Annales de l'Institut H. Poincaré
Probabilités et Statistiques **39** *(2003) 687–701*

Conditional probabilities and permutahedron

*František Matúš**

Abstract. The conditional probabilities of finite conditional probability spaces are considered for points of a smooth manifold of conditional charges. A linear diffeomorphism on the manifold is constructed so that the conditional probabilities map bijectively onto a permutahedron. The facial structure of the permutahedron corresponds to the ways conditional probability spaces decompose. A new global inversion lemma is devised.

1. Introduction

For a finite set N, let N^* be the set of ordered couples $(i|J)$ where $J \subseteq N$ and $i \in J$. A real function P on N^* is called *c-probability* on N if it is nonnegative,

$$\sum_{i \in J} P(i|J) = 1 \qquad \text{for } J \subseteq N \text{ nonempty,}$$

and

$$(*) \qquad P(i|K) = P(i|J) \sum_{j \in J} P(j|K) \qquad \text{for } i \in J \subseteq K \subseteq N\,.$$

The expression $P(i|J)$ with $J = \{i, j\}$ is simplified to $P(i|i, j)$.

This paper studies the family $\mathcal{P}_N$ of all c-probabilities on N. Obviously, $\mathcal{P}_N$ is a semialgebraic compact subset of the Euclidean space $\mathbb{R}^{N^*}$. Writing $P(J|K) = \sum_{j \in J \cap K} P(j|K)$, any c-probability extends to a function on the set of couples $(J|K)$ where $J, K \subseteq N$ and K is nonempty. This extension coincides with the conditional probability in a conditional probability space, see Remark 1.

Let $\varPi_d \subseteq \mathbb{R}^{\langle d \rangle}$ be the d-dimensional permutahedron defined as convex hull of the vectors $(\pi(0), \dots, \pi(d))$ where π runs over all permutations of the set $\langle d \rangle = \{0, 1, \dots, d\}$, $d \geqslant 0$.

This work was supported by the grants GA AV A1075104 and GA ČR 201/01/1482.

2000 Math. Subject Class. Primary 60A05; secondary 52B11, 53A07, 57R50, 52B70, 14P05, 05A18.

Key words and phrases. Conditional probability space, permutahedron, lattice of faces, ordered partition, manifolds in Euclidean spaces, algebraic variety, global inverse theorem.

*F. Matúš is with Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic; matus@utia.cas.cz

Theorem 1. *The linear mapping* $\boldsymbol{W} = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_d)$ *between* $\mathbb{R}^{\langle d\rangle^*}$ *and* $\mathbb{R}^{\langle d\rangle}$ *given by*

$$\boldsymbol{W}_i(P) = \sum_{\substack{j=0\\ j\neq i}}^{d} P(i|i,j)\,, \qquad 0 \leqslant i \leqslant d,$$

restricts to a homeomorphism between $\mathcal{P}_{\langle d\rangle}$ *and* Π_d.

This assertion follows from three lemmas. First, $\boldsymbol{W}$ is restricted to the family of *trivial* c-probabilities P on $N = \langle d\rangle$, defined by requiring $P(i|N) = x_i > 0$ for each i. Such a function P is uniquely determined by the vector $x = (x_0, \ldots, x_d)$ because eq. $(*)$ implies $P(i|J) = P(i|N)/P(J|N)$ where $i \in J$ and $P(J|N) = \sum_{j\in J} x_j$ is positive. This vector x has positive coordinates satisfying $\sum_{i=0}^{d} x_i = 1$, and thus belongs to the relative interior $\mathit{ri}(\Sigma_d)$ of the standard d-dimensional simplex Σ_d. On the other hand, given $x \in \mathit{ri}(\Sigma_d)$, the formula $P(i|J) = x_i/\sum_{j\in J} x_j$ obviously defines a trivial c-probability P. Therefore, the family of trivial c-probabilities on $\langle d\rangle$ is homeomorphic to $\mathit{ri}(\Sigma_d)$. For a trivial P, $\boldsymbol{W}(P)$ equals $\boldsymbol{T}(x)$ where $\boldsymbol{T} = (\boldsymbol{T}_0, \ldots, \boldsymbol{T}_d)$ is given by

$$\boldsymbol{T}_i(x) = x_i \sum_{\substack{j=0\\ j\neq i}}^{d} \frac{1}{x_i + x_j}\,, \qquad 0 \leqslant i \leqslant d\,.$$

Lemma 1. *The rational mapping* $\boldsymbol{T}$ *is a* C^∞*-diffeomorphism between* $\mathit{ri}(\Sigma_d)$ *and* $\mathit{ri}(\Pi_d)$.

Hence, $\boldsymbol{W}$ restricts to a homeomorphism between the family of trivial c-probabilities on $\langle d\rangle$ and $\mathit{ri}(\Pi_d)$. Our proof of Lemma 1 relies on Lemma 2 which is a topological instance of the global inversion theorems, see Section 2.

Actually, there is more structure transformed between $\mathcal{P}_{\langle d\rangle}$ and Π_d by means of $\boldsymbol{W}$, than claimed in Theorem 1. This is detailed in Lemma 3 of Section 3 which relates the faces of Π_d to decompositions of conditional probability spaces. Consequently, Theorem 1 is obtained and continualization of conditional probabilities is discussed in Remark 4.

For investigation of smoothness of the family $\mathcal{P}_N$ in neighbourhoods of nontrivial c-probabilities, a family $\mathcal{S}_N$ of all c*-charges* on N is introduced. The c-charges are defined by omitting the nonnegativity constraints in the definition of c-probabilities. Obviously, the family $\mathcal{S}_N$ is an algebraic subset of the Euclidean space $\mathbb{R}^{N^*}$. Lemma 4 in Section 4 claims that, moreover, $\mathcal{S}_N$ is a C^∞-manifold; for the manifolds and diffeomorphisms see e.g. [3].

Theorem 2. *The mapping* $\boldsymbol{W}$ *restricts to a* C^∞*-diffeomorphism between an open neighbourhood of* $\mathcal{P}_{\langle d\rangle}$ *in* $\mathcal{S}_{\langle d\rangle}$ *and a neighbourhood of* Π_d *in the hyperplane of* $\mathbb{R}^{\langle d\rangle}$ *given by* $\sum_{i=0}^{d} x_i = \binom{d+1}{2}$.

For the proof of this theorem see Section 5. By Remark 6, $\boldsymbol{W}$ is not injective on the whole family $\mathcal{S}_{\langle d\rangle}$.

An equivalent formula for $\boldsymbol{W}$ restricted to $\mathcal{S}_N$ is presented in Section 6.

To conclude, while the family of usual probability measures on $\langle d\rangle$ is obviously identified with the d-dimensional simplex Σ_d, the family of c-probabilities on $\langle d\rangle$ is, by the above theorems, a 'twisted d-dimensional permutahedron', obtained from Π_d by an inversion of $\boldsymbol{W}$.

Remark 1. A *conditional probability space* is a quadruple $(\Omega, \mathcal{A}, \mathcal{B}, Q)$ where $\mathcal{A}$ is a σ-algebra of subsets of Ω, $\mathcal{B} \subseteq \mathcal{A}$ is nonempty, and Q is a nonnegative function on $\mathcal{A} \times \mathcal{B}$ so that $(\Omega, \mathcal{A}, Q(\cdot|B))$ is a probability space with $Q(B|B) = 1$ for every $B \in \mathcal{B}$, and

$$Q(A \cap B|C) = Q(A|B \cap C) \cdot Q(B|C) \,, \qquad A, B \in \mathcal{A}, \;\; B \cap C, C \in \mathcal{B}\,.$$

Here, Q is the *conditional probability* of the space. Given a triple $(\Omega, \mathcal{A}, \mathcal{B})$, a natural problem is to describe the family of all mappings Q that turn $(\Omega, \mathcal{A}, \mathcal{B}, Q)$ into a conditional probability space. The above results solve the problem when $\Omega = N$ is finite, $\mathcal{A}$ is the power set of N and $\mathcal{B}$ contains all nonempty subsets of N (note that $\emptyset \notin \mathcal{B}$ in every conditional probability space). Indeed, it is not difficult to see that $(N, 2^N, 2^N \setminus \{\emptyset\}, Q)$ is a conditional probability space if and only if Q restricted to N^* is a c-probability on N.

Backgrounds [18] of the conditional probability spaces go back to A. Rényi [15] where a remark on p. 287 witnesses that the concept was conceived independently by A.N. Kolmogorov earlier. Further development can be found in [16, 17, 5]. Yet, the independent philosophical work of B. de Finetti [6, 7] precedes Rényi's recognition of axioms. For this line of research see the recent monograph [4].

Remark 2. The permutahedron, or sometimes also permut<u>o</u>hedron, is a classical polytope, for references see [2] and [22]. The lattice of faces appears in [2, p. 96] and [10] where also applications to VLSI circuits emerge. Convex hulls of the points $(\pi(1), \ldots, \pi(n))$ for a set of permutations π related to a poset, and their facial structure, are recently studied in [19] and [1] where motivation originates from a scheduling problem. Another line of generalization leads to the generalized permutahedra of [12] defined as subsets of the extension lattice of a poset.

2. Global inversion lemma and proof of Lemma 1

The proof of Lemma 1 below is based on the fact that $\boldsymbol{T}$ is a local diffeomorphism on the relative interior $\mathit{ri}(\Sigma_d)$ of Σ_d. To prove that a local diffeomorphism is bijective, a global inversion theorem is usually evoked. The following simple lemma, tailored for this purpose in a topological setting, seems to be of more general interest.

Lemma 2. *Let X and Y be Hausdorff spaces and $B \subseteq X$ be a nonempty, open, connected, and relatively compact set. Let $f\colon B \to Y$ be a local homeomorphism, i.e. each $x \in B$ has an open neighbourhood $U_x \subseteq B$ such that $f(U_x)$ is open in Y and f restricts to a homeomorphism between U_x and $f(U_x)$. The two conditions*

(i) *there exists $x \in B$ such that $f(x) = f(y)$ implies $x = y$ for $y \in B$,*
(ii) *if a net x_λ in B converges to $x \in X \setminus B$ then the net $f(x_\lambda)$ has an accumulation point in $Y \setminus f(B)$,*

are sufficient for the mapping f to be a homeomorphism between B and $f(B)$. They are also necessary when $f(B)$ is relatively compact.

Proof. The set $A = \{x \in B;\ \forall y \in B\colon f(x) = f(y) \Rightarrow x = y\}$ is nonempty by (i). If A and $B \setminus A$ were both open then $A = B$ by the connectivity of B. In turn, f is injective and, since every local homeomorphism is obviously continuous and open, the mapping f makes B and $f(B)$ homeomorphic.

If A were not open then there would exist $x \in A$ and a net x_λ out of A that converges to x. The net can be supposed to be in $U_x \subseteq B$. By construction of A, one can write $f(x_\lambda) = f(y_\lambda)$ for some $y_\lambda \in B$ different from x_λ. Since f is bijective on U_x the net y_λ is out of U_x. As B is relatively compact there exists $y \in X$ such that a subnet $y_{\lambda'}$ of the net y_λ converges to y. The continuity of f at x implies $f(x_\lambda) \to f(x)$, and since Y is Hausdorff $f(x)$ is the unique limit point of the net $f(y_\lambda)$. Thus $f(y_{\lambda'})$ converges to $f(x) \in f(B)$. From (ii) applied to the net $y_{\lambda'}$ one deduces that $y \in B$. By continuity of f at y, one has $f(y_{\lambda'}) \to f(y)$. Hence $f(x) = f(y)$, and then $x = y$ because $x \in A$. Therefore $y_{\lambda'}$ is eventually in U_x, a contradiction.

If $B \setminus A$ were not open then there would exist $x \in B \setminus A$ and a net x_λ out of $B \setminus A$ that converges to x. The net can be supposed to be in the open set B, and hence in A. There is $y \in B$ different from x giving $f(x) = f(y)$. Since X is Hausdorff x and y possess disjoint open neighbourhoods, say $x \in C$ and $y \in D$. The continuity of f at x implies $f(x_\lambda) \to f(x)$, and then the net $f(x_\lambda)$ is eventually in the open set $f(D)$. Once $f(x_\lambda) \in f(D)$ one can find $y_\lambda \in D$ such that $f(x_\lambda) = f(y_\lambda)$. From some moment $x_\lambda \in B \cap C$ and $y_\lambda \in B \cap D$, and as C and D are disjoint $x_\lambda \neq y_\lambda$. This contradicts $x_\lambda \in A$.

For the necessity, let $f\colon B \to f(B)$ be a homeomorphism and $f(B)$ be relatively compact in Y. Then (i) holds trivially. If a net x_λ in B converges to $x \in X \setminus B$, then the net $f(x_\lambda)$ has at least one accumulation point $y \in Y$. Passing to a subnet if necessary, one can suppose that $f(x_\lambda) \to y$ and $x_\lambda \to x$. When $y \in f(B)$ the continuity of f^{-1} implies $x_\lambda \to f^{-1}(y)$. Since X is Hausdorff one has $x = f^{-1}(y) \in B$, a contradiction. Therefore $y \in Y \setminus f(B)$ and (ii) is established. □

Remark 3. There is a considerable number of global univalence and inverse function theorems that bear similarities to Lemma 2. For an overview and further references see [11, 13, 14, 21]. What makes Lemma 2 different from known assertions of this type is combination of the conditions (i) and (ii).

Proof of Lemma 1. Let H_1 be the hyperplane given by $\sum_{i=0}^d x_i = 1$ and H_2 be the parallel hyperplane given by $\sum_{i=0}^d x_i = \binom{d+1}{2}$. The relative interior of Σ_d consists of the points of H_1 with all coordinates positive. It is well known that the permutahedron Π_d, viewed as a polytope, is equivalently described by $\Pi_d \subseteq H_2$ and the inequalities $\sum_{i\in I} x_i \geqslant \binom{|I|}{2}$ for $I \subsetneq \langle d\rangle$. Then $\mathit{ri}(\Pi_d)$ consists of the points $x \in H_2$ satisfying $\sum_{i\in I} x_i > \binom{|I|}{2}$ for all $\emptyset \neq I \subsetneq \langle d\rangle$.

Lemma 2 will be applied to the open, connected and bounded set $B = \mathit{ri}(\Sigma_d)$ in the space $X = H_1$ and the mapping $f = \boldsymbol{T}$ of B into $Y = H_2$. The mapping $\boldsymbol{T}$ maps $\mathit{ri}(\Sigma_d)$ into $\mathit{ri}(\Pi_d)$ because for $x \in \mathit{ri}(\Sigma_d)$

$$\sum_{i\in I} \boldsymbol{T}_i(x) = \binom{|I|}{2} + \sum_{i\in I} \sum_{j\in\langle d\rangle\setminus I} \frac{x_i}{x_i + x_j} \geqslant \binom{|I|}{2}, \qquad |I| \subseteq \langle d\rangle,$$

with the strict inequality for I nonempty and different from $\langle d\rangle$. Another consequence of this inequality to be used in the sequel is

(iii) if a sequence x^n in $\mathit{ri}(\Sigma_d)$ converges to x out of $\mathit{ri}(\Sigma_d)$ then the sequence $\boldsymbol{T}(x^n)$ has all accumulation points out of Π_d.

Indeed, if the set $I \subseteq \langle d\rangle$ labels the zero coordinates of the limit point x then $\emptyset \neq I \subsetneq \langle d\rangle$ and each accumulation point z of $\boldsymbol{T}(x^n)$ satisfies $\sum_{i\in I} z_i = \binom{|I|}{2}$, the double sum vanishing in limit. Thus, $z \in \Pi_d \setminus \mathit{ri}(\Pi_d)$.

To show that $\boldsymbol{T}$ is a local homeomorphism on $\mathit{ri}(\Sigma_d)$, observe that the Jacobi matrix

$$\left(\frac{\partial \boldsymbol{T}_i}{\partial x_j}\right)_{i,j=0}^{d} \quad \text{has} \quad \sum_{\substack{k=0\\k\neq i}}^{d} \frac{x_k}{(x_i+x_k)^2} \quad \text{and} \quad -\frac{x_i}{(x_i+x_j)^2}$$

for diagonal and off-diagonal elements, respectively. The column sums equal zero, the diagonal entries are positive, and the off-diagonal entries are negative. Each principal proper submatrix of the matrix is strictly diagonally dominant up to the transposition, and by [8, Theorem 6.1.10, p. 349] regular. Therefore, $\boldsymbol{T}$ is a local homeomorphism of $\mathit{ri}(\Sigma_d)$ into H_2.

To verify (i), note that $x = (\frac{1}{d+1}, \dots, \frac{1}{d+1}) \in \mathit{ri}(\Sigma_d)$ is transformed to $\boldsymbol{T}(x) = (\frac{d}{2}, \dots, \frac{d}{2})$. If $\boldsymbol{T}(y) = \boldsymbol{T}(x)$ for some $y \in \mathit{ri}(\Sigma_d)$ then

$$0 = \boldsymbol{T}_i(y) - \boldsymbol{T}_j(y) = \frac{y_i - y_j}{y_i + y_j} + \sum_{\substack{k=0\\k\notin\{i,j\}}}^{d} \frac{y_i}{y_i+y_k} - \frac{y_j}{y_j+y_k}, \qquad 0 \leqslant i,j \leqslant d.$$

This entails $y_i = y_j$, and thus $y = x$.

The validity of (ii) follows directly from (iii).

Now, Lemma 2 implies that $\boldsymbol{T}$ is a homeomorphism of $\mathit{ri}(\Sigma_d)$ onto its image. The image is open in the topology of H_2. From (iii) it follows that the closure of $\boldsymbol{T}(\mathit{ri}(\Sigma_d))$ is contained in $\boldsymbol{T}(\mathit{ri}(\Sigma_d)) \cup \big[\,\Pi_d \setminus \mathit{ri}(\Pi_d)\,\big]$. Then $\mathit{ri}(\Pi_d) \setminus \boldsymbol{T}(\mathit{ri}(\Sigma_d))$, the relative complement of this closure in $\mathit{ri}(\Pi_d)$, is open in H_2. Since $\mathit{ri}(\Pi_d)$ is connected and $\boldsymbol{T}(\mathit{ri}(\Sigma_d))$ with the complement are both open the latter must be empty. Hence, $\boldsymbol{T}$ maps $\mathit{ri}(\Sigma_d)$ onto $\mathit{ri}(\Pi_d)$.

Obviously, both $\mathit{ri}(\Sigma_d)$ and $\mathit{ri}(\Pi_d)$ are C^∞-manifolds and $\boldsymbol{T}$ is a C^∞-diffeomorphism between them, see [3, p. 67]. □

3. Faces of permutahedron and the mapping $\boldsymbol{W}$

Where N is a finite nonempty set, say $N = \langle d\rangle$, let ρ be an ordered partition of N, i.e. a sequence $N_1, \dots, N_k$, $1 \leqslant k \leqslant |N|$, of nonempty disjoint subsets, blocks, of N that cover N. Let $\rho \preccurlyeq \rho'$ for two ordered partitions of N if each block in ρ' is union of $N_i, N_{i+1}, \dots, N_j$ for some $1 \leqslant i \leqslant j \leqslant k$ and if the order of blocks in ρ' is induced from the order of blocks in ρ. The ordered partitions of N with the relation $\preccurlyeq$ form a poset that becomes a lattice when a bottom element is added.

This lattice is isomorphic to the lattice of faces of the permutahedron Π_d. Given an ordered partition ρ of $N = \langle d\rangle$, up to symmetry with the blocks $\langle j_1\rangle \setminus \langle j_0\rangle, \dots, \langle j_k\rangle \setminus \langle j_{k-1}\rangle$

where $-1 = j_0 < j_1 < \ldots < j_k = d$, $\langle j_0 \rangle = \emptyset$ and $1 \leqslant k \leqslant d+1$, the corresponding nonempty face F_ρ of Π_d is defined by the equalities

$$\sum_{i=j_\ell+1}^{d} x_i = \binom{d-j_\ell}{2}, \qquad 0 \leqslant \ell \leqslant k \,.$$

In fact, any face of Π_d is obtained by prescribing equalities in some of the inequalities $\sum_{i \in I} x_i \geqslant \binom{|I|}{2}$, $I \subseteq \langle d \rangle$. The equalities indexed by I, J and satisfied by x from a face provide

$$\binom{|I|}{2} + \binom{|J|}{2} = \sum_{i \in I \cap J} x_i + \sum_{j \in I \cup J} x_j \geqslant \binom{|I \cap J|}{2} + \binom{|I \cup J|}{2}$$

what implies $0 \geqslant |I \setminus J||J \setminus I|$. Thus, I and J are in inclusion. Prescribed equalities are therefore indexed by $\langle d \rangle \setminus \langle j_0 \rangle, \ldots, \langle d \rangle \setminus \langle j_{k-1} \rangle$ up to symmetry.

The face F_ρ can be shifted in $\mathbb{R}^{d+1}$ to coincide with Cartesian product of the permutahedra $\Pi_{j_1-j_0-1}, \ldots, \Pi_{j_k-j_{k-1}-1}$. The appropriate shift moves the coordinate x_i of $x \in \mathbb{R}^{\langle d \rangle}$ to $x_i - (d - j_\ell)$ for $j_{\ell-1} < i \leqslant j_\ell$, $1 \leqslant \ell \leqslant k$.

An ordered partition of N can be identified within each c-charge on N. To describe the partition, compositions and decompositions of c-charges are discussed below. These correspond to the construction of conditional probability spaces from an ordered set of measures that goes back to [16] (on p. 57 this idea is attributed to E. Marczewski) and to [5, 9].

Having two c-charges $P \in \mathcal{S}_N$, $Q \in \mathcal{S}_M$, and N, M disjoint, a c-charge $R \in \mathcal{S}_{N \cup M}$ can be defined by

$$R(i|I) = \begin{cases} Q(i|I), & I \subseteq M \,, \\ P(i|I \cap N), & i \in N \,, \\ 0, & i \in M,\ I \cap N \neq \emptyset \,, \end{cases}$$

where $(i|I) \in (N \cup M)^*$. Obviously, P and Q are the restrictions of R to N^* and M^*, respectively. Writing $R = P \ltimes Q$, the operation $\ltimes$ is associative but not commutative.

It follows from $(*)$ that any $R \in \mathcal{S}_N$ can be written as $P_1 \ltimes Q$ with P_1 equal to the restriction of R to N_1^* where the set $N_1 = \{i \in N; P(i|N) \neq 0\}$ is not empty. Here, the c-charge $P_1 \in \mathcal{S}_{N_1}$ is *trivial* in the sense that the values $P_1(i|N_1)$, $i \in N_1$, are nonzero. Repeating this kind of decomposition one has $R = P_1 \ltimes \ldots \ltimes P_k$ for $1 \leqslant k \leqslant |N|$ and trivial c-charges $P_1 \in \mathcal{S}_{N_1}$, …, $P_k \in \mathcal{S}_{N_k}$. The sequence of sets $N_1, \ldots, N_k$ is an ordered partition of N induced by R.

Given an ordered partition ρ of N, the notation $\mathcal{P}_N^\rho$ is reserved for the set of all c-probabilities R on N such that ρ is the ordered partition induced by R.

Lemma 3. *For an ordered partition ρ of $\langle d \rangle$ the mapping $\boldsymbol{W}$ restricts to a C^∞-diffeomorphism between $\mathcal{P}_{\langle d \rangle}^\rho$ and the relative interior of the face F_ρ of Π_d.*

Proof. Let ρ have the blocks $\langle j_1 \rangle \setminus \langle j_0 \rangle, \ldots, \langle j_k \rangle \setminus \langle j_{k-1} \rangle$ as above and $\boldsymbol{\iota}^\rho$ be the coordinate projection of $\mathbb{R}^{\langle d \rangle^*}$ to $\mathbb{R}^{\langle d \rangle}$ keeping the coordinates $P(i|\langle j_\ell \rangle \setminus \langle j_{\ell-1} \rangle)$ for $j_{\ell-1} < i \leqslant j_\ell$ and $1 \leqslant \ell \leqslant k$. Since the c-probabilities from $\mathcal{P}_{\langle d \rangle}^\rho$ decompose uniquely into trivial c-probabilities on the blocks of ρ the projection $\boldsymbol{\iota}^\rho$ is a C^∞-diffeomorphism between $\mathcal{P}_{\langle d \rangle}^\rho$ and the Cartesian product of $\mathit{ri}(\Sigma_{j_1-j_0-1}), \ldots, \mathit{ri}(\Sigma_{j_k-j_{k-1}-1})$, to be denoted by $\mathit{ri}(\Sigma_d^\rho)$.

For $P \in \mathcal{P}^{\rho}_{\langle d\rangle}$ and $j_{\ell-1} < i \leqslant j_\ell$ one has $P(i|i,m) \neq 0$ if and only if $j_{\ell-1} < m \leqslant d$, and $P(i|i,m) = 1$ if and only if $j_\ell < m \leqslant d$ or $m = i$. Therefore

$$\boldsymbol{W}_i(P) = (d - j_\ell) + \sum_{\substack{m=j_{\ell-1}+1 \\ m\neq i}}^{j_\ell} P(i|i,m) = (d - j_\ell) + \boldsymbol{T}^{\rho}_i(\boldsymbol{\iota}^{\rho}(P))$$

where $\boldsymbol{T}^{\rho}$ is the mapping from $ri(\Sigma^{\rho}_d)$ to $\mathbb{R}^{\langle d\rangle}$ given by

$$\boldsymbol{T}^{\rho}_i(x) = x_i \sum_{\substack{m=j_{\ell-1}+1 \\ m\neq i}}^{j_\ell} \frac{1}{x_i + x_m} \ .$$

By Lemma 1, $\boldsymbol{T}^{\rho}$ is a C^{∞}-diffeomorphism between $ri(\Sigma^{\rho}_d)$ and the Cartesian product of $ri(\Pi_{j_1-j_0-1}), \ldots, ri(\Pi_{j_k-j_{k-1}-1})$, to be denoted by $ri(\Pi^{\rho}_d)$. Then the composition $\boldsymbol{T}^{\rho}\boldsymbol{\iota}^{\rho}$ is a C^{∞}-diffeomorphism between $\mathcal{P}^{\rho}_{\langle d\rangle}$ and $ri(\Pi^{\rho}_d)$. Since $\boldsymbol{W}$ differs from this mapping only by the shift that moves $ri(\Pi^{\rho}_d)$ to $ri(F_\rho)$, the assertion follows. □

Proof of Theorem 1. The set $\mathcal{P}_{\langle d\rangle}$ partitions into $\mathcal{P}^{\rho}_{\langle d\rangle}$ and the permutahedron Π_d partitions into $ri(F_\rho)$ where ρ runs in both cases over the ordered partitions of $\langle d\rangle$. By Lemma 3, $\boldsymbol{W}$ maps injectively $\mathcal{P}^{\rho}_{\langle d\rangle}$ onto $ri(F_\rho)$. Thus, $\boldsymbol{W}$ is a bijection between $\mathcal{P}_{\langle d\rangle}$ and Π_d. Obviously, $\boldsymbol{W}$ is continuous and closed on the compact $\mathcal{P}_{\langle d\rangle}$, and hence a homeomorphism. □

As a consequence, the family of trivial c-probabilities on N is dense in $\mathcal{P}_N$.

Remark 4. In [20] a sophisticated homeomorphism ϕ of $ri(\Sigma_d)$ into $\mathbb{R}^d$ was constructed so that the closure of $U = \phi(ri(\Sigma_d))$ is homeomorphic to a d-dimensional simplex, and for any $i \in J \subseteq \langle d\rangle$ the real function $f_{i|J}(y) = \phi^{-1}(y)_i / \sum_{j\in J} \phi^{-1}(y)_j$, $y \in U$, is uniformly continuous. This result, interpreted as a simultaneous continualization of conditional probabilities, is a simple consequence of Theorem 1. In fact, the choice $\phi = \boldsymbol{T}$ leads to $U = ri(\Pi_d)$ with its closure Π_d obviously homeomorphic to a simplex. For $y \in U$ one has $y = \boldsymbol{T}(x)$ for a unique $x \in ri(\Sigma_d)$ and then a unique c-probability P exists such that $x_i = P(i|\langle d\rangle) > 0$ for each $0 \leqslant i \leqslant d$. In turn, $\boldsymbol{W}(P) = \boldsymbol{T}(x) = y$ and thus the number $f_{i|J}(y) = x_i / \sum_{j\in J} x_j = P(i|J)$ is a coordinate of $\boldsymbol{W}^{-1}(y)$. By Theorem 1, $\boldsymbol{W}^{-1}$ is continuous on the compact Π_d and thus uniformly continuous on U. Consequently, each function $f_{i|J}$ is uniformly continuous on U.

4. Charts of the manifold $\mathcal{S}_N$

First, a new construction of c-charges is presented. Given $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$, let

$$s_{i,j,k}(t) = \prod_{m=i+1}^{j} t_m \prod_{n=j+1}^{k} (1 - t_n) \,, \qquad 0 \leqslant i \leqslant j \leqslant k \leqslant d \,,$$

and let $V_d \subseteq \mathbb{R}^d$ be the set of t's rendering $\sum_{i\in I} s_{\wedge I,i,\vee I}(t) \neq 0$ for every $\emptyset \neq I \subseteq \langle d\rangle$. Here, $\wedge I$ is the smallest element and $\vee I$ the greatest element of I. Let a mapping $\boldsymbol{Z}$ from the open set V_d into $\mathbb{R}^{\langle d\rangle^*}$ be given by

$$\boldsymbol{Z}t(i|J) = \frac{s_{\wedge J,i,\vee J}(t)}{\sum_{j\in J} s_{\wedge J,j,\vee J}(t)} \ , \quad i \in J \subseteq \langle d\rangle.$$

Then $\boldsymbol{Z}t$ is a c-charge on $\langle d\rangle$ for every $t \in V_d$. In fact, $\sum_{i\in J} \boldsymbol{Z}t(i|J) = 1$ and for $i \in J \subseteq K \subseteq \langle d\rangle$

$$\boldsymbol{Z}t(i|J)\cdot\boldsymbol{Z}t(J|K) = \frac{s_{\wedge J,i,\vee J}(t)}{\sum_{j\in J} s_{\wedge J,j,\vee J}(t)}\cdot\frac{\sum_{j\in J} s_{\wedge K,j,\vee K}(t)}{\sum_{k\in K} s_{\wedge K,k,\vee K}(t)} = \boldsymbol{Z}t(i|K)\,,$$

using the obvious identity $s_{\wedge J,i,\vee J}\cdot s_{\wedge K,j,\vee K} = s_{\wedge J,j,\vee J}\cdot s_{\wedge K,i,\vee K}$. Due to $\boldsymbol{Z}t(i|i,i-1) = t_i$, $1 \leqslant i \leqslant d$, the mapping $\boldsymbol{Z}$ is one-to-one.

Lemma 4. *The family $\mathcal{S}_{\langle d\rangle}$ of c-charges on $\langle d\rangle$ is a C^∞-manifold of dimension d.*

Proof. The idea is to compose the mapping $\boldsymbol{Z}$ with the isometries of $\mathbb{R}^{\langle d\rangle^*}$ that transform P to $(P(\pi(i)|\pi(J))\,;\,(i|J)\in N^*)$ where π is a permutation of $\langle d\rangle$ and $\pi(J) = \{\pi(j)\,;\, j \in J\}$. These compositions will provide an atlas of charts.

The case $d = 0$ is trivial. Let $d \geqslant 1$ and P be a c-charge on $\langle d\rangle$. The vector $x \in \mathbb{R}^{d+1}$ with the coordinates $x_i = P(i|\langle d\rangle)$ is nonzero and, due to $(*)$, solves the homogeneous linear equations $x_{i\oplus 1} = P(i\oplus 1|i\oplus 1, i)(x_{i\oplus 1} + x_i)$, $0 \leqslant i \leqslant d$. Here, $\oplus$ is the addition modulo $d+1$. Hence the determinant

$$\begin{vmatrix} P(1|1,0) & -P(0|1,0) & 0 & \dots & 0 \\ 0 & P(2|2,1) & -P(1|2,1) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -P(d|0,d) & 0 & 0 & \dots & P(0|0,d) \end{vmatrix} = \prod_{i=0}^{d} P(i|i,i\ominus 1) - \prod_{i=0}^{d} P(i|i,i\oplus 1)$$

vanishes. This implies

$$P(d|0,d)\Big[\prod_{m=1}^{d} t_m + \prod_{n=1}^{d}(1-t_n)\Big] = \prod_{m=1}^{d} t_m\,.$$

where $t_i = P(i|i,i-1)$, $1 \leqslant i \leqslant d$. Along the same guidelines

$$(\dagger)\qquad P(j|i,j)\Big[\prod_{m=i+1}^{j} t_m + \prod_{n=i+1}^{j}(1-t_n)\Big] = \prod_{m=i+1}^{j} t_m\,,\qquad 0 \leqslant i < j \leqslant d\,.$$

Since the vector x solves also the $d+1$ equations $x_i = P(i|i,i-1)(x_i + x_{i-1})$, $1 \leqslant i \leqslant d$, and $\sum_{i=0}^{d} x_i = 1$, by Cramer's rule

$$x_i\begin{vmatrix} P(1|1,0) & -P(0|1,0) & 0 & \dots & 0 & 0 \\ 0 & P(2|2,1) & -P(1|2,1) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & P(d|d,d-1) & -P(d-1|d,d-1) \\ 1 & 1 & 1 & \dots & 1 & 1 \end{vmatrix} = \prod_{m=1}^{i} t_m \prod_{n=i+1}^{d}(1-t_n)$$

and thus

$$P(i|\langle d\rangle)\cdot\sum_{k=0}^{d} s_{0,k,d}(t) = s_{0,i,d}(t)\,.$$

The same argument with $J = \{i_0,\dots,i_\ell\}$ (where $i_0 < \dots < i_\ell$ and $1 \leqslant \ell \leqslant d$) implies

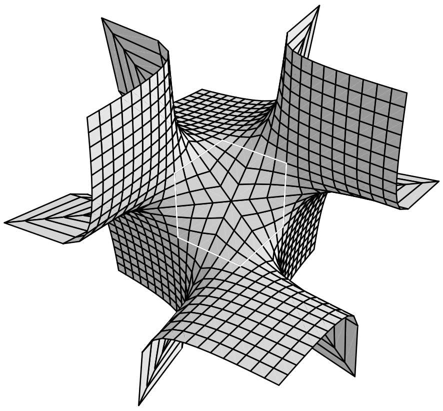
$$P(i_k|J)\sum_{j=0}^{\ell}\prod_{m=1}^{j} P(i_m|i_m,i_{m-1})\prod_{n=j+1}^{\ell} P(i_{n-1}|i_{n-1},i_n) = \prod_{m=1}^{k} P(i_m|i_m,i_{m-1})\prod_{n=k+1}^{\ell} P(i_{n-1}|i_{n-1},i_n)$$

for $0 \leqslant k \leqslant \ell$. Eliminating $P(j|i,j)$ by means of (†)

$$(\ddagger) \qquad P(i|J) \cdot \sum_{j\in J} s_{\wedge J,j,\vee J}(t) = s_{\wedge J,i,\vee J}(t) \;, \qquad i \in J \subseteq \langle d \rangle \;.$$

Let the c-charge P be decomposed into $P = P_1 \ltimes \ldots \ltimes P_k$, $1 \leqslant k \leqslant d+1$, with all components trivial. Up to a permutation of $\langle d \rangle$, the underlying ordered partition has the blocks $\langle j_1 \rangle \setminus \langle j_0 \rangle, \ldots, \langle j_k \rangle \setminus \langle j_{k-1} \rangle$, see Section 3. Then all numbers $t_i = P(i|i, i-1)$, $1 \leqslant i \leqslant d$, differ from 1 and therefore $s_{\wedge J,\wedge J,\vee J}(t)$ does not vanish for nonempty J. On account of (‡), $\sum_{j\in J} s_{\wedge J,j,\vee J}(t)$ in nonzero. Hence, $t \in V_d$, $P = \boldsymbol{Z}t$ and $\boldsymbol{Z}$ maps a neighbourhood of t onto a neighbourhood of P in $\mathcal{S}_{\langle d \rangle}$. This proves that $\boldsymbol{Z}V_d$ is a chart of $\mathcal{S}_{\langle d \rangle}$ and every c-charge belongs to the chart up to a permutation. □

Remark 5. The above arguments imply that each c-charge P is uniquely determined by its values $P(i|i,j)$ (for c-probabilities this follows independently from Theorem 1). Thus, $P \in \mathcal{S}_{\langle 2 \rangle}$ is determined by the triple $(P(1|0,1), P(2|1,2), P(0|2,0))$. Projecting the family $\mathcal{S}_{\langle 2 \rangle}$ to these coordinates one obtains an unbounded surface in $\mathbb{R}^3$. In a neighbourhood of the origin the surface can be depicted as



The six white segments in the figure correspond to the edges of the unit cube which belong to the surface. They border a twisted hexagon which is the projection of the family $\mathcal{P}_{\langle 2 \rangle}$, homeomorphic to the hexagon $\mathit{\Pi}_2$ by Theorem 1. It is not difficult to show that $\mathcal{S}_{\langle 2 \rangle}$ has four connected components, each one C^∞-diffeomorphic to $\mathbb{R}^2$.

5. Proof of Theorem 2

The case $d = 0$ is trivial. First, let us show that for $d \geqslant 1$ each c-probability $P \in \mathcal{P}_{\langle d \rangle}$ has a neighbourhood U_P in $\mathcal{S}_{\langle d \rangle}$ such that $\boldsymbol{W}$ is a C^∞-diffeomorphism between U_P and an open subset of the hyperplane H_2.

For a trivial c-probability P one can take for U_P the family of all trivial c-probabilities. Then $\boldsymbol{W}$ maps U_P onto $ri(\mathit{\Pi}_d)$ by Lemma 3. Let P be a c-probability which is not trivial. Up to a permutation, P can be decomposed and expressed as $P = \boldsymbol{Z}t$ where $t \in \mathbb{R}^d$ has components $0 \leqslant t_i < 1$, $1 \leqslant i \leqslant d$ (see end of the previous proof). To conclude that P has a desired neighbourhood, it is necessary and sufficient to verify that the Jacobi matrix C of the composition $\boldsymbol{WZ}$ at the point t has rank d. An induction argument on the dimension d will be applied.

Using the computations

$$\frac{\partial s_{i,j,k}}{\partial t_\ell} = \frac{\partial}{\partial t_\ell} \prod_{m=i+1}^{j} t_m \prod_{n=j+1}^{k} (1-t_n) = \begin{cases} 0\,, & \ell \leqslant i \text{ or } \ell \geqslant k+1\,, \\ s_{i,\ell-1,\ell-1} \cdot s_{\ell,j,j} \cdot s_{j,k,k}\,, & i+1 \leqslant \ell \leqslant j\,, \\ -s_{i,j,j} \cdot s_{j,j,\ell-1} \cdot s_{\ell,\ell,k}\,, & j+1 \leqslant \ell \leqslant k\,, \end{cases}$$

and

$$(\boldsymbol{WZ})_i = \sum_{k=0}^{i-1} \frac{s_{k,i,i}}{s_{k,k,i}+s_{k,i,i}} + \sum_{k=i+1}^{d} \frac{s_{i,i,k}}{s_{i,i,k}+s_{i,k,k}}\,, \qquad 0 \leqslant i \leqslant d\,,$$

the matrix C has the elements

$$c_{i,j} = \frac{\partial(\boldsymbol{WZ})_i}{\partial t_j} = s_{j,i,i} \cdot s_{j,j,i} \sum_{k=0}^{j-1} \frac{s_{k,j-1,j-1} \cdot s_{k,k,j-1}}{[s_{k,k,i}+s_{k,i,i,}]^2}\,, \qquad 1 \leqslant j \leqslant i \leqslant d\,,$$

and

$$c_{i,j} = \frac{\partial(\boldsymbol{WZ})_i}{\partial t_j} = -s_{i,i,j-1} \cdot s_{i,j-1,j-1} \sum_{k=j}^{d} \frac{s_{j,j,k} \cdot s_{j,k,k}}{[s_{i,i,k}+s_{i,k,k}]^2}\,, \qquad 0 \leqslant i < j \leqslant d\,.$$

Since P is not trivial $t_\ell = 0$ for some $1 \leqslant \ell \leqslant d$. Then the element $c_{i,j}$ vanishes if $1 \leqslant j < \ell \leqslant i \leqslant d$ or $0 \leqslant i < \ell < j \leqslant d$. In other words, the nonzero elements of C belong to the submatrix A indexed by $0 \leqslant i < \ell$ and $1 \leqslant j < \ell$, or to the submatrix B indexed by $\ell \leqslant i \leqslant d$ and $\ell < j \leqslant d$, or to the ℓ-th column of C. If $\ell > 1$ then, in the dimension $\ell - 1 < d$, A coincides with the Jacobi matrix of the composition $\boldsymbol{WZ}$ at the point $(t_1, \dots, t_{\ell-1})$. By induction, A has rank $\ell - 1$. Analogous argument shows that the rank of B is $d - \ell$ when $\ell < d$. Since not all t_i's are positive every $s_{i,j,k}(t)$ is nonnegative and every $s_{i,i,k}(t)$ positive. Now, $c_{i,\ell} \leqslant 0$ for $0 \leqslant i < \ell - 1$ and $c_{\ell-1,\ell} \leqslant -1$. In turn, $(c_{0,\ell}, \dots, c_{\ell-1,\ell})$ cannot be a linear combination of the columns of A as each column sums to zero. Therefore the first ℓ columns of C are linearly independent. Remembering that the rank of B is $d-\ell$, the Jacobi matrix C has rank d. This means that for every $P \in \mathcal{P}_{\langle d \rangle}$ the mapping $\boldsymbol{W}$ is a homeomorphism, and in turn a C^∞-diffeomorphism, of an open set $U_P \subseteq \mathcal{S}_{\langle d \rangle}$ into H_2.

Having the desired neighbourhoods U_P, suppose, by contradiction, that $\boldsymbol{W}$ is not injective on every neighbourhood of $\mathcal{P}_{\langle d \rangle}$ in $\mathcal{S}_{\langle d \rangle}$. By compactness of $\mathcal{P}_{\langle d \rangle}$, in $\mathcal{S}_{\langle d \rangle}$ there exist two convergent sequences $P_n \to P$ and $Q_n \to Q$ with limits in $\mathcal{P}_{\langle d \rangle}$ such that $\boldsymbol{W}(P_n) = \boldsymbol{W}(Q_n)$ and $P_n \neq Q_n$. Due to the continuity of $\boldsymbol{W}$, one has $\boldsymbol{W}(P) = \boldsymbol{W}(Q)$, and thus $P = Q$, by Theorem 1. Then, the sequences P_n and Q_n are eventually in the set U_P where $\boldsymbol{W}$ is injective, a contradiction. This closes the proof of Theorem 2.

Remark 6. The mapping $\boldsymbol{W}$ is not injective on the whole family $\mathcal{S}_{\langle 2 \rangle}$. In fact, the two c-charges given by rows of the table

$P(0\|\langle 2 \rangle)$	$P(1\|\langle 2 \rangle)$	$P(2\|\langle 2 \rangle)$	$P(0\|0,1)$	$P(1\|0,1)$	$P(0\|0,2)$	$P(2\|0,2)$	$P(1\|1,2)$	$P(2\|1,2)$
-1	-1	3	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$
0	-1	2	0	1	0	1	-1	2

are mapped by $\boldsymbol{W}$ to the single point $(0, 0, 3) \notin \mathit{\Pi}_2$.

6. Alternative formula for $\boldsymbol{W}$ restricted to $\mathcal{S}_N$

The mapping $\boldsymbol{W}$ can be written, modulo the polynomials defining $\mathcal{S}_N$, in a different form that seems to be interesting per se. To this end, let S be the permutation group of the set $\langle d\rangle$. For $\pi \in S$ and $P\colon \mathbb{R}^{\langle d\rangle^*} \to \mathbb{R}$ the symbol $P\big(\frac{\pi(k)}{\pi(k\ldots d)}\big)$ abbreviates $P(\pi(k)|\{\pi(k),\ldots,\pi(d)\})$, $0 \leqslant k \leqslant d$.

Lemma 5. *For every c-charge P on $\langle d\rangle$*

$$\boldsymbol{W}_i(P) = d - \sum_{\pi\in S} \pi^{-1}(i) \prod_{k=0}^{d} P\big(\tfrac{\pi(k)}{\pi(k\ldots d)}\big)\,, \qquad 0 \leqslant i \leqslant d\,.$$

Proof. Let

$$\alpha_{i,j,k}^{m\to n} = \sum_{\substack{\pi\in S\\ \pi(m)=n}} P\big(\tfrac{n}{\pi(k\ldots d)}\big) \prod_{\ell=i}^{j-1} P\big(\tfrac{\pi(\ell)}{\pi(\ell\ldots d)}\big)\,, \quad 0 \leqslant i \leqslant j \leqslant k \leqslant m \leqslant d,\ 0 \leqslant n \leqslant d\,.$$

It is straightforward that $\alpha_{i,j,k}^{m\to n} = \alpha_{i,j,k}^{d\to n}$. Hence, keeping n fixed, the superindices can be omitted. The above formula for $\alpha_{i,j+1,k}$ with $j<k$ contains $P\big(\frac{\pi(j)}{\pi(j\ldots d)}\big)$ as the last term of a product. By symmetry, this term can be replaced by $P\big(\frac{\pi(\ell)}{\pi(j\ldots d)}\big)$ for $j \leqslant \ell < k$ and hence

$$\alpha_{i,j+1,k} = \sum_{\substack{\pi\in S\\ \pi(d)=n}} P\big(\tfrac{n}{\pi(k\ldots d)}\big)\, \tfrac{1}{k-j}\, P\big(\tfrac{\pi(j\ldots k-1)}{\pi(j\ldots d)}\big) \prod_{\ell=i}^{j-1} P\big(\tfrac{\pi(\ell)}{\pi(\ell\ldots d)}\big)\,.$$

Using $P\big(\frac{\pi(j\ldots k-1)}{\pi(j\ldots d)}\big) = 1 - P\big(\frac{\pi(k\ldots d)}{\pi(j\ldots d)}\big)$ and $(*)$

$$\alpha_{i,j+1,k} = \tfrac{1}{k-j}\,[\,\alpha_{i,j,k} - \alpha_{i,j,j}\,]\,.$$

This recurrence is used to prove

$$\alpha_{i,j,j} = \sum_{\ell=0}^{j-i} (-1)^{j-i-\ell}\, \frac{\alpha_{i,i,i+\ell}}{\ell\,!(j-i-\ell)!}$$

by induction on $j-i = 0,1,\ldots,d$. In fact, the case $i=j$ is obvious. For $j+1 \leqslant d$

$$\alpha_{i,j+1,j+1} = \alpha_{i,j,j+1} - \alpha_{i,j,j} = \ldots = \frac{\alpha_{i,i,j+1}}{(j+1-i)!} - \sum_{k=0}^{j-i} \frac{\alpha_{i,j-k,j-k}}{(k+1)!}$$

where $\alpha_{i,j-k,j-k}$ is rewritten by the induction hypothesis

$$\alpha_{i,j+1,j+1} = \frac{\alpha_{i,i,j+1}}{(j+1-i)!} - \sum_{k=0}^{j-i} \frac{1}{(k+1)!} \sum_{\ell=0}^{j-k-i} (-1)^{j-k-i-\ell}\, \frac{\alpha_{i,i,i+\ell}}{\ell\,!(j-k-i-\ell)!}\,.$$

After changing the order of summations and manipulations

$$\alpha_{i,j+1,j+1} = \frac{\alpha_{i,i,j+1}}{(j+1-i)!} + \sum_{\ell=0}^{j-i} (-1)^{j+1-i-\ell}\, \frac{\alpha_{i,i,i+\ell}}{\ell\,!(j+1-i-\ell)!} \sum_{k=0}^{j-i-\ell} (-1)^k \binom{j+1-i-\ell}{k+1}$$

which proves the induction step.

Further, let

$$\beta_{j,k}^{m\to n} = \sum_{\substack{\pi\in S\\ \pi(m)=n}} \prod_{\ell=j}^{k} P\big(\tfrac{\pi(\ell)}{\pi(\ell\ldots d)}\big)\,, \quad 0 \leqslant j \leqslant m \leqslant k \leqslant d,\ 0 \leqslant n \leqslant d,$$

and $\gamma_{j,k}$ be given by the same expression but the summation over the whole group S. Immediately

$$\gamma_{j,j} = \sum_{|I|=d-j+1} \sum_{i\in I} P(i|I)\cdot j!(d-j)! = \tfrac{(d+1)!}{(d-j+1)}\ .$$

For $k<d$

$$\gamma_{j,k+1} = \sum_{|I|=d-k} \sum_{i\in I} P(i|I) \sum_{\substack{\pi(k+1)=i\\ \pi(k+1\dots d)=I}} \prod_{\ell=j}^{k} P\big(\tfrac{\pi(\ell)}{\pi(\ell\dots k)\cup I}\big)$$

where the product does not depend on i. Omitting the constraint $\pi(k+1)=i$ in the summation on right can be compensated by dividing by $d-k$. Hence, $(d-k)\cdot\gamma_{j,k+1}=\gamma_{j,k}$. This implies

$$\gamma_{j,k} = (d+1)!\ \tfrac{(d-k)!}{(d-j+1)!}\ .$$

For $m<k$, $\beta^{m\to n}_{j,k}$ rewrites as

$$\sum_{\substack{|I|=d-m\\ I\not\ni n}} \sum_{\substack{\pi(m)=n\\ \pi(m+1\dots d)=I}} \prod_{\ell=j}^{m} P\big(\tfrac{\pi(\ell)}{\pi(\ell\dots m)\cup I}\big)\cdot \prod_{i=m+1}^{k} P\big(\tfrac{\pi(i)}{\pi(i\dots d)}\big)\ .$$

The summation over $\pi\in S$ is further restricted to fixed values $\pi(0),\dots,\pi(m)$ whereby the product over i is summed over all injections from $\langle d\rangle\setminus\langle m\rangle$ to I. This gives $\gamma_{0,k-m-1}$ in the dimension $d-m-1$, i.e. $(d-k)!$, and

$$\beta^{m\to n}_{j,k} = \frac{(d-k)!}{(d-m)!}\,\beta^{m\to n}_{j,m}\ .$$

With the above notations, the aim is to show that $d-\boldsymbol{W}_n(P)$ equals

$$\sum_{\pi\in S}\pi^{-1}(n)\prod_{k=0}^{d} P\big(\tfrac{\pi(k)}{\pi(k\dots d)}\big) = \sum_{m=0}^{d} m\,\beta^{m\to n}_{0,d} = \sum_{m=0}^{d}\tfrac{m}{(d-m)!}\,\beta^{m\to n}_{0,m}\ .$$

The sum involving $\beta^{m\to n}_{0,m}=\alpha^{m\to n}_{0,m,m}$ takes the form

$$\sum_{m=0}^{d}\frac{m}{(d-m)!}\sum_{\ell=0}^{m}(-1)^{m-\ell}\,\frac{\alpha^{d\to n}_{0,0,\ell}}{\ell\,!(m-\ell)!} = \sum_{\ell=0}^{d}\frac{\alpha^{d\to n}_{0,0,\ell}}{\ell\,!(d-\ell)!}\sum_{m=\ell}^{d}(-1)^{m-\ell}\binom{d-\ell}{m-\ell}\,m$$

using the formula for $\alpha_{0,j,j}$ through $\alpha_{0,0,\ell}$. Here, the sum involving combinatorial coefficients is equal to d for $\ell=d$, to -1 for $\ell=d-1$, and to zero otherwise. Hence, the right-hand side has only two terms

$$\frac{\alpha^{d\to n}_{0,0,d}}{(d-1)!}-\frac{\alpha^{d\to n}_{0,0,d-1}}{(d-1)!} = d-\frac{1}{(d-1)!}\sum_{\pi(d)=n} P(n|\pi(d-1),\pi(d))$$

and coincides with $d-\boldsymbol{W}_n(P)$ as desired. □

Acknowledgement. The author wants to express his gratitude to Barbara Vantaggi for several discussions about the conditional independence spaces, to Jiří Outrata for introducing him to the global inversion theorems, and to Imre Csiszár for bringing his attention to the work [20]. The figure was created in Mathematica 4.1.

References

[1] A. von Arnim, R. Schrader and Y. Wang (1996) The permutahedron of N-sparse posets. *Mathematical Programming* **75**, 1-18.

[2] A. Björner, M. Las Vergnas, B. Sturmfels, W. White and G.M. Ziegler (1993) *Oriented Matroids.* Cambridge University Press, Cambridge.

[3] W.M. Boothy (1975) *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, New York, San Francisco, London.

[4] G. Coletti and R. Scozzafava (2002) *Probabilistic Logic in a Coherent Setting.* Kluwer Academic Publishers, Dordrecht.

[5] A. Császár (1955) Sur la structure des espaces de probabilité conditionnelle. *Acta Math. Acad. Sci. Hung.* **6**, 337-361.

[6] B. de Finetti (1949) Sull'impostazione assiomatica del calcolo delle probabilità, *Annali Univ. Trieste* **19**, 3-55.

[7] B. de Finetti (1972) *Probability, Induction and Statistics.* John Wiley & Sons, London, New York, Sydney, Toronto.

[8] R.A. Horn and Ch.R. Johnson (1985) *Matrix Analysis.* Cambridge University Press, Cambridge.

[9] P.H. Krauss (1968) Representation of conditional probability measures on Boolean algebras. *Acta Math. Acad. Sci. Hung.* **19**, 229-241.

[10] M. Maes and B. Kappen (1992) On the permutahedron and the quadratic placement problem. *Philips J. Res.* **46**, 267-292.

[11] T. Parthasarathy (1983) *On Global Univalence Theorems.* Springer-Verlag. New York.

[12] M. Pouzet, K. Reuter, I. Rival and N. Zaguia (1995) A generalized permutahedron. *Algebra Universalis* **34**, 496-509.

[13] M. Radulescu and S. Radulescu (1980) Global inversion theorems and applications to differential equations. *Nonlinear Analysis, Theory, Methods and Appl.* **4**, 951-965.

[14] S. Radulescu and M. Radulescu (1989) Global univalence and global inversion theorems in Banach spaces. *Nonlinear Analysis, Theory, Methods and Appl.* **13**, 539-553.

[15] A. Rényi (1955) On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hung.* **6**, 285-335.

[16] A. Rényi (1956) On conditional probability spaces generated by a dimensionally ordered set of measures. *Theory of Probab. and Appl.* **1**, 55-64.

[17] A. Rényi (1964) Sur les espace simples des Probabilités conditionnelles. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques* **1**, 3-21.

[18] A. Rényi (1970) *Probability theory.* Akadémiai Kiadó, Budapest.

[19] A.S. Schulz (1995) The permutahedron of series-parallel posets. *Discrete Applied Math.* **57**, 85-90.

[20] N.N. Vorobjev and D.K. Faddeyev (1961) Continualization of conditinal probabilities. *Theory of Probab. and Appl.* **6** 116-118.

[21] F.F. Wu and Ch.A. Desoer (1972) Global inverse function theorem. *IEEE Trans. on Circuit Theory* **19**, 199-201.

[22] G.M. Ziegler (1995) *Lectures on Polytopes.* Springer-Verlag, New York.