The Burbea-Rao and Bhattacharyya centroids

Frank Nielsen

www.informationgeometry.org

École Polytechnique, LIX, France

Sony Computer Science Laboratories, FRL, Japan

(joint work with Sylvain Boltz)

Léon Brillouin seminar

May 28, 2010 (Fri.)

Means and centroids

In Euclidean geometry, centroid c of a point set $\mathcal{P} = \{p_1, ..., p_n\}$:
Center of mass (also known as center of gravity):

$$\frac{1}{n}\sum_{i=1}^{n} p_i$$

Unique minimizer of average *squared* Euclidean distances

$$c = \arg\min_p \sum_{i=1}^{n} \frac{1}{n}\|p - p_i\|^2.$$

Two major ways to define means:

- by axiomatization, or
- by optimization (means defined by distances or penalty functions)

Means by axiomatization

Axioms for mean function $M(x_1, x_2)$:

- Reflexivity. $M(x, x) = x$,
- Symmetry. $M(x_1, x_2) = M(x_2, x_1)$,
- Continuity and strict monotonicity. $M(\cdot, \cdot)$ continuous and $M(x_1, x_2) < M(x'_1, x_2)$ for $x_1 < x'_1$, and
- Anonymity. $M(M(x_{11}, x_{12}), M(x_{21}, x_{22})) = M(M(x_{11}, x_{21}), M(x_{12}, x_{22}))$

x_{11}	x_{12}
x_{21}	x_{22}

Yields unique function f (up to an additive constant):

$$M(x_1, x_2) = f^{-1}\left(\frac{f(x_1) + f(x_2)}{2}\right) \overset{\text{equal}}{=} M_f(x_1, x_2)$$

f: continuous, strictly monotonous and increasing function.
(1930: Kolmogorov, Nagumo, + Aczél 1966)

Means by axiomatization: Quasi-arithmetic means

- arithmetic mean $\frac{x_1+x_2}{2} \longleftarrow f(x) = x$
- geometric mean $\sqrt{x_1 x_2} \longleftarrow f(x) = \log x$
- harmonic mean $\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} \longleftarrow f(x) = \frac{1}{x}$

Arithmetic barycenter on the f-representation ($y = f(x)$) :

$$M_f(x_1, ..., x_n; w_1, ..., w_n) = f^{-1}\left(\sum_{i=1}^n w_i f(x_i) = \bar{x}\right)$$

$$f(\bar{x}) = \sum_{i=1}^n w_i f(x_i)$$

$$\bar{y} = \sum_{i=1}^n w_i y_i$$

Dominance and interness of means

Dominance property:

$$M_f(x_1, ..., x_n; w_1, ..., w_n) < M_g(x_1, ..., x_n; w_1, ..., w_n),$$

if and only if g dominates f: $\forall x, g(x) > f(x)$.

Interness property:

$$\min(x_1, ..., x_n) \leq M_f(x_1, ..., x_n) \leq \max(x_1, ..., x_n),$$

limit cases $p \rightarrow \pm\infty$ of *power means* for $f(x) = x^p, p \in \mathbb{R}_*$.

$$M_p(x_1, ..., x_n) = \left(\sum_{i=1}^n w_i x_i^p\right)^{\frac{1}{p}}$$

name of power mean	value of p	formula
maximum	$\rightarrow +\infty$	$\max_i x_i$
quadratic mean (root mean square)	2	$\sqrt{\sum_i w_i x_i^2}$
arithmetic mean	1	$\sum_i w_i x_i$
geometric mean	$\rightarrow 0$	$\prod_i x_i^{w_i}$
harmonic mean	$\rightarrow -1$	$\frac{1}{\sum_i \frac{w_i}{x_i}}$
minimum	$\rightarrow -\infty$	$\min_i x_i$

also called Hölder means.

Means by optimization

$$(\mathrm{OPT}) : \min_x \sum_{i=1}^n w_i d(x, p_i) = \min_x L(x; \mathcal{P}, d),$$

Entropic means (Ben-Tal et al., 1989)

$$I_f(x, p) = p f\left(\frac{x}{p}\right),$$

$f(\cdot)$: strictly convex differentiable function with $f(1) = 0$ and $f'(1) = 0$.
entropic means: linear scale-invariant (homogeneous degree 1):

$$M(\lambda p_1, ..., \lambda p_n; I_f) = \lambda M(p_1, ..., p_n; I_f)$$

Bregman means

$$B_F(x,p) = F(x) - F(p) - (x-p)F'(p),$$

$F(\cdot)$: strictly convex and differentiable function.
(OPT) is convex $\rightarrow$ admits a unique minimizer:

$$M(p_1,...,p_n; B_F) = M_{F'}(p_1,...,p_n) = F'^{-1}\left(\sum_{i=1}^n w_i F'(p_i)\right)$$

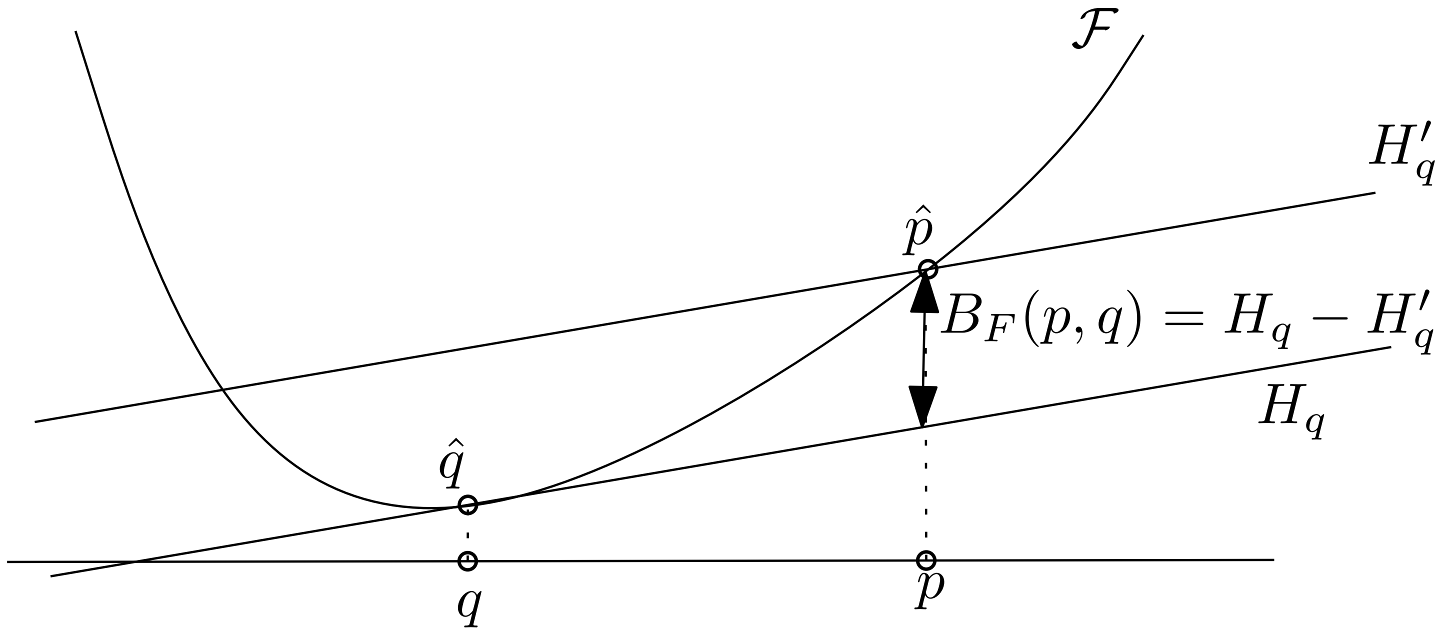
quasi-arithmetic mean for F', the derivative of F.

Since $d(x,p) \neq d(p,x)$, define a *right-sided* centroid M'

$$(\mathrm{OPT}') : \min_x \sum_{i=1}^n w_i d(p_i, x),$$

Visualizing Bregman divergences

$$B_F(p,q) = F(p) - F(q) - \langle p-q, \nabla F(q)\rangle,$$



- Kullback-Leibler ($F(x) = x\log x$): $\mathrm{KL}(p,q) = \sum_{i=1}^d p^{(i)} \log \frac{p^{(i)}}{q^{(i)}}$
- Squared Euclidean L_2^2 ($F(x) = x^2$):
  $L_2^2(p,q) = \sum_{i=1}^d (p^{(i)} - q^{(i)})^2 = \|p-q\|^2$

Information-theoretic sided means

Reference duality

- f-divergences

$$I_f(x,p) = I_{f*}(p,x),$$

  for $f^*(x) = xf(1/x)$.
  Any f-divergence can be symmetrized and stay in the class

- Bregman divergences

$$B_F(x,p) = B_{F^*}(F'(p), F'(x))$$

  for $F^*(\cdot)$ the Legendre convex conjugate ($F^{*\prime} = (F')^{-1}$)
  Only the squared Mahanalobis distances are symmetric Bregman divergences

Separable divergence and means as projections

Separable divergence:

$$d(x, p) = \sum_{i=1}^{d} d_i(x^{(i)}, p^{(i)}),$$

with $x^{(i)}$ denoting the i-th coordinate, and d_i's univariate divergences.
Typical non separable divergence : squared Mahalanobis distance (or other matrix trace divergences)

$$d(x, p) = (x - p)^T Q(x - p)$$

View means of separable divergence as a projection

$$(\mathrm{PROJ}) : \inf_{u \in U} d(u, p)$$

with $u_1 = ... = u_{d \times n} > 0$, and p the $(n \times d)$-dimensional point obtained by stacking the d coordinates of each of the n points.

Burbea-Rao divergences

Based on Jensen's inequality for a convex function F:

$$d(x,p) = \frac{F(x)+F(p)}{2} - F\left(\frac{x+p}{2}\right) \stackrel{\text{equal}}{=} \mathrm{BR}_F(x,p) \geq 0.$$

strictly convex function $F(\cdot)$.

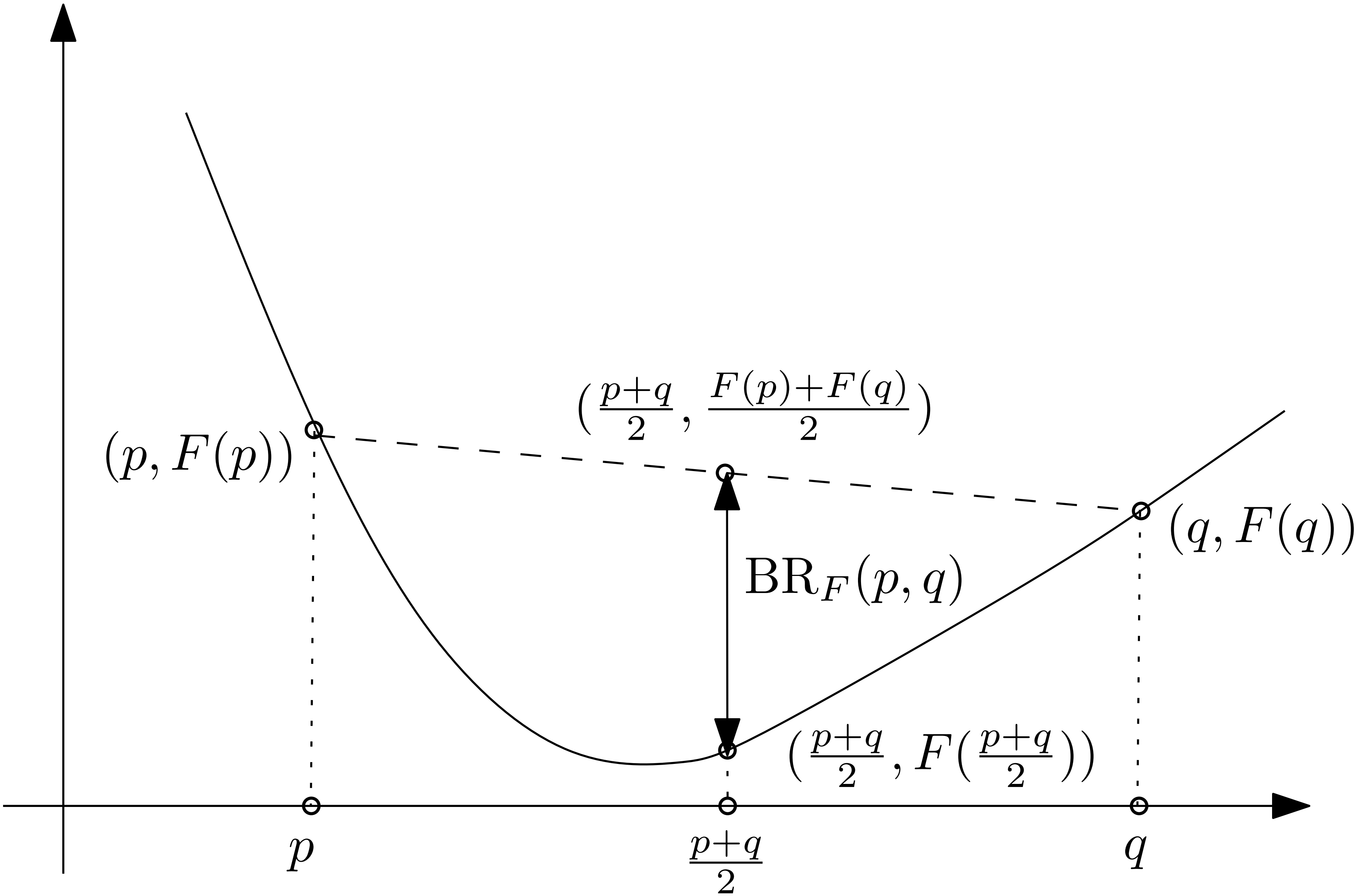
$$\mathrm{BR}_F(p,q) = \sum_{i=1}^{d} \mathrm{BR}_F(p^{(i)}, q^{(i)}),$$

Includes the special case of Jensen-Shannon divergence:

$$\mathrm{JS}(p,q) = H\left(\frac{p+q}{2}\right) - \frac{H(p)+H(q)}{2}$$

$F(x) = -H(x)$, the negative Shannon entropy $H(x) = -x\log x$.
$\rightarrow$ generators are convex and entropies are concave (negative generators)

Visualizing Burbea-Rao divergences



Burbea-Rao divergences: Squared Mahalanobis

$$\begin{aligned}
\mathrm{BR}_F(p,q) &= \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right) \\
&= \frac{2\langle Qp,p\rangle + 2\langle Qq,q\rangle - \langle Q(p+q), p+q\rangle}{4} \\
&= \frac{1}{4}(\langle Qp,p\rangle + \langle Qq,q\rangle - 2\langle Qp,q\rangle) \\
&= \frac{1}{4}\langle Q(p-q), p-q\rangle = \frac{1}{4}\|p-q\|_Q^2.
\end{aligned}$$

(Not a metric. square root of Jensen-Shannon is a metric but not the square roots of all Burbea-Rao divergences.)

Symmetrizing Bregman divergences

- **Jeffreys-Bregman divergences.**

$$
\begin{aligned}
S_F(p;q) &= \frac{B_F(p,q)+B_F(q,p)}{2} \\
&= \frac{1}{2}\langle p-q, \nabla F(p)-\nabla F(q)\rangle,
\end{aligned}
$$

- **Jensen-Bregman divergences (diversity index).**

$$
\begin{aligned}
J_F(p;q) &= \frac{B_F(p,\frac{p+q}{2})+B_F(q,\frac{p+q}{2})}{2} \\
&= \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right) = \mathrm{BR}_F(p,q)
\end{aligned}
$$

Skew Burbea-Rao divergences

$$\begin{aligned} \mathrm{BR}_F^{(\alpha)} &: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+ \\ \mathrm{BR}_F^{(\alpha)}(p,q) &= \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q) \end{aligned}$$

$$\begin{aligned} \mathrm{BR}_F^{(\alpha)}(p,q) &= \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q) \\ &= \mathrm{BR}_F^{(1-\alpha)}(q,p) \end{aligned}$$

Skew symmetrization of Bregman divergences:

$$\alpha B_F(p, \alpha p + (1-\alpha)q) + (1-\alpha) B_F(q, \alpha p + (1-\alpha)q) \stackrel{\text{equal}}{=} \mathrm{BR}_F^{(\alpha)}(p,q)$$

= skew Jensen-Bregman divergences.

Bregman as asymptotic skewed Burbea-Rao

$$
\begin{aligned}
B_F(p,q) &= \lim_{\alpha\rightarrow 1} \frac{1}{1-\alpha} \mathrm{BR}_F^{(\alpha)}(p,q) \\
B_F(q,p) &= \lim_{\alpha\rightarrow 0} \frac{1}{\alpha} \mathrm{BR}_F^{(\alpha)}(p,q)
\end{aligned}
$$

Proof: $F(\alpha p + (1-\alpha)q) = F(p + (1-\alpha)(q-p)) \simeq_{\alpha\simeq 1} F(p) + (1-\alpha)(q-p)\nabla F(p)$

$F(\alpha p+(1-\alpha)q)-\alpha F(p)-(1-\alpha)F(q) \overset{\text{Taylor}}{\simeq}_{\alpha\rightarrow 1} (1-\alpha)F(p)+(1-\alpha)(q-p)\nabla F(p)-(1-\alpha)F(q)$

$\simeq_{\alpha\rightarrow 1} (1-\alpha)\left(F(p)-F(q)-(p-q)\nabla F(p)\right)$

$\lim_{\alpha\rightarrow 1} \mathrm{BR}_F^{(\alpha)}(p,q) = (1-\alpha)B_F(p,q)$

For $0 < \alpha < 1$, swap arguments by setting $\alpha \rightarrow 1-\alpha$:

$$
\mathrm{BR}_F^{(\alpha)}(p,q) = \mathrm{BR}_F^{(1-\alpha)}(q,p)
$$

Burbea-Rao centroids

$$\mathrm{OPT}: c = \arg\min_x \sum_{i=1}^n w_i \mathrm{BR}_F^{(\alpha_i)}(x, p_i) = \arg\min_x L(x)$$

Wlog., equivalent to minimize

$$E(c) = (\sum_{i=1}^n w_i \alpha_i) F(c) - \sum_{i=1}^n w_i F(\alpha_i c + (1-\alpha_i) p_i)$$

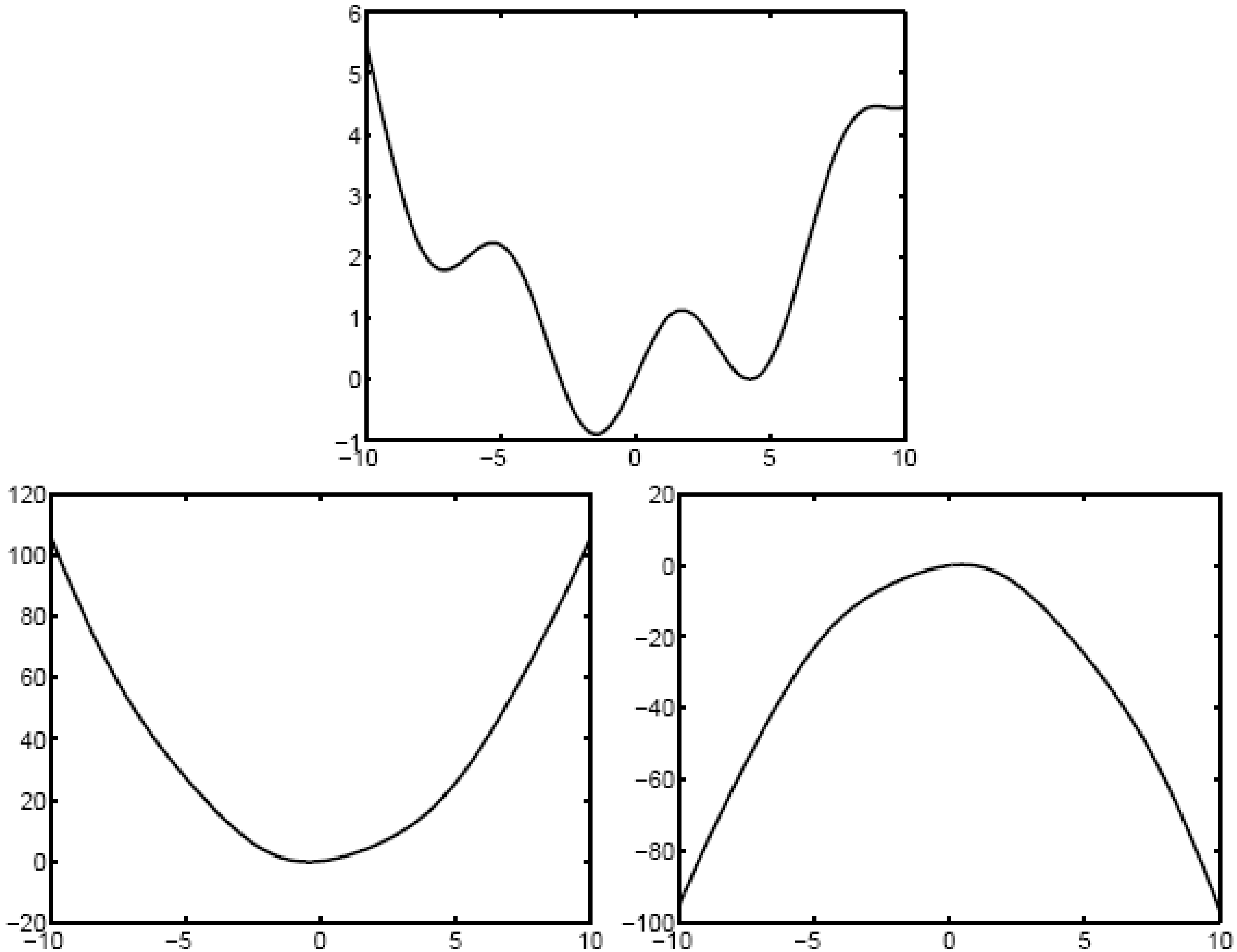
Sum $E = F + G$ of convex F + concave G function $\Rightarrow$ Convex-ConCave Procedure (CCCP, NIPS*01)
Start from arbitrary c_0, and iteratively update as:

$$\nabla F(c_{t+1}) = -\nabla G(c_t)$$

$\Rightarrow$ guaranteed convergence to a local minimum.

ConCave Convex Procedure (CCCP)

$$\min_x E(x) = F(x) + G(x)$$
$$\nabla F(c_{t+1}) = -\nabla G(c_t)$$



Iterative algorithm for Burbea-Rao centroids

Apply CCCP scheme

$$\nabla F(c_{t+1}) = \frac{1}{\sum_{i=1}^n w_i \alpha_i} \sum_{i=1}^n w_i \alpha_i \nabla F\left(\alpha_i c_t + (1-\alpha_i) p_i\right)$$

$$c_{t+1} = \nabla F^{-1} \left(\frac{1}{\sum_{i=1}^n w_i \alpha_i} \sum_{i=1}^n w_i \alpha_i \nabla F\left(\alpha_i c_t + (1-\alpha_i) p_i\right) \right)$$

Get arbitrarily fine approximations of the (skew) Burbea-Rao centroids and barycenters.

Special cases: Closed-form Burbea-Rao centroids

Consider $F(x) = \langle x, x \rangle$.

$$\min E(x) = \frac{F(x)}{2} - \sum_{i=1}^{n} w_i F\left(\frac{p_i + x}{2}\right),$$
$$= \quad \min \frac{\langle x, x \rangle}{2} - \frac{1}{4} \sum_{i=1}^{n} w_i \left(\langle x, x \rangle + 2\langle x, p_i \rangle + \langle p_i, p_i \rangle\right)$$

The minimum obtained when $\nabla E(x) = 0$

$$x = \bar{p} = \sum_{i=1}^{n} w_i p_i$$

Extremal skew cases (for $\alpha \to 0$ or $\alpha \to 1$):
Bregman sided centroids in closed-forms: $\bar{x} = \sum_{i=1}^{n} w_i p_i$ (right-sided) or
$\bar{x} = (\nabla F)^{-1} \left(\sum_{i=1}^{n} w_i \nabla F(p_i)\right)$ (left-sided)
But usually only approximation using CCCP iterations.

Bhattacharyya coefficients/distances

Bhattacharyya coefficient and non-metric distance:

$$C_((p,q) = \int \sqrt{p(x)q(x)}\mathrm{d}x, \qquad 0 < C(p,q) \leq 1, \qquad B(p,q) = -\ln C(p,q).$$

(coefficient is always strictly positive)
Hellinger metric

$$H(p,q) = \sqrt{\frac{1}{2}\int (\sqrt{p(x)} - \sqrt{q(x)})^2 \mathrm{d}x},$$

such that $0 \leq H(p,q) \leq 1$.

$$\begin{aligned} H(p,q) &= \sqrt{\frac{1}{2}\left(\int p(x)\mathrm{d}x + \int q(x)\mathrm{d}x - 2\int \sqrt{p(x)}\sqrt{q(x)}\mathrm{d}x\right)} \\ &= \sqrt{1 - C(p,q)}. \end{aligned}$$

Chernoff coefficients/α-divergences

Skew Bhattacharrya divergences based on Chernoff α-coefficients.

$$
\begin{aligned}
B_\alpha(p,q) &= -\ln \int_x p^\alpha(x) q^{1-\alpha}(x) \mathrm{d}x = -\ln C_\alpha(p,q) \\
&= -\ln \int_x q(x) \left(\frac{p(x)}{q(x)}\right)^\alpha \mathrm{d}x \\
&= -\ln E_q[L^\alpha(x)]
\end{aligned}
$$

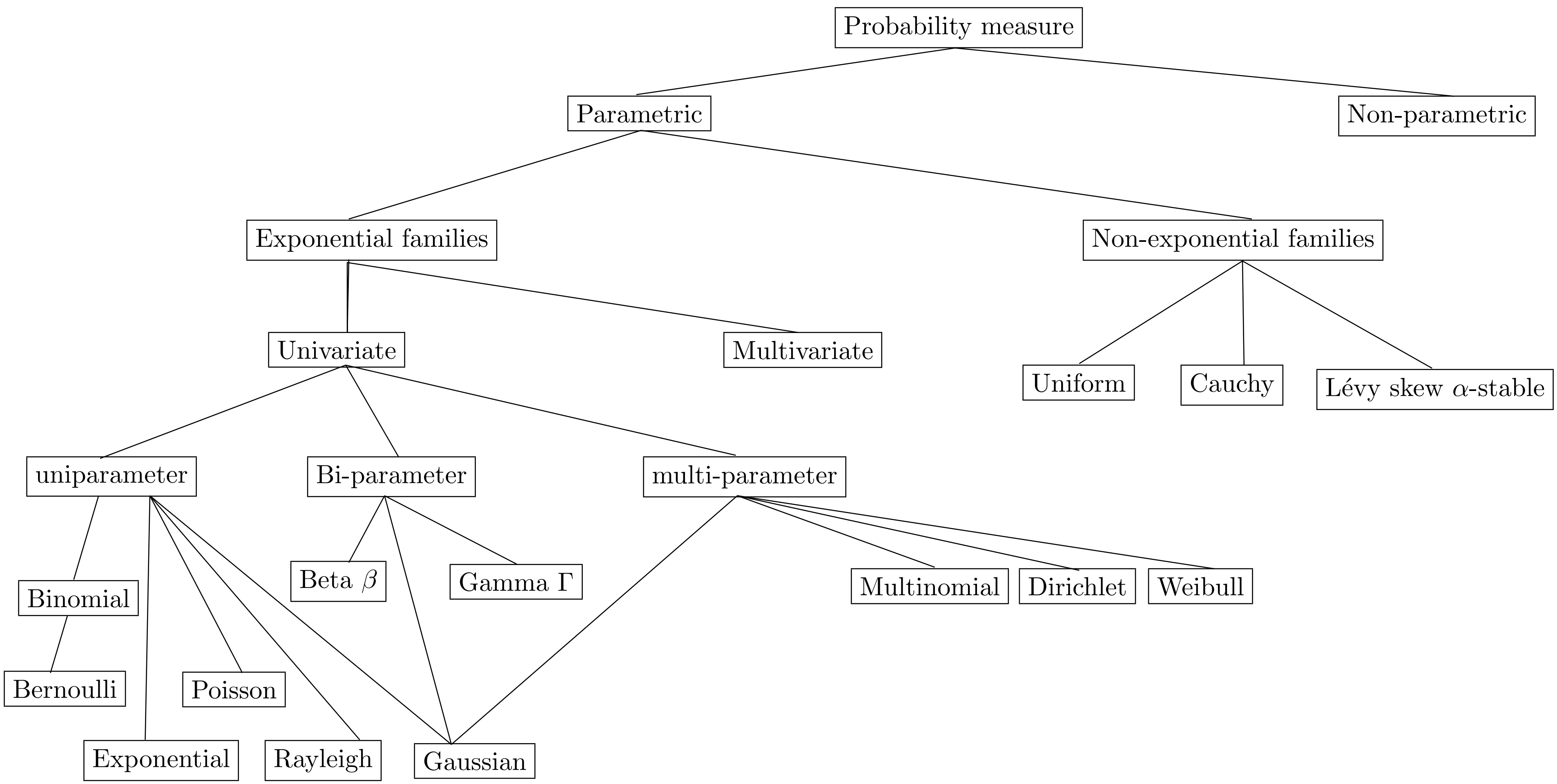
Amari α-divergence:

$$
D_\alpha(p||q) = \begin{cases} \frac{4}{1-\alpha^2}\left(1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} \mathrm{d}x\right), & \alpha \neq \pm 1, \\ \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x = \mathrm{KL}(p,q), & \alpha = -1, \\ \int q(x) \log \frac{q(x)}{p(x)} \mathrm{d}x = \mathrm{KL}(q,p), & \alpha = 1, \end{cases}
$$

$$
D_\alpha(p||q) = D_{-\alpha}(q||p)
$$

Remapping $\alpha' = \frac{1-\alpha}{2}$ ($\alpha = 1 - 2\alpha'$) to get Chernoff α'-divergences

Exponential families in statistics



Exponential families in statistics

Gaussian, Poisson, Bernoulli/multinomial, Gamma/Beta, etc.:

$$p(x;\lambda) = p_F(x;\theta) = \exp\left(\langle t(x), \theta\rangle - F(\theta) + k(x)\right).$$

Example: Poisson distribution

$$p(x;\lambda) = \frac{\lambda^x}{x!}\exp(-\lambda),$$

- the sufficient statistic $t(x) = x$,
- $\theta = \log \lambda$, the natural parameter,
- $F(\theta) = \exp \theta$, the log-normalizer,
- and $k(x) = -\log x!$ the carrier measure (with respect to the counting measure).

Gaussians as an exponential family

$$p(x;\lambda) = p(x;\mu,\Sigma) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu))}{2}\right)$$

- $\theta = (\Sigma^{-1}\mu, \frac{1}{2}\Sigma^{-1}) \in \Theta = \mathbb{R}^d \times \mathbb{K}_{d\times d}$, with $\mathbb{K}_{d\times d}$ cone of positive definite matrices,
- $F(\theta) = \frac{1}{4}\mathrm{tr}(\theta_2^{-1}\theta_1\theta_1^T) - \frac{1}{2}\log\det\theta_2 + \frac{d}{2}\log\pi$,
- $t(x) = (x, -x^T x)$,
- $k(x) = 0$.

Inner product : composite, sum of a dot product and a matrix trace :

$$\langle\theta,\theta'\rangle = \theta_1^T\theta_1' + \mathrm{tr}(\theta_2^T\theta_2').$$

The coordinate transformation $\tau : \Lambda \to \Theta$ is given for $\lambda = (\mu, \Sigma)$ by

$$\tau(\lambda) = \left(\lambda_2^{-1}\lambda_1, \frac{1}{2}\lambda_2^{-1}\right), \qquad \tau^{-1}(\theta) = \left(\frac{1}{2}\theta_2^{-1}\theta_1, \frac{1}{2}\theta_2^{-1}\right)$$

Bhattacharyya/Chernoff of exponential families

Equivalence with skew Burbea-Rao distances:

$$B_\alpha(p_F(x;\theta_p), p_F(x;\theta_q)) = \mathrm{BR}_F^{(\alpha)}(\theta_p,\theta_q) = \alpha F(\theta_p) + (1-\alpha)F(\theta_q) - F(\alpha\theta_p + (1-\alpha)\theta_q)$$

Proof: Chernoff coefficients $C_\alpha(p,q)$ of members $p = p_F(x;\theta_p)$ and $q = p_F(x;\theta_q)$ of the *same* exponential family $\mathcal{E}_F$:

$C_\alpha(p,q) = \int p^\alpha(x) q^{1-\alpha}(x)\mathrm{d}x = \int p_F^\alpha(x;\theta_p) p_F^{1-\alpha}(x;\theta_q)\mathrm{d}x$
$= \int \exp(\alpha(\langle x,\theta_p\rangle - F(\theta_p))) \times \exp((1-\alpha)(\langle x,\theta_q\rangle - F(\theta_q)))\mathrm{d}x$
$= \int \exp\left(\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right)\mathrm{d}x$
$= \exp -(\alpha F(\theta_p) + (1-\alpha)F(\theta_q)) \times$
$\int \exp\left(\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle - F(\alpha\theta_p + (1-\alpha)\theta_q) + F(\alpha\theta_p + (1-\alpha)\theta_q)\right)\mathrm{d}x$
$= \exp\left(F(\alpha\theta_p + (1-\alpha)\theta_q) - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right) \times \int \exp\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle - F(\alpha\theta_p + (1-\alpha)\theta_q)\mathrm{d}x$

$$= \exp\left(F(\alpha\theta_p + (1-\alpha)\theta_q) - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right) \times \underbrace{\int p_F(x;\alpha\theta_p + (1-\alpha)\theta_q)\mathrm{d}x}_{=1}$$

$= \exp(-\mathrm{BR}_F^{(\alpha)}(\theta_p,\theta_q)) > 0.$ Coefficient is always strictly positive. For $\theta_p = \theta_q$, $C_\alpha(\theta_p,\theta_q) = \exp -0 = 1$ and $B_\alpha(\theta_p,\theta_q) = 0$.

α-div./Kullback-Leibler $\leftrightarrow$ Burbea-Rao/Bregman

Skew Bhattacharyya distances on members of the same exponential family is equivalent to skew Burbea-Rao divergences on the natural parameters (without swapping order).

$$B_\alpha(p_F(x;\theta_p), p_F(x;\theta_q)) = \mathrm{BR}_F^{(\alpha)}(\theta_p, \theta_q)$$

For $\alpha = \pm 1$, Kullback-Leibler of exp. fam. = *Bregman divergence* (limit as $\alpha \to 1$ or $\alpha \to 0$).

$$\begin{aligned}
\mathrm{KL}(p,q) &= \mathrm{KL}(p_F(x;\theta_p), p_F(x;\theta_q)) \\
&= \lim_{\alpha' \to 1} D_{\alpha'}(p_F(x;\theta_p), p_F(x;\theta_q)) \\
&= \lim_{\alpha' \to 1} \frac{1}{\alpha'(1-\alpha')} (1 - \underbrace{C_\alpha(p_F(x;\theta_p), p_F(x;\theta_q)))}_{\text{since } \exp x \simeq_{x \simeq 0} 1+x} \\
&= \lim_{\alpha' \to 1} \frac{1}{\alpha'(1-\alpha')} \underbrace{\mathrm{BR}_F^{\alpha'}(\theta_p, \theta_q)}_{(1-\alpha')B_F(\theta_q, \theta_p)} \\
&= \lim_{\alpha' \to 1} \frac{1}{\alpha'} B_F(\theta_q, \theta_p) = B_F(\theta_q, \theta_p)
\end{aligned}$$

Closed-form Bhattacharyya distances for exp. fam.

Exp. fam.	$F(\theta)$ (up to a constant)	Bhattacharyya/Burbea-Rao $\mathrm{BR}_F(\lambda_p, \lambda_q) = \mathrm{BR}_F(\tau(\lambda_p)$
Multinomial	$\log(1 + \sum_{i=1}^{d-1} \exp\theta_i)$	$-\ln\sum_{i=1}^{d} \sqrt{p_i q_i}$
Poisson	$\exp\theta$	$\frac{1}{2}(\sqrt{\mu_p} - \sqrt{\mu_q})^2$
Gaussian	$-\frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\log(-\frac{\pi}{\theta_2})$	$\frac{1}{4}\frac{(\mu_p-\mu_q)^2}{\sigma_p^2+\sigma_q^2} + \frac{1}{2}\ln\frac{\sigma_p^2+\sigma_q^2}{2\sigma_p\sigma_q}$
Gaussian	$\frac{1}{4}\mathrm{tr}(\Theta^{-1}\theta\theta^T) - \frac{1}{2}\log\det\Theta$	$\frac{1}{8}(\mu_p-\mu_q)^T\left(\frac{\Sigma_p+\Sigma_q}{2}\right)^{-1}(\mu_p-\mu_q) + \frac{1}{2}\ln\frac{\det\frac{\Sigma_p+\Sigma_q}{2}}{\det\Sigma_p\det\Sigma_q}$

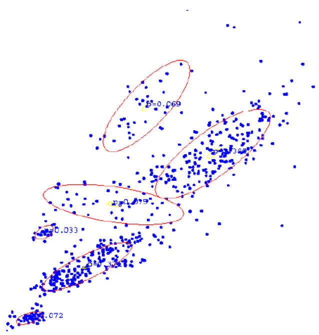
Bhattacharyya, Burbea-Rao, Tsallis, Rényi, $\alpha-, \beta$-divergences are in closed forms for members of the same exponential family.

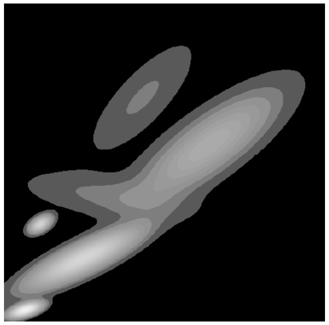
Application: statistical mixtures

cdefinitionGaussian mixture models (GMMs, MoGs: mixture of Gaussians):
Probabilistic modeling of data:

$\mathrm{Pr}(X=x)=\sum_{i=1}^{k} w_i \mathrm{Pr}(X=x|\mu_i,\Sigma_i)$ (with $\sum_i w_i = 1$ and all $w_i \geq 0$).

$$\mathrm{Pr}(X=x|\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Sigma}}\exp-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu).$$

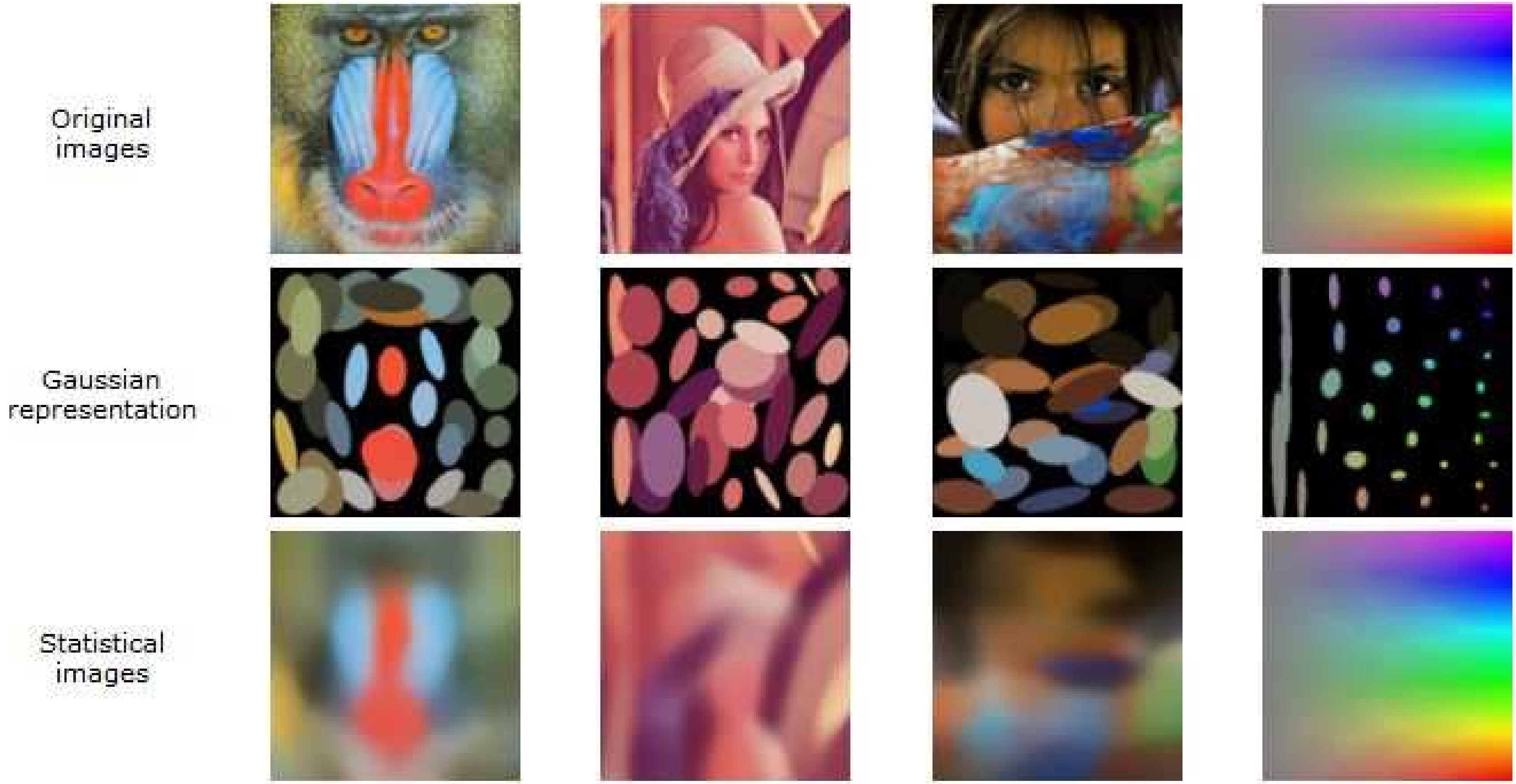




Similar to k-means, soft clustering wrt. to log-likelihood is minimized by the expectation-maximization (EM) algorithm [Dempster'77]

Application: Statistical images and Gaussians

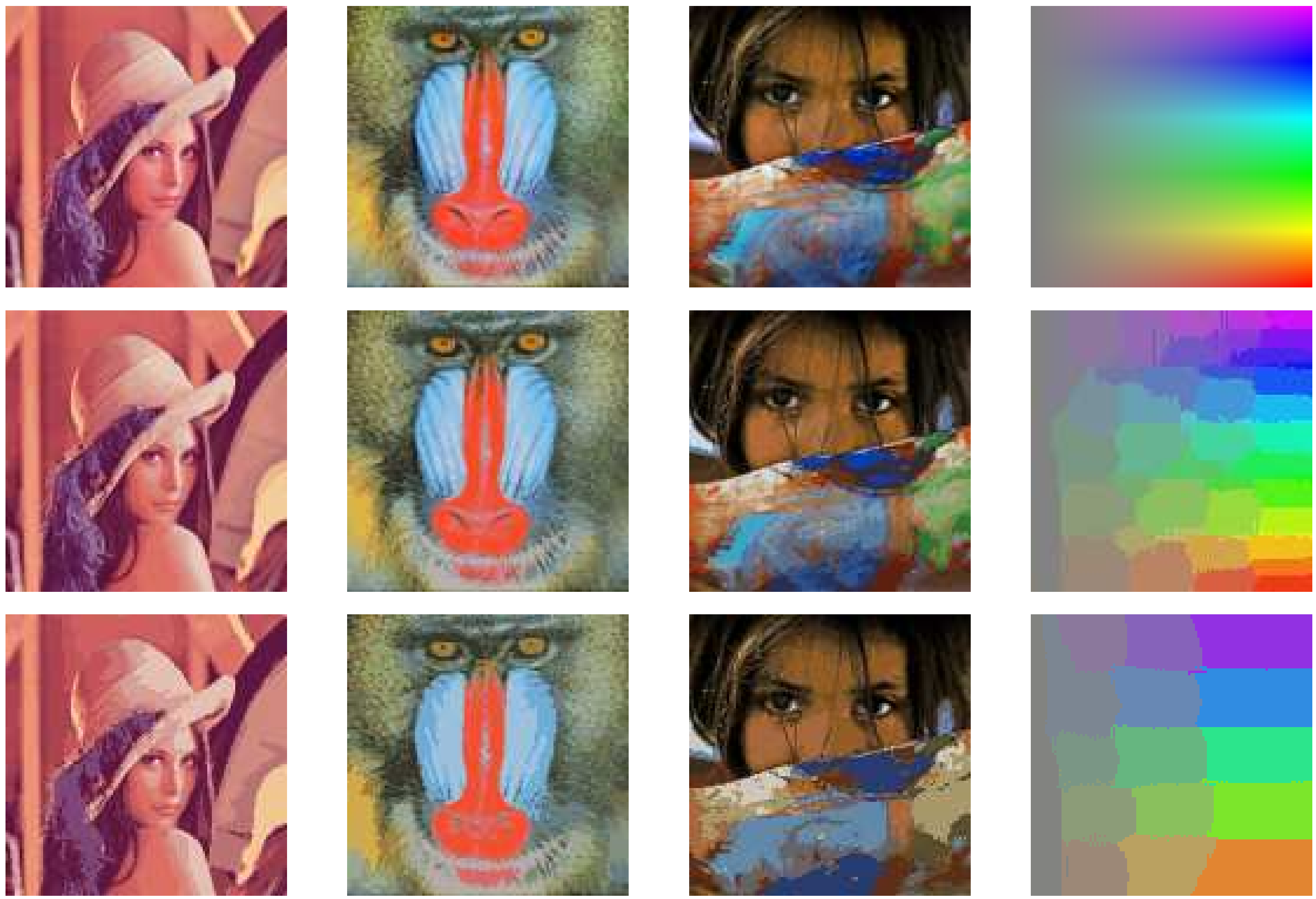
Consider 5D Gaussian Mixture Models (GMMs) of color images (image=RGBxy point set)



Get open source Java(TM) jMEF library:
`www.informationgeometry.org/MEF/`

Hierarchical clustering of GMMs

Hierarchical clustering of GMMs wrt. Bhattacharyya distance. Simplify the number of components of an initial GMM.



(a) source

(b) $k = 48$

(c) $k = 16$

Summary of results

- Skew Burbea-Rao divergences occur when
  - Symmetrizing skew Bregman divergences: Jensen-Bregman divergences
  - Bhattacharyya/Chernoff coefficients/distances of exponential families
- Apply ConCave-Convex procedure (CCCP) for computing Burbea-Rao centroids
- Skewed Burbea-Rao yields *in the limit* Bregman divergences
- Application: Hierarchical clustering of Gaussian mixtures
- (In arXiv:1004.5049, alternative tailored matrix method generalizing ICASSP 2000 but not so efficient as the general scheme)

`www.informationgeometry.org/BurbeaRao/`

References

- "Bhattacharyya clustering with applications to mixture simplifications," ICPR 2010.
- "Sided and symmetrized Bregman centroids," IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2048-2059, June 2009.
- "Bregman Voronoi diagrams," Discrete & Computational Geometry, 2010.
- "On the convexity of some divergence measures based on entropy functions," IEEE Transactions on Information Theory, vol. 28, no. 3, pp. 489-495, 1982.
- "Statistical exponential families: A digest with flash cards," 2009, arXiv.org:0911.4863
- "An optimal Bhattacharyya centroid algorithm for gaussian clustering with applications in automatic speech recognition," ICASSP 2000.
- A. Yuille and A. Rangarajan, "The concave-convex procedure," Neural Computation, vol. 15, no. 4, pp. 915-936, 2003.

References & Acknowledgments

Michèle Basseville (IRISA), Richard Nock (UAG CEREGMIA)

M. Basseville, J.F. Cardoso, "On entropies, divergences and mean values," IEEE International Symposium on Information Theory (ISIT), p.330, 1995.

F. Nielsen, S. Boltz, "The Burbea-Rao and Bhattacharyya centroids," arXiv, 2010. `http://arxiv.org/abs/1004.5049`