*The Annals of Probability* **33** (2005) 582–600

# Closures of exponential families

*Imre Csiszár* and František Matúš†*

**Abstract.** The variation distance closure of an exponential family with a convex set of canonical parameters is described, not assuming any regularity conditions. The tools are the concepts of convex core of a measure and extension of an exponential family the authors introduced previously, and a new concept of accessible faces of a convex set. Two other closures related to the information divergence are also characterized.

## 1. Introduction

Exponential families of probability measures (pm's) include many of the parametric families frequently used in statistics, probability and information theory. Their mathematical theory has been worked out to considerable extent [3, 1, 2, 11]. Though limiting considerations are important and do appear in the literature, less attention has been paid to determining closures of exponential families.

For families supported by a finite or countable set, closures were considered by [1, pp. 154-156] and [2, pp. 191-201], respectively, the latter with regularity conditions. In the general case, different closure concepts come into account. Our main result, Theorem 2 in Section 3, determines the closure in variation distance (variation closure) of a full exponential family and, more generally, of any subfamily with a convex set of canonical parameters. Weak closures appear much harder to describe in general, but Theorem 1 in Section 3 is a step in that direction.

Other closure concepts are based on Kullback-Leibler $I$-divergence (information divergence or relative entropy),

$$D(P\|Q) \triangleq \begin{cases} \int \ln \frac{dP}{dQ}\, dP\ , & \text{if } P \ll Q, \\ +\infty\ , & \text{otherwise.} \end{cases}$$

With the terminology of [7], these are the $I$-closure and reverse $I$-closure ($rI$-closure); early work related to the latter appears in [3]. The $I$-closure of a convex set $\mathcal{S}$ of pm's is

This work was supported by the Hungarian National Foundation for Scientific Research under Grants T 26041, T 32323, TS 40719, and by Grant Agency of Academy of Sciences of the Czech Republic under Grant A1075104.

AMS 2000 Mathematics Subject Classification. Primary 60A10; secondary 60B10, 62B10, 52A20.

Key words and phrases. Accessible face, convex core, convex support, exponential family, extension, information divergence, variation distance, weak convergence.

*I. Csiszár is with A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1364 Budapest, P.O.Box 127, Hungary (`csiszar@renyi.hu`).

†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`).

relevant, for example, in large deviations theory where the conditional limit theorem for i.i.d. sequences on the condition that the empirical distribution belongs to $\mathcal{S}$, involves the "generalized $I$-projection" to $\mathcal{S}$ which is in the $I$-closure of $\mathcal{S}$, see [4]. In the context of exponential families, the $rI$-closure is of major statistical interest, for example, when the likelihood function is bounded but its maximum is not attained a "generalized maximum likelihood estimate" can be introduced as a pm belonging to the $rI$-closure, see [7].

Formally, the variation closure $\mathit{cl}_v(\mathcal{S})$, respectively the $I$-closure $\mathit{cl}_I(\mathcal{S})$ and $rI$-closure $\mathit{cl}_{rI}(\mathcal{S})$ of a set $\mathcal{S}$ of pm's on a given measurable space consists of all pm's $P$ to which there exists a sequence $Q_n$ in $\mathcal{S}$ such that the total variation $|P - Q_n|$, respectively the $I$-divergence $D(Q_n\|P)$ or $D(P\|Q_n)$, goes to zero as $n \to \infty$. The Pinsker inequality $|P-Q|^2 \leqslant 2D(P\|Q)$ implies that both $\mathit{cl}_I(\mathcal{S})$ and $\mathit{cl}_{rI}(\mathcal{S})$ are contained in $\mathit{cl}_v(\mathcal{S})$. For exponential families, the last inclusion gives a good approximation to $\mathit{cl}_{rI}(\mathcal{S})$, for example, all pm's in $\mathit{cl}_v(\mathcal{S})$ with mean belong to $\mathit{cl}_{rI}(\mathcal{S})$. This is one motivation for our study of variation closures, in addition to intrinsic mathematical interest. Theorem 3 in Section 3 characterizes those pm's in the variation closure that belong also to the $rI$-closure. The $I$-closure is much easier to describe than the other closures, see Corollary 2, in particular, full exponential families are $I$-closed. It should be mentioned that the $I$- and $rI$-closures are not topological closure operations as they are not idempotent. An example of an exponential family $\mathcal{E}$ with $\mathit{cl}_{rI}(\mathit{cl}_{rI}(\mathcal{E}))$ strictly larger than $\mathit{cl}_{rI}(\mathcal{E})$ is given in [8]. On the other hand, the $I$- and $rI$-closures are sequential closures in suitable topologies, see [9].

Our attention is focused on exponential families consisting of pm's on $\mathbb{R}^d$ and having a canonical statistic equal to the identity mapping. Clearly, determining their variation, $I$- or $rI$-closures, the same problems are solved for general exponential families of pm's on any measurable space, with $d$-dimensional canonical statistic, by mapping the family to one on $\mathbb{R}^d$ via the canonical statistic.

A crucial construction is that of the extension $\mathit{ext}(\mathcal{E})$ of a full exponential family $\mathcal{E}$, introduced by the authors [5, 7] based on their concept of convex core of a measure on $\mathbb{R}^d$ [6], see the definitions in Section 2. The inclusion $\mathcal{E} \subseteq \mathit{ext}(\mathcal{E})$ is strict unless no nontrivial supporting hyperplane of the (common) convex support of the pm's in $\mathcal{E}$ has positive probability under these pm's, by [7, Remark 3]. By Lemma 6 below, the variation closure of $\mathcal{E}$ is contained in $\mathit{ext}(\mathcal{E})$. A stronger result announced in [5], the variation closedness of $\mathit{ext}(\mathcal{E})$, follows as Corollary 3 from Theorem 4 that deals with variation convergent sequences in $\mathit{ext}(\mathcal{E})$.

The inclusion $\mathit{cl}_v(\mathcal{E}) \subseteq \mathit{ext}(\mathcal{E})$ implies that if the subset $\mathit{cl}_{rI}(\mathcal{E})$ of $\mathit{cl}_v(\mathcal{E})$ is equal to $\mathit{ext}(\mathcal{E})$ (for example, if the domain of canonical parameters is the whole $\mathbb{R}^d$, see [7, Lemma 6(ii)]) then also $\mathit{cl}_v(\mathcal{E}) = \mathit{ext}(\mathcal{E})$. Moreover, since $\mathcal{E}$ is $rI$-closed if and only if $\mathcal{E} = \mathit{ext}(\mathcal{E})$ [7, Corollary 2], the last condition is necessary and sufficient also for the variation closedness of $\mathcal{E}$.

The cases just mentioned, though frequent in practice, are of secondary interest for our purposes. This paper is primarily devoted to the general case when all common regularity conditions are absent, although the assumption of steepness or even regularity (see [1, pp. 116-117]) would not lead to significant simplifications. The typical situation we have in mind is when the pm's in $\mathcal{E}$ have both discrete and continuous components.

## 2. Preliminaries

2.1. **Convex sets and faces.** The closure and affine hull of a set $B \subseteq \mathbb{R}^d$ are denoted by $\mathit{cl}(B)$ and $\mathit{aff}(B)$, and the relative interior (interior in the relative topology of $\mathit{aff}(B)$) by $\mathit{ri}(B)$. The linear subspace of $\mathbb{R}^d$ obtained by shifting $\mathit{aff}(B)$ to contain the origin is denoted by $\mathit{lin}(B)$. Orthogonal projections to subspaces of the form $\mathit{lin}(C)$, where $C \subseteq \mathbb{R}^d$ is a convex set, are often needed in the sequel; they are denoted briefly by $\pi_C$ rather than $\pi_{\mathit{lin}(C)}$.

A *face* of a nonempty convex set $C \subseteq \mathbb{R}^d$ is a nonempty convex subset $F$ of $C$ such that whenever $tx + (1-t)y \in F$ for some $x, y$ in $C$ and $0 < t < 1$ then $x, y$ are in $F$. As in [10] but unlike in [12], the empty set is not considered a face. The *proper* faces are those different from $C$, and the *exposed* faces are $C$ itself and its intersections with the supporting hyperplanes of $C$. A proper exposed face $F$ of $C$ is thus represented as $F = C \cap \{x\colon \langle \tau, x - a\rangle = 0\}$ where $a \in C$ and $\tau \in \mathbb{R}^d$ is a unit vector such that $\langle \tau, x - a\rangle \leqslant 0$ for each $x \in C$. Obviously, there is no loss of generality in assuming $\tau \in \mathit{lin}(C)$. Such a vector $\tau$ *exposes* $F$ in $C$.

2.2. **Convex support and core.** A *measure* will always mean a finite Borel measure on $\mathbb{R}^d$. The *convex support* $\mathit{cs}(\mu)$ and *convex core* $\mathit{cc}(\mu)$ of $\mu$ are defined as the intersection of those convex closed, respectively convex Borel subsets $C$ of $\mathbb{R}^d$ which have full $\mu$-measure, $\mu(C) = \mu(\mathbb{R}^d)$. While the former is a standard concept, the latter is of recent origin [6]. Let us recall from [6] the key facts

$$\mathit{cs}(\mu) = \mathit{cl}(\mathit{cc}(\mu)) \quad \text{and} \quad \mathit{cc}(\mu^{\mathit{cl}(F)}) = F \text{ for faces } F \text{ of } \mathit{cc}(\mu)\,, \tag{1}$$

where the restriction of a measure $\mu$ to a Borel subset $B$ of $\mathbb{R}^d$ is denoted by $\mu^B$.

Note that the convex closed set $\mathit{cs}(\mu)$ is of full $\mu$-measure but the convex set $\mathit{cc}(\mu)$, though measurable by [6, Theorem 1], need not be.

For brevity, $\mathit{lin}(\mu)$ is written instead of $\mathit{lin}(\mathit{cs}(\mu))$ and similarly with $\mathit{ri}(\mu)$ and $\mathit{aff}(\mu)$.

**Lemma 1.** *A supporting hyperplane $H$ of $\mathit{cs}(\mu)$ is of positive $\mu$ measure if and only if $F = H \cap \mathit{cc}(\mu)$ is non-empty. Moreover, $\mu(H \setminus \mathit{cl}(F)) = 0$.*

*Proof.* Using [6, Lemma 2(ii)], $F = \mathit{cc}(\mu^H)$. This and (1) give $\mathit{cl}(F) = \mathit{cs}(\mu^H)$, whence both assertions follow. □

2.3. **Exponential families.** The term *exponential family* without any adjective will mean below a full exponential family based on a (non-zero) measure $\mu$ on $\mathbb{R}^d$, with a canonical statistic equal to the identity mapping. This family $\mathcal{E} = \mathcal{E}_\mu$ consists of the pm's $Q_\vartheta$ with $\mu$-densities

$$\frac{dQ_\vartheta}{d\mu}(x) \triangleq e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)} \tag{2}$$

where

$$\Lambda(\vartheta) = \Lambda_\mu(\vartheta) \triangleq \ln \int_{\mathbb{R}^d} e^{\langle \vartheta, x\rangle}\, \mu(dx)$$

and the canonical parameter $\vartheta$ belongs to $\mathsf{dom}(\Lambda) = \{\vartheta\colon \Lambda(\vartheta) < \infty\}$. Note that $\mu$ is not uniquely determined by the family $\mathcal{E}$, in particular, any member of $\mathcal{E}$ could play the role of $\mu$; in this paper, however, $\mu$ will be regarded as given.

Clearly, if $\vartheta \in \mathsf{dom}(\Lambda)$ and $\theta - \vartheta$ is orthogonal to $\mathsf{lin}(\mu)$ for some $\theta \in \mathbb{R}^d$ then also $\theta \in \mathsf{dom}(\Lambda)$ and $Q_\vartheta = Q_\theta$. A bijective parametrization can be given as

$$\text{(3)} \qquad \mathcal{E} = \{Q_\vartheta\colon \vartheta \in \Theta\} \;\text{ where }\; \Theta = \Theta_\mu \triangleq \mathsf{dom}(\Lambda_\mu) \cap \mathsf{lin}(\mu) = \pi_{\mathsf{lin}(\mu)}(\mathsf{dom}(\Lambda_\mu))\,.$$

Here, $\Theta$ equals $\mathsf{dom}(\Lambda)$ if and only if $\mathsf{lin}(\mu) = \mathbb{R}^d$ (when (2) is called a minimal representation). For the purposes below, it is essential not to require that condition, and not to require $\mu$ to be a pm, either.

Of main interest will be subfamilies

$$\mathcal{E}_\Xi = \{Q_\vartheta\colon \vartheta \in \Xi\}, \qquad \Xi \subseteq \Theta\,,$$

of the full family $\mathcal{E}$, primarily when $\Xi$ is convex. In this case, $\mathcal{E}_\Xi$ is called *canonically convex exponential family.*

The function $\Lambda$ is known to be convex and lower semicontinuous, thus continuous on closed segments contained in $\mathsf{dom}(\Lambda)$. The following lemma is a minor improvement of [3, Lemma 23.5].

**Lemma 2.** *Let $\vartheta_n$ be a sequence in $\mathsf{dom}(\Lambda)$ converging to some $\vartheta \in \mathbb{R}^d$.*

*(i) If $Q_{\vartheta_n}$ converges weakly then $\vartheta \in \mathsf{dom}(\Lambda)$ and $\Lambda(\vartheta_n)$ converges to $\Lambda(\vartheta)$.*

*(ii) If $\vartheta \in \mathsf{dom}(\Lambda)$ and $\Lambda(\vartheta_n) \to \Lambda(\vartheta)$ then $Q_{\vartheta_n}$ converges to $Q_\vartheta$ in variation.*

*Proof.* *(i)* If $Q_{\vartheta_n}$ converges weakly to some pm $P$ then for each continuity set $B$ of $P$

$$P(B) = \lim_{n\to\infty} Q_{\vartheta_n}(B) = \lim_{n\to\infty} e^{-\Lambda(\vartheta_n)} \int_B e^{\langle \vartheta_n, x\rangle}\, \mu(dx)$$

where if $B$ is compact

$$\int_B e^{\langle \vartheta_n, x\rangle}\, \mu(dx) \to \int_B e^{\langle \vartheta, x\rangle}\, \mu(dx)\,,$$

by dominated convergence. If also $P(B) > 0$ then $e^{-\Lambda(\vartheta_n)}$ converges to a positive number $c$. Hence

$$P(B) = c \cdot \int_B e^{\langle \vartheta, x\rangle}\, \mu(dx)$$

for each compact continuity set $B$ of $P$, and consequently, for all Borel sets $B$. When $B = \mathbb{R}^d$ it follows that $c = e^{-\Lambda(\vartheta)}$. Hence, $\vartheta \in \mathsf{dom}(\Lambda)$ and $\Lambda(\vartheta_n)$ converges to $\Lambda(\vartheta)$.

*(ii)* Under the assumptions, the $\mu$-densities $e^{\langle \vartheta_n, x\rangle - \Lambda(\vartheta_n)}$ of $Q_{\vartheta_n}$ converge to the $\mu$-density $e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}$ of $Q_\vartheta$ pointwise, which is known to imply $Q_{\vartheta_n} \to Q_\vartheta$ in variation. □

2.4. **Extensions of exponential families.** The restriction of $\mu$ to the closure of a face $F$ of $\mathsf{cc}(\mu)$ is a non-zero measure by (1). The exponential family based on this restriction $\mu^{\mathsf{cl}(F)}$ is denoted by $\mathcal{E}^F$. It consists of the pm's $Q_{F,\vartheta}$ defined as in (2) with $\mu$ and $\Lambda$ replaced by $\mu^{\mathsf{cl}(F)}$ and

$$\Lambda_F(\vartheta) \triangleq \ln \int_{\mathsf{cl}(F)} e^{\langle \vartheta, x\rangle}\, \mu(dx)\,.$$

Obviously, $\mathsf{dom}(\Lambda) \subseteq \mathsf{dom}(\Lambda_F)$. The family $\mathcal{E}^F$ is bijectively parameterized as

$$\mathcal{E}^F = \{Q_{F,\vartheta}\colon \vartheta \in \Theta_F\} \;\text{ where }\; \Theta_F \triangleq \mathsf{dom}(\Lambda_F) \cap \mathsf{lin}(F) = \pi_F(\mathsf{dom}(\Lambda_F))$$

similarly to (3), since $\mathit{lin}(\mu^{\mathit{cl}(F)}) = \mathit{lin}(F)$ by (1). For each $Q_\vartheta \in \mathcal{E}$ with $\vartheta \in \mathit{dom}(\Lambda)$, its conditioning $Q_\vartheta(\cdot|\mathit{cl}(F))$, equal to the restriction $Q_\vartheta^{\mathit{cl}(F)}$ divided by $Q_\vartheta(\mathit{cl}(F))$, belongs to $\mathcal{E}^F$. The simple fact that $Q_\vartheta(\cdot|\mathit{cl}(F))$ coincides with the pm $Q_{F,\theta}$ where $\theta = \pi_F(\vartheta)$ is in $\Theta_F$ will be repeatedly used in the sequel. These conditionings of the pm's in $\mathcal{E}$ exhaust $\mathcal{E}^F$ if and only if the inclusion $\pi_F(\Theta) \subseteq \Theta_F$ holds with equality.

The *extension* $\mathit{ext}(\mathcal{E})$ of an exponential family $\mathcal{E} = \mathcal{E}_\mu$ is the union of the families $\mathcal{E}^F$ over all faces $F$ of $\mathit{cc}(\mu)$. Each $\mathcal{E}^F$ is called a *component* of $\mathit{ext}(\mathcal{E})$. A similar construction of a "boundary at infinity" of $\mathcal{E}$ which uses faces of $\mathit{cs}(\mu)$ rather than of $\mathit{cc}(\mu)$ was proposed earlier in [3]. Some crucial assertions on an exponential family completed by its "boundary" were found erroneous, but their analogues for $\mathit{ext}(\mathcal{E})$ valid, in [5, 7]. The reason is that $\mathit{ext}(\mathcal{E})$ may be strictly larger than $\mathcal{E}$ completed by its "boundary": By [6, Lemma 11], the latter consists of those components $\mathcal{E}^F$ that correspond to the proper exposed faces $F$ of $\mathit{cc}(\mu)$.

2.5. **Accessible faces.** For any face $F$ of a convex set $C \subseteq \mathbb{R}^d$ there exists a chain

$$C = F_0 \supset F_1 \supset \cdots \supset F_m = F\,,$$

not necessarily unique, such that $F_i$ is an exposed face of $F_{i-1}$, $1 \leqslant i \leqslant m$. If for every $1 \leqslant i \leqslant m$ a unit vector $\tau_i \in \mathit{lin}(F_{i-1})$ exposes $F_i$ in $F_{i-1}$ then $\tau_1, \ldots, \tau_m$ is called an *access sequence* to the face $F$ of $C$; the access sequence to $F = C$ is empty. Since $\tau_i \in \mathit{lin}(F_{i-1})$ is orthogonal to $\mathit{lin}(F_i)$, the vectors of any nonempty access sequence are orthonormal.

Let $C$ and $\Xi$ be two nonempty convex subsets of $\mathbb{R}^d$. For our main result, where the role of $C$ is played by $\mathit{cc}(\mu)$ and the role of $\Xi$ by a convex subset of $\Theta_\mu$, a new concept of $\Xi$-accessible faces of $C$ is suitable. This concept involves a constraint on access sequences in terms of recession cones of projections of $\mathit{ri}(\Xi)$. Recall that the *recession cone* of a convex set $\Xi \subseteq \mathbb{R}^d$ is

$$\mathit{rec}(\Xi) = \{\tau\colon \vartheta + t\,\tau \in \Xi \ \text{ for all } \ \vartheta \in \Xi,\, t \geqslant 0\}\,.$$

By [12, Theorem 8.2 and Corollary 8.3.1], $\mathit{rec}(\mathit{ri}(\Xi)) = \mathit{rec}(\mathit{cl}(\Xi))$, and this is a closed cone containing $\mathit{rec}(\Xi)$. Now, a face $F$ of the convex set $C$ is *$\Xi$-accessible* if an access sequence $\tau_1, \ldots, \tau_m$ to $F$ exists such that

$$\tau_i \in \mathit{rec}(\pi_{F_{i-1}}(\mathit{ri}(\Xi)))\,, \qquad 1 \leqslant i \leqslant m\,. \tag{4}$$

An access sequence to $F$ satisfying (4) will be called *adapted* to $\Xi$. It may seem artificial that these notions depend on $\Xi$ only through its relative interior but if $\mathit{ri}$ were omitted in (4), some later assertions would not hold, see Example 3 in Section 3. Note that the empty sequence is trivially adapted, thus $C$ itself is always a $\Xi$-accessible face of $C$.

**Lemma 3.** *If $\Xi \subseteq \mathit{lin}(C)$, an access sequence $\tau_1, \ldots, \tau_m$ to a proper face $F$ of $C$ is adapted to $\Xi$ if and only if $\tau_1 \in \mathit{rec}(\mathit{ri}(\Xi))$ and for the face $F_1$ of $C$ exposed by $\tau_1$ the access sequence $\tau_2, \ldots, \tau_m$ to the face $F$ of $F_1$ is adapted to $\pi_{F_1}(\Xi)$.*

*Proof.* By the hypotheses $\Xi \subseteq \mathit{lin}(C)$, the set $\pi_{F_0}(\mathit{ri}(\Xi))$ in the condition (4) for $i = 1$ is equal to $\mathit{ri}(\Xi)$. In the conditions for $2 \leqslant i \leqslant m$, the sets $\pi_{F_{i-1}}(\mathit{ri}(\Xi))$ are equal to the sets $\pi_{F_{i-1}}(\mathit{ri}(\pi_{F_1}(\Xi)))$ appearing in the analogue of (4) for the adaptedness of $\tau_2, \ldots, \tau_m$ to

$\pi_{F_1}(\Xi)$, since the operation *ri* interchanges with orthogonal projections [12, Theorem 6.6], and $\pi_{F_{i-1}}\pi_{F_1} = \pi_{F_{i-1}}$ if $i \geqslant 2$. $\square$

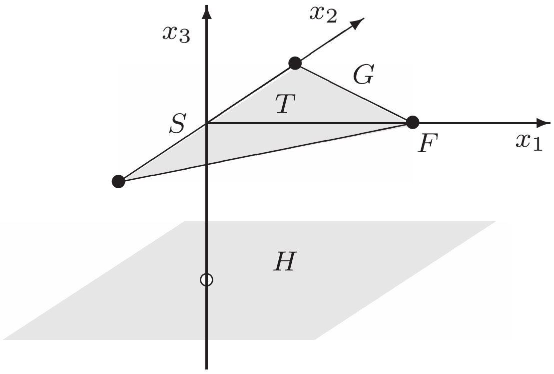
**Example 1.** Let $C \subset \mathbb{R}^3$ be the convex hull of the union of the plane $H = \mathbb{R}^2 \times \{-1\}$ and the triangle $T$ with vertices $(1,0,0)$, $(0,\pm 1,0)$, and let $\Xi$ consist of those $\vartheta \in \mathbb{R}^3$ whose second coordinate is strictly between $-1$ and 1. In this example, $\Xi$-accessibility of the nine faces of $C$ is discussed. The proper exposed faces of $C$ are $H$ and $T$. Of the six nonexposed faces of $C$, equal to faces of $T$, consider

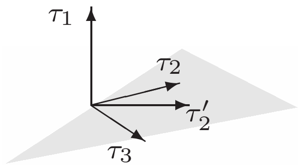
$$F = \{(1,0,0)\}\,,\ G = \{(t,1-t,0)\colon 0 \leqslant t \leqslant 1\}\,,\ S = \{(0,t,0)\colon |t| \leqslant 1\}\,.$$

Since $\Xi$ is open, the relative interiors in (4) can be ignored. Note that the recession cone of $\pi_C(\Xi) = \Xi$ is $\mathbb{R} \times \{0\} \times \mathbb{R}$ and the recession cone of $\pi_T(\Xi)$ is $\mathbb{R} \times \{(0,0)\}$. Let

$$\tau_1 = (0,0,1)\,,\quad \tau_2 = (1/\sqrt{2}\,, 1/\sqrt{2}\,, 0)\,,\quad \tau_3 = (1/\sqrt{2}\,, -1/\sqrt{2}\,, 0) \ \text{ and } \ \tau_2' = (1,0,0)\,.$$

Since $\mathit{rec}(\Xi)$ contains both $-\tau_1$ and $\tau_1$, the faces $H$ and $T$ are $\Xi$-accessible. Both $\tau_1, \tau_2, \tau_3$ and $\tau_1, \tau_2'$ are access sequences to $F$, with corresponding chains $C \supset T \supset G \supset F$ and $C \supset T \supset F$, respectively. Since $\mathit{rec}(\pi_T(\Xi))$ contains $\tau_2'$ but not $\tau_2$, the access sequence $\tau_1, \tau_2'$ is adapted to $\Xi$ whereas $\tau_1, \tau_2, \tau_3$ is not. Due to the former, $F$ is a $\Xi$-accessible face of $C$. On the other hand, $G$ is not $\Xi$-accessible, as the only access sequence $\tau_1, \tau_2$ to $G$, with chain $C \supset T \supset G$, has $\tau_2 \notin \mathit{rec}(\pi_T(\Xi))$. Similarly, the segment $S$ is $\Xi$-accessible but its endpoints are not.




2.6. **Partial means.** When studying $rI$-closures of exponential families, pm's that do not have a mean require a special attention. The following simple concept will be useful: for any pm $P$ on $\mathbb{R}^d$, write

$$\mathsf{M}(P) = \{\tau \in \mathbb{R}^d\colon \langle \tau, \cdot \rangle \text{ is } P\text{-integrable}\}$$

and define the *partial mean* $\mathsf{m}(P)$ as the unique element of the linear space $\mathsf{M}(P)$ with

$$\int_{\mathbb{R}^d} \langle \tau, x \rangle\, P(dx) = \langle \tau, \mathsf{m}(P) \rangle\,, \quad \tau \in \mathsf{M}(P)\,.$$

Note that $\mathsf{M}(P) = \mathbb{R}^d$ if and only if $P$ has a mean, in which case $\mathsf{m}(P)$ equals the mean.

The following lemma is well-known but usually stated in less generality.

**Lemma 4.** *For $\vartheta \in \mathit{dom}(\Lambda)$ and a unit vector $\tau$ such that $\vartheta + t\,\tau \in \mathit{dom}(\Lambda)$ for some $t > 0$, the integral $\int \langle \tau, x \rangle\, Q_\vartheta(dx)$ exists, either finite or $-\infty$. This integral equals the directional derivative of $\Lambda$ at $\vartheta$ in the direction $\tau$.*

*Proof.* The directional derivative, that is, the right derivative of the function $t \mapsto \Lambda(\vartheta + t\tau)$ at $t = 0$, equals

$$\frac{1}{\int e^{\langle\vartheta,x\rangle}\,\mu(dx)}\ \lim_{t\downarrow 0} \int \frac{e^{t\langle\tau,x\rangle}-1}{t}\, e^{\langle\vartheta,x\rangle}\mu(dx)\,.$$

The ratio $(e^{t\langle\tau,x\rangle}-1)/t$ decreases to $\langle\tau,x\rangle$ when $t$ decreases to zero, and the assertion follows by the monotone convergence. □

The following lemma shows the relevance of partial means for exponential families.

**Lemma 5.** *The I-divergence $D(Q_\theta\|Q_\vartheta)$ of pm's in $\mathcal{E}$ is finite if and only if $\theta-\vartheta \in \mathsf{M}(Q_\theta)$, in which case*

$$D(Q_\theta\|Q_\vartheta) = \langle\theta-\vartheta, \mathsf{m}(Q_\theta)\rangle - \Lambda(\theta) + \Lambda(\vartheta)\,.$$

*Proof.* By definition,

$$D(Q_\theta\|Q_\vartheta) = \int_{\mathbb{R}^d} \ln \frac{e^{\langle\theta,x\rangle-\Lambda(\theta)}}{e^{\langle\vartheta,x\rangle-\Lambda(\vartheta)}}\, Q_\theta(dx) = \int_{\mathbb{R}^d} \langle\theta-\vartheta,x\rangle\, Q_\theta(dx)\ - \Lambda(\theta) + \Lambda(\vartheta)\,.$$

The last integral is finite if and only if $\theta-\vartheta$ belongs to $\mathsf{M}(Q_\theta)$, in which case it equals $\langle\theta-\vartheta, \mathsf{m}(Q_\theta)\rangle$. □

This lemma is also used with $\mathcal{E}$ replaced by a component $\mathcal{E}^F$ of $\mathsf{ext}(\mathcal{E})$, combined with the obvious identity

$$D(P\|Q) = D(P\|Q(\cdot|\mathsf{cl}(F))) - \ln\, Q(\mathsf{cl}(F))\,, \qquad P \in \mathcal{E}^F, Q \in \mathcal{E}\,. \tag{5}$$

## 3. Main results

Below, $\mathcal{E} = \mathcal{E}_\mu$ always denotes an exponential family in the sense of Section 2.3, $Q_\vartheta$ with $\vartheta \in \Theta = \mathsf{dom}(\Lambda) \cap \mathsf{lin}(\mu)$ denotes a pm in $\mathcal{E}$, and $\mathcal{E}_\Xi$ with $\Xi \subseteq \Theta$ denotes a subfamily of $\mathcal{E}$.

**Theorem 1.** *If a sequence of pm's $Q_{\vartheta_n}$ with $\vartheta_n$ in $\Theta$ converges weakly to a pm $P$ then one of the following two alternatives takes place.*

*(i) The sequence $\vartheta_n$ converges to an element $\vartheta$ of $\Theta$, $P = Q_\vartheta$ and $Q_{\vartheta_n} \to P$ even in variation distance.*

*(ii) The norm of $\vartheta_n$ goes to $\infty$, a proper exposed face $G$ of $\mathsf{cs}(\mu)$ exists such that $P(G) = 1$, and the limit of any convergent subsequence of $\vartheta_n/\|\vartheta_n\|$ exposes such a face of $\mathsf{cs}(\mu)$.*

The proof is given in Section 4. It follows from Theorem 1 that the weak convergence of a sequence $Q_{\vartheta_n}$ to some $Q_\vartheta$ in $\mathcal{E}$ implies $\vartheta_n \to \vartheta$, see [1, Theorem 8.3] for a direct proof. A consequence of this and Lemma 2 is stated for reference purposes.

**Corollary 1.** *For a sequence $\vartheta_n$ in $\Theta$ and $\vartheta \in \Theta$ the following assertions are equivalent.*

*(i) Convergence $Q_{\vartheta_n} \to Q_\vartheta$ weakly.*

*(ii) Convergence $Q_{\vartheta_n} \to Q_\vartheta$ in variation.*

*(iii) Convergences $\vartheta_n \to \vartheta$ and $\Lambda(\vartheta_n) \to \Lambda(\vartheta)$.*

**Corollary 2.** *For any subset $\Xi$ of $\Theta$*

$$\mathcal{E}_{\mathit{cl}(\Xi)\cap\Theta} \supseteq \mathit{cl}_v(\mathcal{E}_\Xi) \cap \mathcal{E} \supseteq \mathit{cl}_I(\mathcal{E}_\Xi)\,.$$

*When $\Xi$ is convex the equalities take place.*

*Proof.* The first inclusion follows from Corollary 1. For the second one, if $D(Q_{\vartheta_n}\|P) \to 0$, the sequence $Q_{\vartheta_n}$ converges in variation to $P$, thus $\mathit{cl}_v(\mathcal{E}_\Xi)$ contains $\mathit{cl}_I(\mathcal{E}_\Xi)$. Moreover, the alternative *(i)* holds in Theorem 1 since otherwise $D(Q_{\vartheta_n}\|P) = +\infty$ for all $n$. This proves that $\mathcal{E}$ contains $\mathit{cl}_I(\mathcal{E}_\Xi)$.

Supposing $\Xi$ is convex, let $\vartheta \in \mathit{cl}(\Xi) \cap \Theta$ and let $\tau$ be a unit vector such that $\vartheta + t\,\tau$ belongs to $\mathit{ri}(\Xi)$ for some $t > 0$. It suffices to show that $D(Q_{\vartheta_n}\|Q_\vartheta) \to 0$ for $\vartheta_n = \vartheta + t_n\,\tau$ with $t_n$ decreasing to zero. By Lemma 4, $\tau \in M(Q_{\vartheta_n})$, and then Lemma 5 implies

$$D(Q_{\vartheta_n}\|Q_\vartheta) = t_n\,\langle \tau, m(Q_{\vartheta_n})\rangle - \Lambda(\vartheta_n) + \Lambda(\vartheta)\,.$$

Since $\Lambda(\vartheta_n) \to \Lambda(\vartheta)$ and the sequence $\langle \tau, m(Q_{\vartheta_n})\rangle$ is decreasing by Lemma 4, the claim follows. $\square$

The following technical assertion, crucial for Theorem 2, is proved in Section 5. The notion of $\Xi$-accessibility enters the scene.

**Lemma 6.** *If a sequence of pm's in a canonically convex exponential family $\mathcal{E}_\Xi$ converges in variation distance to a pm $P$ then there exists a $\Xi$-accessible face $F$ of $\mathit{cc}(\mu)$ such that $P$ belongs to $\mathcal{E}^F$.*

Lemma 6 enables us to conclude that the variation closure of a canonically convex family $\mathcal{E}_\Xi$ is contained in the union of those components $\mathcal{E}^F$ of $\mathit{ext}(\mathcal{E})$ which correspond to the $\Xi$-accessible faces $F$ of $\mathit{cc}(\mu)$. For our main result still another tool is needed, namely a special convergence concept: A sequence of pm's $Q_n$ will be said to converge *neatly* to a pm $P$ if the $P$-dominated component of $Q_n$ has constant $P$-density $c_n$ and $c_n \to 1$, or equivalently, if $Q_n(\cdot|\mathit{cs}(P))$ equals $P$ when defined and $Q_n(\mathit{cs}(P)) \to 1$. Obviously, the neat convergence implies variation convergence, and even $rI$-convergence, due to $D(P\|Q_n) = -\ln c_n$.

**Lemma 7.** *For a convex subset $\Xi$ of $\Theta$ and $\Xi$-accessible face $F$ of $\mathit{cc}(\mu)$, each pm $Q_{F,\theta}$ with $\theta \in \pi_F(\mathit{ri}(\Xi))$ is the neat limit of a sequence from $\mathcal{E}_{\mathit{ri}(\Xi)}$.*

This lemma is proved in Section 5.

**Theorem 2.** *The variation closure of a canonically convex exponential family $\mathcal{E}_\Xi$ is*

$$\mathit{cl}_v(\mathcal{E}_\Xi) = \bigcup \{Q_{F,\theta} \colon \theta \in \mathit{cl}(\pi_F(\Xi)) \cap \Theta_F\}$$

*where the union runs over all $\Xi$-accessible faces $F$ of $\mathit{cc}(\mu)$.*

*Proof.* Suppose $P$ is the limit in variation distance of a sequence $Q_n$ in $\mathcal{E}_\Xi$. By Lemma 6, $P \in \mathcal{E}^F$ for a $\Xi$-accessible face $F$ of $\mathit{cc}(\mu)$. Then $P(\mathit{cl}(F)) = 1$ implies $Q_n(\mathit{cl}(F)) \to 1$, and thus the sequence $Q_n(\cdot|\mathit{cl}(F))$ in $\mathcal{E}^F$ also converges to $P$ in variation distance. These conditioned pm's are of form $Q_{F,\theta_n}$ with $\theta_n \in \pi_F(\Xi)$, thus Corollary 2 implies $P = Q_{F,\theta}$ for some $\theta$ in $\mathit{cl}(\pi_F(\Xi)) \cap \Theta_F$. This proves the inclusion $\subseteq$.

Conversely, if $F$ is a $\Xi$-accessible face of $\mathsf{cc}(\mu)$, Lemma 7 implies that $\mathsf{cl}_v(\mathcal{E}_\Xi)$ contains the pm's $Q_{F,\theta}$ with $\theta \in \pi_F(\mathsf{ri}(\Xi))$. Thus, by Corollary 2 and the convexity of $\pi_F(\mathsf{ri}(\Xi))$

$$\mathsf{cl}_v(\mathcal{E}_\Xi) \supseteq \{Q_{F,\theta} \colon \theta \in \mathsf{cl}(\pi_F(\mathsf{ri}(\Xi))) \cap \Theta_F\}\,.$$

Since $\mathsf{ri}$ interchanges with projections, the inclusion $\supseteq$ follows. □

**Example 2.** Using the notation of Example 1, let $\mu$ be the sum of the measure giving mass 1 to each vertex of $T$ and of the pm on the plane $H$ with density $e^{-x_1^2}e^{-|x_2|}$ w.r.t. the Lebesgue measure on $H$. Then $\mathsf{cc}(\mu) = C$, and $\Theta_\mu = \mathsf{dom}(\Lambda_\mu)$ coincides with $\Xi$ of Example 1, defined by restricting the second coordinate to be between $-1$ and 1. By Theorem 2, the variation closure of the full family $\mathcal{E} = \mathcal{E}_\mu$ intersects five out of the nine components of $\mathsf{ext}(\mathcal{E})$. In addition to the full families $\mathcal{E}$, $\mathcal{E}^H$ and $\mathcal{E}^F$, the latter consisting of the point mass at $(1,0,0)$, $\mathsf{cl}_v(\mathcal{E})$ contains also some pm's from $\mathcal{E}^T$ and $\mathcal{E}^S$. Note that $\mathsf{cl}_v(\mathcal{E})$ intersects $\mathcal{E}^F$ but not $\mathcal{E}^G$, although $F \subset G$.

Part *(ii)* of the following Theorem 3 gives a necessary and sufficient condition for a pm $P$ in $\mathsf{cl}_v(\mathcal{E}_\Xi)$ to belong also to $\mathsf{cl}_{rI}(\mathcal{E}_\Xi)$, even when $\Xi \subseteq \Theta$ is not convex. When $\Xi$ is convex, this condition can be effectively verified by Theorem 2. Part *(iii)* provides a simple sufficient condition which has a direct proof. A trivial consequence of Theorem 3 is that $\mathsf{cl}_{rI}(\mathcal{E}_\Xi)$ contains all pm's in $\mathsf{cl}_v(\mathcal{E}_\Xi)$ that have a mean.

**Theorem 3.** *(i) Suppose a sequence $Q_{\vartheta_n}$, $\vartheta_n \in \Theta$, of pm's in $\mathcal{E}$ converges in variation to a pm $P = Q_{F,\theta}$, $\theta \in \Theta_F$, in a component $\mathcal{E}^F$ of $\mathsf{ext}(\mathcal{E})$. Then $D(P\|Q_{\vartheta_n})$ goes to zero if it is eventually finite which is equivalent to $\vartheta_n \in \theta + \mathsf{M}(P)$ eventually.*

*(ii) A pm $P = Q_{F,\theta}$ belongs to the rI-closure of a subfamily $\mathcal{E}_\Xi$ of $\mathcal{E}$ if and only if it belongs to the variation closure of $\mathcal{E}_{\Xi\cap(\theta+\mathsf{M}(P))}$.*

*(iii) The rI-closure of a canonically convex exponential family $\mathcal{E}_\Xi$ contains all pm's $P$ in the variation closure of $\mathcal{E}_\Xi$ that satisfy $D(P\|Q) < \infty$ for some $Q \in \mathcal{E}_{\mathsf{ri}(\Xi)}$.*

*Proof.* *(i)* Consider first the case when $\mathcal{E}^F = \mathcal{E}$, that is $P = Q_\vartheta$ for some $\vartheta \in \Theta$. Then $\vartheta_n \to \vartheta$ and $\Lambda(\vartheta_n) \to \Lambda(\vartheta)$, by Corollary 1. Thus, the assertion for this case follows by Lemma 5.

When $F$ is a proper face of $\mathsf{cc}(\mu)$, note that the variation convergence assumption implies $Q_{\vartheta_n}(\mathsf{cl}(F)) \to P(\mathsf{cl}(F)) = 1$. Hence, eq. (5) shows that $D(P\|Q_{\vartheta_n})$ goes to zero if and only if $D(P\|Q_{\vartheta_n}(\cdot|\mathsf{cl}(F)))$ does. Moreover, $Q_{\vartheta_n}(\cdot|\mathsf{cl}(F))$, equal to $Q_{F,\theta_n} \in \mathcal{E}^F$ with $\theta_n = \pi_F(\vartheta_n)$, also converges in variation to $P$. It follows, applying the result in the first case to $\mathcal{E}^F$ in the role of $\mathcal{E}$, that $D(P\|Q_{\vartheta_n}) \to 0$ if it is eventually finite. Also by (5), the finiteness of $D(P\|Q_{\vartheta_n})$ is equivalent to that of $D(P\|Q_{\vartheta_n}(\cdot|\mathsf{cl}(F)))$, and therefore by Lemma 5, to $\theta - \theta_n \in \mathsf{M}(P)$. Since $\vartheta_n - \theta_n$ is orthogonal to $\mathsf{lin}(F)$, it belongs to $\mathsf{M}(P)$, thus the last condition is equivalent to $\vartheta_n \in \theta + \mathsf{M}(P)$.

*(ii)* This follows directly from *(i)*.

*(iii)* Suppose $P = Q_{F,\theta} \in \mathsf{cl}_v(\mathcal{E}_\Xi)$ where $F$ is a $\Xi$-accessible face of $\mathsf{cc}(\mu)$ and $\theta$ belongs to $\mathsf{cl}(\pi_F(\Xi)) \cap \Theta_F$, see Theorem 2. If $D(P\|Q_{\vartheta_0})$ is finite for some $\vartheta_0 \in \mathsf{ri}(\Xi)$ then as in the proof of part *(i)*, $D(P\|Q_{F,\theta_0})$ is also finite, where $\theta_0 = \pi_F(\vartheta_0)$, $Q_{F,\theta_0} = Q_{\vartheta_0}(\cdot|\mathsf{cl}(F))$.

Then for $\theta_n = t_n\theta_0 + (1-t_n)\theta$ with $t_n \downarrow 0$, $D(P\|Q_{F,\theta_n})$ is also finite by Lemma 5, and $Q_{F,\theta_n} \to Q_{F,\theta}$ in variation by Corollary 1. It follows by part *(i)* that $D(P\|Q_{F,\theta_n}) \to 0$.

Since $\theta_0 \in \pi_F(\mathsf{ri}(\Xi)) = \mathsf{ri}(\pi_F(\Xi))$ and $\theta \in \mathsf{cl}(\pi_F(\Xi))$ imply $\theta_n \in \mathsf{ri}(\pi_F(\Xi))$, Lemma 7 gives that each $Q_{F,\theta_n}$ is the neat limit of a sequence in $\mathcal{E}_{\mathsf{ri}(\Xi)}$. Thus to each $\theta_n$ there exists $\vartheta_n \in \mathsf{ri}(\Xi)$ such that $Q_{F,\theta_n} = Q_{\vartheta_n}(\cdot|\mathsf{cl}(F))$ and $Q_{\vartheta_n}(\mathsf{cl}(F))$ is arbitrarily close to 1. From this and $D(P\|Q_{F,\theta_n}) \to 0$ the claim $D(P\|Q_{\vartheta_n}) \to 0$ follows by eq. (5). $\square$

The following example illustrates a use of Theorem 3*(ii)* and Theorem 2 when deciding whether a pm belongs to the $rI$-closure of a canonically convex exponential family. It also illustrates why $\mathsf{ri}(\Xi)$ rather than $\Xi$ appears in the definition (4) of $\Xi$-accessibility and in Lemma 7.

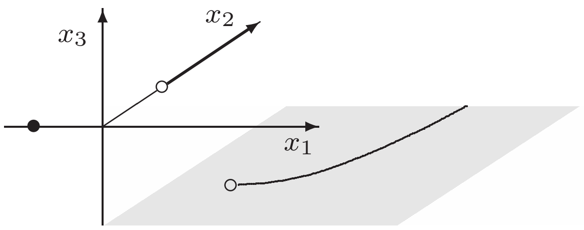
**Example 3.** Let $\mu$ be the measure on $\mathbb{R}^3$ equal to the sum of the point mass at $(-1,0,0)$, the image $P$ under $t \mapsto (0,t,0)$ of the pm with density $dt/t^2$ on the half-line $t>1$, and the image under $t \mapsto (t,t^2,-1)$ of the pm with density $2dt/t^3$ on the same half-line. Then

$$\Lambda(\vartheta) = \ln\left[e^{-\vartheta_1} + \int_1^\infty e^{\vartheta_2 t}\,\frac{dt}{t^2} + e^{-\vartheta_3}\int_1^\infty e^{\vartheta_1 t+\vartheta_2 t^2}\,\frac{2dt}{t^3}\right], \quad \vartheta = (\vartheta_1,\vartheta_2,\vartheta_3)\,,$$

$\mathsf{dom}(\Lambda)$ is given by $\vartheta_2 < 0$ or $\vartheta_2 = 0$, $\vartheta_1 \leqslant 0$, and $\Theta = \mathsf{dom}(\Lambda)$. Consider $\Xi = \Theta$, thus $\mathcal{E}_\Xi = \mathcal{E}$, and the face $F = \{(0,t,0)\colon t>1\}$ of $\mathsf{cc}(\mu)$. This $F$ is not exposed and the unique access sequence to it is $\tau_1 = (0,0,1)$, $\tau_2 = (1,0,0)$. This access sequence is adapted to $\Xi$ but it would not be if $\Xi$ rather than $\mathsf{ri}(\Xi)$ had been used in the definition (4). Since $\pi_F(\Xi) = \{(0,t,0)\colon t \leqslant 0\} = \Theta_F$, and $F$ is $\Xi$-accessible, Theorem 2 gives that $\mathcal{E}^F \subseteq \mathsf{cl}_v(\mathcal{E})$. In particular, $\mathsf{cl}_v(\mathcal{E})$ contains $P$ that equals $Q_{F,\theta}$ with $\theta = (0,0,0) \in \pi_F(\Xi)$. On the other hand, as

$$\mathsf{M}(P) = \mathbb{R}\times\{0\}\times\mathbb{R}\,, \quad \Xi \cap (\theta + \mathsf{M}(P)) = \{(\vartheta_1,0,\vartheta_3)\colon \vartheta_1 \leqslant 0\}$$

and the unique access sequence to $F$ is not adapted to the latter set, $P$ is not in the variation closure of $\mathcal{E}_{\Xi\cap(\theta+\mathsf{M}(P))}$, by Theorem 2. Consequently $P \notin \mathsf{cl}_{rI}(\mathcal{E})$, by Theorem 3*(ii)*. Thus, $P$ cannot be the neat limit of any sequence in $\mathcal{E}$, showing that Lemma 7 is not valid when $\mathsf{ri}(\Xi)$ is replaced by $\Xi$.




Our final results address variation convergence of arbitrary sequences in $\mathsf{ext}(\mathcal{E})$.

**Theorem 4.** *If a sequence $Q_n$ in $\mathsf{ext}(\mathcal{E})$ converges in variation distance to a pm $P$ then $P$ belongs to a component $\mathcal{E}^F$ of $\mathsf{ext}(\mathcal{E})$, for sufficiently large $n$ the face $F_n$ of $\mathsf{cc}(\mu)$ with $Q_n \in \mathcal{E}^{F_n}$ contains $F$, the conditioned pm's $Q_n(\cdot|\mathsf{cl}(F))$ belong to $\mathcal{E}^F$, and also converge to $P$ in variation distance.*

Theorem 4 will be proved in Section 6.

**Corollary 3.** *The extension $\mathsf{ext}(\mathcal{E})$ of an exponential family $\mathcal{E}$ is variation closed.*

Corollary 3 strengthens [7, Theorem 2] on the $rI$-closedness of $\mathsf{ext}(\mathcal{E})$. Note that a family $\mathcal{E}$ completed by its "boundary at infinity" in the sense of [3] is not necessarily $rI$-closed, let alone variation closed, contrary to [3, Lemma 23.7], see [6, Example 3].

**Corollary 4.** *If a sequence $Q_n$ in $\mathsf{ext}(\mathcal{E})$ converges in variation distance to a pm $P$ and $D(P\|Q_n)$ is eventually finite then $D(P\|Q_n) \to 0$.*

The eventual finiteness takes place, in particular, if $P$ has a mean.

*Proof.* By Theorem 4, the variation convergence of $Q_n \in \mathcal{E}^{F_n}$ to $P$ implies $P \in \mathcal{E}^F$, $F \subseteq F_n$ eventually, and $Q_n(\cdot|\mathsf{cl}(F)) \to P$ in variation. Hence, the proof can be completed similarly to that of Theorem 3*(i)*, using eq. (5) with $\mathcal{E}^{F_n}$ playing the role of $\mathcal{E}$. □

## 4. Weak convergence in exponential families

In this section Theorem 1 is proved. For its second alternative, a corollary of the following lemma is needed. Note that

$$\rho(a) \triangleq \inf \{\|y-a\| : y \in \mathsf{aff}(\mu) \setminus \mathsf{cs}(\mu)\}$$

is obviously positive for any element $a$ of $\mathsf{ri}(\mu)$.

**Lemma 8.** *For $a \in \mathsf{ri}(\mu)$ and $0 < s < \rho(a)$ there exists a positive constant $C$ such that the inequality*

$$\langle \vartheta, b-a\rangle \geqslant r\,\|\vartheta\| - C\,e^{-s\|\vartheta\|}\big[\,r\,\|\vartheta\|e^{r\|\vartheta\|}+1\,\big] \tag{6}$$

*holds for all $0<r<s$, $\vartheta \in \Theta$, and $b \in \mathbb{R}^d$ satisfying $\langle \vartheta, b\rangle = \int_{\mathbb{R}^d} \langle \vartheta, x\rangle\, Q_\vartheta(dx)$.*

In particular, (6) holds when $b$ is the mean of $Q_\vartheta$. What will be actually used is the following consequence of Lemma 8.

**Corollary 5.** *Let $\vartheta_n$ be a sequence in $\Theta$ such that each pm $Q_{\vartheta_n}$ has a mean $b_n$. If $\|\vartheta_n\| \to \infty$, $\vartheta_n/\|\vartheta_n\| \to \tau$ and $b_n \to b$ then $\langle \tau, b-a\rangle \geqslant \rho(a)$ for each $a \in \mathsf{ri}(\mu)$.*

*Proof.* For $a \in \mathsf{ri}(\mu)$ and $0<r<s<\rho(a)$, (6) implies

$$\langle \vartheta_n, b_n - a\rangle \geqslant r\,\|\vartheta_n\| - C\big[\,r\,\|\vartheta_n\|e^{-(s-r)\|\vartheta_n\|} + e^{-s\|\vartheta_n\|}\,\big]\,.$$

Dividing both sides by $\|\vartheta_n\|$ and letting $n \to \infty$, it follows that $\langle \tau, b-a\rangle \geqslant r$ for all $0<r<s<\rho(a)$. □

For the proof of Lemma 8 the following auxiliary lemma, a simple refinement of known assertions, see [3, Lemma 21.8] and [11, Proof of Theorem 3.1], is needed. Denote

$$A^{\vartheta}_{a,s} \triangleq \{x \in \mathbb{R}^d : \langle \vartheta, x-a\rangle \geqslant s\|\vartheta\|\}\,, \qquad \vartheta, a \in \mathbb{R}^d,\ s>0\,,$$

which is a closed half-space of distance $s$ from $a$ when $\vartheta \neq 0$.

**Lemma 9.** *If $a \in \mathsf{ri}(\mu)$ and $0<s<\rho(a)$ then*

$$c_{a,s} \triangleq \inf \{\mu(A^{\vartheta}_{a,s}) : \vartheta \in \mathsf{lin}(\mu)\}$$

*is positive and*

$$\Lambda(\vartheta) - \langle \vartheta, a\rangle - s\,\|\vartheta\| \geqslant \ln c_{a,s}\,, \qquad \vartheta \in \mathsf{lin}(\mu). \tag{7}$$

*Proof.* Let $\mu(A^{\vartheta_n}_{a,s}) \downarrow c_{a,s}$ for a sequence $\vartheta_n$ in $\mathit{lin}(\mu)$, that can be supposed to consist of unit vectors converging to some $\tau \in \mathit{lin}(\mu)$. Each $x \in \mathbb{R}^d$ with $\langle \tau, x-a\rangle > s$ belongs to $A^{\vartheta_n}_{a,s}$ eventually, thus the open halfspace given by the last inequality is covered by the union over all $m$ of the intersections $\bigcap_{n\geqslant m} A^{\vartheta_n}_{a,s}$. Since the $\mu$-measure of the halfspace if positive due to $s<\rho(a)$, one of these intersections has positive $\mu$-measure. Thus, $c_{a,s}>0$. Finally, the inequalities

$$e^{\Lambda(\vartheta)-\langle\vartheta,a\rangle} = \int_{\mathbb{R}^d} e^{\langle\vartheta,x-a\rangle}\,\mu(dx) \geqslant \int_{A^{\vartheta}_{a,s}} e^{\langle\vartheta,x-a\rangle}\,\mu(dx) \geqslant e^{s\|\vartheta\|}\cdot\mu(A^{\vartheta}_{a,s})\,,$$

are valid for any $\vartheta \in \mathit{lin}(\mu)$ and imply (7). ☐

*Proof of Lemma 8.* Abbreviate $A^{\vartheta}_{a,r}$ to $A$ and its complement to $B$. Then

$$\langle\vartheta, b-a\rangle = \int_{\mathbb{R}^d} \langle\vartheta, x-a\rangle\, Q_\vartheta(dx) \geqslant \int_B \langle\vartheta, x-a\rangle\, Q_\vartheta(dx) + r\,\|\vartheta\|\, Q_\vartheta(A)$$

and, in turn,

$$\begin{aligned}\langle\vartheta, b-a\rangle - r\,\|\vartheta\| &\geqslant \int_B \big[\langle\vartheta, x-a\rangle - r\|\vartheta\|\big]\, Q_\vartheta(dx)\\ &= e^{\langle\vartheta,a\rangle-\Lambda(\vartheta)} \int_B \big[\langle\vartheta, x-a\rangle - r\|\vartheta\|\big]\, e^{\langle\vartheta,x-a\rangle}\,\mu(dx)\,.\end{aligned}$$

Since $te^t > -1$ for each $t\in\mathbb{R}$ and $e^{\langle\vartheta,x-a\rangle} < e^{r\|\vartheta\|}$ for $x\in B$, the last integral is bounded below by

$$-\big[1 + r\|\vartheta\|\,e^{r\,\|\vartheta\|}\big]\,\mu(B)\,.$$

It follows that

$$\langle\vartheta, b-a\rangle - r\,\|\vartheta\| \geqslant -\,e^{\langle\vartheta,a\rangle-\Lambda(\vartheta)+s\,\|\vartheta\|}\,\mu(\mathbb{R}^d)\cdot e^{-s\,\|\vartheta\|}\,\big[r\,\|\vartheta\|\,e^{r\|\vartheta\|}+1\big]\,.$$

This completes the proof, on account of (7). (The constant $C$ in (6) can be chosen as $\mu(\mathbb{R}^d)/c_{a,s}$.) ☐

*Proof of Theorem 1.* Let a sequence of pm's $Q_{\vartheta_n}$ with $\vartheta_n$ in $\Theta = \mathit{dom}(\Lambda)\cap\mathit{lin}(\mu)$ converge weakly to a pm $P$. Note that $P(\mathit{cs}(\mu))=1$ because $Q(\mathit{cs}(\mu))=1$ for every $Q\in\mathcal{E}$.

In the case when the sequence $\vartheta_n$ is bounded, let $\vartheta$ be the limit of an arbitrary convergent subsequence. By Lemma 2*(i)*, the weak convergence of $Q_{\vartheta_n}$ along this subsequence implies $\vartheta\in\mathit{dom}(\Lambda)$, and since clearly $\vartheta\in\mathit{lin}(\mu)$, also $\vartheta\in\Theta$. Moreover, $P=Q_\vartheta$ by Lemma 2, and since the parametrization (3) of $\mathcal{E}$ is bijective, it follows that each convergent subsequence of $\vartheta_n$ has the same limit. Thus $\vartheta_n$ converges to $\vartheta$. Applying Lemma 2, $\Lambda(\vartheta_n)\to\Lambda(\vartheta)$ follows from the weak convergence of $Q_{\vartheta_n}$ by *(i)*, and $Q_{\vartheta_n}\to P$ in variation by *(ii)*.

In the case when $\|\vartheta_n\|\to\infty$, assume with no loss of generality that $\vartheta_n/\|\vartheta_n\|$ converges to some $\tau$, clearly in $\mathit{lin}(\mu)$. Suppose first that $\mathit{cs}(\mu)$ is compact. Then the mean $b_n$ of $Q_{\vartheta_n}$ exists and converges to the mean $b$ of $P$. By Corollary 5, $\langle\tau, b-a\rangle\geqslant\rho(a)$ for each $a\in\mathit{ri}(\mu)$, and thus $\langle\tau, x-b\rangle\leqslant 0$ for each $x\in\mathit{cs}(\mu)$. Since obviously $b\in\mathit{cs}(\mu)$, it follows that $H_b=\{x\colon\langle\tau,x-b\rangle=0\}$ is a supporting hyperplane of $\mathit{cs}(\mu)$, and $G=H_b\cap\mathit{cs}(\mu)$ is a proper face of $\mathit{cs}(\mu)$, exposed by $\tau$. Then $P(\mathit{cs}(\mu))=1$ implies $P(G)=P(H_b)$ and this equals 1 because the mean $b$ of $P$ belongs to $H_b$.

Turning to the situation when $\mathit{cs}(\mu)$ is not compact, there exists a continuity set $B$ of $P$ such that $P(B)>0$ and $\mathit{lin}(\mu^B)=\mathit{lin}(\mu)$. Then the conditioned pm's $Q_{\vartheta_n}(\cdot|B)$ belong to the exponential family based on $\mu^B$, the parameters $\vartheta_n$ in $\Theta = \Theta_\mu$ belong also to $\Theta_{\mu^B}$, and $Q_{\vartheta_n}(\cdot|B)$ converges weakly to $P(\cdot|B)$. If, in addition, $B$ is compact then $P(\cdot|B)$ has a mean $b$ and the result proved above gives that $H_b$ is a nontrivial supporting hyperplane of $\mathit{cs}(\mu^B)$. Taking another compact continuity set $C$ of $P$ with $C \supseteq B$, let $c$ be the mean of $P(\cdot|C)$ and $H_c$ the corresponding supporting hyperplane of $\mathit{cs}(\mu^C)$. Since $P(H_b|B) = 1$, and $P(H_c|C) = 1$ together with $C \supseteq B$ imply $P(H_c|B) = 1$, the parallel hyperplanes $H_b$ and $H_c$ coincide. This proves that $H_b$ is a nontrivial supporting hyperplane of $\mathit{cs}(\mu^C)$ satisfying $P(H_b|C) = 1$ for those compact continuity sets $C$ of $P$ which contain $B$. Then $P(H_b) = 1$ and $H_b$ is a nontrivial supporting hyperplane to $\mathit{cs}(\mu)$, as well, because each $x \in \mathit{ri}(\mu)$ belongs to $\mathit{cs}(\mu^C)$ for some $C$ as above. Thus, $G = H_b \cap \mathit{cs}(\mu)$ is a proper face of $\mathit{cs}(\mu)$ exposed by $\tau$, and $P(G) = 1$, the same conclusions as when $\mathit{cs}(\mu)$ was compact.

Finally, $\vartheta_n$ cannot have both a convergent subsequence and another subsequence with norms tending to infinity. Indeed, by the above arguments, the former implies $P \in \mathcal{E}$ and the latter implies $P(G) = 1$ for a proper face of $\mathit{cs}(\mu)$, a contradiction. □

## 5. Proofs of Lemma 6 and Lemma 7

*Proof of Lemma 6.* By induction on the dimension of $\mathit{cs}(\mu)$. The case of zero dimension is trivial. The induction hypothesis assumes the assertion is true for canonically convex exponential families based on measures whose convex support has smaller dimension than that of the given $\mu$.

Given $Q_{\vartheta_n}$ with $\vartheta_n \in \Xi$ converging in variation distance to $P$, if the alternative *(i)* in Theorem 1 takes place then $P \in \mathcal{E}^F$ holds with $F = \mathit{cc}(\mu)$, obviously a $\Xi$-accessible face of $\mathit{cc}(\mu)$. Otherwise, by Theorem 1*(ii)*, $\|\vartheta_n\| \to +\infty$ and the limit $\tau_1$ of any convergent subsequence of $\vartheta_n/\|\vartheta_n\|$ exposes a proper face $G$ of $\mathit{cs}(\mu)$ with $P(G) = 1$. Note that $\vartheta_n \in \Theta \subseteq \mathit{lin}(\mu)$ implies $\tau_1 \in \mathit{lin}(\mu)$. The supporting hyperplane $H = \{x\colon \langle \tau_1, x - a\rangle = 0\}$ containing $G$ has positive $\mu$-measure, since $P(H) = P(G) = 1$ and $\mu$ dominates the variation limit $P$ of the pm's $Q_{\vartheta_n}$. It follows by Lemma 1 that $F_1 = H \cap \mathit{cc}(\mu)$ is a face of $\mathit{cc}(\mu)$, clearly exposed by $\tau_1$, and $\mu(H \setminus \mathit{cl}(F_1)) = 0$. Thus $P(H) = 1$ implies $P(\mathit{cl}(F_1)) = 1$. By the variation convergence $Q_{\vartheta_n} \to P$, it follows that $Q_{\vartheta_n}(\mathit{cl}(F_1)) \to 1$ and the conditioned pm's $Q_{\vartheta_n}(\cdot|\mathit{cl}(F_1))$ also converge to $P$ in variation distance. These pm's belong to $\mathcal{E}^{F_1}$, thus can be written as $Q_{F_1,\theta_n}$ with $\theta_n = \pi_{F_1}(\vartheta_n) \in \pi_{F_1}(\Xi)$. Hence, by the induction hypothesis applied to the canonically convex exponential family

$$\{Q_{F_1,\theta} \in \mathcal{E}^{F_1}\colon \theta \in \pi_{F_1}(\Xi)\}\,,$$

their variation limit $P$ belongs to $\mathcal{E}^F$ for a face $F$ of $\mathit{cc}(\mu^{\mathit{cl}(F_1)}) = F_1$ which is $\pi_{F_1}(\Xi)$-accessible, that is, an access sequence $\tau_2, \ldots, \tau_m$ to the face $F$ of $F_1$ is adapted to $\pi_{F_1}(\Xi)$. Since $\tau_1$ is the limit of a convergent subsequence of $\vartheta_n/\|\vartheta_n\|$ with $\vartheta_n \in \Xi$ it belongs to $\mathit{rec}(\mathit{ri}(\Xi))$ by [12, Theorem 8.2]. This and the adaptedness of $\tau_2, \ldots, \tau_m$ to $\pi_{F_1}(\Xi)$ imply by Lemma 3 that the access sequence $\tau_1, \ldots, \tau_m$ to the face $F$ of $\mathit{cc}(\mu)$ is adapted to $\Xi$. □

The following simple auxiliary assertion resembles [3, Lemma 21.7] and [7, Lemmas 6 and 7*(i)*].

**Lemma 10.** *If $\tau$ exposes a face $G$ of $\mathit{cc}(\mu)$ then $\tau$ belongs to the recession cone of $\mathit{dom}(\Lambda_\mu)$ and for every $\vartheta \in \mathit{dom}(\Lambda_\mu)$ the sequence $Q_{\vartheta+n\tau}$ converges neatly to $Q_\vartheta(\cdot|\mathit{cl}(G)) \in \mathcal{E}^G$.*

*Proof.* Since $\tau$ exposes $G$, $\langle\tau, x\rangle \leqslant \langle\tau, a\rangle$ for all $x \in \mathit{cc}(\mu)$ and $a \in G$, with equality if and only if $x \in G$. Then, for $t \geqslant 0$ the function $e^{\langle\vartheta+t\tau,x\rangle}$ of $x \in \mathit{cs}(\mu)$ is bounded above by $e^{\langle\vartheta,x\rangle} \cdot e^{t\langle\tau,a\rangle}$. This implies that $\vartheta + t\tau \in \mathit{dom}(\Lambda_\mu)$ whenever $\vartheta \in \mathit{dom}(\Lambda_\mu)$ and $t \geqslant 0$, proving $\tau \in \mathit{rec}(\mathit{dom}(\Lambda_\mu))$.

Knowing from Lemma 1 that $\mu(\mathit{cl}(G))$ equals the $\mu$-measure of the supporting hyperplane $H = \{x \colon \langle\tau, x - a\rangle = 0\}$ with $a \in G$, one has for any $\vartheta \in \mathit{dom}(\Lambda_\mu)$

$$\begin{aligned}\Lambda_\mu(\vartheta + n\tau) - \langle\vartheta + n\tau, a\rangle &= \ln\Big[\int_{\mathbb{R}^d} e^{\langle\vartheta+n\tau,x-a\rangle}\,\mu(dx)\Big]\\ &= \ln\Big[\int_{\mathit{cl}(G)} e^{\langle\vartheta,x-a\rangle}\,\mu(dx) + \int_{\mathit{cs}(\mu)\setminus H} e^{\langle\vartheta+n\tau,x-a\rangle}\,\mu(dx)\Big].\end{aligned}$$

When $n$ tends to infinity $\Lambda_\mu(\vartheta + n\tau) - \langle\vartheta + n\tau, a\rangle$ decreases to $\Lambda_G(\vartheta) - \langle\vartheta, a\rangle$, since the integral over $\mathit{cs}(\mu) \setminus H$ decreases to zero by dominated convergence. This fact and $e^{\langle\vartheta+n\tau,x\rangle} = e^{\langle\vartheta,x\rangle+n\langle\tau,a\rangle}$ for $x \in \mathit{cl}(G)$ imply

$$Q_{\vartheta+n\tau}(\mathit{cl}(G)) = e^{\langle\vartheta+n\tau,a\rangle-\Lambda(\vartheta+n\tau)}\int_{\mathit{cl}(G)} e^{\langle\vartheta,x-a\rangle}\,\mu(dx) \to \int_{\mathit{cl}(G)} e^{\langle\vartheta,x\rangle-\Lambda_G(\vartheta)}\,\mu(dx) = 1$$

and $Q_{\vartheta+n\tau}(\cdot|\mathit{cl}(G)) = Q_\vartheta(\cdot|\mathit{cl}(G))$, thus the neat convergence follows. □

*Proof of Lemma 7.* By induction. As in the proof of Lemma 6, the induction hypothesis assumes the assertion is true for exponential families based on measures whose convex support has smaller dimension than that of $\mu$.

The assertion trivially holds if $F = \mathit{cc}(\mu)$, thus suppose $F$ is proper face of $\mathit{cc}(\mu)$, and let $\tau_1, \ldots, \tau_m$ be an access sequence to $F$ adapted to $\Xi$. Let $F_1$ be the the face of $\mathit{cc}(\mu)$ exposed by $\tau_1$. Then, by Lemma 3, $\tau_1 \in \mathit{rec}(\mathit{ri}(\Xi))$ and the access sequence $\tau_2, \ldots, \tau_m$ to the face $F$ of $F_1$ is adapted to $\pi_{F_1}(\Xi)$, thus $F$ is a $\pi_{F_1}(\Xi)$-accessible face of $F_1$.

To prove that for $\vartheta \in \mathit{ri}(\Xi)$ there exists a sequence of pm's in $\mathcal{E}_{\mathit{ri}(\Xi)}$ that converges neatly to $Q_\vartheta(\cdot|\mathit{cl}(F))$, apply the induction hypothesis to the exponential family $\mathcal{E}^{F_1}$ based on $\mu^{\mathit{cl}(F_1)}$ with convex core $F_1$, to the $\pi_{F_1}(\Xi)$-accessible face $F$ of $F_1$, and to $\theta = \pi_{F_1}(\vartheta)$ in $\pi_{F_1}(\mathit{ri}(\Xi)) = \mathit{ri}(\pi_{F_1}(\Xi))$. It follows that some sequence $Q_{F_1,\theta_n}$ in $\mathcal{E}^{F_1}$ with $\theta_n \in \pi_{F_1}(\mathit{ri}(\Xi))$ converges neatly to the conditioning on $\mathit{cl}(F)$ of $Q_{F_1,\theta}$, which equals $Q_\vartheta(\cdot|\mathit{cl}(F))$ since $\theta = \pi_{F_1}(\vartheta)$. Here, on account of $\theta_n \in \pi_{F_1}(\mathit{ri}(\Xi))$, the pm $Q_{F_1,\theta_n}$ equals $Q_{\vartheta_n}(\cdot|\mathit{cl}(F_1))$ for some $\vartheta_n \in \mathit{ri}(\Xi)$.

Since $\tau_1 \in \mathit{rec}(\mathit{ri}(\Xi))$, Lemma 10 gives that the sequence $Q_{\vartheta_n+k\tau_1}$ converges neatly to $Q_{\vartheta_n}(\cdot|\mathit{cl}(F_1)) = Q_{F_1,\theta_n}$ where each $Q_{\vartheta_n+k\tau_1}$ is in $\mathcal{E}_{\mathit{ri}(\Xi)}$ due to $\tau_1 \in \mathit{rec}(\mathit{ri}(\Xi))$. The last assertion and the neat convergence of $Q_{F_1,\theta_n}$ to $Q_\vartheta(\cdot|\mathit{cl}(F))$ implies that for a suitable sequence $k_n \to \infty$, the pm's $Q_{\vartheta_n+k_n\tau_1}$ in $\mathcal{E}_{\mathit{ri}(\Xi)}$ converge neatly to $Q_\vartheta(\cdot|\mathit{cl}(F))$. □

## 6. Variation convergence in $\mathit{ext}(\mathcal{E})$

In this section Theorem 4 is proved. An auxiliary lemma is sent forward.

**Lemma 11.** *If $\mu$ dominates a pm $P$ then* *(i) there exists a face $F$ of* $\mathsf{cc}(\mu)$ *with* $P(\mathsf{cl}(F)) = 1$ *such that all faces $G$ of* $\mathsf{cc}(\mu)$ *with* $P(\mathsf{cl}(G)) = 1$ *contain $F$*

*(ii) if* $P(\mathsf{cl}(F_n)) \to 1$ *for a sequence $F_n$ of proper faces of* $\mathsf{cc}(\mu)$ *then the face $F$ of (i) is proper.*

*Proof. (i)* The closure $\mathsf{cs}(\mu)$ of $\mathsf{cc}(\mu)$ has full $\mu$-measure hence also full $P$-measure due to domination. Thus, the face $G = \mathsf{cc}(\mu)$ of $\mathsf{cc}(\mu)$ satisfies $P(\mathsf{cl}(G)) = 1$. Consider any face $G$ with the last property, and let $F$ be a face with that property whose dimension is smallest. Then both $\mu^{\mathsf{cl}(G)}$ and $\mu^{\mathsf{cl}(F)}$ dominate $P$, hence so does also the restriction of $\mu$ to $\mathsf{cl}(F) \cap \mathsf{cl}(G)$. By [6, Corollary 4], this intersection has the same $\mu$-measure as $\mathsf{cl}(F \cap G)$. Therefore, the restriction of $\mu$ to $\mathsf{cl}(F \cap G)$ dominates $P$, and thus $P(\mathsf{cl}(F \cap G)) = 1$. The minimality of the dimension of $F$ implies $F \subseteq G$.

*(ii)* The proper faces $F_n$ in the hypotheses can be supposed to be exposed, thus let a unit vector $\tau_n$ from $\mathsf{lin}(\mu)$ expose $F_n$ of $\mathsf{cc}(\mu)$. Then for $a \in \mathsf{ri}(\mu)$ the closed halfspace $\{x\colon \langle \tau_n, x - a\rangle \leqslant 0\}$ is disjoint with $\mathsf{cl}(F_n)$, thus its $P$-measure is at most $1 - P(\mathsf{cl}(F_n))$. It can be assumed that $\tau_n \to \tau$, and then as in the proof of Lemma 9,

$$\{x\colon \langle \tau, x - a\rangle < 0\} \subseteq \bigcup_{m\geqslant 1} \bigcap_{n\geqslant m} \{x\colon \langle \tau_n, x - a\rangle \leqslant 0\}\,.$$

Since $P(\{x\colon \langle \tau_n, x - a\rangle \leqslant 0\}) \leqslant 1 - P(\mathsf{cl}(F_n))$ and $P(\mathsf{cl}(F_n)) \to 1$, the open halfspace on the left hand side has $P$-measure zero whenever $a \in \mathsf{ri}(\mu)$. Hence, on account of $P(\mathsf{cs}(\mu)) = 1$, $\tau$ exposes a proper face of $\mathsf{cs}(\mu)$ that has full $P$-measure. Thus, there exists a nontrivial supporting hyperplane $H$ of $\mathsf{cs}(\mu)$ with $\mu(H) > 0$. By Lemma 1, $G = H \cap \mathsf{cc}(\mu)$ is a proper face of $\mathsf{cc}(\mu)$ and $\mu(H \setminus \mathsf{cl}(G)) = 0$. It follows that $P(\mathsf{cl}(G)) = P(H) = 1$. Hence, $G$ contains the face $F$ of *(i)* which implies that $F$ is proper. $\square$

*Proof of Theorem 4.* The variation limit $P$ of pm's $Q_n$ in $\mathsf{ext}(\mathcal{E})$ is obviously dominated by $\mu$, thus Lemma 11*(i)* applies to this $P$: let $F$ be the smallest face of $\mathsf{cc}(\mu)$ with closure of full $P$-measure. The variation convergence $Q_n \to P$ implies $Q_n(\mathsf{cl}(F)) \to P(\mathsf{cl}(F)) = 1$, and $Q_n \in \mathcal{E}^{F_n}$ implies $Q_n(\mathsf{cl}(F_n)) \to 1$. Since $Q_n(\mathsf{cl}(F_n) \cap \mathsf{cl}(F))$ equals $Q_n(\mathsf{cl}(F_n \cap F))$ by [6, Corollary 4] it follows that $Q_n(\mathsf{cl}(F_n \cap F)) \to 1$. Thus, again by the variation convergence, also $P(\mathsf{cl}(F_n \cap F)) \to 1$. If a subsequence of $F \cap F_n$ consisted of proper faces of $F$ the last limit relation would imply by Lemma 11*(ii)*, applied to $\mu^{\mathsf{cl}(F)}$ in the role of $\mu$, the existence of a proper face of $F = \mathsf{cc}(\mu^{\mathsf{cl}(F)})$ with closure of full $P$-measure, a contradiction to the choice of $F$. This proves that $F_n$ eventually contains $F$. The last inclusion implies that the conditioning $Q_n(\cdot|\mathsf{cl}(F))$ of $Q_n \in \mathcal{E}^{F_n}$ belongs to $\mathcal{E}^F$, and since $Q_n(\mathsf{cl}(F)) \to 1$, these conditionings also converge to $P$ in variation distance. Finally, applying Theorem 1 to the pm's $Q_n(\cdot|\mathsf{cl}(F))$ in $\mathcal{E}^F$, the alternative *(ii)* is ruled out by $F$ not having proper faces with closure of full $P$-measure, and it follows that $P \in \mathcal{E}^F$. $\square$

## References


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

[2] Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families.* Inst. of Math. Statist. Lecture Notes – Monograph Series, Vol. 9.

[3] Chentsov, N.N. (1982). *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs, Amer. Math. Soc., Providence – Rhode Island (Russian original: Nauka, Moscow, 1972).
[4] Csiszár, I. (1984). Sanov property, generalized $I$-projections, and a conditional limit theorem, *Ann. Probab.* **12**, 768–793.
[5] Csiszár, I. and Matúš, F. (2000). Information projections revisited, in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000, 490.
[6] Csiszár, I. and Matúš, F. (2001). Convex cores of measures on $\mathbb{R}^d$, *Studia Sci. Math. Hungar.* **38**, 177–190.
[7] Csiszár, I. and Matúš, F. (2003). Information projections revisited. *IEEE Trans. Inform. Theory*, vol. **49**, 1474–1490.
[8] Csiszár, I. and Matúš, F. (2003). On information closures of exponential families: a counterexample. Submitted to *IEEE Trans. Inform. Theory.*
[9] Harremoës, P. (2002). The information topology, in *Proc. 2002 IEEE Int. Symp. on Information Theory*, Lausanne, Switzerland, June/July 2002, 431.
[10] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2001). *Fundamentals of Convex Analysis.* Springer-Verlag, Berlin.
[11] Letac, G. (1992). *Lectures on Natural Exponential Families and their Variance Functions.* Monografias de Matemática **50**, Instituto de Matemática Pura e Aplicada, Rio de Janeiro.
[12] Rockafellar, R.T. (1970). *Convex Analysis.* Princeton University Press.