Around boundaries of exponential families

František Matúš

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
E-mail: matus@utia.cas.cz

April 16-18, 2012, Fields Institute, Toronto

μ ... a nonzero Borel measure on $\mathbb{R}^d$

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu : \mathbb{R}^d \to (-\infty, +\infty]$$

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

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu : \mathbb{R}^d \to (-\infty, +\infty]$$

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

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu : \mathbb{R}^d \to (-\infty, +\infty]$$

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

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$

(the cumulant generating function of μ)

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu : \mathbb{R}^d \to (-\infty, +\infty]$$
$$\vartheta \mapsto \ln \int_{\mathbb{R}^d} e^{\langle \vartheta, x\rangle} \mu(dx)$$

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$

(the cumulant generating function of μ)

$\Lambda = \Lambda_\mu$ ... is convex, lower-semicontinuous

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu \colon \mathbb{R}^d \to (-\infty, +\infty]$$

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

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$

(the cumulant generating function of μ)

$\Lambda = \Lambda_\mu$ ... is convex, lower-semicontinuous

$dom(\Lambda) = \{\vartheta \in \mathbb{R}^d \colon \Lambda(\vartheta) < +\infty\}$ ... the effective domain of Λ

μ ... a nonzero Borel measure on $\mathbb{R}^d$

The log-Laplace transform Λ_μ of μ

$$\Lambda_\mu \colon \mathbb{R}^d \to (-\infty, +\infty]$$

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

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$

(the cumulant generating function of μ)

$\Lambda = \Lambda_\mu$ ... is convex, lower-semicontinuous

$dom(\Lambda) = \{\vartheta \in \mathbb{R}^d \colon \Lambda(\vartheta) < +\infty\}$ ... the effective domain of Λ

From now on it is assumed that Λ is finite on a ball.

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \{ Q_\vartheta : \vartheta \in dom(\Lambda) \}$$

where $$\frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$$

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \left\{ Q_\vartheta \colon \vartheta \in dom(\Lambda) \right\}$$

where $\frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$

ϑ ... the canonical parameter

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \{ Q_\vartheta : \vartheta \in dom(\Lambda) \}$$

$$\text{where} \quad \frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$$

ϑ ... the canonical parameter

From now on it is assumed that $\vartheta \mapsto Q_\vartheta$ is one-to-one.

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \{ Q_\vartheta : \vartheta \in dom(\Lambda) \}$$

$$\text{where} \quad \frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$$

ϑ ... the canonical parameter

From now on it is assumed that $\vartheta \mapsto Q_\vartheta$ is one-to-one.

Equivalently, Λ is strictly convex,

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \left\{ Q_\vartheta \colon \vartheta \in dom(\Lambda) \right\}$$

where $\frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$

ϑ ... the canonical parameter

From now on it is assumed that $\vartheta \mapsto Q_\vartheta$ is one-to-one.

Equivalently, Λ is strictly convex,

or μ is not supported by a hyperplane,

The (standard) exponential family based on μ

$$\mathcal{E}_\mu = \{ Q_\vartheta : \vartheta \in dom(\Lambda) \}$$

where $\frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$

ϑ ... the canonical parameter

From now on it is assumed that $\vartheta \mapsto Q_\vartheta$ is one-to-one.

Equivalently, Λ is strictly convex,

or μ is not supported by a hyperplane,

or the convex support $cs(\mu)$ of μ has nonempty interior.

The (standard) exponential family based on μ

$\mathcal{E}_\mu = \{ Q_\vartheta : \vartheta \in dom(\Lambda) \}$

where $\frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)}, \quad x \in \mathbb{R}^d.$

ϑ ... the canonical parameter

From now on it is assumed that $\vartheta \mapsto Q_\vartheta$ is one-to-one.

Equivalently, Λ is strictly convex,

or μ is not supported by a hyperplane,

or the convex support $cs(\mu)$ of μ has nonempty interior.

(the smallest closed convex set with μ-negligible complement)

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)} \, \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in \mathit{int}(\mathit{dom}(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in \mathit{dom}(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}\, \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda' \colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle\vartheta,x\rangle - \Lambda(\vartheta)}\, \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda' \colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

Let M denote $\Lambda'(int(dom(\Lambda)))$ and ψ the inverse of Λ'.

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}\,\mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda' \colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

Let M denote $\Lambda'(int(dom(\Lambda)))$ and ψ the inverse of Λ'.

Thus, $Q_{\psi(a)}$ has the mean a, once $a \in M$, and

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)} \mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda' \colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

Let M denote $\Lambda'(int(dom(\Lambda)))$ and ψ the inverse of Λ'.

Thus, $Q_{\psi(a)}$ has the mean a, once $a \in M$, and

$$\{Q_\vartheta \colon \vartheta \in int(dom(\Lambda))\} = \{Q_{\psi(a)} \colon a \in M\}$$

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}\,\mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda' \colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

Let M denote $\Lambda'(int(dom(\Lambda)))$ and ψ the inverse of Λ'.

Thus, $Q_{\psi(a)}$ has the mean a, once $a \in M$, and

$$\{Q_\vartheta \colon \vartheta \in int(dom(\Lambda))\} = \{Q_{\psi(a)} \colon a \in M\}$$

... parametrization via means (N.E.F./F.E.N.)

The gradient of $\Lambda = \Lambda_\mu$ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda'(\vartheta) = \int_{\mathbb{R}^d} x \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}\,\mu(dx) = \int_{\mathbb{R}^d} x \cdot Q_\vartheta(dx)$$

is the mean of Q_ϑ.

For $\vartheta \in dom(\Lambda)$ general, Q_ϑ need not have mean.

Two different pm's in $\mathcal{E}_\mu$ cannot have the same mean.

Hence, $\Lambda'\colon int(dom(\Lambda)) \to int(cs(\mu))$ is injective.

Let M denote $\Lambda'(int(dom(\Lambda)))$ and ψ the inverse of Λ'.

Thus, $Q_{\psi(a)}$ has the mean a, once $a \in M$, and

$$\big\{Q_\vartheta\colon \vartheta \in int(dom(\Lambda))\big\} = \{Q_{\psi(a)}\colon a \in M\}$$

... parametrization via means (N.E.F./F.E.N.)

$M = int(cs(\mu))$ if and only if Λ is essentially smooth ($\mathcal{E}$ is steep).

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

… the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^*\colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^*\colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

The supremum in finite for $a \in dom(\Lambda^*)$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

The supremum in finite for $a \in dom(\Lambda^*)$

$$M \subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu)$$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

The supremum in finite for $a \in dom(\Lambda^*)$

$$M \subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu)$$

$$dom(\Lambda^*) = cc(\mu) + bar(dom(\Lambda)) \qquad \text{[Csi\&Ma 08]}$$

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^*\colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

The supremum in finite for $a \in dom(\Lambda^*)$

$$M \subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu)$$

$$dom(\Lambda^*) = cc(\mu) + bar(dom(\Lambda)) \qquad \text{[Csi\&Ma 08]}$$

$cc(\mu)$ ... the convex core of μ

Given a sample mean $a \in \mathbb{R}^d$, $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$

... the normalized log-likelihood of data w.r.t. $\mathcal{E}$

Maximum likelihood principle advises to maximize over ϑ.

The Fenchel conjugate $\Lambda^* \colon \mathbb{R}^d \to (-\infty, +\infty]$ of Λ

$$\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$$

If $a \in M$ then $\psi(a)$ is a maximizer and $Q_{\psi(a)}$ is the unique MLE.

A maximizer exists if and only if $a \in int(cs(\mu))$.

The supremum in finite for $a \in dom(\Lambda^*)$

$$M \subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu)$$

$$dom(\Lambda^*) = cc(\mu) + bar(dom(\Lambda))$$ [Csi&Ma 08]

$cc(\mu)$ ... the convex core of μ

$bar(C)$ ... the barrier cone of $C \subseteq \mathbb{R}^d$

The Hessian of Λ at $\vartheta \in \mathit{int}(\mathit{dom}(\Lambda))$ exists and

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda''(\vartheta) \text{ is the covariance matrix of } Q_\vartheta.$$

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and

$$\Lambda''(\vartheta) \text{ is the covariance matrix of } Q_\vartheta.$$

$$\Lambda''(\vartheta) = \int_{\mathbb{R}^d} [x - \Lambda'(\vartheta)]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx)$$

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and

$\Lambda''(\vartheta)$ is the covariance matrix of Q_ϑ.

$$\Lambda''(\vartheta) = \int_{\mathbb{R}^d} [x - \Lambda'(\vartheta)]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx)$$

where $y^{[2]} \in \mathbb{R}^{d \times d}$ for $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ denotes the matrix $(y_i \cdot y_j)_{i,j=1}^d$.

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and

$\Lambda''(\vartheta)$ is the covariance matrix of Q_ϑ.

$$\Lambda''(\vartheta) = \int_{\mathbb{R}^d} [x - \Lambda'(\vartheta)]^{[2]} \cdot e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)} \mu(dx)$$

where $y^{[2]} \in \mathbb{R}^{d\times d}$ for $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ denotes the matrix $(y_i \cdot y_j)_{i,j=1}^d$.

The variance function of μ

$$V = V_\mu : M \to \mathbb{R}^{d\times d}$$
$$a \mapsto \Lambda''(\psi(a))$$

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and

$\Lambda''(\vartheta)$ is the covariance matrix of Q_ϑ.

$$\Lambda''(\vartheta) = \int_{\mathbb{R}^d} [x - \Lambda'(\vartheta)]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx)$$

where $y^{[2]} \in \mathbb{R}^{d \times d}$ for $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ denotes the matrix $(y_i \cdot y_j)_{i,j=1}^d$.

The variance function of μ

$$V = V_\mu : M \to \mathbb{R}^{d \times d}$$
$$a \mapsto \Lambda''(\psi(a))$$

$\mathcal{E}_\mu = \mathcal{E}_\nu$ if and only if V_μ coincides with V_ν on a ball.

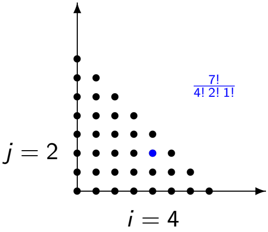
$d = 2$, $n \geqslant 1$

$d = 2$, $n \geqslant 1$

$\mu_n = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\, j!\, (n-i-j)!} \delta_{(i,j)}$ ... a finite measure on $\mathbb{R}^2$

$d = 2$, $n \geqslant 1$

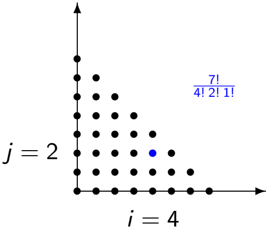
$\mu_n = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\,\delta_{(i,j)}$ ... a finite measure on $\mathbb{R}^2$



$n = 7$

$d = 2$, $n \geqslant 1$

$\mu_n = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\,\delta_{(i,j)}$ … a finite measure on $\mathbb{R}^2$

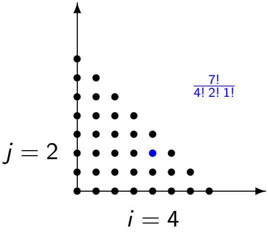


$n = 7$

$\vartheta = (\vartheta_1, \vartheta_2)$

$d = 2,\ n \geqslant 1$

$\mu_n = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\,\delta_{(i,j)}$ … a finite measure on $\mathbb{R}^2$



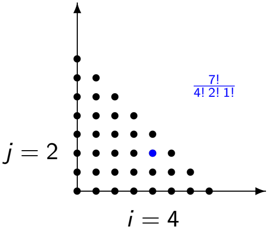
$n = 7$

$\vartheta = (\vartheta_1, \vartheta_2)$

$\Lambda(\vartheta) = \ln\left(e^{\vartheta_1} + e^{\vartheta_2} + 1\right)^n$ … the log-Laplace transform

$d = 2,\ n \geqslant 1$

$\mu_n = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\,\delta_{(i,j)}$ ... a finite measure on $\mathbb{R}^2$



$n = 7$

$\vartheta = (\vartheta_1, \vartheta_2)$

$\Lambda(\vartheta) = \ln\left(e^{\vartheta_1} + e^{\vartheta_2} + 1\right)^n$ ... the log-Laplace transform

$dom(\Lambda) = \mathbb{R}^2$ ... the effective domain of Λ

$\mathcal{E}_\mu = \{ Q_\vartheta \colon \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \}$ ... the exponential family

$\mathcal{E}_\mu = \{ Q_\vartheta \colon \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \}$ ... the exponential family

where $Q_\vartheta = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\, e^{\vartheta_1 i + \vartheta_2 j - \Lambda(\vartheta)} \delta_{(i,j)}$.

$\mathcal{E}_\mu = \{ Q_\vartheta \colon \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \}$ ... the exponential family

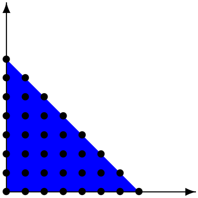
where $Q_\vartheta = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\, e^{\vartheta_1 i + \vartheta_2 j - \Lambda(\vartheta)} \delta_{(i,j)}$.

the convex support $cs(\mu)$ of μ is

$\mathcal{E}_\mu = \left\{ Q_\vartheta \colon \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \right\}$ ... the exponential family

where $Q_\vartheta = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!}\, e^{\vartheta_1 i + \vartheta_2 j - \Lambda(\vartheta)} \delta_{(i,j)}$.

the convex support $cs(\mu)$ of μ is



$n = 7$

$$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$$

... the mean of Q_ϑ

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ … the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

Its inverse at $a = (a_1, a_2) \in int(cs(\mu))$ is

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

Its inverse at $a = (a_1, a_2) \in int(cs(\mu))$ is

$$\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2} , \ln \frac{a_2}{n-a_1-a_2} \right)$$

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

Its inverse at $a = (a_1, a_2) \in int(cs(\mu))$ is

$$\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2} , \ln \frac{a_2}{n-a_1-a_2} \right)$$

$$Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!} a_1^i\, a_2^j\, (n - a_1 - a_2)^{n-i-j} \delta_{(i,j)}.$$

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1} , e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

Its inverse at $a = (a_1, a_2) \in int(cs(\mu))$ is

$$\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2} , \ln \frac{a_2}{n-a_1-a_2} \right)$$

$$Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\, j!\, (n-i-j)!} a_1^i\, a_2^j \,(n - a_1 - a_2)^{n-i-j} \delta_{(i,j)}.$$

$\mathcal{E}_\mu = \left\{ Q_{\psi(a)} \colon a \in int(cs(\mu)) \right\}$ ... 'multinomial family'

$\Lambda'(\vartheta) = \frac{n}{e^{\vartheta_1}+e^{\vartheta_2}+1} \left(e^{\vartheta_1}, e^{\vartheta_2} \right)$ ... the mean of Q_ϑ

Λ' is a bijection between $\mathbb{R}^2$ and

M ... the interior of the triangle $cs(\mu)$

Its inverse at $a = (a_1, a_2) \in int(cs(\mu))$ is

$$\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2}, \ln \frac{a_2}{n-a_1-a_2} \right)$$

$$Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i!\, j!\, (n-i-j)!}\, a_1^i\, a_2^j\, (n - a_1 - a_2)^{n-i-j} \delta_{(i,j)}.$$

$\mathcal{E}_\mu = \left\{ Q_{\psi(a)} \colon a \in int(cs(\mu)) \right\}$ ... 'multinomial family'

(of the dimension $d = 2$ with the parameter n)

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1) \end{bmatrix}$$

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1) \end{bmatrix}$$

... the variance of Q_ϑ

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1) \end{bmatrix}$$

... the variance of Q_ϑ

The variance function is matrix-valued,

$$\Lambda''(\vartheta) = \frac{n\cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2}\begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1)\end{bmatrix}$$

... the variance of Q_ϑ

The variance function is matrix-valued,

$$V(a) = \Lambda''(\psi(a)) = \frac{1}{n}\begin{bmatrix} a_1(n-a_1) & -a_1a_2 \\ -a_1a_2 & a_2(n-a_2)\end{bmatrix} = diag(a) - \frac{1}{n}a^{[2]}$$

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1) \end{bmatrix}$$

... the variance of Q_ϑ

The variance function is matrix-valued,

$$V(a) = \Lambda''(\psi(a)) = \frac{1}{n} \begin{bmatrix} a_1(n-a_1) & -a_1a_2 \\ -a_1a_2 & a_2(n-a_2) \end{bmatrix} = diag(a) - \frac{1}{n}a^{[2]}$$

... the variance of $Q_{\psi(a)}$

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1}+e^{\vartheta_2}+1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2}+1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1}+1) \end{bmatrix}$$

... the variance of Q_ϑ

The variance function is matrix-valued,

$$V(a) = \Lambda''(\psi(a)) = \frac{1}{n} \begin{bmatrix} a_1(n-a_1) & -a_1a_2 \\ -a_1a_2 & a_2(n-a_2) \end{bmatrix} = diag(a) - \frac{1}{n}a^{[2]}$$

... the variance of $Q_{\psi(a)}$

each entry is a bivariate polynomial in a_1, a_2
of the degree $\leqslant 2$ (V is quadratic)

In the topology of the total variation on pm's on $\mathbb{R}^d$,

In the topology of the total variation on pm's on $\mathbb{R}^d$,

what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?

In the topology of the total variation on pm's on $\mathbb{R}^d$,

what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?

($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?
($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?
($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

For a sequence $a_n \in M$ converging to $a \in dom(\Lambda^*)$,

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?

($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

For a sequence $a_n \in M$ converging to $a \in dom(\Lambda^*)$,
if $\Lambda^*(a_n) \to \Lambda^*(a)$ then $Q_{\psi(a_n)}$ converges [Csi&Ma 08, Thm 5.6]

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?
($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

For a sequence $a_n \in M$ converging to $a \in dom(\Lambda^*)$,
if $\Lambda^*(a_n) \to \Lambda^*(a)$ then $Q_{\psi(a_n)}$ converges [Csi&Ma 08, Thm 5.6]

In particular,

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?
($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

For a sequence $a_n \in M$ converging to $a \in dom(\Lambda^*)$,
if $\Lambda^*(a_n) \to \Lambda^*(a)$ then $Q_{\psi(a_n)}$ converges [Csi&Ma 08, Thm 5.6]

In particular,

if $a \in int(cs(\mu))$ then $Q_{\psi(a_n)}$ has the limit in $\mathcal{E}$;

In the topology of the total variation on pm's on $\mathbb{R}^d$,
what is behavior of $Q_{\psi(a_n)}$ for a convergent sequence $a_n \in M$?
($Q_{\psi(a)}$ is the MLE if a is an empirical mean)

$cl_v(\mathcal{E})$ matters, described in [Csi&Ma 05] via $cc(\mu)$

For a sequence $a_n \in M$ converging to $a \in dom(\Lambda^*)$,
if $\Lambda^*(a_n) \to \Lambda^*(a)$ then $Q_{\psi(a_n)}$ converges [Csi&Ma 08, Thm 5.6]

In particular,

if $a \in int(cs(\mu))$ then $Q_{\psi(a_n)}$ has the limit in $\mathcal{E}$;
if $a \in M$ then $Q_{\psi(a_n)} \to Q_{\psi(a)}$

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$

$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$

$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$

$$\varepsilon \mapsto V(a + \varepsilon(b - a))$$

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$

$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$

$$\varepsilon \mapsto V(a + \varepsilon(b - a))$$

...

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$

$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$

$$\varepsilon \mapsto V(a + \varepsilon(b - a))$$

...

Jørgensen, Martínez, Tsao (1994) V when $d = 1$

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b-a))$$

$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$

$$\varepsilon \mapsto V(a + \varepsilon(b-a))$$

...

Jørgensen, Martínez, Tsao (1994) V when $d = 1$

Masmoudi (1999) V when $d \geqslant 1$, under many restrictions

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$.

For $\varepsilon \downarrow 0$ what is behavior of

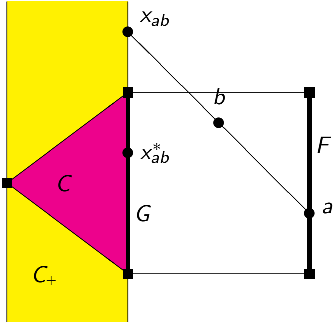
$$\varepsilon \mapsto \Lambda^*(a + \varepsilon(b - a))$$
$$\varepsilon \mapsto Q_{\psi(a+\varepsilon(b-a))}$$
$$\varepsilon \mapsto V(a + \varepsilon(b - a))$$
$$\dots$$

Jørgensen, Martínez, Tsao (1994) V when $d = 1$

Masmoudi (1999) V when $d \geqslant 1$, under many restrictions

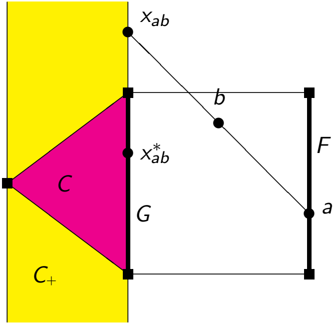
Matúš (2007) Λ^* when the support $s(\mu)$ of μ is finite

In the figure, μ is concentrated on the five black squares.



In the figure, μ is concentrated on the five black squares.

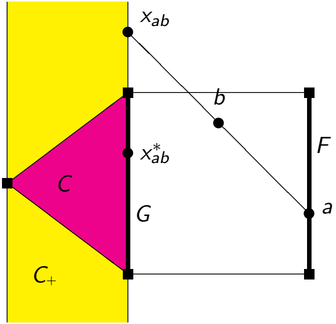
$cs(\mu)$ the pentagon



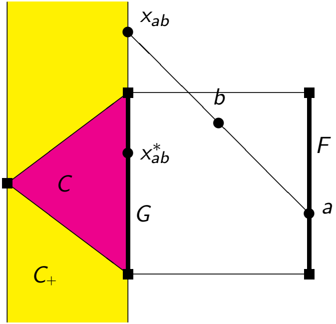
In the figure, μ is concentrated on the five black squares.

$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$



In the figure, μ is concentrated on the five black squares.

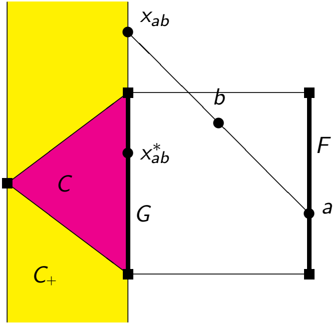


$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$

$b \in int(cs(\mu))$

In the figure, μ is concentrated on the five black squares.



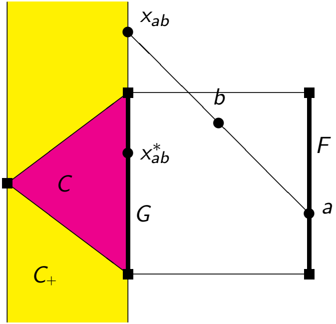
$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$

$b \in int(cs(\mu))$

C ... the convex hull of $s(\mu) \setminus F$

In the figure, μ is concentrated on the five black squares.



$cs(\mu)$ the pentagon

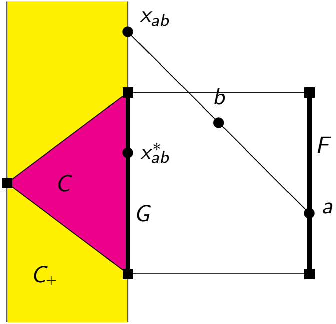
a inside a unique face F of $cs(\mu)$

$b \in int(cs(\mu))$

C ... the convex hull of $s(\mu) \setminus F$

$C_+ = C + lin(F)$ ... the strip

In the figure, μ is concentrated on the five black squares.



$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$

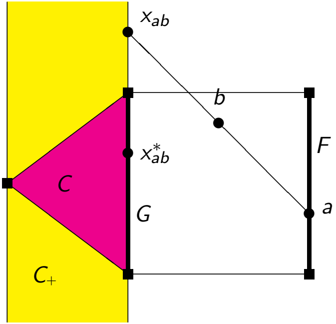
$b \in int(cs(\mu))$

C ... the convex hull of $s(\mu) \setminus F$

$C_+ = C + lin(F)$ ... the strip

x_{ab} ... a nearest point of C_+

In the figure, μ is concentrated on the five black squares.



$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$

$b \in int(cs(\mu))$

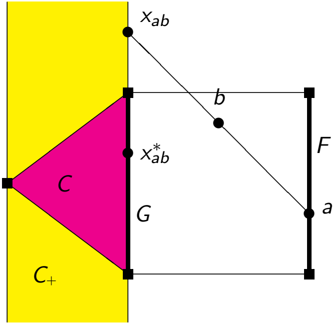
C ... the convex hull of $s(\mu) \setminus F$

$C_+ = C + lin(F)$ ... the strip

x_{ab} ... a nearest point of C_+

G ... a face of C

In the figure, μ is concentrated on the five black squares.



$cs(\mu)$ the pentagon

a inside a unique face F of $cs(\mu)$

$b \in int(cs(\mu))$

C … the convex hull of $s(\mu) \setminus F$

$C_+ = C + lin(F)$ … the strip

x_{ab} … a nearest point of C_+

G … a face of C

x^*_{ab} … a special point inside G

Theorem (M07)

If μ is a pm on $\mathbb{R}^d$ concentrated on a finite set, $a \in ri(F)$ for a proper face F of $cs(\mu)$, $b \in int(cs(\mu))$ and $\varepsilon > 0$ then

$$\Lambda^*(a + \varepsilon(x_{ab} - a)) = \Lambda^*(a) + \varepsilon \ln \varepsilon + \varepsilon \left[\Psi^*_{C,\Xi}(x_{ab}) - 1 - \Lambda^*(a) \right] + o(\varepsilon)$$

Theorem (M07)

If μ is a pm on $\mathbb{R}^d$ concentrated on a finite set, $a \in ri(F)$ for a proper face F of $cs(\mu)$, $b \in int(cs(\mu))$ and $\varepsilon > 0$ then

$$\Lambda^*(a + \varepsilon\,(x_{ab} - a)) = \Lambda^*(a) + \varepsilon \ln \varepsilon + \varepsilon \left[\Psi^*_{C,\Xi}(x_{ab}) - 1 - \Lambda^*(a) \right] + o(\varepsilon)$$

This approximation was applied to complete the first order conditions for a probability measure to be a maximizer of the divergence from an exponential family (Ay (2002)).

Theorem (unpubl)

Under the above assumptions,
if $x \in s(\mu)$ *then* $Q_{\psi(a+\varepsilon\,(x_{ab}-a))}(x)$ *equals*

$$\begin{cases} (1-\varepsilon)\cdot Q_{F,\psi_F(a)}(x) + \varepsilon \cdot \langle x_{ab} - x^*_{ab}, Q'_{F,\psi_F(a)}(x)\rangle\,, & x \in F\,, \\ \varepsilon \cdot Q_{G,\psi_G(x^*_{ab})}(x)\,, & x \in G\,, \\ 0\,, & \text{otherwise.} \end{cases}$$

up to $o(\varepsilon)$*-terms.*

Theorem (unpubl)

Under the above assumptions, $V(a+\varepsilon(x_{ab}-a))$ *equals*

$$(1-\varepsilon)\,V_F(a)+\varepsilon\left[\,(x_{ab}-x^*_{ab})V'_F(a)+V_G(x^*_{ab})+[x^*_{ab}-a]^{[2]}\,\right]$$

up to an $o(\varepsilon)$*-term.*

$\mathcal{F}_{1,r} = \{\mathsf{N}(a, r) \colon a \in \mathbb{R}\}$ $\quad V(a) = r \quad$ $r > 0$

$$\mathcal{F}_{1,r} = \{\mathrm{N}(a, r) \colon a \in \mathbb{R}\} \qquad V(a) = r \qquad r > 0$$

$$\mathcal{F}_2 = \{\mathrm{Poi}(a) \colon a > 0)\} \qquad V(a) = a$$

$\mathcal{F}_{1,r} = \{\text{N}(a, r) \colon a \in \mathbb{R}\}$ $\quad V(a) = r \quad r > 0$

$\mathcal{F}_2 = \{\text{Poi}(a) \colon a > 0)\}$ $\quad V(a) = a$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$$\mathcal{F}_{1,r} = \{\mathrm{N}(a, r) \colon a \in \mathbb{R}\} \qquad V(a) = r \qquad r > 0$$

$$\mathcal{F}_2 = \{\mathrm{Poi}(a) \colon a > 0)\} \qquad V(a) = a$$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$$\mathcal{F}_{3,n} = \{\mathrm{Bi}(a, n) \colon 0 < a < n\} \qquad V(a) = \frac{1}{n}\, a(n - a) \qquad n \geqslant 1$$

$$\mathcal{F}_{1,r} = \{\mathrm{N}(a,r)\colon a \in \mathbb{R}\} \qquad V(a) = r \qquad r > 0$$

$$\mathcal{F}_2 = \{\mathrm{Poi}(a)\colon a > 0)\} \qquad V(a) = a$$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$$\mathcal{F}_{3,n} = \{\mathrm{Bi}(a,n)\colon 0 < a < n\} \qquad V(a) = \tfrac{1}{n}\, a(n-a) \qquad n \geqslant 1$$

$$\mathcal{F}_{4,r} = \{\mathrm{NBi}(a,r)\colon a > 0\} \qquad V(a) = \tfrac{1}{r}\, a(n+a) \qquad r > 0$$

$\mathcal{F}_{1,r} = \{\text{N}(a,r)\colon a \in \mathbb{R}\}$ $\quad V(a) = r$ $\quad r > 0$

$\mathcal{F}_{2} = \{\text{Poi}(a)\colon a > 0)\}$ $\quad V(a) = a$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$\mathcal{F}_{3,n} = \{\text{Bi}(a,n)\colon 0 < a < n\}$ $\quad V(a) = \frac{1}{n}\, a(n-a)$ $\quad n \geqslant 1$

$\mathcal{F}_{4,r} = \{\text{NBi}(a,r)\colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a(n+a)$ $\quad r > 0$

$\mathcal{F}_{5,r} = \{\text{Ga}(a,r)\colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a^2$ $\quad r > 0$

$\mathcal{F}_{1,r} = \{\text{N}(a,r)\colon a \in \mathbb{R}\}$ $\quad V(a) = r$ $\quad r > 0$

$\mathcal{F}_2 = \{\text{Poi}(a)\colon a > 0)\}$ $\quad V(a) = a$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$\mathcal{F}_{3,n} = \{\text{Bi}(a,n)\colon 0 < a < n\}$ $\quad V(a) = \frac{1}{n}\, a(n-a)$ $\quad n \geqslant 1$

$\mathcal{F}_{4,r} = \{\text{NBi}(a,r)\colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a(n+a)$ $\quad r > 0$

$\mathcal{F}_{5,r} = \{\text{Ga}(a,r)\colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a^2$ $\quad r > 0$

$\mathcal{F}_{6,r} = \{\text{Ghs}(a,r)\colon a \in \mathbb{R}\}$ $\quad V(a) = \frac{1}{r}\, a^2 + r$ $\quad r > 0$

$\mathcal{F}_{1,r} = \{\text{N}(a, r) \colon a \in \mathbb{R}\}$ $\quad V(a) = r$ $\quad r > 0$

$\mathcal{F}_2 = \{\text{Poi}(a) \colon a > 0)\}$ $\quad V(a) = a$

Theorem (Morris 82)

If the variance function of an exponential family $\mathcal{E}_\mu$ on $\mathbb{R}$ equals a quadratic polynomial on the open interval M_μ then $\mathcal{E}_\mu$ is one of the families $\mathcal{F}_{1,\cdot} - \mathcal{F}_{6,\cdot}$ up to an affine transform.

$\mathcal{F}_{3,n} = \{\text{Bi}(a, n) \colon 0 < a < n\}$ $\quad V(a) = \frac{1}{n}\, a(n - a)$ $\quad n \geqslant 1$

$\mathcal{F}_{4,r} = \{\text{NBi}(a, r) \colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a(n + a)$ $\quad r > 0$

$\mathcal{F}_{5,r} = \{\text{Ga}(a, r) \colon a > 0\}$ $\quad V(a) = \frac{1}{r}\, a^2$ $\quad r > 0$

$\mathcal{F}_{6,r} = \{\text{Ghs}(a, r) \colon a \in \mathbb{R}\}$ $\quad V(a) = \frac{1}{r}\, a^2 + r$ $\quad r > 0$

... generalized hyperbolic secant

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

$d = 1$, VF cubic,   Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d \geqslant 1$, VF simple quadratic, Casalis (1996)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d \geqslant 1$, VF simple quadratic, Casalis (1996)

$d \geqslant 1$, VF simple cubic, Hassairi&Zarai (2006)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d \geqslant 1$, VF simple quadratic, Casalis (1996)

$d \geqslant 1$, VF simple cubic, Hassairi&Zarai (2006)

$d \geqslant 1$, Letac&Wesołowski (2008)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d \geqslant 1$, VF simple quadratic, Casalis (1996)

$d \geqslant 1$, VF simple cubic, Hassairi&Zarai (2006)

$d \geqslant 1$, Letac&Wesołowski (2008)

$d = 2$, each $V_{i,i}$ affine in a_i, Chachulska (2010)

$d = 1$, VF cubic, Letac&Mora (1990), ten types

$d = 1$, Λ' has a meromorphic extension,
Bar-Lev&Bshouty&Enis (1991)

Letac (1992) Lectures on NEF's and their VF's

The case $d = 2$ and VF quadratic is open.

$d \geqslant 1$, each diagonal element $V_{i,i}$ is a function of a_i
Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994)

$d \geqslant 1$, VF simple quadratic, Casalis (1996)

$d \geqslant 1$, VF simple cubic, Hassairi&Zarai (2006)

$d \geqslant 1$, Letac&Wesołowski (2008)

$d = 2$, each $V_{i,i}$ affine in a_i, Chachulska (2010)

...

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$

Theorem (unpubl)

If an EF has a quadratic VF and finite support
then it coincides with the product of multinomial families
up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$

intersect in an extreme point

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$ intersect in an extreme point

the restrictions of μ to F or E are known by induction

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$ intersect in an extreme point

the restrictions of μ to F or E are known by induction

the approximation of V around $a \in ri(F)$ is applied

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$ intersect in an extreme point

the restrictions of μ to F or E are known by induction

the approximation of V around $a \in ri(F)$ is applied

and combined with the assumption

Theorem (unpubl)

If an EF has a quadratic VF and finite support then it coincides with the product of multinomial families up to an affinity.

Proof by induction on d:

$d = 1$ by Morris classification

$d \geqslant 2$:

a facet F and edge E of the polytope $cs(\mu)$ intersect in an extreme point

the restrictions of μ to F or E are known by induction

the approximation of V around $a \in ri(F)$ is applied and combined with the assumption on the quadratic behaviour of VF.