Closed-Form Information-Theoretic Divergences for Statistical Mixtures

Frank NIELSEN
Sony Computer Science Laboratories, Inc.
http://www.informationgeometry.org

Consider two statistical mixtures $m(x)=\sum_{i=1}^{k} w_i p_F(x;\theta_i)$ and $m(x)=\sum_{i=1}^{k'} w'_i p_F(x;\theta'_i)$.

Kullback-Leibler divergence (relative entropy) *does not admit* a closed-form formula:

$$\mathrm{KL}(m:m')=\int_{x\in\mathbb{X}} m(x)\log\frac{m(x)}{m'(x)}\mathrm{d}x$$

$\longrightarrow$ Stochastic approximation using Monte-Carlo integration (costly)

$$\tilde{\mathrm{KL}}(m:m')\approx\frac{1}{n}\sum_{\substack{x_i\sim m\\1\leq i\leq n}}\left(\log\frac{m(x_i)}{m'(x_i)}+\frac{m'(x_i)}{m(x_i)}-1\right)$$

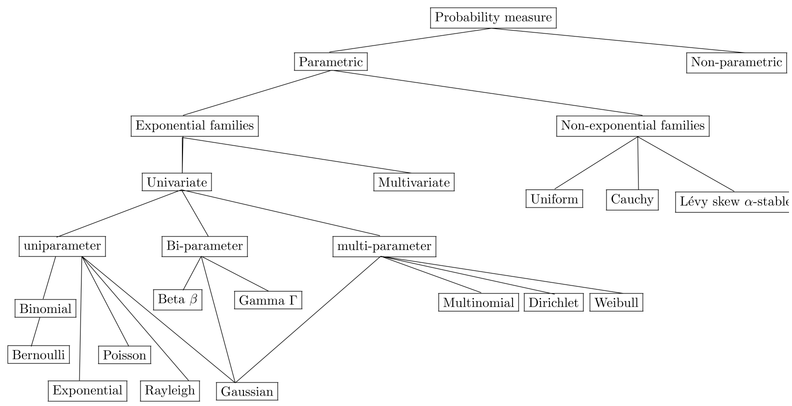
Consider alternative **Cauchy-Schwarz divergence**:

$$\boxed{\mathrm{CS}(P:Q)=-\log\frac{\int p(x)q(x)\mathrm{d}x}{\sqrt{\int p(x)^2\mathrm{d}x\int q(x)^2\mathrm{d}x}}}$$

Consider mixture components from the same exponential families

$$p_F(x;\theta)=e^{\langle t(x),\theta\rangle-F(\theta)+k(x)}$$

- $t(x)$ = sufficient statistics,
- θ = natural parameter,
- $F(\theta)$ = log-normalizer,
- $k(x)$ = auxiliary carrier measure



The integral of product of mixtures yields:

$$\int m(x)m'(x)\mathrm{d}x=\sum_{i=1}^{k}\sum_{j=1}^{k'} w_i w'_j\int p_F(x;\theta_i)p_F(x;\theta'_j)\mathrm{d}x$$

$$\boxed{\int m(x)m'(x)\mathrm{d}x=\sum_{i=1}^{k}\sum_{j=1}^{k'} w_i w'_j e^{\Delta_F(\theta_i,\theta'_j)}}\qquad\boxed{\Delta_F(\theta_i,\theta'_j)=F(\theta_i+\theta'_j)-(F(\theta_i)+F(\theta'_j))}$$

$\longleftarrow$ when natural parameter space $\Theta=\{\theta\mid\int p_F(x;\theta)\mathrm{d}x<\infty\}$ is a **convex cone**. Generic Cauchy-Schwarz divergence formula applies for $(k(x)=0)$:

- mixtures of Bernoulli,
- mixtures of zero-centered Laplacian,
- mixtures of Wishart,
- mixtures of Gaussian,
- etc.

Besides Cauchy-Schwarz divergence, also closed-form solution for:

- Square loss: $\int(m(x)-m'(x))^2\mathrm{d}x=\int m^2(x)\mathrm{d}x-2\int m(x)m'(x)\mathrm{d}x+\int m'^2(x)\mathrm{d}x$
- total Square loss (total Bregman divergence for $F(x)=\langle x,x\rangle$ generator)
- Jensen-quadratic Rényi divergence: $\mathrm{JR}_2(m,m')=H_2\left(\frac{m+m'}{2}\right)-\frac{H_2(m)+H_2(m')}{2}$ with $H_2(m)=-\log\sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j e^{\Delta_F(\theta_i,\theta_j)}$