On Information Closures of Exponential Families: A Counterexample

Imre Csiszár *Fellow, IEEE,* and František Matúš

***Abstract*— A closure of an exponential family in reversed information divergence may be neither reverse information-closed nor log-convex.**

***Index Terms*— Exponential family, information closure, log-convexity.**

I. Introduction

In the context of information projections and maximum likelihood estimation in exponential families, two information theoretical closures of sets of probability measures (PMs) are relevant, see [1] and [3], [4], [5]. The reverse information closure (*rI*-closure) $cl_{rI}(\mathcal{S})$ of a family $\mathcal{S}$ of PMs on a measurable space consists of those PMs P for which $D(P\|\mathcal{S})$, the infimum of the information divergence (relative entropy) $D(P\|Q)$ subject to $Q \in \mathcal{S}$, is equal to zero. Changing the role of arguments, the collection of PMs P with $D(\mathcal{S}\|P)$ equal to zero is called the *I*-closure of $\mathcal{S}$. Neither *rI*- nor *I*-closure is a topological closure operation, but both are sequential closures in suitable topologies, see the discussion and references in [5].

In this correspondence two problems on the *rI*-closure of exponential families, left open in [5, Remark 7], are settled in negative. In Section 2 we give an example of a standard exponential family of PMs on $\mathbb{R}^3$ whose *rI*-closure is neither *rI*-closed nor log-convex. Here, a family $\mathcal{S}$ is *rI-closed* if $cl_{rI}(\mathcal{S}) = \mathcal{S}$, and *log-convex* if it contains all log-convex combinations of pairs of not mutually singular PMs in $\mathcal{S}$. These log-convex combinations are defined, for not mutually singular PMs P and Q with densities p and q w.r.t. a dominating measure μ, as the PMs with μ-densities $p^t q^{1-t} / \int p^t q^{1-t} d\mu$, $0 < t < 1$.

In Section 3 we show that $d = 3$ is the smallest dimension admitting such counterexamples. Section 2 is self-contained while Section 3 relies upon results of [5]. In the following, the information divergence is defined in terms of logarithms to the base e rather than 2.

Manuscript received December 17, 2002; revised October 13,2003. This work was supported by the Hungarian National Foundation for Scientific Research under Grants T 26041, T 32323, TS 40719, by Grant Agency of Academy of Sciences of the Czech Republic under Grant A1075104, and by Grant Agency of the Czech Republic under Grant 201/01/1482. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Yokohama, June/July 2003.

I. Csiszár is with A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1364 Budapest, P.O.Box 127, Hungary (e-mail: 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 (e-mail: matus@utia.cas.cz).

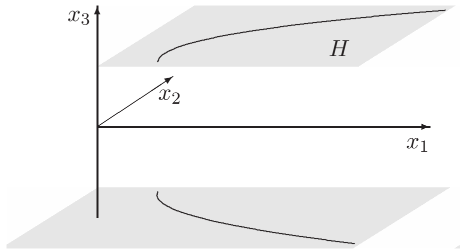

Fig. 1. Two curves support the f-image μ of λ.

II. The counterexample

Throughout this correspondence, only finite Borel measures on $\mathbb{R}^d$ are considered. The (standard, full) *exponential family* $\mathcal{E}_\mu$ based on a nonzero measure μ on $\mathbb{R}^d$ consists of the PMs $Q_{\mu,\vartheta}$ with μ-densities

$$\frac{dQ_{\mu,\vartheta}}{d\mu}(x) = e^{\langle\vartheta,x\rangle - \Lambda_\mu(\vartheta)}$$

where $\langle\vartheta, x\rangle = \sum_{i=1}^d \vartheta_i x_i$ denotes the inner product

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

and ϑ ranges over

$$dom(\Lambda_\mu) = \{\vartheta \colon \Lambda_\mu(\vartheta) < +\infty\}.$$

Obviously, $\mathcal{E}_\mu$ is log-convex.

Let λ be the measure on $\mathbb{R}$ absolutely continuous with respect to the Lebesgue measure, with density equal to t^{-2} when $|t| > 1$ and 0 otherwise. Let further μ be the image of λ on $\mathbb{R}^3$ under the mapping

$$f(t) = (|t|^3, \operatorname{sgn} t \cdot \ln|t|, \operatorname{sgn} t)$$

supported by two curves in $\mathbb{R}^3$ contained in two parallel planes, see Fig. 1.

Theorem 1: With μ as above, the *rI*-closure of $\mathcal{E}_\mu$ is neither *rI*-closed nor log-convex.

The proof is based on properties of the intersection of $cl_{rI}(\mathcal{E}_\mu)$ with the family $\mathcal{E}_\nu$ based on the restriction ν of μ to the plane $H = \{(x_1, x_2, x_3) \colon x_3 = 1\}$. Since

$$\begin{aligned}\Lambda_\nu(\tau) &= \ln \int_H e^{\langle \tau, x\rangle}\, \nu(dx) \\ &= \ln \left[e^{\tau_3} \int_1^{+\infty} e^{\tau_1 t^3} \cdot t^{\tau_2 - 2}\, dt \right]\end{aligned} \tag{1}$$

$\tau = (\tau_1, \tau_2, \tau_3)$ belongs to $dom(\Lambda_\nu)$ if and only if $\tau_1 < 0$ or $\tau_1 = 0$, $\tau_2 < 1$; note that the measure $Q_{\nu,\tau}$ does not depend on τ_3. Similarly

$$\begin{aligned}\Lambda_\mu(\vartheta) = \ln \Big[& e^{\vartheta_3} \int_1^{+\infty} e^{\vartheta_1 t^3} \cdot t^{\vartheta_2 - 2}\, dt \\ & + e^{-\vartheta_3} \int_1^{+\infty} e^{\vartheta_1 t^3} \cdot t^{-\vartheta_2 - 2}\, dt \Big]\end{aligned} \tag{2}$$

and ϑ is an element of $dom(\Lambda_\mu)$ if and only if $\vartheta_1 < 0$ or $\vartheta_1 = 0$, $|\vartheta_2| < 1$.

Lemma 1: For $\tau \in dom(\Lambda_\nu)$

$$D(Q_{\nu,\tau} \| \mathcal{E}_\mu) = \inf_{\vartheta \in dom(\Lambda_\mu)} D(Q_{\nu,\tau} \| Q_{\nu,\vartheta})\,. \tag{3}$$

Note that $D(Q_{\nu,\tau} \| \mathcal{E}_\mu)$ equals by definition the right-hand side of (3) with $Q_{\nu,\vartheta}$ replaced by $Q_{\mu,\vartheta}$.

Proof: The divergence $D(Q_{\nu,\tau} \| Q_{\mu,\vartheta})$ for ϑ in $dom(\Lambda_\mu)$ is the integral of

$$\ln \frac{dQ_{\nu,\tau}}{dQ_{\mu,\vartheta}} = \ln \frac{dQ_{\nu,\tau}}{dQ_{\nu,\vartheta}} + \ln \frac{dQ_{\nu,\vartheta}}{dQ_{\mu,\vartheta}}$$

w.r.t. the PM $Q_{\nu,\tau}$. Since the right-most term equals $\Lambda_\mu(\vartheta) - \Lambda_\nu(\vartheta)$ on H, where the PM $Q_{\nu,\tau}$ is concentrated, it follows that

$$D(Q_{\nu,\tau} \| Q_{\mu,\vartheta}) = D(Q_{\nu,\tau} \| Q_{\nu,\vartheta}) + \Lambda_\mu(\vartheta) - \Lambda_\nu(\vartheta)\,. \tag{4}$$

Equations (1) and (2) show that $\Lambda_\mu(\vartheta) \geqslant \Lambda_\nu(\vartheta)$, and for fixed ϑ_1, ϑ_2 ($\vartheta_1 < 0$ or $\vartheta_1 = 0$, $|\vartheta_2| < 1$) and $\vartheta_3 \to +\infty$ the difference $\Lambda_\mu(\vartheta) - \Lambda_\nu(\vartheta)$ goes to zero. Hence, (4) implies (3), recalling that $Q_{\nu,\vartheta}$ does not depend on ϑ_3. $\square$

Lemma 2: For $\tau = (\tau_1, \tau_2, 0)$ and $\vartheta = (\vartheta_1, \vartheta_2, 0)$ in $dom(\Lambda_\nu)$,

$$\begin{aligned}D(Q_{\nu,\tau} \| Q_{\nu,\vartheta}) &= \Lambda_\nu(\vartheta) - \Lambda_\nu(\tau) + e^{-\Lambda_\nu(\tau)} \\ &\cdot \int_1^{+\infty} \left[(\tau_1 - \vartheta_1)\, t^3 + (\tau_2 - \vartheta_2) \ln t \right] e^{\tau_1 t^3} \cdot t^{\tau_2 - 2}\, dt\,.\end{aligned} \tag{5}$$

Proof: Since ν is the restriction of μ to H,

$$\begin{aligned}& D(Q_{\nu,\tau} \| Q_{\nu,\vartheta}) \\ &= \int_H \ln \frac{e^{\langle \tau, x\rangle - \Lambda_\nu(\tau)}}{e^{\langle \vartheta, x\rangle - \Lambda_\nu(\vartheta)}} \cdot e^{\langle \tau, x\rangle - \Lambda_\nu(\tau)}\, \mu(dx) \\ &= \Lambda_\nu(\vartheta) - \Lambda_\nu(\tau) + \\ &\quad \int_1^{+\infty} \langle \tau - \vartheta, f(t)\rangle \cdot e^{\tau_1 t^3 + \tau_2 \ln t - \Lambda_\nu(\tau)} \cdot t^{-2}\, dt\end{aligned}$$

which immediately rewrites to (5). $\square$

Proposition 1: A PM $Q_{\nu,\tau} \in \mathcal{E}_\nu$ belongs to $cl_{rI}(\mathcal{E}_\mu)$ unless $\tau_1 = 0$, $\tau_2 \in [-2, -1)$.

Proof: Let $\tau = (\tau_1, \tau_2, 0) \in dom(\Lambda_\nu)$. If τ belongs to $dom(\Lambda_\mu)$ then the infimum of (3) is obviously zero, and hence Lemma 1 implies $Q_{\nu,\tau} \in cl_{rI}(\mathcal{E}_\mu)$. Otherwise, $\tau_1 = 0$ and $\tau_2 \leqslant -1$. If $\tau_2 < -2$ then (5) shows that for $\vartheta = (\vartheta_1, \tau_2, 0)$ with $\vartheta_1 \nearrow \tau_1 = 0$, the divergence $D(Q_{\nu,\tau} \| Q_{\nu,\vartheta})$ tends to zero. Thus again by Lemma 1, $Q_{\nu,\tau} \in cl_{rI}(\mathcal{E}_\mu)$.

In the remaining case $\tau_1 = 0$ and $-2 \leqslant \tau_2 \leqslant -1$, $D(Q_{\nu,\tau} \| Q_{\nu,\vartheta})$ equals $+\infty$ whenever $\vartheta_1 < 0$, by (5). Hence, the right-hand side of (3) equals the infimum of $D(Q_{\nu,\tau} \| Q_{\nu,\vartheta})$ subject to $\vartheta_1 = 0$ and $|\vartheta_2| < 1$. For $\tau_1 = \vartheta_1 = 0$, (5) shows that $D(Q_{\tau,\nu} \| Q_{\nu,\vartheta})$ is a continuous, convex function of $\vartheta_2 < 1$, attaining its global minimum 0 at $\vartheta_2 = \tau_2$. Therefore, its infimum subject to $|\vartheta_2| < 1$ is equal to $D(Q_{\nu,\tau} \| Q_{\nu,(0,-1,0)})$. This proves that the right-hand side of (3) is equal to $D(Q_{\nu,\tau} \| Q_{\nu,(0,-1,0)})$. Thus, $Q_{\nu,\tau}$ with $\tau_1 = 0$ and $-2 \leqslant \tau_2 \leqslant -1$ belongs to $cl_{rI}(\mathcal{E}_\mu)$ if and only if $\tau_2 = -1$, by Lemma 1. $\square$

Proof of Theorem 1: By Proposition 1, $Q_{\nu,(0,-2,0)}$ does not belong to $cl_{rI}(\mathcal{E}_\mu)$ while it is a log-convex combination of the two PMs $Q_{\nu,(0,-1,0)}$ and $Q_{\nu,(0,-3,0)}$ in the closure.

Further, $Q_{\nu,(0,-2,0)}$ belongs to $cl_{rI}(cl_{rI}(\mathcal{E}_\mu))$ because the infimum of $D(Q_{\nu,(0,-2,0)} \| Q_{\nu,\tau})$ subject to $\tau_1 = 0$ and $\tau_2 < -2$ equals 0, by (5). $\square$

Remark 1: A more general concept of exponential family admits a directional statistic different from the identity mapping. With λ and f as at the beginning of this section, the exponential family $\mathcal{E}_{\lambda,f}$ with directional statistic f consists of the PMs $Q_{\lambda,f,\vartheta}$ with λ-densities

$$\frac{dQ_{\lambda,f,\vartheta}}{d\lambda}(t) = e^{\langle \vartheta, f(t)\rangle - \Lambda_{\lambda,f}(\vartheta)}$$

where

$$\Lambda_{\lambda,f}(\vartheta) = \ln \int_{\mathbb{R}} e^{\langle \vartheta, f(t)\rangle}\, \lambda(dt)\,.$$

Clearly, $\Lambda_{\lambda,f} = \Lambda_\mu$, and the family $\mathcal{E}_{\lambda,f}$ is statistically equivalent to the standard family $\mathcal{E}_\mu$. Thus, an equivalent statement of Theorem 1 is that the *rI*-closure of $\mathcal{E}_{\lambda,f}$ is neither *rI*-closed nor log-convex.

III. Additional results

In this section we show that $d = 3$ is the smallest dimension for which a counterexample as in Section 2 exists. To this end, some results of [5] are recalled.

The *convex core* $cc(\mu)$ of a finite measure μ on $\mathbb{R}^d$ is the intersection of all convex Borel sets of full μ-measure. For a face F of $cc(\mu)$, $\mathcal{E}_\mu^F$ denotes the exponential family based on the restriction of μ to $cl(F)$, the closure of F. The *extension* of $\mathcal{E}_\mu$ is defined as

$$ext(\mathcal{E}_\mu) = \bigcup \left\{\mathcal{E}_\mu^F : F \text{ face of } cc(\mu)\right\}. \tag{6}$$

By [5, Theorem 2], $ext(\mathcal{E}_\mu)$ is both *rI*-closed and log-convex. In particular,

$$cl_{rI}(\mathcal{E}_\mu) \subseteq ext(\mathcal{E}_\mu). \tag{7}$$

Remark 2: The inclusion (7) and Proposition 1 enable us to describe explicitly $cl_{rI}(\mathcal{E}_\mu)$ for μ as in Section 2. The faces of $cc(\mu)$ are $cc(\mu)$ itself and (the proper faces) $F = cc(\mu) \cap H$, $F' = cc(\mu) \cap H'$, where $H = \{(x_1, x_2, x_3): x_3 = 1\}$, $H' = \{(x_1, x_2, x_3): x_3 = -1\}$. Here, F is the relatively open subset of the plane H bounded by the halfline $\{(t, 0, 1): t \geqslant 1\}$ and the curve $\{(t^3, \ln t, 1): t \geqslant 1\}$, see Fig. 1. The restrictions of μ to to $cl(F)$ and $cl(F')$, equal those to H and H', denoted by ν and ν', thus,

$$ext(\mathcal{E}_\mu) = \mathcal{E}_\mu \cup \mathcal{E}_\nu \cup \mathcal{E}_{\nu'}.$$

This provides the claimed description of $cl_{rI}(\mathcal{E}_\mu)$, since Proposition 1 determines the intersection $\mathcal{E}_\nu \cap cl_{rI}(\mathcal{E}_\mu)$, and $\mathcal{E}_{\nu'} \cap cl_{rI}(\mathcal{E}_\mu)$ is similarly determined by symmetry.

The auxiliary results in Lemmas 3 and 4 below are effectively known, but not formulated in the literature exactly as needed in the sequel. Lemma 3 is an extension of Lemma 1 to measures on $\mathbb{R}^d$. An *exposed face* of $cc(\mu)$ is the intersection of $cc(\mu)$ with one of its supporting hyperplanes. When Q belongs to $\mathcal{E}_\mu$, the PM $Q(\cdot|cl(F))$ obtained from Q by conditioning on $cl(F)$ belongs to $\mathcal{E}_\mu^F$, and $\{Q(\cdot|cl(F)): Q \in \mathcal{E}_\mu\}$ is a log-convex subfamily of $\mathcal{E}_\mu^F$.

Lemma 3: Let F be an exposed face of $cc(\mu)$. Then for each $P \in \mathcal{E}_\mu^F$

$$D(P\|\mathcal{E}_\mu) = \inf_{Q \in \mathcal{E}_\mu} D(P\|Q(\cdot|cl(F))). \tag{8}$$

Proof: As $P \in \mathcal{E}_\mu^F$ is concentrated on $cl(F)$

$$\begin{aligned} D(P\|Q) &= \int_{cl(F)} \ln \tfrac{dP}{dQ}\, dP \\ &= \int_{cl(F)} \ln \left[\tfrac{dP}{dQ(\cdot|cl(F))} \cdot \tfrac{dQ(\cdot|cl(F))}{dQ}\right] dP \\ &= D(P\|Q(\cdot|cl(F))) - \ln Q(cl(F)). \end{aligned} \tag{9}$$

By [5, Lemma 7], to any $Q \in \mathcal{E}_\mu$ there exists a sequence Q_n in $\mathcal{E}_\mu$ satisfying $Q_n(\cdot|cl(F)) = Q(\cdot|cl(F))$ for $n = 1, 2, \ldots$ and $Q_n(cl(F)) \to 1$. Then, using (9),

$$\begin{aligned} D(P\|Q_n) &= D(P\|Q_n(\cdot|cl(F))) - \ln Q_n(cl(F)) \\ &= D(P\|Q(\cdot|cl(F))) - \ln Q_n(cl(F)) \end{aligned}$$

and, therefore,

$$D(P\|Q(\cdot|cl(F))) = \lim_{n\to\infty} D(P\|Q_n). \tag{10}$$

Clearly, (9) and (10) imply (8). □

Lemma 4: Let μ be a finite measure on $\mathbb{R}$ such that both $cc(\mu)$ and $dom(\Lambda_\mu)$ are nontrivial intervals. For a PM P, and a sequence ϑ_n in $dom(\Lambda_\mu)$ not eventually constant, Q_{μ,ϑ_n} *rI*-converges to P if and only if either of the following holds.

(i) $P \in \mathcal{E}_\mu$, P has a (finite) mean, and $\vartheta_n \to \vartheta$ where ϑ parameterizes P;

(ii) $cc(\mu)$ contains its right/left endpoint a, P is the point mass at a, and $\vartheta_n \to \pm\infty$.

Proof: Sufficiency: When $P = Q_{\mu,\vartheta}$ has mean a then

$$D(P\|Q_{\mu,\vartheta_n}) = (\vartheta - \vartheta_n)\, a - \Lambda_\mu(\vartheta) + \Lambda_\mu(\vartheta_n) \tag{11}$$

and $\vartheta_n \to \vartheta$ implies the *rI*-convergence because, in the one dimensional case, the function Λ_μ is continous on $dom(\Lambda_\mu)$, see [1, Lemma 18.2]. If P is the point mass at a point a with $\mu(\{a\}) > 0$ then

$$\begin{aligned} D(P\|Q_{\mu,\vartheta_n}) &= -\ln Q_{\mu,\vartheta_n}(\{a\}) \\ &= -\vartheta_n a + \Lambda_\mu(\vartheta_n) - \ln \mu(\{a\}). \end{aligned} \tag{12}$$

If this a is the right/left endpoint of $cc(\mu)$ then $\vartheta_n \to \pm\infty$ implies that $\vartheta_n a - \Lambda_\mu(\vartheta_n)$ goes to $-\ln \mu(\{a\})$. Hence, the claimed *rI*-convergence follows.

Necessity: Suppose Q_{μ,ϑ_n} *rI*-converges to P. Then the divergences below are finite for n, m sufficiently large, and

$$D(P\|Q_{\mu,\vartheta_n}) - D(P\|Q_{\mu,\vartheta_m}) = \int_{\mathbb{R}} (\vartheta_m - \vartheta_n)\, x\, P(dx) - \Lambda_\mu(\vartheta_m) + \Lambda_\mu(\vartheta_n).$$

Hence, unless ϑ_n is eventually constant, P has a mean, say $a = \int x\, P(dx)$. Then, on account of $P \in ext(\mathcal{E}_\mu)$, either $P \in \mathcal{E}_\mu$ or P is the point mass at an endpoint a of $cc(\mu)$ with $\mu(\{a\}) > 0$.

In the first case, $P = Q_{\mu,\vartheta}$ for some $\vartheta \in dom(\Lambda_\mu)$ and its mean a is an interior point of the interval $cc(\mu)$. Since $D(P\|Q_{\mu,\vartheta_n})$ tends to zero, (11) implies that $\vartheta_n a - \Lambda_\mu(\vartheta_n)$ converges to $\vartheta a - \Lambda_\mu(\vartheta)$, which is known to be the maximum of $\tau a - \Lambda_\mu(\tau)$ subject to $\tau \in dom(\Lambda_\mu)$. Hence, by strict concavity of the last function, $\vartheta_n \to \vartheta$ follows.

In the second case, $D(P\|Q_{\mu,\vartheta_n}) \to 0$ and (12) imply that $\vartheta_n a - \Lambda_\mu(\vartheta_n)$ goes to $-\ln \mu(\{a\})$. When a is the right/left endpoint of $cc(\mu)$, then $dom(\Lambda_\mu)$ is unbounded from above/below and $-\ln \mu(\{a\})$ equals the supremum of the strictly increasing/decreasing function $\tau a - \Lambda_\mu(\tau)$. Hence, $\vartheta_n \to \pm\infty$ follows. □

Lemma 5: If $\mathcal{F}$ is a log-convex subfamily of an exponential family $\mathcal{E}_\mu$ of PMs on $\mathbb{R}$ then $cl_{rI}(\mathcal{F})$ is *rI*-closed and log-convex. The same is true also for log-convex subfamilies of $\mathcal{E}_\mu$ based on a measure μ on $\mathbb{R}^d$ with one-dimensional convex core.

Proof: Both assertions obviously follow from the first one with $\mathcal{F}$ consisting of more than one PM. Such a subfamily $\mathcal{F} = \{Q_{\mu,\vartheta} \colon \vartheta \in \Xi\}$ of $\mathcal{E}_\mu$, is log-convex if and only if Ξ is a nontrivial subinterval of $\mathit{dom}(\Lambda_\mu)$. Lemma 4 implies that a PM P belongs to $\mathit{cl}_{rI}(\mathcal{F}) \setminus \mathcal{F}$ if and only if either (i) $P = Q_{\mu,\vartheta}$ for an endpoint ϑ of Ξ such that $\vartheta \in \mathit{dom}(\Lambda_\mu) \setminus \Xi$ and $Q_{\mu,\vartheta}$ has a mean, or (ii) P is the point mass at the right/left endpoint of $\mathit{cc}(\mu)$ that belongs to $\mathit{cc}(\mu)$ and Ξ is unbounded from the right/left. Therefore, $\mathit{cl}_{rI}(\mathcal{F}) \setminus \mathcal{F}$ contains at most two PMs. In turn, if $D(P\|Q_n) \to 0$ for Q_n in $\mathit{cl}_{rI}(\mathcal{F})$ then this sequence is either eventually in $\mathcal{F}$ or contains a constant subsequence. Hence, $\mathit{cl}_{rI}(\mathcal{F})$ is *rI*-closed.

It follows from the above reasoning on endpoints of Ξ that $\mathit{cl}_{rI}(\mathcal{F}) \cap \mathcal{E}_\mu$ is parametrized by an interval and thus is log-convex. Since log-convex combinations involving a point mass equal the point mass, the log-convexity of $\mathit{cl}_{rI}(\mathcal{E}_\mu)$ follows. □

Theorem 2: The *rI*-closure of an exponential family of PMs on $\mathbb{R}^2$ is both *rI*-closed and log-convex.

Proof: *rI*-closedness: The aim is to prove that if a PM P is the *rI*-limit of a sequence of PMs Q_n in $\mathit{cl}_{rI}(\mathcal{E}_\mu)$ then $P \in \mathit{cl}_{rI}(\mathcal{E}_\mu)$. This is obvious if $P \in \mathcal{E}_\mu$ or if a subsequence of Q_n is in $\mathcal{E}_\mu$. Therefore we may suppose that $P \notin \mathcal{E}_\mu$, $Q_n \notin \mathcal{E}_\mu$ and $D(P\|Q_n)$ is finite for each n. Since the *rI*-closedness of $\mathit{ext}(\mathcal{E}_\mu)$ implies $P \in \mathit{ext}(\mathcal{E}_\mu)$ and $Q_n \in \mathit{ext}(\mathcal{E}_\mu)$, there are proper faces F and G_n of $\mathit{cc}(\mu)$ such that $P \in \mathcal{E}_\mu^F$ and $Q_n \in \mathcal{E}_\mu^{G_n}$. If F is a singleton, say $F = \{a\}$, then P is the unit mass at a, thus $D(P\|Q_n) = \ln Q_n(\{a\})$. Then the *rI*-convergence of Q_n to P is equivalent to $Q_n(\{a\}) \to 1$, and as $Q_n \in \mathit{cl}_{rI}(\mathcal{E}_\mu)$ implies existence of $Q'_n \in \mathcal{E}_\mu$ with $Q_n(\{a\}) - Q'_n(\{a\}) \to 0$, the assertion follows. In the remaining case when F is one dimensional, necessarily $G_n = F$ for all n, on account of $D(P\|Q_n) < +\infty$. Since any face of dimension one of $\mathit{cc}(\mu) \subseteq \mathbb{R}^2$ is exposed, Lemma 3 gives $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{E}_\mu) = \mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{F})$ where $\mathcal{F} = \{Q(\cdot|\mathit{cl}(F)) \colon Q \in \mathcal{E}_\mu\}$. Thus the *rI*-limit P of Q_n belongs to the *rI*-closure of $\mathit{cl}_{rI}(\mathcal{F})$. As $\mathit{cl}_{rI}(\mathcal{F})$ is *rI*-closed by Lemma 5, the assertion holds also in this case.

Log-convexity: On account of (6) and (7)

$$\mathit{cl}_{rI}(\mathcal{E}_\mu) = \bigcup \left\{ \mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{E}_\mu) \colon F \text{ face of } \mathit{cc}(\mu) \right\}.$$

Each intersection $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{E}_\mu)$ with one dimensional F equals $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{F})$ where $\mathcal{F} = \{Q(\cdot|\mathit{cl}(F)) \colon Q \in \mathcal{E}_\mu\}$, by Lemma 3. As $\mathcal{F}$ is log-convex, $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{F})$ is log-convex by Lemma 5. This implies that each intersection $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{E}_\mu)$ is log-convex. It remains to prove that $\mathit{cl}_{rI}(\mathcal{E}_\mu)$ contains the log-convex combinations of any PMs P and Q from the closure that belong to different components $\mathcal{E}_\mu^F$, $\mathcal{E}_\mu^G$ of $\mathit{ext}(\mathcal{E}_\mu)$ and are not mutually singular. The latter implies $\mu(\mathit{cl}(F) \cap \mathit{cl}(G)) > 0$ which is equivalent to $F \cap G \neq \emptyset$, by [2, Corollaries 3,4]. If F or G is a singleton, the log-convex combinations of P and Q are equal to P or Q. When F and G are both one-dimensional, then $F \cap G$ equals a singleton $\{a\}$, and the log-convex combinations equal the point mass at a. This belongs to $\mathit{cl}_{rI}(\mathcal{E}_\mu)$ by [5, Lemma 7] since $\{a\}$ is necessarily an exposed face of $\mathit{cc}(\mu)$. Finally, if F is one-dimensional and $G = \mathit{cc}(\mu)$ then the log-convex combinations of P and Q are equal to those of P and $Q(\cdot|\mathit{cl}(F))$, both in the log-convex family $\mathcal{E}_\mu^F \cap \mathit{cl}_{rI}(\mathcal{E}_\mu)$. □

References

[1] N.N. Chentsov, *Statistical Decision Rules and Optimal Inference*, ser. Translations of Mathematical Monographs. Providence, RI: Amer. Math. Soc., 1982. Original publication (in Russian). Moscow, U.S.S.R.: Nauka, 1972.

[2] I. Csiszár and F. Matúš, “Convex cores of measures on $\mathbb{R}^d$,” *Stud. Sci. Math. Hungar.*, vol. 38, pp. 177–190, 2001.

[3] ——, “Information projections revisited,” in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 940.

[4] ——, “Information closure of exponential families and generalized maximum likelihood estimates,” in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002, p. 434.

[5] ——, Information projections revisited. *IEEE Trans. Inform. Theory*, vol. 49, pp. 1474–1490, June 2003.