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

Log-Laplace transform Exponential family Means Variances

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

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln\,\int_{\mathbb{R}^d}\;e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln \,\int_{\mathbb{R}^d} \, e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

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

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln\,\int_{\mathbb{R}^d}\,e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

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

(the cumulant generating function of μ)

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln\,\int_{\mathbb{R}^d}\,e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

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

(the cumulant generating function of μ)

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

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln\,\int_{\mathbb{R}^d}\,e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

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

(the cumulant generating function of μ)

 $\Lambda = \Lambda_{\mu} \dots$ is convex, lower-semicontinuous $dom(\Lambda) = \{ \vartheta \in \mathbb{R}^d : \Lambda(\vartheta) < +\infty \} \dots$ the effective domain of Λ

Log-Laplace transform Exponential family Means Variances

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

The log-Laplace transform Λ_{μ} of μ

$$egin{aligned} &\Lambda_\mu\colon \mathbb{R}^d o (-\infty,+\infty] \ &artheta\mapsto \ln\,\int_{\mathbb{R}^d}\,e^{\langleartheta,x
angle}\,\mu(dx) \end{aligned}$$

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

(the cumulant generating function of μ)

 $\Lambda = \Lambda_{\mu} \dots$ is convex, lower-semicontinuous $dom(\Lambda) = \{ \vartheta \in \mathbb{R}^d : \Lambda(\vartheta) < +\infty \} \dots$ 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

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathsf{dom}(\Lambda) \big\}$

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

The (standard) exponential family based on μ

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathsf{dom}(\Lambda) \big\}$

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

 ϑ ... the canonical parameter

The (standard) exponential family based on μ

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

where
$$rac{dQ_{artheta}}{d\mu}(x)=e^{\langle artheta,x
angle -\Lambda(artheta)},\ x\in\mathbb{R}^{d}.$$

 ϑ ... the canonical parameter

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

v

The (standard) exponential family based on μ

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathit{dom}(\Lambda) \big\}$

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

 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

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathit{dom}(\Lambda) \big\}$

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

 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 $\boldsymbol{\mu}$ is not supported by a hyperplane,

The (standard) exponential family based on μ

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathit{dom}(\Lambda) \big\}$

where
$$rac{dQ_{artheta}}{d\mu}(x)=e^{\langleartheta,x
angle-\Lambda(artheta)},\ x\in\mathbb{R}^{d}.$$

 ϑ ... the canonical parameter

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

W

Equivalently, Λ is strictly convex,

or $\boldsymbol{\mu}$ is not supported by a hyperplane,

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

The (standard) exponential family based on μ

 $\mathcal{E}_{\mu} = \big\{ \mathcal{Q}_{\vartheta} \colon \vartheta \in \mathit{dom}(\Lambda) \big\}$

where
$$rac{dQ_{artheta}}{d\mu}(x)=e^{\langleartheta,x
angle-\Lambda(artheta)},\ x\in\mathbb{R}^{d}.$$

 ϑ ... the canonical parameter

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

w

Equivalently, Λ is strictly convex,

or $\boldsymbol{\mu}$ 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)

Log-Laplace transform Exponential family Means Variances

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} . Exponential families Multinomial family Around boundary Quadratic VF Log-Laplace transform Exponential family Means Variances

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.

Exponential families Multinomial family Around boundary Quadratic VF Log-Laplace transform Exponential family Means Variances

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 $\vartheta \in \textit{dom}(\Lambda)$ general, \mathcal{Q}_{ϑ} need not have mean.

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

Exponential families Multinomial family Around boundary Quadratic VF Log-Laplace transform Exponential family Means Variances

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}_μ cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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}_{μ} cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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}_{μ} cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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}_{μ} cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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}_{μ} cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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 $\{Q_{\vartheta} : \vartheta \in int(dom(\Lambda))\} = \{Q_{\psi(a)} : a \in M\}$... 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}_μ cannot have the same mean. Hence, $\Lambda' : int(dom(\Lambda)) \rightarrow int(cs(\mu))$ 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 $\{Q_{\vartheta} : \vartheta \in int(dom(\Lambda))\} = \{Q_{\psi(a)} : a \in M\}$... 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).

Log-Laplace transform Exponential family Means Variances

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

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

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

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

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

Maximum likelihood principle advises to maximize over ϑ .

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

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

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. $\ensuremath{\mathcal{E}}$

Maximum likelihood principle advises to maximize over ϑ .

The Fenchel conjugate $\Lambda^* : \mathbb{R}^d \to (-\infty, +\infty]$ of Λ $\Lambda^*(a) = \sup_{\vartheta \in \mathbb{R}^d} \left[\langle \vartheta, a \rangle - \Lambda(\vartheta) \right]$

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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.

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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 $a \in int(cs(\mu))$.

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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 $a \in dom(\Lambda^*)$

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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 $a \in dom(\Lambda^*)$

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

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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]

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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

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 $a \in dom(\Lambda^*)$

$$\begin{split} M &\subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu) \\ dom(\Lambda^*) &= cc(\mu) + bar(dom(\Lambda)) \quad [\mathsf{Csi}\&\mathsf{Ma} \ \mathsf{08}] \\ &\quad cc(\mu) \ ... \ \text{the convex core of } \mu \end{split}$$

Exponential families	Log-Laplace transform
Multinomial family	
Around boundary	Means
Quadratic VF	

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

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

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 $a \in dom(\Lambda^*)$

$$\begin{split} M &\subseteq int(cs(\mu)) \subseteq dom(\Lambda^*) \subseteq cs(\mu) \\ dom(\Lambda^*) &= cc(\mu) + bar(dom(\Lambda)) \quad [\text{Csi}\&\text{Ma 08}] \\ & cc(\mu) \dots \text{ the convex core of } \mu \\ & bar(C) \dots \text{ the barrier cone of } C \subseteq \mathbb{R}^d \end{split}$$

Exponential families Multinomial family Around boundary Quadratic VF Variances Log-Laplace transfor Exponential family Means Variances

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

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and $\Lambda''(\vartheta)$ is the covariance matrix of Q_ϑ .



The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and $\Lambda''(\vartheta)$ is the covariance matrix of Q_{ϑ} . $\Lambda''(\vartheta) = \int_{\mathbb{R}^d} [x - \Lambda'(\vartheta)]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \mu(dx)$



The Hessian of arLambda at $artheta\in {\it int}({\it dom}(arLambda))$ exists and

$$\begin{split} \Lambda''(\vartheta) &\text{ is the covariance matrix of } Q_{\vartheta}.\\ \Lambda''(\vartheta) &= \int_{\mathbb{R}^d} \left[x - \Lambda'(\vartheta) \right]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx) \\ &\text{ where } y^{[2]} \in \mathbb{R}^{d \times d} \text{ for } y = (y_1, \dots, y_d) \in \mathbb{R}^d \\ &\text{ denotes the matrix } (y_i \cdot y_j)_{i,j=1}^d. \end{split}$$

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

$$\begin{split} \Lambda''(\vartheta) \text{ is the covariance matrix of } Q_{\vartheta}.\\ \Lambda''(\vartheta) &= \int_{\mathbb{R}^d} \left[x - \Lambda'(\vartheta) \right]^{[2]} \cdot e^{\langle \vartheta, x \rangle - \Lambda(\vartheta)} \, \mu(dx)\\ \text{ where } y^{[2]} \in \mathbb{R}^{d \times d} \text{ for } y = (y_1, \dots, y_d) \in \mathbb{R}^d\\ \text{ denotes the matrix } (y_i \cdot y_j)_{i, i=1}^d. \end{split}$$

The variance function of μ

$$V = V_{\mu} \colon M \to \mathbb{R}^{d \times d}$$

 $a \mapsto \Lambda''(\psi(a))$

The Hessian of Λ at $\vartheta \in int(dom(\Lambda))$ exists and $\Lambda''(\vartheta)$ is the covariance matrix of Ω

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

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

The variance function of μ

$$V = V_{\mu} \colon M \to \mathbb{R}^{d \times a}$$

 $a \mapsto \Lambda''(\psi(a))$

 $\mathcal{E}_{\mu} = \mathcal{E}_{\nu}$ if and only if V_{μ} coincides with V_{ν} on a ball.

Exponential families Multinomial family Around boundary Quadratic VF Log-Laplace transform Exponential family Means Variances

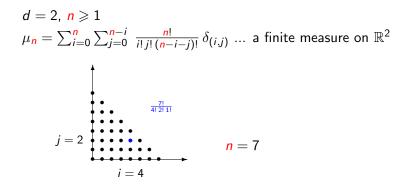
 $d = 2, n \ge 1$

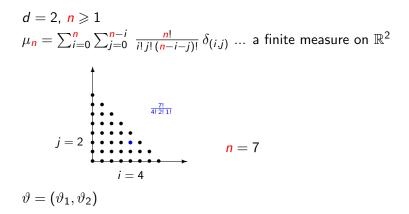
Exponential families Multinomial family Around boundary Quadratic VF

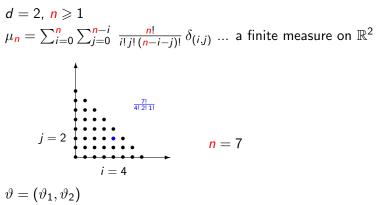
Log-Laplace transform Exponential family Means Variances

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

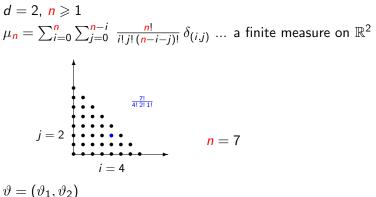
$$\mu_n = \sum_{i=0}^n \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} \delta_{(i,j)} \dots \text{ a finite measure on } \mathbb{R}^2$$







 $arLambda(artheta) = {\sf ln}\,\left(\,e^{artheta_1} + e^{artheta_2} + 1\,
ight)^n\,...\,$ the log-Laplace transform



 $\Lambda(\vartheta) = \ln \left(e^{\vartheta_1} + e^{\vartheta_2} + 1 \right)^n \dots \text{ the log-Laplace transform}$ $dom(\Lambda) = \mathbb{R}^2 \dots \text{ the effective domain of } \Lambda$

$$\mathcal{E}_\mu=\left\{ \mathcal{Q}_artheta\colonartheta=(artheta_1,artheta_2)\in\mathbb{R}^2
ight\}$$
 ... the exponential family

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

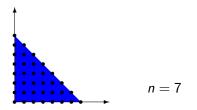
$$\text{ where } Q_{\vartheta} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} e^{\vartheta_1 i + \vartheta_2 j - \Lambda(\vartheta)} \delta_{(i,j)}.$$

$$\begin{aligned} \mathcal{E}_{\mu} &= \left\{ Q_{\vartheta} \colon \vartheta = \left(\vartheta_{1}, \vartheta_{2}\right) \in \mathbb{R}^{2} \right\} \ ... \ \text{the exponential family} \\ & \text{where} \ \ Q_{\vartheta} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \ \frac{n!}{i! j! (n-i-j)!} \ e^{\vartheta_{1}i + \vartheta_{2}j - \Lambda(\vartheta)} \delta_{(i,j)}. \end{aligned}$$

the convex support $cs(\mu)$ of μ is

$$\begin{aligned} \mathcal{E}_{\mu} &= \left\{ Q_{\vartheta} \colon \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \right\} \ ... \ \text{the exponential family} \\ \text{where} \ \ Q_{\vartheta} &= \sum_{i=0}^n \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} \ e^{\vartheta_1 i + \vartheta_2 j - \Lambda(\vartheta)} \delta_{(i,j)}. \end{aligned}$$

the convex support $cs(\mu)$ of μ is



Exponential families Multinomial family Around boundary Quadratic VF Variances

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

... the mean of Q_ϑ

Exponential families Multinomial family Around boundary Quadratic VF Variances Log-Laplace t Exponential family Means Variances

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

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

... 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

 \varLambda' 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
$$\mathcal{Q}_artheta$$

arLambda' 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

$$\Lambda'(artheta) = rac{n}{e^{artheta_1} + e^{artheta_2} + 1} \left(\, e^{artheta_1} \, , \, e^{artheta_2} \,
ight)$$

... the mean of
$$\mathcal{Q}_artheta$$

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

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

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

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

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

Its inverse at
$$a = (a_1, a_2) \in int(cs(\mu))$$
 is
 $\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2}, \ln \frac{a_2}{n-a_1-a_2} \right)$
 $Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} a_1^i a_2^j (n-a_1-a_2)^{n-i-j} \delta_{(i,j)}.$

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

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

Its inverse at
$$a = (a_1, a_2) \in int(cs(\mu))$$
 is
 $\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2}, \ln \frac{a_2}{n-a_1-a_2} \right)$
 $Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} a_1^i a_2^j (n-a_1-a_2)^{n-i-j} \delta_{(i,j)}.$

$$\mathcal{E}_{\mu} = ig\{ Q_{\psi({\sf a})} \colon {\sf a} \in {\it int}({\it cs}(\mu)) ig\} \ ... \ `multinomial family'$$

$$\Lambda'(\vartheta) = rac{n}{e^{\vartheta_1} + e^{\vartheta_2} + 1} \left(e^{\vartheta_1}, e^{\vartheta_2} \right) \qquad \qquad \dots \text{ the mean of } Q_{\vartheta}$$

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

Its inverse at
$$a = (a_1, a_2) \in int(cs(\mu))$$
 is
 $\psi(a) = \left(\ln \frac{a_1}{n-a_1-a_2}, \ln \frac{a_2}{n-a_1-a_2} \right)$
 $Q_{\psi(a)} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{n!}{i! j! (n-i-j)!} a_1^i a_2^j (n-a_1-a_2)^{n-i-j} \delta_{(i,j)}.$

$$\mathcal{E}_{\mu} = \left\{ Q_{\psi(a)} \colon a \in int(cs(\mu)) \right\} \dots \text{ `multinomial family'}$$
(of the dimension $d = 2$ with the parameter n)

Exponential families Multinomial family Around boundary Quadratic VF

Log-Laplace transform Exponential family Means Variances

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

Exponential families Multinomial family Around boundary Quadratic VF Variances

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

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

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

Exponential families Multinomial family Around boundary Quadratic VF Log-Laplace transform Exponential family Means Variances

$$\Lambda''(\vartheta) = \frac{n \cdot e^{\vartheta_1}}{(e^{\vartheta_1} + e^{\vartheta_2} + 1)^2} \begin{bmatrix} e^{\vartheta_1}(e^{\vartheta_2} + 1) & -e^{\vartheta_1}e^{\vartheta_2} \\ -e^{\vartheta_1}e^{\vartheta_2} & e^{\vartheta_2}(e^{\vartheta_1} + 1) \end{bmatrix}$$
... 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 ≤ 2 (*V* is quadratic) Exponential families Multinomial family Around boundary Quadratic VF Limiting with means Limiting along segments Approximation of A^* Approximation of Q_{ϑ} and V

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

Exponential families Limiting with means Multinomial family Around boundary Quadratic VF

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 $a_n \in M$?
 Exponential families
 Limiting with means

 Multinomial family
 Limiting along segments

 Around boundary
 Approximation of A^*

 Quadratic VF
 Approximation of Q_{ij} and V

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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_v(\mathcal{E})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_v(\mathcal{E})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_{v}(\mathcal{E})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_{v}(\mathcal{E})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_v(\mathcal{E})$ 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 $a_n \in M$? $(Q_{\psi(a)} \text{ is the MLE if } a \text{ is an empirical mean})$

 $cl_{v}(\mathcal{E})$ 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,

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

Exponential families Limiting with means Multinomial family Around boundary Quadratic VF Approximation of Q_{ij} and

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

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

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$. For $\varepsilon \downarrow 0$ what is behavior of

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

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

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$. For $\varepsilon \downarrow 0$ what is behavior of

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

. . .

Assume $a \in dom(\Lambda^*) \setminus int(cs(\mu))$ and $b \in int(cs(\mu))$. For $\varepsilon \downarrow 0$ what is behavior of

$$egin{aligned} arepsilon &\mapsto \Lambda^*(a+arepsilon(b-a))\ arepsilon &\mapsto Q_{\psi(a+arepsilon(b-a))}\ arepsilon &\mapsto V(a+arepsilon(b-a)) \end{aligned}$$

. . .

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

$$egin{aligned} &arepsilon\mapsto\Lambda^*(a+arepsilon(b-a))\ &arepsilon\mapsto Q_{\psi(a+arepsilon(b-a))}\ &arepsilon\mapsto V(a+arepsilon(b-a)) \end{aligned}$$

. . .

Jørgensen, Martínez, Tsao (1994) V when d = 1Masmoudi (1999) V when $d \ge 1$, under many restrictions

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

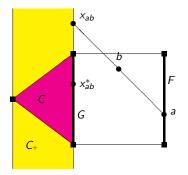
For $\varepsilon \downarrow 0$ what is behavior of

. . .

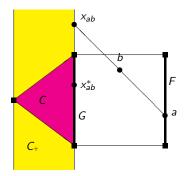
$$egin{array}{l} arepsilon\mapsto\Lambda^*({\sf a}+arepsilon(b-{\sf a}))\ arepsilon\mapsto Q_{\psi({\sf a}+arepsilon(b-{\sf a}))}\ arepsilon\mapsto V({\sf a}+arepsilon(b-{\sf a})) \end{array}$$

Jørgensen, Martínez, Tsao (1994) V when d = 1Masmoudi (1999) V when $d \ge 1$, 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.



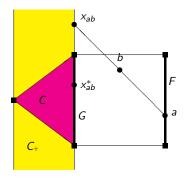
In the figure, μ is concentrated on the five black squares. $cs(\mu) \mbox{ the pentagon}$



In the figure, $\boldsymbol{\mu}$ is concentrated on the five black squares.

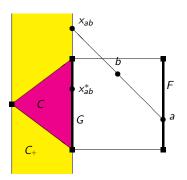
 $\mathit{cs}(\mu)$ the pentagon

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





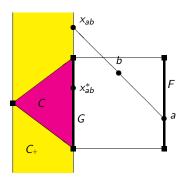
In the figure, $\boldsymbol{\mu}$ is concentrated on the five black squares.



 $cs(\mu)$ 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.

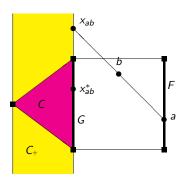


 $cs(\mu)$ the pentagon a inside a unique face F of $cs(\mu)$ $b \in int(cs(\mu))$

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

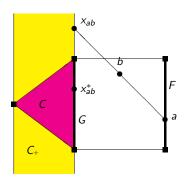


In the figure, $\boldsymbol{\mu}$ is concentrated on the five black squares.



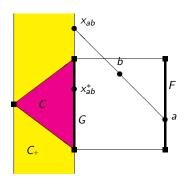
 $cs(\mu)$ the pentagon a inside a unique face F of $cs(\mu)$ $b \in int(cs(\mu))$ C ... the convex hull of $s(\mu) \setminus F$ $C_+ = C + lin(F)$... the strip

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



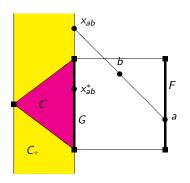
 $cs(\mu)$ the pentagon *a* inside a unique face *F* of $cs(\mu)$ $b \in int(cs(\mu))$ *C* ... the convex hull of $s(\mu) \setminus F$ $C_+ = C + lin(F)$... the strip x_{ab} ... a nearest point of C_+

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

$$\begin{split} \Lambda^*(\mathbf{a} + \varepsilon \left(\mathbf{x}_{ab} - \mathbf{a} \right)) &= \Lambda^*(\mathbf{a}) + \varepsilon \ln \varepsilon \\ &+ \varepsilon \left[\Psi^*_{\mathcal{C},\Xi}(\mathbf{x}_{ab}) - 1 - \Lambda^*(\mathbf{a}) \right] + o(\varepsilon) \end{split}$$

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

$$egin{aligned} &\Lambda^*(\mathbf{a} + arepsilon \left(\mathbf{x}_{\mathbf{a}\mathbf{b}} - \mathbf{a}
ight)
ight) = \Lambda^*(\mathbf{a}) + arepsilon \ln arepsilon \ &+ arepsilon \left[\Psi^*_{\mathcal{C}, arepsilon}(\mathbf{x}_{\mathbf{a}\mathbf{b}}) - 1 - \Lambda^*(\mathbf{a})
ight] + o(arepsilon) \end{aligned}$$

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

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

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

Exponential families	Limiting with means
Multinomial family	Limiting along segments
Around boundary	Approximation of Λ^*
Quadratic VF	Approximation of Q_{a} and V

Theorem (unpubl)

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

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

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

Morris classification Further classifications Multinomial families

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

r > 0

Morris classification Further classifications Multinomial families

$$\mathcal{F}_{1,r} = \{ \mathsf{N}(a,r) \colon a \in \mathbb{R} \}$$
 $V(a) = r$
 $\mathcal{F}_2 = \{ \mathsf{Poi}(a) \colon a > 0 \}$ $V(a) = a$

Morris classification Further classifications Multinomial families

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

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

Theorem (Morris 82)

Morris classification Further classifications Multinomial families

$$\begin{aligned} \mathcal{F}_{1,r} &= \{\mathsf{N}(a,r) \colon a \in \mathbb{R}\} & V(a) = r & r > 0 \\ \mathcal{F}_2 &= \{\mathsf{Poi}(a) \colon a > 0\} \} & V(a) = a \end{aligned}$$

Theorem (Morris 82)

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

Morris classification Further classifications Multinomial families

$$\begin{aligned} \mathcal{F}_{1,r} &= \{ \mathsf{N}(a,r) \colon a \in \mathbb{R} \} \qquad V(a) = r \qquad r > 0 \\ \mathcal{F}_2 &= \{ \mathsf{Poi}(a) \colon a > 0 \} \} \qquad V(a) = a \end{aligned}$$

Theorem (Morris 82)

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

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

Morris classification Further classifications Multinomial families

$$\begin{aligned} \mathcal{F}_{1,r} &= \{\mathsf{N}(a,r) \colon a \in \mathbb{R}\} & V(a) = r & r > 0 \\ \mathcal{F}_2 &= \{\mathsf{Poi}(a) \colon a > 0\} \} & V(a) = a \end{aligned}$$

Theorem (Morris 82)

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

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

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

Morris classification Further classifications Multinomial families

$$\begin{aligned} \mathcal{F}_{1,r} &= \{ \mathsf{N}(a,r) \colon a \in \mathbb{R} \} \qquad V(a) = r \qquad r > 0 \\ \mathcal{F}_2 &= \{ \mathsf{Poi}(a) \colon a > 0 \} \} \qquad V(a) = a \end{aligned}$$

Theorem (Morris 82)

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

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

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

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

Morris classification Further classifications Multinomial families

$$\begin{aligned} \mathcal{F}_{1,r} &= \{ \mathsf{N}(a,r) \colon a \in \mathbb{R} \} \qquad V(a) = r \qquad r > 0 \\ \mathcal{F}_2 &= \{ \mathsf{Poi}(a) \colon a > 0 \} \} \qquad V(a) = a \end{aligned}$$

Theorem (Morris 82)

 $\mathcal{F}_{6,r} = \{ \mathsf{Ghs}(a,r) \colon a \in \mathbb{R} \}$

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

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

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

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

$$V(a) = \frac{1}{r}a^2 + r \qquad r > 0$$

... generalized hyperbolic secant

Morris classification Further classifications Multinomial families

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

Morris classification Further classifications Multinomial families

d = 1, VF cubic, Letac&Mora (1990), ten types d = 1, Λ' has a meromorphic extension, Bar-Lev&Bshouty&Enis (1991)

Morris classification Further classifications Multinomial families

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

Morris classification Further classifications Multinomial families

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.

Morris classification Further classifications Multinomial families

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.

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

Morris classification Further classifications Multinomial families

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.

 $d \ge 1$, each diagonal element $V_{i,i}$ is a function of a_i Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994) $d \ge 1$, VF simple quadratic, Casalis (1996)

Morris classification Further classifications Multinomial families

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.

 $d \ge 1$, each diagonal element $V_{i,i}$ is a function of a_i Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994) $d \ge 1$, VF simple quadratic, Casalis (1996) $d \ge 1$, VF simple cubic, Hassairi&Zarai (2006)

Morris classification Further classifications Multinomial families

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.

 $d \ge 1$, each diagonal element $V_{i,i}$ is a function of a_i Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994) $d \ge 1$, VF simple quadratic, Casalis (1996) $d \ge 1$, VF simple cubic, Hassairi&Zarai (2006) $d \ge 1$, Letac&Wesołowski (2008)

Morris classification Further classifications Multinomial families

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.

 $d \ge 1$, each diagonal element $V_{i,i}$ is a function of a_i Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994) $d \ge 1$, VF simple quadratic, Casalis (1996) $d \ge 1$, VF simple cubic, Hassairi&Zarai (2006) $d \ge 1$, Letac&Wesołowski (2008) d = 2, each $V_{i,i}$ affine in a_i , Chachulska (2010)

Morris classification Further classifications Multinomial families

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.

. . .

 $d \ge 1$, each diagonal element $V_{i,i}$ is a function of a_i Bar-Lev&Bshouty&Enis&Letac&Lu&Richards (1994) $d \ge 1$, VF simple quadratic, Casalis (1996) $d \ge 1$, VF simple cubic, Hassairi&Zarai (2006) $d \ge 1$, Letac&Wesołowski (2008) d = 2, each $V_{i,i}$ affine in a_i , Chachulska (2010)

Morris classification Further classifications Multinomial families

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.

Morris classification Further classifications Multinomial families

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*:

Morris classification Further classifications Multinomial families

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

Morris classification Further classifications Multinomial families

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 $d \ge 2$:

Morris classification Further classifications Multinomial families

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 $d \ge 2$: a facet *F* and edge *E* of the polytope $cs(\mu)$

Morris classification Further classifications Multinomial families

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

 $d \ge 2$:

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

intersect in an extreme point

Morris classification Further classifications Multinomial families

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

 $d \ge 2$:

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

intersect in an extreme point

the restrictions of μ to ${\it F}$ or ${\it E}$ are known by induction

Morris classification Further classifications Multinomial families

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

 $d \ge 2$:

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

Morris classification Further classifications Multinomial families

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

 $d \ge 2$:

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 Exponential families Multinomial family Quadratic VE

Multinomial families

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

 $d \ge 2$:

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.