Proceedings of WUPES 2009, Sept. 19–23, Liblice, Czech Republic, 176–185

CONDITIONAL PROBABILITY SPACES AND CLOSURES OF EXPONENTIAL FAMILIES

František Matúš
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
matus@utia.cas.cz

Abstract

A set of conditional probabilities is introduced by conditioning in the probability measures from an exponential family. A closure of the set is found, using previous results on the closure of another exponential family in the variational distance. The conditioning in the exponential family of all positive probabilities on a finite space is discussed and related to the permutahedra.

1 Introduction

A *conditional probability space* consists of a measurable space $(\Omega, \mathcal{A})$, nonempty set $\mathcal{B} \subseteq \mathcal{A}$ and family $\boldsymbol{P}$ of probability measures (pm's) $\boldsymbol{P}(\cdot|B)$, $B \in \mathcal{B}$, on $(\Omega, \mathcal{A})$ that satisfy $\boldsymbol{P}(B|B) = 1$ whenever $B \in \mathcal{B}$, and

$$\boldsymbol{P}(A|C) = \boldsymbol{P}(A|B) \cdot \boldsymbol{P}(B|C) \quad \text{whenever } A \in \mathcal{A},\ B, C \in \mathcal{B} \text{ and } A \subseteq B \subseteq C.$$

When viewed alternatively as a nonnegative function on $\mathcal{A} \times \mathcal{B}$, the family $\boldsymbol{P}$ is called the *conditional probability* (cp) on $(\Omega, \mathcal{A}, \mathcal{B})$ [12, 13, 8, 9, 3]. In this work, the set $\mathcal{B}$ is assumed to be finite.

Let μ be a finite nonzero measure on $(\Omega, \mathcal{A})$, $f\colon \Omega \to \mathbb{R}^d$ an $\mathcal{A}$-measurable function and $f\mu$ the image of μ under f, $f\mu(D) = \mu(f^{-1}(D))$, $D \subseteq \mathbb{R}^d$ Borel. The log-Laplace transform $\Lambda_{\mu,f}$ of $f\mu$,

$$\Lambda_{\mu,f}(\vartheta) = \ln \int_\Omega e^{\langle \vartheta, f\rangle}\, d\mu = \ln \int_{\mathbb{R}^d} e^{\langle \vartheta, x\rangle}\, f\mu(dx)\,, \qquad \vartheta \in \mathbb{R}^d\,,$$

is a convex function, finite on its nonempty domain $\mathit{dom}(\Lambda_{\mu,f})$ [14]. The full *exponential family* $\mathcal{E}_{\mu,f}$ determined by μ and f consists of the pm's $Q_{\mu,f,\vartheta}$ with

This work was supported by Grant Agency of Academy of Sciences of the Czech Republic, Grant IAA 100750603, and by Grant Agency of the Czech Republic, Grant 201/08/0539.

the μ-density $e^{\langle\vartheta,f\rangle-\Lambda_{\mu,f}(\vartheta)}$ and $\vartheta \in \mathit{dom}(\Lambda_{\mu,f})$ [1, 2]. The family is endowed here with the topology of the variational distance $|P-Q|$ of pm's P and Q.

This work proposes to study sets of cp's that are analogous to the exponential families. Let μ be a measure on $(\Omega,\mathcal{A})$ that is positive and finite on $\mathcal{B}$, thus $0<\mu(B)<+\infty$, $B\in\mathcal{B}$, and μ^B be the restriction of μ to B, $\mu^B(A)=\mu(A\cap B)$, $A\in\mathcal{A}$. If ϑ belongs to $\mathit{dom}(\Lambda_{\mu^B,f})$ for every $B\in\mathcal{B}$ then the family $\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}$ of pm's defined by

$$\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}(A|B)=Q_{\mu^B,f,\vartheta}(A)\,,\qquad A\in\mathcal{A}\,,B\in\mathcal{B}\,,$$

is a cp on $(\Omega,\mathcal{A},\mathcal{B})$ by Remark 2.1. The main object of interest here is the set

$$\mathfrak{E}^{\mathcal{B}}_{\mu,f}=\left\{\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}\colon \vartheta\in\textstyle\bigcap_{B\in\mathcal{B}}\mathit{dom}(\Lambda_{\mu^B,f})\right\}.$$

When $\mathcal{B}=\{\Omega\}$ this set is effectively the same as $\mathcal{E}_{\mu,f}$. It contains a cp $\boldsymbol{P}$ if and only if $\boldsymbol{P}(\cdot|\Omega)=Q_{\mu,f,\vartheta}$ for some $\vartheta\in\mathit{dom}(\Lambda_{\mu,f})$. Sets of cp's are endowed with the topology of the sum distance $\sum_{B\in\mathcal{B}}|\boldsymbol{P}(\cdot|B)-\boldsymbol{Q}(\cdot|B)|$ of cp's $\boldsymbol{P}$ and $\boldsymbol{Q}$.

Basic observations on the sets $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ are collected in Section 2. The main idea is to transform a cp $\boldsymbol{P}$ to the product of $\boldsymbol{P}(\cdot|B)$ over $B\in\mathcal{B}$, denoted by $\boldsymbol{\Pi P}$. The image $\boldsymbol{\Pi}\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ is recognized to be a full exponential family, see Lemma 2.3. This family is then reduced in two steps, see Lemma 2.4. A one-to-one canonical parametrization of the set $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ is described in Remark 2.7. Another parametrization follows from Lemma 2.9.

The closure of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ is found in Theorem 3.3 applying the results of [6]. Under some assumptions on μ and f it is homeomorphic to a convex set, see Corollary 3.5.

Section 4 presents the special case of a finite Ω and the family $\mathcal{E}_{\mu,f}$ of all positive pm's on Ω. Relations to the algebraic approach of [11] are discussed. If $\mathcal{B}$ is the family of all nonempty subsets of Ω then the closure of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ exhausts all cp's and can be parameterized by the points of a permutahedron of the dimension $|\Omega|-1$, as found earlier in [10].

2 Basic observations

If a measure μ on $(\Omega,\mathcal{A})$ is positive and finite on $\mathcal{B}$ then the mapping

$$(A|B)\mapsto\frac{\mu(A\cap B)}{\mu(B)}=\frac{\mu^B(A)}{\mu^B(\Omega)}\,,\qquad A\in\mathcal{A}\,,\;B\in\mathcal{B}\,,$$

gives rise to a cp. For a necessary and sufficient condition on a cp to be generated from a measure as above see [4, (6.3), p. 351].

Remark 2.1. The assumption that μ is finite on $\mathcal{B}$ is equivalent to the finiteness of $\mu(\bigcup\mathcal{B})$ where $\bigcup\mathcal{B}=\bigcup_{B\in\mathcal{B}}B$, using that $\mathcal{B}$ is finite. If ν denotes the restriction of μ to $\bigcup\mathcal{B}$ then $\mathit{dom}(\Lambda_{\nu,f})$ is equal to the intersection of $\mathit{dom}(\Lambda_{\mu^B,f})$ over $B\in\mathcal{B}$. For ϑ in this domain

$$\frac{Q_{\nu,f,\vartheta}(A\cap B)}{Q_{\nu,f,\vartheta}(B)}=\int_A e^{\langle\vartheta,f\rangle}\,d\mu^B\Big/\int_\Omega e^{\langle\vartheta,f\rangle}\,d\mu^B=Q_{\mu^B,f,\vartheta}(A)\,,\qquad A\in\mathcal{A}\,,$$

thus $\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}$ is the cp on $(\Omega,\mathcal{A},\mathcal{B})$ generated from $Q_{\nu,f,\vartheta}$. Hence, the set $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ can be constructed alternatively from $\mathcal{E}_{\nu,f}$ by conditioning to the sets $B\in\mathcal{B}$.

Lemma 2.2. *The mapping* $\mathbf{\Pi}$ *is a homeomorphism of the family of cp's into the family of product pm's on* $(\Omega^{\mathcal{B}},\mathcal{A}^{\mathcal{B}})$.

Proof (sketch). The sum distance between cp's $\boldsymbol{P}$, $\boldsymbol{Q}$ majorizes the variational distance between $\mathbf{\Pi}\boldsymbol{P}$ and $\mathbf{\Pi}\boldsymbol{Q}$. The variational distance between two products of pm's majorizes the variational distance between any two marginal pm's. $\square$

For a finite measure μ on $(\Omega,\mathcal{A})$, let $\mu_{\mathcal{B}}$ be product of the restrictions μ^B over $B\in\mathcal{B}$. For a function $f\colon\Omega\to\mathbb{R}^d$ let $f_{\mathcal{B}}$ map an element $\omega_{\mathcal{B}}=(\omega_B)_{B\in\mathcal{B}}$ of $\Omega^{\mathcal{B}}$ to $(f(\omega_B))_{B\in\mathcal{B}}$, an element of $(\mathbb{R}^d)^{\mathcal{B}}$. The function f is always assumed to be $\mathcal{A}$-measurable. Let Σ map $(x_B)_{B\in\mathcal{B}}\in(\mathbb{R}^d)^{\mathcal{B}}$ to $\sum_{B\in\mathcal{B}}x_B\in\mathbb{R}^d$.

Lemma 2.3. *If* μ *is positive and finite on* $\mathcal{B}$ *then*
(i) $\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}=\sum_{B\in\mathcal{B}}\Lambda_{\mu^B,f}$
(ii) $\mathbf{\Pi}\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}=Q_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}},\vartheta}$ *for* $\vartheta\in\mathsf{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$
(iii) $\mathbf{\Pi}$ *restricts to a homeomorphism between* $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ *and* $\mathcal{E}_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}$.

Proof. For $\vartheta\in\mathbb{R}^d$

$$\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}(\vartheta)=\ln\int_{\Omega^{\mathcal{B}}}e^{\langle\vartheta,\Sigma f_{\mathcal{B}}\rangle}\,d\mu_{\mathcal{B}}=\ln\int_{\Omega^{\mathcal{B}}}\prod_{B\in\mathcal{B}}e^{\langle\vartheta,f(\omega_B)\rangle}\,\mu_{\mathcal{B}}(d\omega_{\mathcal{B}})$$

using $\langle\vartheta,\Sigma f_{\mathcal{B}}(\omega_{\mathcal{B}})\rangle=\sum_{B\in\mathcal{B}}\langle\vartheta,f(\omega_B)\rangle$. Hence,

$$\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}(\vartheta)=\ln\prod_{B\in\mathcal{B}}\int_{\Omega}e^{\langle\vartheta,f(\omega)\rangle}\,\mu^B(d\omega)=\sum_{B\in\mathcal{B}}\Lambda_{\mu^B,f}(\vartheta)$$

which proves *(i)*. It follows that $\mathsf{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$ is the intersection of $\mathsf{dom}(\Lambda_{\mu^B,f})$ over $B\in\mathcal{B}$. For ϑ in the domain the product pm $\mathbf{\Pi}\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}$ is absolutely continuous w.r.t. $\mu_{\mathcal{B}}$ and by *(i)* has the density

$$\begin{aligned}\prod_{B\in\mathcal{B}}d\boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}(\cdot|B)\Big/d\mu^B\ (\omega_{\mathcal{B}})&=\prod_{B\in\mathcal{B}}\exp\left[\langle\vartheta,f(\omega_B)\rangle-\Lambda_{\mu^B,f}(\vartheta)\right]\\&=\exp[\langle\vartheta,\Sigma f_{\mathcal{B}}(\omega_{\mathcal{B}})\rangle-\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}(\vartheta)]=dQ_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}},\vartheta}\big/d\mu_{\mathcal{B}}\ (\omega_{\mathcal{B}})\,,\end{aligned}$$

thus *(ii)* holds. Then *(iii)* follows by Lemma 2.2. $\square$

Lemma 2.4. *If* μ *is positive and finite on* $\mathcal{B}$ *then*
(i) $\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}=\Lambda_{f_{\mathcal{B}}\mu_{\mathcal{B}},\Sigma}=\Lambda_{\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}},\mathsf{id}}$
where id *denotes the identity mapping on* $\mathbb{R}^d$*, and for* $\vartheta\in\mathsf{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$
(ii) $f_{\mathcal{B}}Q_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}},\vartheta}=Q_{f_{\mathcal{B}}\mu_{\mathcal{B}},\Sigma,\vartheta}$
(iii) $\Sigma f_{\mathcal{B}}Q_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}},\vartheta}=\Sigma Q_{f_{\mathcal{B}}\mu_{\mathcal{B}},\Sigma,\vartheta}=Q_{\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}},\mathsf{id},\vartheta}$.

A proof is standard and omitted.

The *convex core* $\mathsf{cc}(\nu)$ of a finite Borel measure ν on $\mathbb{R}^d$ is intersection of the convex Borel sets $D\subseteq\mathbb{R}^d$ with $\nu(\mathbb{R}^d\setminus D)=0$ [5]. Let $\mathsf{ri}(\nu)$ denote the relative interior of $\mathsf{cc}(\nu)$.

Lemma 2.5. *If μ is finite on $\mathcal{B}$ then*
(i) $\mathit{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}}) = \prod_{B\in\mathcal{B}} \mathit{cc}(f\mu^B)$
(ii) $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}}) = \Sigma \mathit{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}}) = \sum_{B\in\mathcal{B}} \mathit{cc}(f\mu^B)$.

Proof. Since $f_{\mathcal{B}}\mu_{\mathcal{B}}$ is the product of the measures $f\mu^B$ over $B\in\mathcal{B}$ the first equality follows from [5, Lemma 7]. The Σ-image of a product measure is the convolution of marginals. Hence, the second assertion is a consequence of [5, Corollary 8]. □

Corollary 2.6. *Lemma 2.5 remains valid when cc is replaced by ri.*

For a convex set $D\subseteq\mathbb{R}^d$ let $\mathit{lin}(D)$ denote the linear space generated by the differences $x-y$ with $x,y\in D$ and π_D the orthogonal projection onto $\mathit{lin}(D)$. In the case $D=\mathit{cc}(\nu)$ the abbreviations $\mathit{lin}(\nu)$ and π_ν are used.

Remark 2.7. If a measure μ on $(\Omega,\mathcal{A})$ is nonzero and finite, and $f\colon\Omega\to\mathbb{R}^d$ then $\vartheta\in\mathit{dom}(\Lambda_{\mu,f})$ and $\pi_{f\mu}(\vartheta-\theta)=0$ imply $\theta\in\mathit{dom}(\Lambda_{\mu,f})$. The exponential family $\mathcal{E}_{\mu,f}$ is bijectively parameterized by $\pi_{f\mu}(\mathit{dom}(\Lambda_{\mu,f}))$. If follows on account of Lemma 2.3 that $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ is bijectively parameterized by $\pi_{\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}}}(\mathit{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}))$. Here, the projection is onto $\mathit{lin}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$ which is the sum of $\mathit{lin}(f\mu^B)$ over $B\in\mathcal{B}$, by Lemma 2.5 *(ii)*.

Remark 2.8. The log-Laplace transform $\Lambda_{\mu,f}$ is differentiable at any ϑ from the interior of its domain and $\nabla\Lambda_{\mu,f}(\vartheta)=\int_\Omega f\,dQ_{\mu,f,\vartheta}$ [1, 2]. If the domain is open then $\nabla\Lambda_{\mu,f}$ gives rise to a diffeomorphism between the relatively open sets $\pi_{f\mu}(\mathit{dom}(\Lambda_{\mu,f}))$ and $\mathit{ri}(f\mu)$. Thus, the mapping $P\mapsto\int_\Omega f\,dP$ is defined for every $P\in\mathcal{E}_{\mu,f}$, and it is a homeomorphism between $\mathcal{E}_{\mu,f}$ and $\mathit{ri}(f\mu)$, see also [6, Corollary 1].

Let $\mathbf{M}_f$ denote the composition of two mappings

$$\boldsymbol{P}\mapsto\mathbf{\Pi}\boldsymbol{P}\mapsto\int_{\Omega^{\mathcal{B}}}\Sigma f_{\mathcal{B}}\,d\,\mathbf{\Pi}\boldsymbol{P}$$

defined at any cp $\boldsymbol{P}$ such that the integral exists. Rewriting the integral to

$$\int_{\Omega^{\mathcal{B}}}\sum_{B\in\mathcal{B}} f(\omega_B)\cdot\prod_{B\in\mathcal{B}}\boldsymbol{P}(d\omega_B|B)$$

the existence is equivalent to $\boldsymbol{P}(\cdot|B)$-integrability of f for $B\in\mathcal{B}$, in which case

$$\mathbf{M}_f\boldsymbol{P}=\sum_{B\in\mathcal{B}}\int_\Omega f(\omega)\,\boldsymbol{P}(d\omega|B)\,.$$

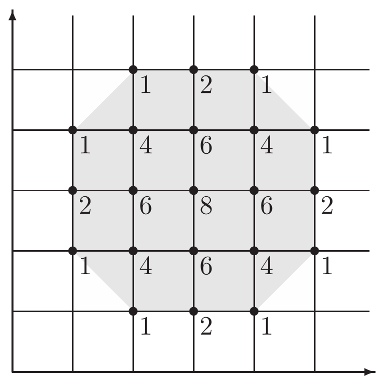
Lemma 2.9. *If $\mathit{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$ is open then $\mathbf{M}_f$ restricts to a homeomorphism between $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ and $\mathit{ri}(\Sigma f_{\mathcal{B}}\,\mu_{\mathcal{B}})=\sum_{B\in\mathcal{B}}\mathit{ri}(f\mu^B)$.*

Proof. The restriction is a composition of two homeomorphisms. The first one comes from Lemma 2.3 *(iii)*, between $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ and $\mathcal{E}_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}$. The second one makes homeomorphic $\mathcal{E}_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}$ and $\mathit{ri}(\Sigma f_{\mathcal{B}}\,\mu_{\mathcal{B}})$, by Remark 2.8. It remains to refer to Corollary 2.6. □

Example 2.10. Let $\Omega = \{0,1\}^2$, $\mathcal{A}$ be the algebra of all subsets of Ω and $\mathcal{B} = \binom{\Omega}{2}$ consist of all two-element subsets of Ω. Let μ be the counting measure on Ω and f the embedding of Ω to $\mathbb{R}^2$. The family $\mathcal{E}_{\mu,f}$ consists of all positive product pm's on Ω and the set $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ of the cp's that are generated from these pm's, see Remark 2.1. Denoting by δ_x the Borel pm on $\mathbb{R}^2$ that is supported by $x \in \mathbb{R}^2$, the $f_{\mathcal{B}}$-image of $\mu_{\mathcal{B}}$ is the product

$$[\delta_{(0,0)}+\delta_{(1,0)}]\times[\delta_{(0,0)}+\delta_{(0,1)}]\times[\delta_{(0,0)}+\delta_{(1,1)}]\times[\delta_{(1,0)}+\delta_{(0,1)}]\times[\delta_{(1,0)}+\delta_{(1,1)}]\times[\delta_{(0,1)}+\delta_{(1,1)}]\,.$$

Then, $\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}}$ is the convolution of the six measures. It is equal to the linear combination of δ_x's where x runs over the points of the configuration below and the coefficients in the combination correspond to the labels of the points.



The shaded hexagon is the convex core of $\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}}$. Lemma 2.5 expresses the hexagon as the sum of the edges and diagonals of the unit square. Further,

$$\begin{aligned}
\mathbf{M}_f \boldsymbol{P} &= \textstyle\sum_{B\in\mathcal{B}} \sum_{\omega\in\Omega} f(\omega)\, \boldsymbol{P}(\omega|B) \\
&= (1,0)[\boldsymbol{P}(10|00,10) + \boldsymbol{P}(10|10,01) + \boldsymbol{P}(10|10,11)] \\
&\quad + (0,1)[\boldsymbol{P}(01|00,01) + \boldsymbol{P}(01|10,01) + \boldsymbol{P}(01|01,11)] \\
&\quad + (1,1)[\boldsymbol{P}(11|00,11) + \boldsymbol{P}(11|10,11) + \boldsymbol{P}(11|01,11)]
\end{aligned}$$

where e.g. $\boldsymbol{P}(10|00,10)$ is an abbreviation for $\boldsymbol{P}\big(\{(1,0)\}|\{(0,0),(1,0)\}\big)$. By Lemma 2.9, the mapping $\mathbf{M}_f$ restricts to a homeomorphism between $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ and the interior of the hexagon.

3 Closures of the families $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$

Given a convex set D in a Euclidean space, its nonempty convex subset F is a *face* if each segment contained in D with an interior point in F is contained in F.

Lemma 3.1. *If μ is finite on $\mathcal{B}$ and F is a face of $\mathsf{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$ then*
(i) $F_\Sigma = \Sigma^{-1}(F) \cap \mathsf{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}})$ is a face of $\mathsf{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}})$
(ii) $F_\Sigma = \prod_{B\in\mathcal{B}} F_{\Sigma,B}$ where $F_{\Sigma,B}$ is a unique face of $\mathsf{cc}(f\mu^B)$
(iii) $\Sigma F_\Sigma = F$.

Proof. The assertions follow from Lemma 2.5 and basic convex geometry. □

Let $\mu_{\mathcal{B},F} = \prod_{B\in\mathcal{B}} \mu^{B,F}$ where $\mu^{B,F}$ is the restriction of μ to $B \cap f^{-1}(\mathit{cl}(F_{\Sigma,B}))$.

Lemma 3.2. *If μ is finite on $\mathcal{B}$ and F is a face of $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$ then $\mu_{\mathcal{B},F}$ is nonzero and finite and $\Sigma f_{\mathcal{B}}\mu_{\mathcal{B},F} = (\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})^{\mathit{cl}(F)}$.*

Proof (sketch). Since F is a face, thus a nonempty set, every $F_{\Sigma,B}$ is a face of $\mathit{cc}(f\mu^B)$ by Lemma 3.1. Therefore $f\mu^B(\mathit{cl}(F_{\Sigma,B})) = \mu^{B,F}(\Omega)$ is positive by [5, Corolary 3]. Thus, $\mu_{\mathcal{B},F}$ is nonzero. Since μ is finite on $\mathcal{B}$ every $f\mu^B$ is finite, and the finiteness of $\mu_{\mathcal{B},F}$ follows.

The equality is a consequence of $f_{\mathcal{B}}\mu_{\mathcal{B},F} = (f_{\mathcal{B}}\mu_{\mathcal{B}})^{\Sigma^{-1}(\mathit{cl}(F))}$. Since $f_{\mathcal{B}}\mu_{\mathcal{B},F}$ is the restriction of $f_{\mathcal{B}}\mu_{\mathcal{B}}$ to $\prod_{B\in\mathcal{B}} \mathit{cl}(F_{\Sigma,B}) = \mathit{cl}(F_\Sigma)$ the aim is to prove that $f_{\mathcal{B}}\mu_{\mathcal{B}}(\Sigma^{-1}(\mathit{cl}(F)) \setminus \mathit{cl}(F_\Sigma)) = 0$, using that the two sets are in inclusion.

If $F_\Sigma = \mathit{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}})$ then $\mathit{cl}(F_\Sigma)$ has the complement of $f_{\mathcal{B}}\mu_{\mathcal{B}}$-measure zero by [5, Lemma 1]. Otherwise, F is not equal to $D = \mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$. Assume first that F is exposed, thus a nontrivial supporting hyperplane H to D exists such that $F = H \cap D$. Then, $\Sigma^{-1}(H)$ is a supporting hyperplane of $\mathit{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}})$ and $\Sigma^{-1}(H) \cap \mathit{cc}(f_{\mathcal{B}}\mu_{\mathcal{B}}) = F_\Sigma$. By [6, Lemma 1], $f_{\mathcal{B}}\mu_{\mathcal{B}}(\Sigma^{-1}(H) \setminus \mathit{cl}(F_\Sigma)) = 0$ and the equality holds. If F is not exposed then it can be approached by a chain of exposed faces and the equation obtains from the corresponding equations in the chain. Details are omitted. □

Where D and Ξ are nonempty convex subsets in a Euclidean space, the concept of *Ξ-accessible* face of D was introduced in [6, Subsection 2.5]. The definition is rather technical and not repeated here, using later only the simple facts that D is always a Ξ-accessible face of D and every face is $\mathbb{R}^d$-accessible.

For a face F of $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$ the family $\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}$ of pm's given by

$$\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}(\cdot|B) = Q_{\mu^{B,F},f,\vartheta}\,, \qquad B \in \mathcal{B}\,,$$

is a cp on $(\Omega, \mathcal{A}, \mathcal{B})$ by Remark 2.1 where $\{B \cap f^{-1}(\mathit{cl}(F_{\Sigma,B}))\colon B \in \mathcal{B}\}$ plays the role of $\mathcal{B}$. In particular, if F equals the convex core then $\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta} = \boldsymbol{Q}^{\mathcal{B}}_{\mu,f,\vartheta}$.

Theorem 3.3. *If μ is positive and finite on $\mathcal{B}$ then the closure of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ is the union of the families*

$$\mathfrak{E}^{\mathcal{B},F}_{\mu,f} = \big\{\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}\colon\ \vartheta \in \mathit{cl}(\pi_F(\mathit{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}))) \cap \mathit{dom}(\Lambda_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}}})\big\}$$

over the $\mathit{dom}(\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$-accessible faces F of $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$.

Proof. By assumption $\nu = \Sigma f_{\mathcal{B}}\mu_{\mathcal{B}}$ is nonzero and finite, thus [6, Theorem 2] applies to the full standard exponential family $\mathcal{E}_{\nu,\mathit{id}}$ with $\Xi = \mathit{dom}(\Lambda_{\nu,\mathit{id}})$ and implies

$$\mathit{cl}(\mathcal{E}_{\nu,\mathit{id}}) = \bigcup \big\{Q_{\nu_F,\mathit{id},\vartheta}\colon\ \vartheta \in \mathit{cl}(\pi_F(\Xi)) \cap \mathit{dom}(\Lambda_{\nu_F,\mathit{id}})\big\}$$

where the union is over the Ξ-accessible faces F of $\mathit{cc}(\nu)$ and ν_F denotes the restriction of ν to $\mathit{cl}(F)$. By Lemma 2.4 *(i)*, $\Lambda_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}$ equals $\Lambda_{\nu,\mathit{id}}$ so that the above union is over the same family of faces as in the assertion of the theorem.

Lemma 3.2 implies that ν_F is the $\Sigma f_{\mathcal{B}}$-image of the nonzero and finite product measure $\mu_{\mathcal{B},F}$. Hence, $\Lambda_{\nu_F,\mathit{id}}$ equals $\Lambda_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}}}$ by Lemma 2.4 *(i)*. It follows that in the above union ϑ ranges over the same parameter set as in the assertion of the theorem. Since $Q_{\nu_F,\mathit{id},\vartheta}$ is the $\Sigma f_{\mathcal{B}}$-image of $Q_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}},\vartheta}$, it is possible to conclude by Lemma 2.4 *(iii)* that

$$\mathit{cl}(\mathcal{E}_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}}) = \bigcup \left\{ Q_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}},\vartheta} \colon\ \vartheta \in \mathit{cl}(\pi_F(\mathit{dom}(\Xi))) \cap \mathit{dom}(\Lambda_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}}}) \right\}.$$

On account of Lemma 2.2, it suffices to prove that $Q_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}},\vartheta}$ equals $\mathbf{\Pi}\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}$ but this follows from Lemma 2.3 *(ii)*. □

Corollary 3.4. *If $\Lambda_{\mu^B,f}$ is everywhere finite for all $B \in \mathcal{B}$ then*

$$\mathfrak{E}^{\mathcal{B},F}_{\mu,f} = \left\{ \boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta} \colon\ \vartheta \in \mathit{lin}(F) \right\} \quad \text{and} \quad \mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f}) = \bigcup \mathfrak{E}^{\mathcal{B},F}_{\mu,f}$$

where the union is over all faces F of $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$. The mapping $\mathbf{M}_f$ restricts to a bijection between $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ and $\sum_{B\in\mathcal{B}} \mathit{cc}(f\mu^B)$.

Proof. The assumption implies that $\mathit{dom}(\Lambda_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}}}) = \mathbb{R}^d$ for all faces F and that all faces are accessible. To prove the second assertion, Lemma 2.3 *(iii)* is applied to $\mu_{\mathcal{B},F}$ in the role of $\mu_{\mathcal{B}}$. Then, $\mathbf{\Pi}$ restricts to a bijection between $\mathfrak{E}^{\mathcal{B},F}_{\mu,f}$ and $\mathcal{E}_{\mu_{\mathcal{B},F},\Sigma f_{\mathcal{B}}}$. By Remark 2.8, $\mathbf{M}_f$ maps $\mathfrak{E}^{\mathcal{B},F}_{\mu,f}$ bijectively onto $\mathit{ri}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B},F})$. This set equals $\mathit{ri}(F)$ by Lemma 3.2. It follows from Theorem 3.3 that $\mathbf{M}_f$ maps $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ bijectively onto the union of $\mathit{ri}(F)$. The union is equal to $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$, and thus to the sum of $\mathit{cc}(f\mu^B)$ by Lemma 2.5 *(ii)*. □

Corollary 3.5. *If $\sum_{B\in\mathcal{B}} \mathit{cc}(f\mu^B)$ is bounded and locally simplicial then $\mathbf{M}_f$ restricts to a homeomorphism between $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ and this sum.*

Proof. The boundedness implies that the mapping $P \mapsto \int_{\Omega^{\mathcal{B}}} \Sigma f_{\mathcal{B}}\, dP$ is continuous on $\mathit{cl}(\mathcal{E}_{\mu_{\mathcal{B}},\Sigma f_{\mathcal{B}}})$. Its inverse is continuous due to the second assumption, see [7, Remark 5.9]. By Lemma 2.2, the assertion follows. □

Example 3.6. Let $(\Omega, \mathcal{A}, \mathcal{B})$, μ and f be as in Example 2.10. The segment $F = \{(t,1)\colon 2 \leqslant t \leqslant 4\}$ is a face of the hexagon $\mathit{cc}(\Sigma f_{\mathcal{B}}\mu_{\mathcal{B}})$. Then F_Σ is the square

$$\{((t,0),(0,0),(0,0),(1,0),(1,0),(r,1))\colon 0 \leqslant t, r \leqslant 1\}$$

and $\Sigma f_{\mathcal{B}}\mu_{\mathcal{B},F}$ is the convolution

$$[\delta_{(0,0)}+\delta_{(1,0)}] * \delta_{(0,0)} * \delta_{(0,0)} * \delta_{(1,0)} * \delta_{(1,0)} * [\delta_{(0,1)}+\delta_{(1,1)}] = \delta_{(2,1)} + 2\delta_{(3,1)} + \delta_{(4,1)}\,.$$

The cp $\boldsymbol{P} = \boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}(\cdot|B)$, $\vartheta = (t,0) \in \mathit{lin}(F)$, from $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ is given by

$$\boldsymbol{P}(10|00,10) = \boldsymbol{P}(11|01,11) = \tfrac{e^t}{1+e^t}$$
$$\boldsymbol{P}(00|00,01) = \boldsymbol{P}(00|00,11) = \boldsymbol{P}(10|10,01) = \boldsymbol{P}(10|10,11) = 1\,.$$

The closure of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ consists of the family itself and 16 families corresponding to all vertices and edges of the hexagon.

4 Discussion

In this section, the space Ω is finite, $\mathcal{A}$ is the algebra 2^Ω of all subsets of Ω, μ is the counting measure on Ω and f maps Ω to $\mathbb{R}^\Omega$ such that $f(\omega)$ is the vector with the ω-th coordinate equal to 1 and the remaining ones to 0. The family $\mathcal{E}_{\mu,f}$ consists of all pm's P on Ω that are positive in the sense $P(\omega) > 0$, $\omega \in \Omega$.

For $B \subseteq \Omega$ the measure $f\mu^B$ is concentrated on the linearly independent set $f(\Omega)$, and hence $cc(f\mu^B)$ is the simplex Δ_B spanned by the set.

Example 4.1. If $\Omega = \{0, 1, \ldots, m\}$, $m \geqslant 1$, and $\mathcal{B} = \binom{\Omega}{2}$ then $\sum_{B\in\mathcal{B}} \Delta_B$ is the sum of all segments with the endpoints in $f(\Omega)$. This is the polytope known under the name *permutahedron* [16], equivalently defined as the convex hull of all the points $(\rho(0), \rho(1), \ldots, \rho(m))$ where ρ is any permutation of Ω. Assume $A_1, \ldots, A_k$ is an ordered partition of Ω such that $\omega < \omega'$ for $\omega \in A_i$, $\omega' \in A_j$ and $1 \leqslant i < j \leqslant k$. The convex hull of the points $(\rho(m), \ldots, \rho(1), \rho(0))$ where ρ is any permutation of Ω that satisfies $\rho(A_i) = A_i$, $1 \leqslant i \leqslant k$, is a face F of the permutahedron. It is the sum of the faces

$$F_{\Sigma,B} = \begin{cases} \Delta_B\,, & \text{if } B \subseteq A_i \text{ for some } 1 \leqslant i \leqslant k\,,\\ \{f(\omega)\}\,, & \text{otherwise}\,, \end{cases}$$

over $B = \{\omega, \omega'\} \in \binom{\Omega}{2}$ with $\omega < \omega'$. Hence, for $\vartheta = (\vartheta_\omega)_{\omega\in\Omega} \in \mathbb{R}^\Omega$

$$\boldsymbol{Q}^{\mathcal{B},F}_{\mu,f,\vartheta}(\omega|B) = \begin{cases} e^{\vartheta_\omega}/[e^{\vartheta_\omega} + e^{\vartheta_{\omega'}}]\,, & \text{if } B \subseteq A_i \text{ for some } 1 \leqslant i \leqslant k\,,\\ 1\,, & \text{otherwise}\,. \end{cases}$$

Each cp of $cl(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ has this form up to a permutation.

Remark 4.2. Let $\Omega^*_{\mathcal{B}}$ denote the set of ordered couples $(\omega|B)$ with $\omega \in B \in \mathcal{B}$. For $\mathcal{B} \subseteq \mathcal{A}$ nonempty, a cp $\boldsymbol{P}$ on $(\Omega, \mathcal{A}, \mathcal{B})$ is uniquely given by its nonnegative values $\boldsymbol{P}(\omega|B)$, $(\omega|B) \in \Omega^*_{\mathcal{B}}$. They are constrained by $\sum_{\omega\in B} \boldsymbol{P}(\omega|B) = 1$, $B \in \mathcal{B}$, and

$$\boldsymbol{P}(\omega|C) = \boldsymbol{P}(\omega|B) \cdot \textstyle\sum_{\omega'\in B} \boldsymbol{P}(\omega'|C)\,, \qquad \omega \in B \subseteq C \text{ and } B, C \in \mathcal{B}.$$

By Remark 2.1, $\boldsymbol{P} \in \mathfrak{E}^{\mathcal{B}}_{\mu,f}$ if and only if there exists a positive measure on Ω that generates $\boldsymbol{P}$. It follows from the general results of [4, (6.3), p. 351] that this takes place if and only if all $\boldsymbol{P}(\omega|B)$ are positive and $\boldsymbol{P}$ satisfies the polynomial constraints

$$\textstyle\prod_{i=1}^n \boldsymbol{P}(A_i|B_i) = \prod_{i=1}^n \boldsymbol{P}(A_i|B_{i+1})$$

for $n \geqslant 1$, $B_1, \ldots, B_{n+1} \in \mathcal{B}$ with $B_1 = B_{n+1}$ and $A_i \subseteq B_i \cap B_{i+1}$, $1 \leqslant i \leqslant n$. Here, it can be assumed equivalently that all A_i's are singletons $\{\omega_i\}$. Such a constraint, will be referred to as *Császár* one.

Remark 4.3. It was observed in [11] that Császár constraints correspond to cycles in the bipartite graph $\mathcal{G}_{\mathcal{B}}$ between Ω and $\mathcal{B}$ with the edge from each $B \in \mathcal{B}$ to each of its elements ω. Since the incidence matrix of any bipartite graph is unimodular [15, 19.2] Császár constraints play a distinguished role in the toric ideal induced by the incidence matrix of $\mathcal{G}_{\mathcal{B}}$, see [11, Proposition 3.4].

Lemma 4.4. *A cp $\boldsymbol{P}$ on $(\Omega, 2^\Omega, \mathcal{B})$ satisfies Császár constraints if and only if it extends to a cp $\boldsymbol{P}'$ on $(\Omega, 2^\Omega, 2^\Omega \setminus \{\emptyset\})$, in the sense $\boldsymbol{P}'(\cdot|B) = \boldsymbol{P}(\cdot|B)$, $B \in \mathcal{B}$.*

A proof is omitted; it is based on [4, (5.9), p. 349] that establishes a connection between the constraints and the generation of a cp from a family of measures ordered according to dimension.

Corollary 4.5. *The closure of $\mathfrak{E}^{\mathcal{B}}_{\mu,f}$ consists of all cp's on $(\Omega, 2^\Omega, \mathcal{B})$ that satisfy Császár constraints.*

Example 4.6. In the situation of Example 4.1 with $m \geqslant 2$, for $B = \{\omega, \omega'\}$ with $\omega < \omega'$ let $\boldsymbol{P}(\omega|B) = 1$ and $\boldsymbol{P}(\omega'|B) = 0$ with the exception $\boldsymbol{P}(0|\{0, m\}) = 0$ and $\boldsymbol{P}(m|\{0, m\}) = 1$. Then $\boldsymbol{P}$ is a cp on $(\Omega, \mathcal{A}, \binom{\Omega}{2})$ that violates the Császár constraint with $n = m+1$ and $B_1 = \{0, 1\}, \ldots, B_m = \{m-1, m\}, B_n = \{0, m\}$. Thus, $\boldsymbol{P}$ does not belong to $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$.

Remark 4.7. It is not difficult to see that for $\mathcal{B} = \binom{\Omega}{2}$ every $\boldsymbol{P} \in \mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ extends to a cp on $(\Omega, 2^\Omega, 2^\Omega \setminus \{\emptyset\})$ uniquely, see the proof of [10, Lemma 4]. In general, it is only a minor technicality not to admit the singletons of Ω in the sets $\mathcal{B}$.

Remark 4.8. In [11], Császár constraints are interpreted as polynomials and are used to define a multiprojective toric variety. The variety lives in the product of the projective spaces of $\mathbb{C}^B$ over $B \in \mathcal{B}$. A point z of this variety is a $\mathcal{B}$-tuple of points z_B with the projective coordinates $z_{(\omega|B)}$, $\omega \in B$. By [11, Theorem 4.3], the mapping

$$z \mapsto \sum_{B \in \mathcal{B}} \sum_{\omega \in B} f(\omega) \, \frac{|z_{(\omega|B)}|}{\sum_{\omega' \in B} |z_{(\omega'|B)}|}$$

is a bijection between the nonnegative part of the variety and $\sum_{B \in \mathcal{B}} \Delta_B$. (Note that in the original definition of this mapping, denoted by ν, the column $a_{.i|I}$ must be replaced by its projection to the V-coordinates).

The mapping $\mathbf{M}_f$ moves a cp $\boldsymbol{P}$ on $(\Omega, 2^\Omega, \mathcal{B})$ linearly as

$$\boldsymbol{P} \mapsto \sum_{B \in \mathcal{B}} \sum_{\omega \in B} f(\omega) \, \boldsymbol{P}(\omega|B) = \big(\sum_{B \in \mathcal{B}} \boldsymbol{P}(\omega|B) \big)_{\omega \in \Omega}.$$

By Corollary 3.5, $\mathbf{M}_f$ restricts to the homeomorphism between $\mathit{cl}(\mathfrak{E}^{\mathcal{B}}_{\mu,f})$ and the sum of Δ_B over $B \in \mathcal{B}$. On account of Corollary 4.5, the closure corresponds to the nonnegative part of the variety from Remark 4.8. Hence, in the setting of this section, the assertion of Corollary 3.5 is equivalent to the statement of [11, Theorem 4.3].

By Corollary 3.5 and Remark 4.7, the family of cp's on $(\Omega, 2^\Omega, \mathcal{B})$ with $\mathcal{B} = \{B \subseteq \Omega \colon |B| \geqslant 2\}$ is homeomorphic to the permutahedron of Example 4.1 via

$$\boldsymbol{P} \mapsto \big(\sum_{\omega' \in \Omega \setminus \{\omega\}} \boldsymbol{P}(\omega|\{\omega, \omega'\}) \big)_{\omega \in \Omega}$$

which is the content of [10, Theorem 1].

References

[1] Barndorff-Nielsen, O., *Information and Exponential Families in Statistical Theory.* Wiley, New York, 1978.

[2] Chentsov, N.N., *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs, AMS, Providence – Rhode Island, 1982 (Russian original: Nauka, Moscow, 1972).

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

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

[5] Csiszár, I. and Matúš, F., Convex cores of measures on $\mathbb{R}^d$. *Studia Sci. Math. Hungar.* **38** (2001) 177–190.

[6] Csiszár, I. and Matúš, F., Closures of exponential families. *Annals of Probability* **33** (2005) 582–600.

[7] Csiszár, I. and Matúš, F., Generalized maximum likelihood estimates for exponential families. *Probab. Th. and Related Fields* **141** (2008) 213–246.

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

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

[10] Matúš, F., Conditional probabilities and permutahedron. *Annales de l'Institut H. Poincaré, Probabilités et Statistiques* **39** (2003) 687–701.

[11] Morton, J., Relations among conditional probabilities. August 2008, arXiv: 0808.1149v1.

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

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

[14] Rockafellar, R.T., *Convex Analysis.* Princeton Univ. Press, Princeton 1970.

[15] Schrijver, A., *Theory of Integer and Linear Programming.* John Wiley & Sons, New York, 1998.

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