Cross-entropy, differential entropy and Kullback-Leibler divergence between Rayleigh distributions

Frank Nielsen
Frank.Nielsen@acm.org

Abstract

We instantiate the generic formula of information-theoretic quantities [3] (cross-entropy, entropy and Kullback-Leibler divergence) for exponential families to the case of Rayleigh distributions.

A Rayleigh distribution of scale parameter $\sigma > 0$ has probability density function:

$$p_\sigma(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \geq 0.$$

The family $\{p_\sigma(x) \ : \ \sigma > 0\}$ of Rayleigh distributions form a univariate *exponential family* [2]

$$\mathcal{E} = \{p_\theta(x) = \exp(t(x)\theta - F(\theta) + k(x)) \ : \ \theta \in \Theta\}$$

with the following canonical decomposition: natural parameter $\theta(\sigma) = -\frac{1}{2\sigma^2}$ (natural parameter space $\Theta = \mathbb{R}_{--}$), sufficient statistic $t(x) = x^2$, log-normalizer $F(\theta) = -\log(-2\theta)$ ($F_\sigma(\sigma) = 2\log\sigma$), and *non-zero* auxiliary carrier term $k(x) = \log x$.

The cross-entropy [3, 1] $h^\times$ between two distributions p_{θ_1} and p_{θ_2} belonging to the same exponential family is

$$\begin{aligned} h^\times(p_{\theta_1} : p_{\theta_2}) &= -E_{p_{\theta_1}}[\log p_{\theta_2}] = -\int p_{\theta_1}(x)\log p_{\theta_2}(x)\mathrm{d}x, & (1)\\ &= -\int p_{\theta_1}(x)(t(x)\theta_2 - F(\theta_2) + k(x))\mathrm{d}x = F(\theta_2) - \theta_2 F'(\theta_1) - E_{\theta_1}[k(x)], & (2)\end{aligned}$$

since $\int p_{\theta_1}(x)t(x)\mathrm{d}x = F'(\theta_1)$.

Consider the term

$$E_\theta[k(x)] = E_\sigma[k(x)] = E_\sigma[\log x] = \int_0^\infty \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)(\log x)\mathrm{d}x,$$

and make the change of variable $y = \frac{x}{\sigma}$ with $\mathrm{d}y = \frac{\mathrm{d}x}{\sigma}$:

$$E_\sigma[k(x)] = \int_0^\infty y\exp\left(-\frac{y^2}{2}\right)(\log\sigma + \log y)\mathrm{d}y$$

Using the fact that $\int_0^\infty y\exp(-\frac{y^2}{2})\mathrm{d}y = 1$ and $\int_0^\infty (y\log y)\exp(-\frac{y^2}{2})\mathrm{d}y = \frac{1}{2}(\log 2 - \gamma)$ (using a computer algebra system[1], where $\gamma = -\int_0^\infty e^{-x}\log x\mathrm{d}x \simeq 0.5772156649015328606065120900 82$ is the Euler-Mascheroni constant), we find that

$$\boxed{E_\sigma[k(x)] = \frac{1}{2}(\log 2 - \gamma) + \log\sigma.} \qquad (3)$$

[1] For example, using online Wolfram alpha.

It follows that the cross-entropy between two Rayleigh distributions is

$$h^{\times}(p_{\sigma_1} : p_{\sigma_2}) \quad = \quad h^{\times}(p_{\theta(\sigma_1)} : p_{\theta(\sigma_2)}),$$

$$\boxed{h^{\times}(p_{\sigma_1} : p_{\sigma_2}) = 2\log\sigma_2 + \frac{\sigma_1^2}{\sigma_2^2} - \frac{1}{2}(\log 2 - \gamma) - \log\sigma_1} \tag{4}$$

The Shannon's differential entropy [1] is the self cross-entropy:

$$\begin{aligned} h(p_\sigma) &= h^{\times}(p_\sigma : p_\sigma) = -\int p_\theta(x)\log p_\theta(x)\mathrm{d}x, \\ &= F(\theta) - \theta F'(\theta) - E_\theta[k(x)], \end{aligned}$$

$$\boxed{h(p_\sigma) = 1 + \log\frac{\sigma}{\sqrt{2}} + \frac{\gamma}{2}.} \tag{5}$$

The Rayleigh distributions also form a scale family with density $p_\sigma(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, where $f(x) = x\exp(-\frac{x^2}{2}) = p_1(x)$. The differential entropy of a scale density is $h(p_\sigma) = h(p_1) + \log\sigma$. Thus we check that $h(p_1) = 1 - \log\sqrt{2} + \frac{\gamma}{2} \simeq 0.94$.

Notice that when $k(x) = 0$, $E_\theta[k(x)] = 0$, and $h(p_\theta) = F(\theta) - \theta F'(\theta) = -F^*(\eta)$, where $F^*(\eta) = \sup_{\theta\in\Theta}\{\theta\eta - F(\theta)\}$ is the Legendre convex conjugate and $\eta = F'(\theta)$ the dual moment parameterization. Thus $F^*(\eta) = -h(p_\theta)$ is called the *negentropy* in the literature (but requires $k(x) = 0$ like for the Gaussian family).

The Kullback-Leibler divergence [1] is the difference between the cross-entropy and the entropy:

$$\mathrm{KL}(p_{\theta_1} : p_{\theta_2}) \quad = \quad h^{\times}(p_{\theta_1} : p_{\theta_2}) - h(p_{\theta_1}) = \int p_{\theta_1}(x)\log\frac{p_{\theta_1}(x)}{p_{\theta_2}(x)}\mathrm{d}x,$$

$$\boxed{\mathrm{KL}(p_{\theta_1} : p_{\theta_2}) = 2\log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2}{\sigma_2^2}.} \tag{6}$$

Notice that the Kullback-Leibler divergence $\mathrm{KL}(p_{\theta_1} : p_{\theta_2})$ between members of the same univariate exponential family amounts to compute the equivalent univariate Bregman divergence $B_F(\theta_2 : \theta_1)$ with

$$B_F(\theta_2 : \theta_1) = F(\theta_2) - F(\theta_1) - (\theta_2 - \theta_1)F'(\theta_1),$$

with $F(\theta(\sigma)) = 2\log\sigma$ and $F'(\theta(\sigma)) = -\frac{1}{\theta(\sigma)} = 2\sigma^2$.

References

[1] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[2] Frank Nielsen and Vincent Garcia. Statistical exponential families: A digest with flash cards. *arXiv preprint arXiv:0911.4863*, 2009.

[3] Frank Nielsen and Richard Nock. Entropies and cross-entropies of exponential families. In *IEEE International Conference on Image Processing (ICIP)*, pages 3621–3624. IEEE, 2010.