DECOMPOSITION OF UNCERTAINTY IN INVARIANT DISTRIBUTIONS
EXTENDED ABSTRACT

MARCIN PĘSKI

Say that distribution of a sequence of random variables is exchangeable, it is invariant to all permutations of the variables. De Finetti's Theorem says that any exchangeable distribution can be decomposed into an aggregate shock that affects all variables, and independent idiosyncratic shocks that affect each variable individually. A class of related results shows that distributions that are invariant with respect to other, smaller groups of permutations also can be decomposed into independent shocks. In the talk, I will describe a general property of a group of permutations, finite dimensionality, that is responsible for all known and some novel decomposition results.

Consider a joint distribution of random variables $\theta(x) \in \{0,1\}$ indexed with infinitely many elements $x \in X$. Say that the joint distribution ω of $\theta(x)$ is *exchangeable,* if for all bijections $g : X \to X$, all tuples $x_1, ..., x_k X$ and $y_1, ..., y_k, \in Y$,

$$\omega(\theta(x_1) = y_1, ..., \theta(x_k) = y_k) = \omega(\theta(g(x_1)) = y_1, ..., \theta(g(x_k)) = y_k). \tag{0.1}$$

By De Finetti's Theorem, if ω is exchangeable, then there is a measurable function $f : [0,1]^2 \to \{0,1\}$, and a collection of independent random variables u_X and u_x for each $x \in X$, all distributed uniformly on interval $[0,1]$, such that ω is equal to the joint distribution of variables defined as

$$\theta(x) := f(u_X, u_x) \text{ for each } x \in X.$$

Here, u_X is interpreted as aggregate shock that affects variables $\theta(x)$ for all x; u_x is an idiosyncratic shock that affects only variable $\theta(x)$ for a particular x.

Exchangeability has very natural interpretation. If a Bayesian agent does not know anything that allows him to differentiate between elements of space $x \in X$, he has to have exchangeable beliefs. Exchangeability is applicable to many situations. Consider the following example.

Example 1 (One product, many customers). *A company wants to predict whether customers $c \in C = X$ will buy its product. Let $\theta(c) = 1$ if and only if customer c buys the product. If the company does not know of any differences between customers, company's beliefs ω are a*

Date: June 2008.

priori exchangeable. For example, the ex ante probability that customer c buys the product is the same as the probability that c' buys it,

$$\omega\left(\theta\left(c\right)=1\right)=\omega\left(\theta\left(c'\right)=1\right).$$

We say that c and c' are of the same type. Moreover, the conditional probability that c_0 buys product given that $c\neq c_0$ bought it is the same as given that $c'\neq c_0, c$ bought it, $\omega\left(\theta\left(c_0\right)=1|\theta\left(c\right)=1\right)=\omega\left(\theta\left(c_0\right)=1|\theta\left(c_1'\right)=1\right)$. We say that relation between c_0 and c has the same type as relation between c_0 and c'.

De Finetti's Theorem imposes constraints on correlations between exchangeable random variables. In particular, they can be correlated only through an aggregate shock. Because the aggregate shock is constant across all $x\in X$, one can learn it given a sufficient number of observations. On the other hand, idiosyncratic shocks are independent between observations, hence not learnable. Thus, de Finetti's Theorem provides the simplest model of induction, which clearly identifies a priori, learnable information, and not learnable information.

On the other hand, exchangeability is not applicable, when the agent has some prior knowledge that allows him to distinguish between elements of x, and/ or relations between different elements. Such a priori knowledge often comes from logical consequences of the structure of the problem. Consider the following example:

Example 2 (Many products, many customers)**.** *Instead, suppose that company has beliefs about customers buying many products. Let $X=C\times P$, where C is the set of customers, P is the set of products, and X is the set of customer-product pairs. Let $\theta\left(c,p\right)=1$ if and only if customer c buys product p. Suppose that the agent wants to make a prediction whether customer c_0 buys product p_0.*

In the above example, it not natural to assume that all observations lead to the same prediction. For example, $\theta\left(c,p_0\right)=1$ for $c\neq p_0$ indicates that other customers usually like p_0. This is a sign of a good quality of product p_0. If the customers are known to care about the quality, such information raises the probability that customer c_0 will buy the product. On the other hand, $\theta\left(c_0,p\right)=1$ for $p\neq p_0$ means that customer c_0 is not very picky. If all customers have the same known preferences, then information about customer c_0 is a pure noise that does not affect the probability that c_0 will buy product p_0. Thus,

$$\omega\left(\theta\left(c_0,p_0\right)=1|\theta\left(c_0,p\right)=1\right)<\omega\left(\theta\left(c_0,p_0\right)=1|\theta\left(c,p_0\right)=1\right),$$

and ω is not exchangeable. We say that the relation between (c_0,p_0) and (c_0,p) has different type than relation (c_0,p_0) and (c,p_0).

One can modify de Finetti's exchangeability to allow for non-trivial types of relations. For each set A, let $\Pi\left(A\right)$ be the set of all bijections $g_A:A\to A$. Let $G_{C\times P}=\Pi\left(C\right)\times\Pi\left(P\right)$ and for any $g = (g_C, g_P) \in G_{C\times P}$, for any $x = (c, p) \in X$, define

$$g(x) = (g_C(c), g_P(p)).$$

Say that ω is *matrix exchangeable*, if for all $g \in G$, all tuples $x_1, ..., x_k \in X$ and $y_1, ..., y_k, \in Y$, (0.1) holds. In other words, ω is matrix exchangeable if it is invariant under renaming customers and products. (Aldous, 82) shows that if ω is exchangeable, then there is a measurable function $f : [0,1]^4 \to \{0,1\}$, and a collection of independent random variables u_X, u_c for all $c \in C$, u_p for all $p \in P$, and and u_x for each $x \in X$, all distributed uniformly on interval $[0,1]$, such that ω is equal to the joint distribution of variables defined as

$$\theta(c,p) := f\left(u_X, u_c, u_p, u_{(c,p)}\right) \text{ for each } (c,p) \in X.$$

Here, u_X and $u_{(c,p)}$ are aggregate shocks, u_c is a shock to customer c, and u_p is a shock to product p. One can interpret u_c and u_p as hidden fundamentals of a customer and a product respectively; such hidden fundamentals may correspond to tastes, differences in quality, or some other unobservable factors that affect customer's choices. Notice that shock u_c affects infinitely many observations $x \in \{(c, p') : p' \in P\}$ that all refer to the same customer c. In principle, given sufficiently many observations of customer c's choices, such shock is learnable. The same refers to the product-specific shock u_p.

One easily compare de Finetti's Theorem and the above result in a general framework. Let $G \subseteq \Pi(X)$ be a group of bijections on X.[1] Say that distribution ω is invariant to G if all tuples $x_1, ..., x_k \in X$ and $y_1, ..., y_k, \in Y$, (0.1) holds. Then, ω is exchangeable, if it is invariant to the group of all bijections $\Pi(X)$; ω is matrix exchangeable, if it is invariant to $G_{C\times P}$. The two results show that if the joint distribution of variables $\theta(x)$ is invariant with respect to appropriate groups of permutations, then it can be represented as a decomposition into independent shocks.

The purpose of this note is to extend decomposition results to a class of permutation groups. For each tuple $\bar{x} = (x_1, ..., x_n) \in X^n$, define the type of $\bar{x}$ as the set of all tuples that can be obtained as image of $\bar{x}$ under bijections $g \in G$:

$$[\bar{x}] = \{(g(x_1), ..., g(x_n)) : g \in G\} \subseteq X^n.$$

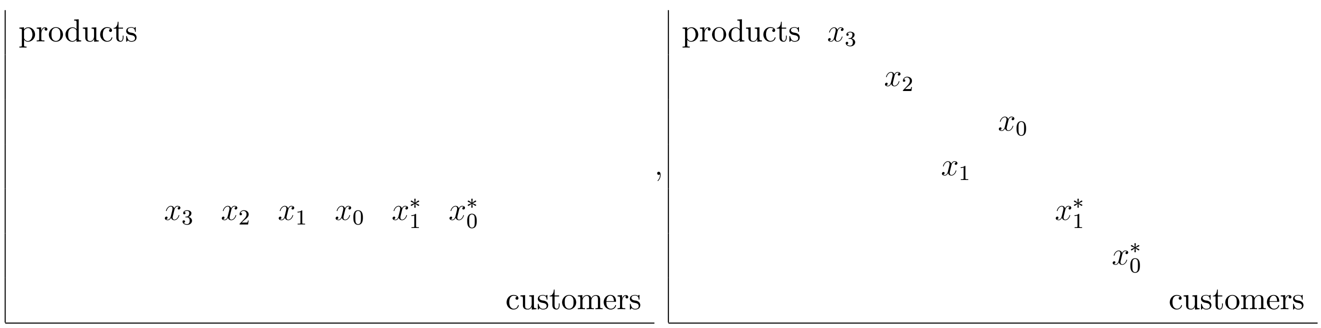
One can interpret a type $[\bar{x}] \subseteq X^n$ as an n-ary relation on X. Say that G has *finitely many types*, if, for each n, the set of different types of tuples $\bar{x} \in X^n$ is finite, $\#\{[\bar{x}] : \bar{x} \in X^n\} < \infty$.

A binary type is a type of a two-element tuple: $[\bar{x}]$ for some $\bar{x} \in X^2$. Say that G has *rich examples of binary types*, if for each binary type $[(x_0^*, x_1^*)] \subseteq X^2$, there is an infinite subset $X' \subseteq X$ such that for all $x_0, x_1 \in X$, $x_0 \neq x_1$,

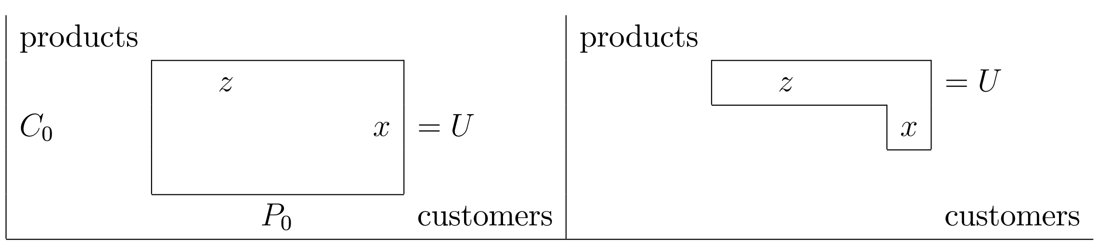
$$(x_0, x_1) \in [(x_0^*, x_1^*)].$$

[1] $G \in \Pi(X)$ is a group of bijections if (a) $\mathrm{id}_X \in G$, (b) for any $g, g'G$, $g \circ g'G$, and (c) for any gG, $g^{-1} \in G$.

G has rich examples of binary types, if, for any binary type, there is an infinite subset of X such that any pair of distinct elements of the subset belongs to the given binary type. Consider Example 2. If x_0^* and x_1^* refer to the same product, then every pair of elements of X that refers to the same product has the same type as tuple (x_0^*, x_1^*) . If x_0^* and x_1^* do not refer to the same product nor to the same customer, then one can find an infinite subset X' such that no pair of elements of X' refers to the same product or customer. See the following figure.



Finite $U \subseteq X$ is *local*, if for any two tuples $\bar{x}, \bar{z} \in U^k$ such that $\bar{x}$ and $\bar{z}$ have the same type, there exists $g \in G$ so that a bijection $g(U) = U$ and $g(\bar{x}) = \bar{z}$. Local sets are sensible finite approximations of $G \longmapsto X$. Consider Example 2. For any finite $C_0 \subseteq C$, $P_0 \subseteq P$, set $U = C_0 \times P_0$ is local. See the left hand of the following figure. On the other hand, the set on the right hand of the following figure is not local. This is because x and z have the same type, but there is no bijection $g \in G$ so that $g(U) = U$ and $g(x) = z$.



Say that $G \longmapsto X$ is *finitely dimensional* if there is a constant $c_0 < \infty$ such that for any finite $V \subseteq X$, there is a local $U \supseteq V$ such that

$$|U| \leq c_0 2^{\frac{1}{28}n} \approx c_0 (1.025)^n .$$

G is finitely dimensional, if the finite approximations do not grow too quickly. In fact, the above definition of finite dimensionality resembles a definition of fractal dimension known from chaos theory. One can also interpret it as a bound on the size of memory of an algorithm that is required to process n points of data.

Consider a collection $\mathcal{S} \subseteq 2^X$ of subsets $S \subseteq X$ for each $S \in \mathcal{S}$. For each $x \in X$, define the set of elements of S that cover x :

$$\mathcal{S}(x) = \{S \in \mathcal{S} : x \in S\}.$$

Say that $\mathcal{S}$ is regular, if there is $m < \infty$ such that for each x, $|\mathcal{S}(x)| = m$. An enumeration of a regular collection $\mathcal{S}$ is a function $j : X \times \{1, ..., m\} \to \mathcal{S}(x)$.

Theorem 1. *Suppose that $G \longmapsto X$ has finitely many types, it is finitely dimensional and it has rich examples of binary relations. Then, there is $m < \infty$, a regular collection $\mathcal{S}$ and enumeration j, such that for any invariant distribution ω, there exists a measurable $f : [0,1]^m \to Y$ such that ω is equal to the joint distribution of variables*

$$\theta(x) = f\left(u_{j_x(1)}, ..., u_{j_x(m)}\right).$$

The Theorem shows that attractive features of de Finetti's Theorem extend to a class of permutation groups. There are constraints on correlations of different variables $\theta(x)$ in invariant distributions. Two variables are correlated if and only if they are affected by the same shock, which happens if and only if they belong to the same set $S \in \mathcal{S}$. Because ex ante correlation implies high probability of ex post similarity, elements of regular collection $S \in \mathcal{S}$ define the similarity structure on S. Additionally, for any set $S \in \mathcal{S}$ such that $|S| = \infty$, there is infinitely many observations that are affected by shock u_S. Thus, such shock is in principle learnable: it can be identified in the long-run, after observing sufficiently many observations $\theta(x)$ for $x \in S$. In other words, the Theorem allows one to distinguish components of invariant distributions that are learnable from those that are not. Additionally, one

A simple heuristic argument allows one to be more precise about the properties of regular collection $\mathcal{S}$ from the Theorem. Suppose that there is a decomposition of invariant distribution ω and set S such that $\theta(x)$ is affected by shock u_S in the same way for each $x \in S$. Because ω is invariant, it must be that $\theta(x)$ is affected in the same way by shock $u_{g(S)}$ for any $g \in G$ such that $x \in g(S)$. Define *index of* S as the number of permutations of S that cover x :

$$i(S) = \#\{g(S) : g \in G \text{ and } x \in g(S)\}.$$

If $i(S) = \infty$, then there is infinitely many permutations $g(S)$ of S that cover x. Thus, infinitely many independent shocks $u_{g(S)}$ affect $\theta(x)$ in the same way, which implies that none of them affects $\theta(x)$. In other words, if $\mathcal{S}$ is a regular collection from the Theorem and $S \in \mathcal{S}$, then it must be that $i(S) < \infty$. This motivates the following definition:

Definition 1. *Set S is a* concept, *if $i(S) < \infty$.*

The name "concept" is motivated by applications below. For example, $S = X, \{x\}$ are (trivial) concepts with index $i(S) = 1$. In general, it depends on group of permutations G whether set S is a concept, it or not. .Let $\mathcal{S}_G$ be a collection of all concepts. One shows that if G has finitely many types, and it has rich examples of binary types, then $\mathcal{S}_G$ is regular. We have the following corollary:

Corollary 1. *Suppose that $G \longmapsto X$ has finitely many types, it is finitely dimensional and it has rich examples of binary relations. Then, there is $m < \infty$, and enumeration $j : X \times \{1, .., m\} \to \mathcal{S}_G$ such that for any invariant distribution ω, there exists a measurable $f : [0,1]^m \to Y$ such that ω is equal to the joint distribution of variables*

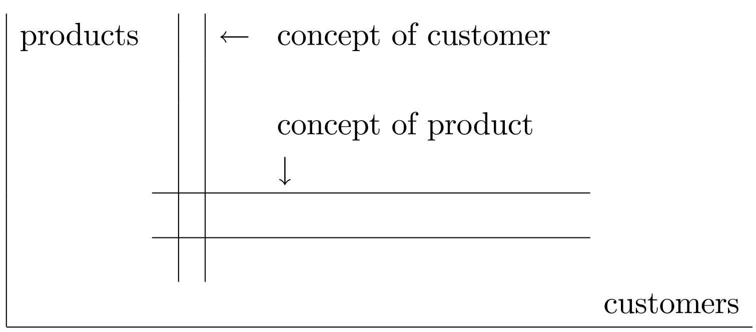
$$\theta(x) = f\left(u_{j_x(1)}, ..., u_{j_x(m)}\right).$$

Consider Example 1. One can easily check that there are only trivial concepts: space X and one-element sets $\{x\}$ for each $x \in X$. Thus, de Finetti's Theorem is a consequence of the Corollary.

Consider Example 2. There are two types of nontrivial concepts: Concept S_c of customer c and concept S_p of product p : .

$$S_c = \{(c, p') : p' \in P\}, S_p = \{(c', p) : c' \in C\}$$

Each concept has index $i(S) = 1$. Such concepts are called *blocks.*



Finally consider an example that is covered by the Corollary, but not by the existing results.

Example 3 (Customers, zipcodes, products, firms)**.** *A marketing company predicts whether customers buy products $p \in P$. The company knows that each customer lives in a unique location characterized by its zipcode $z \in Z$ and the zipcode is known. Also, the company knows that a unique firm produces any given product. Given zipcode, a customer is identified by id cC that is randomly shared by customers across zipcodes. Similarly, given firm, a product is characterized by id $p \in P$ that is randomly shared across firms.*

Let $X = Z \times C \times F \times P$ and let $\theta(z, c, f, p) = 1$ if and only if customer c in zipcode z buys product p of firm f. Consider the following binary relations:

- xR_Zx' if x and x' refer to the same zipcode but different customers,
- $xR_{ZC}x'$ if x and x' refer to the same customer and zipcode,
- xR_Fx' if x and x' refer to the same firm,
- $xR_{FP}x'$ if x and x' refer to the same firm and product.

Let G be the group of **all** permutations that preserve the above relations: for each x, x', each $R \in \{R_Z, R_{ZC}, R_F, R_{FP}\}$, each $g \in G$,

$$\text{if } xRx', \text{ then } g(x)\, Rg(x').$$

Notice that customers across different zipcodes are unrelated and that products across different firms are unrelated.

There are 6 types of non-trivial concepts:

- concept of a zipcode, $S_z = \{(c', \mathbf{z}, f', p')\}$,
- a zipcode and a customer, $S_{zc} = \{(\mathbf{c}, \mathbf{z}, f', p')\}$,
- a firm, $S_f = \{(c', z', \mathbf{f}, p')\}$,
- a firm and a product, $S_{fp} = \{(c', z', \mathbf{f}, \mathbf{p})\}$
- a zipcode, a customer, and a firm, $S_{czf} = \{(\mathbf{c}, \mathbf{z}, \mathbf{f}, p')\}$
- a zipcode, a product and a firm, $S_{zfp} = \{(c', \mathbf{z}, \mathbf{f}, \mathbf{p})\}$.

Additionally, there are two types of trivial concepts: space X and one-element set. The corollary implies that for each invariant ω, there is $f : [0,1]^8 \to R$ such that ω is equal to the joint distribution of random variables

$$\theta(x) := f\left(u_X, u_z, u_f, u_{zc}, u_{fp}, u_{czf}, u_{zfp}, u_{\{x\}}\right).$$

Thus, there are 8 independent sources of uncertainty.

Department of Economics, University of Chicago
E-mail address: mpeski@uchicago.edu