symmetric as well.*

Note that it is a direct consequence of the first part of Proposition 11 and Corollary 2 that any sub-collection of an exchangeable random vector must again be exchangeable. Likewise, if X is i -axially symmetrically distributed, then $(X_i, X_j) \stackrel{d}{=} (-X_i, X_j)$ for any $j \neq i$.⁷

⁷This holds for longer marginals, such as (X_i, X_j, X_m) etc. as well.

Proposition 12 (Conditionals) *If X is ij -exchangeable and $P(X_j \in B) > 0$, then $P(X_i \in A|X_j \in B) = P(X_j \in A|X_i \in B)$ and also $P(X_i \in A|X_k \in B) = P(X_j \in A|X_k \in B)$ if $P(X_k \in B) > 0$ and $k \neq i, j$. Likewise, if X is i -axially symmetric, then $P(X_i \in A|X_j \in B) = P(-X_i \in A|X_j \in B)$ for any $j \neq i$.*

It follows from (the proof of) Proposition 12 and the previous arguments that if X is exchangeable, then any conditional distribution of any sub-collection of the RV's in X must be exchangeable, provided that the conditioning RV's have a positive probability.⁸ Likewise, if X is axially symmetric, hence also zero symmetric, then the conditional distributions obtained by any collection of RV's in X must again be zero symmetric. This does not hold, however, if X is zero symmetric but *not* axially symmetric (see below). We first show that the analogue of Proposition 12 holds for conditional densities.

Proposition 13 *Suppose that $X = (X_1, \dots, X_n)$ is discrete or absolutely continuous wrt Lebesgue measure, such that X has a density $f(x_1, \dots, x_n)$ (wrt to Counting or Lebesgue measure, respectively). If X is ij -exchangeable, then $f_{X_i|X_j}(a|b) = f_{X_j|X_i}(a|b)$ almost everywhere whenever $f_{X_j}(b) > 0$. If X is i -axially symmetrically distributed, then $f_{X_i|X_j}(a|b) = f_{X_i|X_j}(-a|b)$ almost everywhere whenever $f_{X_j}(b) > 0$.*

The above statements hold analogously if the conditioning is for several random variables. Note however, that the statement does *not* hold, in general, if X is zero symmetric. As an example, consider $X = (X_1, X_2)$ with joint density $f(x_1, x_2) = \frac{3}{8}(x_1 + x_2)^2$ on $[-1, 1] \times [-1, 1]$. This random vector is zero symmetric (Corollary 9) but not 1-axially symmetrically distributed, and $f_{X_1|X_2}(x_1|x_2) = \frac{(x_1+x_2)^2}{f_{X_2}(x_2)} \neq \frac{(-x_1+x_2)^2}{f_{X_2}(x_2)} = f_{X_1|X_2}(-x_1|x_2)$ in general,⁹ showing that the RV $X_1|X_2 = x_2$ fails to be zero symmetric if $x_2 < 0$.

3.3 Symmetry-preserving Transformations

The last part of our analysis is interested in transformations that preserve the zero symmetry of a random vector, as is inspired by our example from discrete choice theory. We first note that zero symmetry of a random vector X is not preserved, in general, under axially symmetric transformations. As an example, consider a zero symmetric RV X and the axially symmetric transformation $f(x) = x^2$. Then, the RV $Y = f(X) = X^2$ fails to be zero symmetric (unless X is degenerate).¹⁰ Our main result is that zero symmetric is always preserved under *center symmetric transformations*, which is helpful in that many specific transformations verify this criterion.

⁸For example, if $X = (X_1, X_2, X_3, X_4)$ is exchangeable, then $P((X_1, X_2) \in A|(X_3, X_4) \in B) = P((X_3, X_4) \in A|(X_1, X_2) \in B)$ provided that $P((X_3, X_4) \in B) > 0$.

⁹For example, $(x_1 + x_2)^2 \neq (-x_1 + x_2)^2$ on the entire set $[-1, 0) \times [-1, 0)$, such that for $x_2 \in (-1, 0)$ we get $f_{X_1|X_2}(x_1|x_2) \neq f_{X_1|X_2}(-x_1|x_2)$ for every $x_1 \in [-1, 0)$.

¹⁰To see this, recall that for zero symmetry we require that $F_Z(z) + F_Z(-z) = 1$ everywhere. If X is not degenerate, take any $z_0 > 0$ such that $F_Z(z_0) \in (0, 1)$. But as $F_Z(z) = 0$ for any $z < 0$ we cannot have that $F_Z(z_0) + F_Z(-z_0) = 1$.

Proposition 14 *If $X = (X_1, \dots, X_n)$ is a zero symmetric random vector and $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ a measurable and center symmetric function, then $f(X)$ is a zero symmetric random vector.*

Proof: $f(X) \stackrel{d}{=} f(-X)$ because $X \stackrel{d}{=} -X$ and f is measurable. The claim follows because $f(-X) = -f(X)$ for any center symmetric function f . ■

Theoretical and applied statistics have been frequently interested in the distributions of the sum and the product of two or more RV's. In context of zero symmetric RV's, Hamedani and Walter (1985) show that in case of $n = 2$ independent RV's the zero symmetry of at least one RV is necessary and sufficient for the zero symmetry of the product X_1X_2 , or Rubin and Sellke (1986) establish that any RV who has a zero mean (or no mean at all) can always be written as the sum of two zero symmetric RV's. The following result adds to this literature by exploiting Proposition 14 in that sums and products of RV's from a random vector X are specific examples of center symmetric transformations of X .

Corollary 5 *Let $X = (X_1, \dots, X_n)$ be a zero symmetric random vector. Then $Y = \sum_{i=1}^n \alpha_i X_i$, $\alpha_i \in \mathbb{R} \forall i = 1, \dots, n$, is zero symmetric. Further, any product of the form $Y = X_1 \cdot X_2 \dots \cdot X_k$ consisting of an odd number of factors is zero symmetric.*

Proof: Follows from Proposition 14 as the function $f(x_1, \dots, x_n) \equiv \sum_{i=1}^n \alpha_i x_i$ is center symmetric, and so are products of the type $f(x_1, \dots, x_n) \equiv x_1 \cdot \dots \cdot x_k$ with an odd number of factors. ■

We remark that there is no analogue to Proposition 14, in general, if X is only i -axially symmetrically distributed (instead of zero symmetric). For example, the fact that every RV in X is zero symmetric does, by itself, not allow us to conclude that the sum of these RV's must also be zero symmetric. In fact, this result explains an example provided by Chen and Shepp (1983), who show that the sum of two zero symmetric RV's with Cauchy distribution fails to be zero symmetric.

We remark that the relevant condition $f(X) = -f(X)$ in the proof of Proposition 14 may be satisfied, even though X is not zero symmetric. For example, if $(X_1, X_2) \stackrel{d}{=} (-X_1, X_2)$, then the transformation $f(x_1, x_2) = x_1x_2$ implies that $f(-x_1, x_2) = -x_1x_2 = -f(x_1, x_2)$. This shows that $Y = X_1X_2$ is zero symmetric despite that (X_1, X_2) is not. This argument can be generalized to see that if $X = (X_1, \dots, X_n)$ is zero symmetric *because* X also is axially symmetrically distributed, then *any* product of the RV's in X must again be a zero symmetric RV. In particular, if X_1, \dots, X_n are independent random variables, and each X_i itself is zero symmetric, then $Y = X_1 \cdot \dots \cdot X_n$ is also zero symmetric.

While elementary transformations, such as weighted sums, differences or odd products are center symmetric, this obviously will not hold for every transformation of interest. To mention one example, the *order statistics*

pertaining to a zero symmetric $X = (X_1, \dots, X_n)$ will generally fail to be zero symmetrically distributed.¹¹

Application to Discrete Choice We now resume the discrete choice model outlined in the introduction, generalized to the case of $n \geq 2$ products. The random vector $X = (X_1, \dots, X_n)$ governs how tastes are dispersed and, accordingly, the $(n-1)$ -dimensional random vector $Y_j \equiv (X_1 - X_j, \dots, X_n - X_j)$ captures the distribution of taste differences from the perspective of product j . Recalling that consumer valuation for product $j = 1, \dots, n$ is $U_j = q_j - p_j + X_j$, the choice probability system is determined by $P_j = P(Y_j \leq w_j)$, where $j = 1, \dots, n$ and w_j is the $(n-1)$ vector with i -th entry $w_{ij} = (q_j - q_i) + (p_i - p_j)$ for $i \neq j$. Thus, as with $n = 2$, the difference distribution Y_j captures everything about the taste distribution that is relevant for the choice probability system. However, Y_j is a random vector for $n > 2$, and as $Y_j \neq Y_k$ for $j \neq k$ is not excluded, these distributions may look quite differently for different products. Nevertheless, we now show that zero symmetry is all that is needed to discipline the possible correlations in the RV's of X in that every difference distribution Y_j again must be zero symmetric and, accordingly, the choice probabilities remain unbiased in this sense. Formally, the result is a direct consequence of Proposition 14 by noting that Y_j can be obtained from X via the continuous transformation $f_j : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$, $f_j(x_1, \dots, x_n) = (x_1 - x_j, \dots, x_n - x_j)$.

Corollary 6 (Unbiased Choice Probability System) *Let $X = (X_1, \dots, X_n)$ be zero symmetric. Then, the $n-1$ -dimensional random vector $Y_i = (X_1 - X_i, X_2 - X_i, \dots, X_n - X_i)$, $i = 1, \dots, n$, of differences is zero symmetric as well.*

Proof For $i = 1, \dots, n$, the Borel function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$, $f_i(x_1, \dots, x_n) = (x_1 - x_i, x_2 - x_i, \dots, x_n - x_i)$ is center symmetric and $Y_i = f_i(X)$. The claim therefore follows from Proposition 14.

4 Conclusion

This paper studied random vectors with at least partially symmetric distributions, either in the sense that the random variables can be at least partially reordered without affecting the joint distribution, or in the sense of being symmetrically dispersed around some coordinate axis, or around the origin. In the former case, we established that exchangeability of the RV's in essence is a binary property: The RV's in a random vector $X = (X_1, \dots, X_n)$ can be reordered in any arbitrary way without changing its distribution if and only if this holds for any two RV's that are exchanged with each other. This property can be viewed as a generalization

¹¹Let $X = (X_1, X_2)$ and consider $Y = \max\{X_1, X_2\}$. Note that $f(x_1, x_2) \equiv \max\{x_1, x_2\}$ fails to be center symmetric. If X_1, X_2 are iid with center symmetric density $f(x)$, then $Y = \max\{X_1, X_2\}$ has a density $g(z) = 2F(z)f(z)$. Because $g(z) = g(-z) \Leftrightarrow F(z) = 1/2$, the density $g(z)$ violates the condition $g(z) = g(-z)$ almost everywhere, for which reason Y cannot be zero symmetric.

of the “irrelevance of sampling order” to the case of dependent random variables. Things are slightly more complicated for axially symmetric random vectors, in that axial symmetry is sufficient but not necessary for X to have a zero symmetric distribution: If a random vector X is axially symmetrically distributed for every coordinate $i = 1, \dots, n$, then X must also be zero symmetric, while the converse is false in general. This is also vindicated by the fact that the RV’s of a zero symmetric random vector which is *not* axially symmetric may be correlated with each other, and may have conditional distributions that fail to be themselves zero symmetric. Yet, in our application to discrete choice, we show that the zero symmetry of the idiosyncratic taste distribution is enough to assure that consumer demand cannot be biased towards particular products, despite that tastes could be correlated with each other in the consumer population.

References

- Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT press.
- Chen, R. and Shepp, L. A. (1983). On the sum of symmetric random variables. *The American Statistician*, 37(3):237–237.
- Hamedani, G. and Walter, G. (1985). On the product of symmetric random variables. *Statistics & probability letters*, 3(5):251–253.
- Hefti, A., Liu, S., and Schmutzler, A. (2022). Preferences, confusion and competition. *Economic Journal*.
- Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229.
- Rubin, H. and Sellke, T. (1986). On the distributions of sums of symmetric random variables and vectors. *The Annals of Probability*, 14:247–259.

A Appendix

A.1 Proofs

Proof Proposition 1 We only prove the “if”-part. Consider the function $h(x) = (h^1(x), \dots, h^m(x))$, and suppose that $Dh(x) = Dh(-x)$ on M . Recall that $Dh(x)$ can be represented by an $m \times n$ (Jacobian) matrix with all partial derivative functions $h_j^i(x) \equiv \frac{\partial h^i(x)}{\partial x_j}$, $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$ as its entries. Note that each of these partial derivative functions also verifies $h_j^i(-x) = h_j^i(x)$. The anti-derivative of $h_j^i(x)$ with respect to x_j is $h^i(x) + \varphi_j^i(x_{-j})$ meaning that $h^i(x)$ is determined up to a function $\varphi_j^i(x_{-j})$ that generally could depend on all covariates except for x_j .¹² Because we obtain the same anti-derivative for any $j = 1, \dots, n$, we have that $\varphi_1^i(x_{-1}) = \dots = \varphi_n^i(x_{-n})$ for every $x \in M$, which implies that $\varphi_j^i(x_{-j}) = k_0^i$ (a constant) for every $j = 1, \dots, n$. Repeating the above argument for the function $h_j^i(-x)$ gives $-h^i(-x) + k_1^i$ as its anti-derivative. Because $h_j^i(-x) = h_j^i(x)$ on M , the anti-derivatives of these two functions also coincide, from which we obtain that $h^i(-x) = -h^i(x) + c^i$, $c^i \equiv k_1^i - k_0^i$. Proceeding in this way for each $i = 1, \dots, n$ shows that

$$\begin{pmatrix} h^1(-x) \\ \vdots \\ h^m(-x) \end{pmatrix} = - \begin{pmatrix} h^1(x) \\ \vdots \\ h^m(x) \end{pmatrix} + \begin{pmatrix} c^1 \\ \vdots \\ c^n \end{pmatrix},$$

and hence $h(-x) = -h(x) + c$. ■

Proof Proposition 2 We show the claim for the case where $i = 1$ and $j = 2$, which is plainly due to notational convenience. It shall be evident that the arguments in the following proof are valid for any choice of $i, j \in \{1, \dots, n\}$. Accordingly, let $Y = (X_2, X_1, X_3, \dots, X_n)$.

1) \Leftrightarrow 2): Note that $F_Y(x_1, \dots, x_n) = P(X_2 \leq x_1, X_1 \leq x_2, \dots, X_n \leq x_n) = F_X(x_2, x_1, \dots, x_n)$ on \mathbb{R}^n . If X is 12-exchangeable, then $X \stackrel{d}{=} Y$, and hence $F_X(x) = F_Y(x)$, which shows that $F_X(x_1, x_2, \dots, x_n) = F_X(x_2, x_1, \dots, x_n)$ by the previous equation. If, conversely, $F_X(x_1, x_2, \dots, x_n) = F_X(x_2, x_1, \dots, x_n)$ on \mathbb{R}^n , then also $F_X(x) = F_Y(x) \forall x$, which implies that $X \stackrel{d}{=} Y$.

1) \Leftrightarrow 3): Note that for all $(t_1, \dots, t_n) \in \mathbb{R}^n$

$$\varphi_Y(t_1, t_2, \dots, t_n) = E \left[e^{it_1 X_2} e^{it_2 X_1} \dots e^{it_n X_n} \right] = \varphi_X(t_2, t_1, \dots, t_n). \quad (3)$$

If $X \stackrel{d}{=} Y$, then also $\varphi_X(t_1, t_2, \dots, t_n) = \varphi_Y(t_1, t_2, \dots, t_n)$, and (3) implies that $\varphi_X(t_1, t_2, \dots, t_n) = \varphi_X(t_2, t_1, \dots, t_n)$. Conversely, if $\varphi_X(t_1, t_2, \dots, t_n) = \varphi_X(t_2, t_1, \dots, t_n)$, then also $\varphi_X(t_1, t_2, \dots, t_n) = \varphi_Y(t_1, t_2, \dots, t_n)$ and thus $X \stackrel{d}{=} Y$.

1) \Leftrightarrow 4): Suppose that the MGF m_X exists. Similar to (3), we then have

$$m_X(t_2, t_1, \dots, t_n) = E \left[e^{t_1 X_2} e^{t_2 X_1} \dots e^{t_n X_n} \right] = m_Y(t_1, t_2, \dots, t_n),$$

¹²The anti-derivative can be found by integrating with respect to x_j holding all other covariates fixed, which is well-defined because M is open and convex, and $h_j^i(x)$ is continuous in x_j .

showing that m_Y must also be well-defined. Proceeding as in the proof that 1) \Leftrightarrow 3) shows the equivalence of 1) and 4). ■

Proof Corollary 1 As before, we only consider the case where $(i, j) = (1, 2)$, and let $Y = (X_2, X_1, \dots, X_n)$. We first consider the continuous case. Suppose that f_X is a (Lebesgue-) density for X . Because $f_X(x_1, x_2, \dots, x_n) = f_X(x_2, x_1, \dots, x_n)$ almost everywhere (wrt to Lebesgue measure) by presumption, we have for any $x = (x_1, \dots, x_n)$

$$\begin{aligned} F_X(x_2, x_1, \dots, x_n) &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_n = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(s_2, s_1, \dots, s_n) ds_1 ds_2 \dots ds_n \\ &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(s_2, s_1, \dots, s_n) ds_2 ds_1 \dots ds_n = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n \\ &= F_X(x_1, x_2, \dots, x_n) \end{aligned}$$

Thus, $F_X(x_1, \dots, x_n)$ is 12-exchangeable, which implies that $X \stackrel{d}{=} Y$ by Proposition 2. Moreover, because $F_Y(x_1, x_2, \dots, x_n) = F_X(x_2, x_1, \dots, x_n)$, the previous argument also shows that $f_X(x_1, x_2, \dots, x_n)$ also is a density for Y . Conversely, let $X \stackrel{d}{=} Y$. Note first that

$$\begin{aligned} F_X(x_1, x_2, \dots, x_n) &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_n, \\ F_X(x_2, x_1, \dots, x_n) &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_n \end{aligned}$$

But because by Proposition 2 the equation $F_X(x_1, x_2, \dots, x_n) = F_X(x_2, x_1, \dots, x_n)$ holds for any (x_1, \dots, x_n) , we obtain

$$f_X(x_1, x_2, \dots, x_n) = \frac{\partial F_X(x_1, x_2, \dots, x_n)}{\partial x_1 \dots \partial x_n} = \frac{\partial F_X(x_2, x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} = f_X(x_2, x_1, \dots, x_n).$$

This shows that f_X is a density that is 12-exchangeable, and the claim follows as any other density for X differs from f_X only on zero measure sets. We now turn to the case where X is a discrete random vector. Then X has a density function (wrt Counting measure) $f_X(x_1, \dots, x_n) \equiv dF_X(\{(x_1, \dots, x_n)\}) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ which is strictly positive on $\text{supp}(X)$. If $X \stackrel{d}{=} Y$, then X and Y have the same support, and

$$\begin{aligned} f_X(x_1, x_2, \dots, x_n) &= dF_X(\{(x_1, x_2, \dots, x_n)\}) = dF_Y(\{(x_1, x_2, \dots, x_n)\}) \\ &= P(X_2 = x_1, X_1 = x_2, \dots, X_n = x_n) = f_X(x_2, x_1, \dots, x_n), \end{aligned}$$

showing that $f_X(x)$ must be 12-exchangeable on the support of X . Conversely, if $f_X(x_1, x_2, \dots, x_n) = f_X(x_2, x_1, \dots, x_n)$ on $\text{supp}(X)$ then

$$F_X(x_1, x_2, \dots, x_n) = \sum_{\substack{s \in \text{supp}(X) \\ s \leq x}} f_X(s_1, s_2, \dots, s_n) = \sum_{\substack{s \in \text{supp}(X) \\ s \leq x}} f_X(s_2, s_1, \dots, s_n) = F_X(x_2, x_1, \dots, x_n)$$

must hold for any (x_1, \dots, x_n) , and thus $X \stackrel{d}{=} Y$. ■

Proof Lemma 1 Note first that the trivial binary permutation σ_{ii} simply means that no entry of $(1, \dots, n)$ is permuted. Let $\sigma(1, \dots, n) \equiv (\sigma(1), \dots, \sigma(n))$ be an arbitrary permutation of $(1, \dots, n)$. Then $\sigma_{1, \sigma(1)}(1, \dots, n)$ is the binary permutation of $(1, \dots, n)$ that replaces 1 with $\sigma(1)$. Applying the binary permutation $\sigma_{2, \sigma(2)}$ to this permutation replaces its second entry with $\sigma(2)$, holding the first entry (i.e., $\sigma(1)$) constant. More generally,

for $i = 1, \dots, n-1$, $\sigma_{i\sigma(i)}$ replaces the i -th entry of $\sigma_{(i-1)\sigma(i-1)}$ with $\sigma(i)$, holding $(\sigma(1), \dots, \sigma(i-1))$ constant. Continuing this finite process up to $\sigma_{(n-1)\sigma(n-1)}$ assures that the so obtained list is equally arranged as $\sigma(1, \dots, n)$ up to the first $n-1$ entries, which implies that this must hold for the residual n -th entry as well. ■

Proof Corollary 2 It is evident that an exchangeable random vector must also be ij -exchangeable. For the converse, if X is ij -exchangeable for any pair (i, j) , then so is $F_X(x_1, \dots, x_n)$ with respect to any two coordinates x_i and x_j by Proposition 2 for any given (x_1, \dots, x_n) . Now, fix the values (x_1, \dots, x_n) and let $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ be an arbitrary permutation of (x_1, \dots, x_n) . By Lemma 1 we can obtain this permutation by conducting at most $n-1$ binary permutations. But as $F_X(x_1, \dots, x_n)$ is ij -exchangeable with respect to any two coordinates x_i and x_j (and any values) these binary permutations cannot alter the value of $F(x_1, \dots, x_n)$, and thus $F_X(x_1, \dots, x_n) = F_X(x_{\sigma(1)}, \dots, x_{\sigma(n)})$, which yields $X \stackrel{d}{=} \sigma(X)$ by Proposition 2. ■

Proof Proposition 4 We prove the claim for $i = 1$, which keeps notation simple. It shall be evident that each of the arguments below applies for any $i = 1, \dots, n$. Let $Y = (-X_1, X_2, \dots, X_n)$. Note that $\varphi_X(t_1, \dots, t_n) = E[e^{it_1 X_1} \dots \cdot e^{it_n X_n}]$ and $\varphi_Y(t_1, \dots, t_n) = E[e^{it_1(-X_1)} \dots \cdot e^{it_n X_n}]$.

1) \Leftrightarrow 2): If $X \stackrel{d}{=} Y$ then $\varphi_X(t_1, \dots, t_n) = \varphi_Y(t_1, \dots, t_n) = E[e^{i(-t_1)X_1} \dots \cdot e^{it_n X_n}] = \varphi_X(-t_1, \dots, t_n)$. For the converse, note that $\varphi_X(t_1, \dots, t_n) = \varphi_X(-t_1, \dots, t_n) = E[e^{it_1(-X_1)} \dots \cdot e^{it_n X_n}] = \varphi_Y(t_1, \dots, t_n)$, which implies $X \stackrel{d}{=} Y$.

1) \Leftrightarrow 3): Follows from the previous proof as $m_X(t_1, \dots, t_n) = E[e^{t_1 X_1} \dots \cdot e^{t_n X_n}]$ and $m_Y(t_1, \dots, t_n) = E[e^{t_1(-X_1)} \dots \cdot e^{t_n X_n}]$. ■

Proof Proposition 6 By presumption and Proposition 4 we have for every $i = 1, \dots, n$ an equation

$$\varphi_X(t_1, \dots, t_n) = \varphi_X(t_1, \dots, -t_i, \dots, t_n) \quad (e_i),$$

is valid for every $(t_1, \dots, t_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} \varphi_X(t_1, \dots, t_n) &\stackrel{\text{by } (e_1)}{=} \varphi_X(-t_1, t_2, \dots, t_n) \\ &\stackrel{\text{by } (e_2)}{=} \varphi_X(-t_1, -t_2, \dots, t_n) \\ &\vdots \\ &\stackrel{\text{by } (e_{n-1})}{=} \varphi_X(-t_1, -t_2, \dots, -t_{n-1}, t_n) \\ &\stackrel{\text{by } (e_n)}{=} \varphi_X(-t_1, -t_2, \dots, -t_{n-1}, -t_n) \end{aligned},$$

showing that X is zero symmetric by Proposition 5. It is straightforward to modify the above argument to see that, e.g., $\varphi_X(t_1, \dots, t_n) = \varphi_X(-t_1, t_2, -t_3, \dots, t_n)$, which implies the second claim. ■

Proof Lemma 2 The proof is organized in two steps.

Step 1: For every $A, B \in \mathcal{B}^n$: $\mu(-A \cap B) = \mu(A \cap -B)$. Wlog, we can assume that $A, B \neq \emptyset$. Note that $x \in (-A \cap B) \Leftrightarrow -x \in (A \cap -B)$. Thus if $Z \equiv (-A \cap B)$, then $-Z = (A \cap -B)$. The claim now follows as $\mu(Z) = \mu(-Z)$ by presumption.

Step 2: Let $B \in \mathcal{B}^n$ be nonempty. We first show (2) if f is a simple function. For $A \in \mathcal{B}^n$ let $\mathbb{1}_A(x)$ denote the indicator function of subset A . A simple function is a measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ which takes on only

finitely many different values $\{a_1, \dots, a_K\}$. It is well known that any simple function $f(x)$ can be represented as a finite sum of “disjoint” indicator functions, such that $f(x) = \sum_{k=1}^K a_k \mathbb{I}_{A_k}(x)$, where $A_k = f^{-1}(\{a_k\}) \in \mathcal{B}^n$. The μ -integral for simple functions then is defined as $\int f d\mu = \sum_{k=1}^K a_k \mu(A_k)$. Then

$$\int_{-B} f(x) d\mu(x) = \int_{\mathbb{R}^n} \mathbb{I}_{-B}(x) f(x) d\mu(x) = \sum_{k=1}^K a_k \mu(A_k \cap -B) = \sum_{k=1}^K a_k \mu(-A_k \cap B), \quad (4)$$

where the last equality follows from Step 1. Next, note that always $\mathbb{I}_B(-x) = \mathbb{I}_{-B}(x)$. In particular, this means that $\mathbb{I}_{A_k}(-x) = \mathbb{I}_{-A_k}(x)$. Thus $f(-x) = \sum_{k=1}^K a_k \mathbb{I}_{A_k}(-x) = \sum_{k=1}^K a_k \mathbb{I}_{-A_k}(x)$ and hence

$$\int_B f(-x) d\mu(x) = \int_{\mathbb{R}^n} \mathbb{I}_B(x) f(-x) d\mu(x) = \sum_{k=1}^K a_k \mu(-A_k \cap B). \quad (5)$$

Comparing (4) and (5) shows that $\int_{-B} f(x) d\mu(x) = \int_B f(-x) d\mu(x)$.

We now show the claim if f is an arbitrary measurable and non-negative function. Then, there exists a sequence $(g_n(x))_n$ of simple functions that converge pointwise from below to $f(x)$, and the μ -integral of f is defined as $\int f d\mu = \lim_{n \rightarrow \infty} \int g_n d\mu$. Note that, for any given $B \in \mathcal{B}^n$, $\mathbb{I}_B f$ then also is non-negative and measurable. Moreover, the sequence $(\mathbb{I}_B(x) g_n(x))_n$ converges pointwise from below to $\mathbb{I}_B f$. Thus

$$\begin{aligned} \int_{-B} f(x) d\mu(x) &= \int_{\mathbb{R}^n} \mathbb{I}_{-B}(x) f(x) d\mu(x) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^n} \mathbb{I}_{-B}(x) g_n(x) d\mu \\ \int_B f(-x) d\mu(x) &= \int_{\mathbb{R}^n} \mathbb{I}_B(x) f(-x) d\mu(x) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^n} \mathbb{I}_B(x) g_n(-x) d\mu \end{aligned}$$

By the previous step, the two integrals on the right must coincide for every n , which implies that $\int_{-B} f(x) d\mu(x) = \int_B f(-x) d\mu(x)$. \blacksquare

Proof Proposition 9 By the Radon-Nikodym Theorem $dF_X(B) = \int_B f(x) d\mu(x)$, $B \in \mathcal{B}^n$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is a density function. Moreover, we have $dF_{-X}(B) = P(-X \in B) = P(X \in -B) = \int_{-B} f(x) d\mu(x)$. But, as $\mu(B) = \mu(-B)$, Lemma 2 assures that for every $B \in \mathcal{B}^n$: $dF_{-X}(B) = \int_{-B} f(x) d\mu(x) = \int_B f(-x) d\mu(x)$. Thus, $g(x) \equiv f(-x)$ is a density for $-X$. Now, $X \stackrel{d}{=} -X$ implies that $dF_{-X}(B) = dF_X(B)$ on \mathcal{B}^n . As also $dF_{-X}(B) = \int_B f(-x) d\mu(x)$ and $dF_X(B) = \int_B f(x) d\mu(x)$, we must have $\int_B f(-x) d\mu(x) = \int_B f(x) d\mu(x)$ for any $B \in \mathcal{B}^n$, which assures that $f(x) = f(-x)$ μ -almost everywhere. Conversely, let the density of X verify $f(x) = f(-x)$ μ -almost everywhere. Thus using Lemma 2: $dF_{-X}(B) = P(X \in -B) = \int_{-B} f(x) d\mu(x) = \int_B f(-x) d\mu(x) = \int_B f(x) d\mu(x) = dF_X(B)$, and hence $X \stackrel{d}{=} -X$. \blacksquare

Proof Corollary 4 For any nonempty set $B \in \mathcal{B}^n$ define $B_{-i} \equiv \{(b_1, \dots, -b_i, \dots, b_n) : (b_1, \dots, b_n) \in B\}$ as the sets of all elements in B where the sign of the i -th coordinate is flipped. We confine attention to the case where X is 1-axially symmetrically distributed; all arguments equally apply if X is i -axially symmetrically distributed for any $i \in \{1, \dots, n\}$.

Let $X = (X_1, \dots, X_n)$ be absolutely continuous with respect to some measure μ that satisfies $\mu(B) = \mu(B_{-1})$ for every $B \in \mathcal{B}^n$. We claim that then X has a density $f_X(x)$, and X is 1-symmetrically distributed if and only if X has a density $f_X(x)$ that satisfies $f_X(x_1, x_2, \dots, x_n) = f_X(-x_1, x_2, \dots, x_n)$ almost everywhere. This is analogous to Proposition 9. We first note that $\mu(A \cap B_{-1}) = \mu(A_{-1} \cap B)$ for any nonempty $A, B \in \mathcal{B}^n$. This

follows by exchanging $-B$ with B_{-1} , $-x$ with $(-x_1, x_2, \dots, x_n)$ and Z with Z_{-1} in the proof of Step 1 in Lemma 2. The next step is to show that $\int_{B_{-1}} f(x_1, x_2, \dots, x_n) d\mu(x) = \int_B f(-x_1, x_2, \dots, x_n)$ if f is a measurable and non-negative function. Again, this result is obtained by completely mimicking the proof of Step 2 in Lemma 2. The proof is completed by repeating all the remaining arguments in the proof of Proposition 9.

As for the claim in Corollary 4, we can proceed as in the proof of Corollary 3. The Counting measure γ obviously verifies $\gamma(B_{-1}) = \gamma(B)$, and so does the Lebesgue measure λ : If I_{-1} denotes the $n \times n$ identity matrix with first entry $a_{11} = -1$ and $T(b) = I_{-1}b$, then $T(B) = B_{-1}$ and thus $\lambda(B_{-1}) = \lambda(T(B)) = \lambda(B)$. ■

Proof Proposition 11 First, note that if $X = (X_1, \dots, X_n)$, the marginal distribution dF_{X_i} can be derived from transforming X with the function $f_i(X)$, where $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $f_i(x_1, \dots, x_n) = x_i$. To see this, suppose that $i = 1$, noting that f_1 is measurable. Let $B \in \mathcal{B}$. Then $f_1^{-1}(B) = B \times \mathbb{R}^{n-1}$, and hence $P(f_1(X) \in B) = P(X^{-1}f_1^{-1}(B)) = P(X \in B \times \mathbb{R}^{n-1}) = dF_{X_1}(B)$. A similar argument holds for any $i = 2, \dots, n$. Suppose now that $X = (X_1, \dots, X_n)$ is ij -exchangeable. Let Y be the random vector where X_i and X_j are exchanged. Consider the continuous map $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $f_i(x_1, \dots, x_i, \dots, x_n) = x_i$. Because $X \stackrel{d}{=} Y$ we also have $f_i(X) \stackrel{d}{=} f_i(Y)$, which shows that $X_i \stackrel{d}{=} X_j$. Suppose next that X is i -axially symmetrically distributed. Consider the continuous map $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $f_i(x_1, \dots, x_i, \dots, x_n) = -x_i$. Because $(X_1, \dots, X_i, \dots, X_n) \stackrel{d}{=} (X_1, \dots, -X_i, \dots, X_n)$ we also have $f_i(X_1, \dots, X_i, \dots, X_n) \stackrel{d}{=} f_i(X_1, \dots, -X_i, \dots, X_n)$, which shows that $X_i \stackrel{d}{=} -X_i$. Let $X \stackrel{d}{=} -X$, and consider any sub-collection (X_a, \dots, X_m) , $a \leq m \leq n$, of the RV's in X . Then the corresponding continuous map $f(x_1, \dots, x_n) = (x_a, \dots, x_m)$ verifies $(X_a, \dots, X_m) = f(X) \stackrel{d}{=} f(-X) = -(X_a, \dots, X_m)$, showing the last claim. ■

Proof Proposition 12 By Proposition 12, X_i and X_j have the same marginals, i.e., $P(X_j \in B) = P(X_i \in B)$. Likewise, the marginals (X_i, X_j) and (X_j, X_i) must be the same, i.e., $(X_i, X_j) \stackrel{d}{=} (X_j, X_i)$. Therefore $P(X_i \in A, X_j \in B) = P(X_j \in A, X_i \in B)$, and thus

$$P(X_i \in A | X_j \in B) = \frac{P(X_i \in A, X_j \in B)}{P(X_j \in B)} = \frac{P(X_j \in A, X_i \in B)}{P(X_i \in B)} = P(X_j \in A | X_i \in B).$$

The same type of argument also gives $P(X_i \in A | X_k \in B) = P(X_j \in A | X_k \in B)$ under the presumptions of the proposition. If X is i -axially symmetrically distributed, then $X_i \stackrel{d}{=} -X_i$ and also $(X_i, X_j) \stackrel{d}{=} (-X_i, X_j)$ for $j \neq i$. Therefore

$$P(X_i \in A | X_j \in B) = \frac{P(X_i \in A, X_j \in B)}{P(X_j \in B)} = \frac{P(-X_i \in A, X_j \in B)}{P(X_j \in B)} = P(-X_i \in A | X_j \in B),$$

completing the proof. ■

Proof Proposition 13 Let $f_{X_i, X_j}(a, b)$ denote the marginal density of (X_i, X_j) . As $(X_i, X_j) \stackrel{d}{=} (X_j, X_i)$, Corollary implies 1 that this density is exchangeable almost everywhere. Likewise, we have $f_{X_j}(x) = f_{X_i}(x)$ almost everywhere, so take $f_{X_j}(b) > 0$ such that $f_{X_j}(b) = f_{X_i}(b)$ almost surely. By definition of the conditional density we obtain

$$f_{X_i | X_j}(a | b) = \frac{f_{(X_i, X_j)}(a, b)}{f_{X_j}(b)} = \frac{f_{(X_j, X_i)}(a, b)}{f_{X_i}(b)} = f_{X_j | X_i}(a | b)$$

almost everywhere. As for the last claim, $(X_i, X_j) \stackrel{d}{=} (-X_i, X_j)$ and Corollary 4 imply that $f_{X_i, X_j}(a, b) = f_{X_i, X_j}(-a, b)$ almost everywhere. Hence as in the previous step $f_{X_i|X_j}(a|b) = \frac{f_{(X_i, X_j)}(a, b)}{f_{X_j}(b)} = \frac{f_{(X_i, X_j)}(-a, b)}{f_{X_j}(b)} = f_{X_i|X_j}(-a|b)$ almost everywhere. ■