ENTROPIES AND CROSS-ENTROPIES OF EXPONENTIAL FAMILIES

Frank NIELSEN and Richard NOCK

ICIP 2010 – Wednesday, September 29, 2010 – Hong Kong – WA.PC.6a

1. Introduction

Exponential families

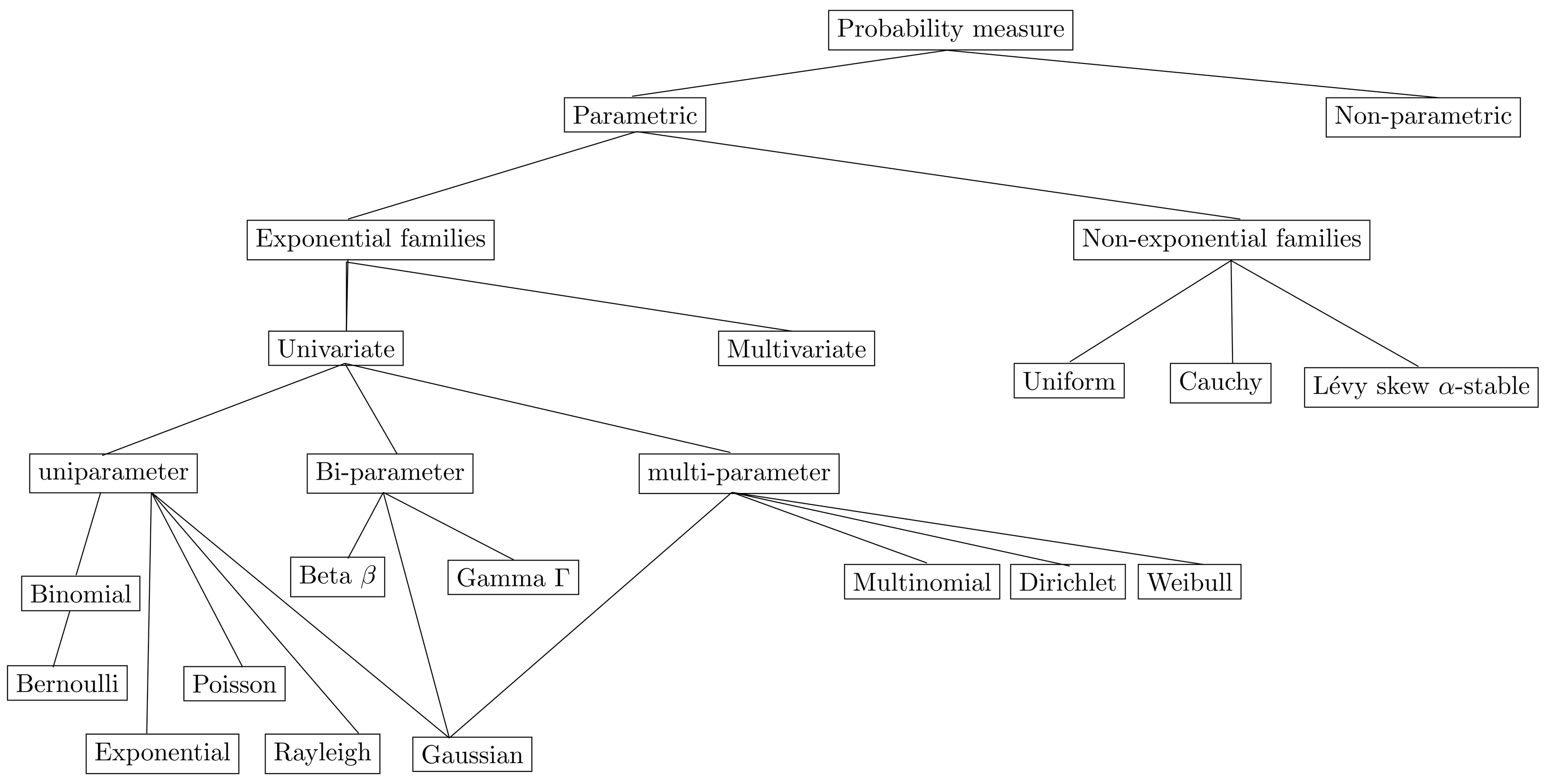
Versatile family of probability densities: Binomial, multinomial (frequency histogram), Poisson, Gaussian, Rayleigh, Gamma, Beta, Dirichlet, Weibull, etc.

Entropy and relative entropy

Shannon entropy $H(P) = \int p(x) \log \frac{1}{p(x)} \mathrm{d}x = -\int p(x) \log p(x) \mathrm{d}x = E_P[-\log p(x)]$

Kullback-Leibler divergence $\mathrm{KL}(p(x)||q(x)) = \int_x p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x = E_P\left[\log \frac{p(x)}{q(x)}\right]$, rewritten as $\mathrm{KL}(p(x)||q(x)) = H^\times(p(x)||q(x)) - H(p(x)) \geq 0$ with $H^\times(p(x)||q(x)) = \int_x p(x) \log \frac{1}{q(x)} \mathrm{d}x$

2. Exponential family



Canonical decomposition: $p_F(x;\theta) = \exp(\langle t(x), \theta\rangle - F(\theta) + k(x))$, with $\langle x, y\rangle = x^T y$

3. Relative entropy: Bregman divergence

For members $p_F(x;\theta_p)$ and $p_F(x;\theta_q)$ of the same exponential family, the relative entropy is always in closed-form.
Computed as a Bregman divergence on swapped natural parameters:

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

$$\mathrm{KL}(p_F(x;\theta_p)||p_F(x;\theta_q)) = \int_x p_f(x|\theta_p) \log \frac{p_F(x|\theta_p)}{p_F(x|\theta_q)} \mathrm{d}x = B_F(\theta_q||\theta_p)$$

Note that finite discrete distributions (say, of d events) are exponential families in disguise: Those distributions are precisely multinomials with $d-1$ degrees of freedom.

4. Entropy

$$\begin{aligned}\mathrm{KL}(p||q) &= B_F(\theta_q||\theta_p) \\ &= F(\theta_q) - F(\theta_p) - \langle \theta_q - \theta_p, \nabla F(\theta_p)\rangle\end{aligned}$$

$$\mathrm{KL}(p||q) = \underbrace{F(\theta_q) - \langle \theta_q, \nabla F(\theta_p)\rangle}_{\sim H_F^\times(\theta_p||\theta_q)} - \underbrace{(F(\theta_p) - \langle \theta_p, \nabla F(\theta_p)\rangle)}_{\sim H_F(\theta_p)}$$

$$H(P) = H_F(\theta_p) = F(\theta_p) - \langle \theta_p, \nabla F(\theta_p)\rangle + b$$

Since $H(p) \geq 0$, we necessarily have $b \geq \langle \theta_p, \nabla F(\theta_p)\rangle - F(\theta_p)$. $rightarrow$ compare exactly the entropy of two members of the same exponential family since the constant term b vanishes.

$$b = -\int k(x) p_F(x;\theta) \mathrm{d}x = -E[k(x)]$$

5. An example

Rayleigh distribution

$$p(x;\sigma^2) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$F(\theta) = -\log(-2\theta)$, natural parameter $\theta = -\frac{1}{2\sigma^2}$, sufficient statistic $t(x) = x^2$, gradient $F'(\theta) = -\frac{1}{\theta}$ and carrier measure $k(x) = \log x$. Let $X \sim \mathrm{Rayleigh}(\sigma^2)$, we have:

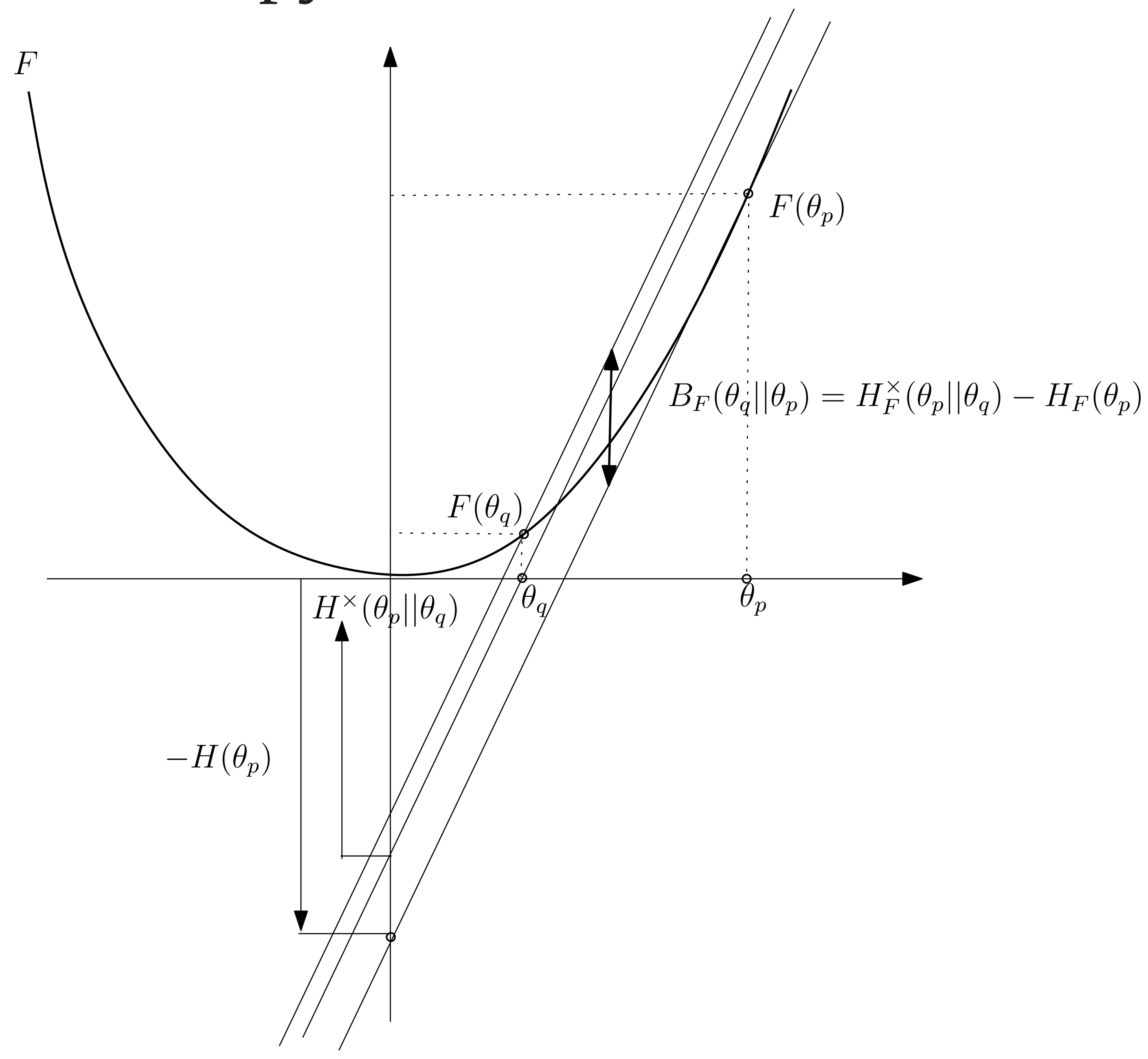
$$H(X) = 1 + \ln \frac{\sigma}{\sqrt{2}} + \frac{\gamma}{2},$$

where $\gamma = 0.57721566...$ stands for the Euler-Mascheroni constant. This is the term related to the carrier measure $\log x$ integrated over the distribution.

Implementations available at:
`http://www.informationgeometry.org/MEF/`

6. Results and geometric interpretation

$\Rightarrow$ compare exactly Shannon entropy of members of the same exponential families.



$\Rightarrow$ Results related to maximum entropy and mixture of exponential families