deBruijn identities: from Shannon, Kullback–Leibler and Fisher to Salicrú, Csizàr and generalized Fisher informations

Steeve Zozor & Jean-Marc Brossier

Laboratoire Grenoblois d'Image, Parole, Signal et Automatique (GIPSA-Lab, CNRS), Grenoble, France

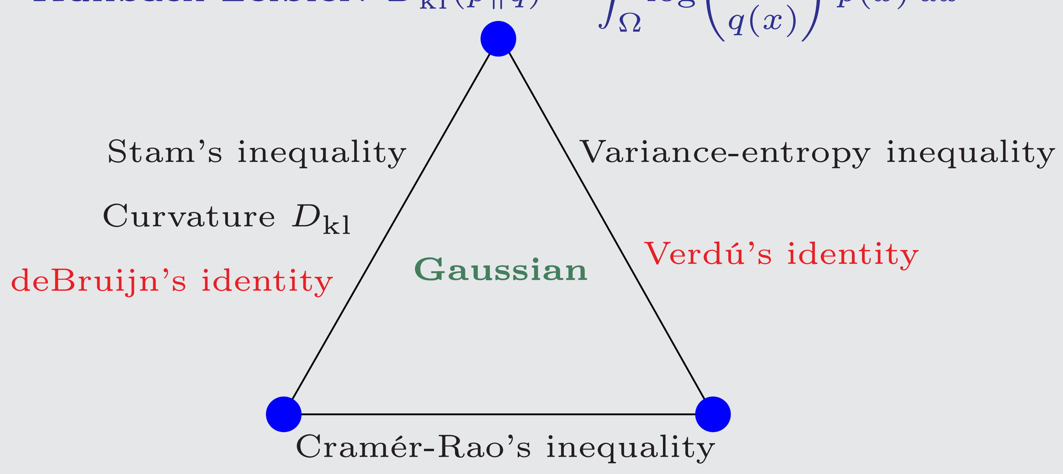
Abstract

We propose a generalization of the usual deBruijn identity that links the Shannon differential entropy (or the Kullback–Leibler divergence) and the Fisher information (or the Fisher divergence) of the output of a Gaussian channel. The generalization makes use of Salicrú entropies on the one hand, and of divergences of the Csizàr class on the other hand, as generalizations of the Shannon entropy and of the Kullback–Leibler divergence respectively. The generalized deBruijn identities induce the definition of generalized Fisher informations and generalized Fisher divergences (some of such generalizations exist in the literature). Moreover, we provide results that go beyond the Gaussian channel: we are then able to characterize a noisy channel using general measures of mutual information, both for Gaussian and nonGaussian channels.

From the classical "quadriptych" to extensions?

Shannon: $H(p) = -\int_\Omega p(x) \log p(x)\, dx$

Kullback-Leibler: $D_{\mathrm{kl}}(p\|q) = \int_\Omega \log\left(\frac{p(x)}{q(x)}\right) p(x)\, dx$

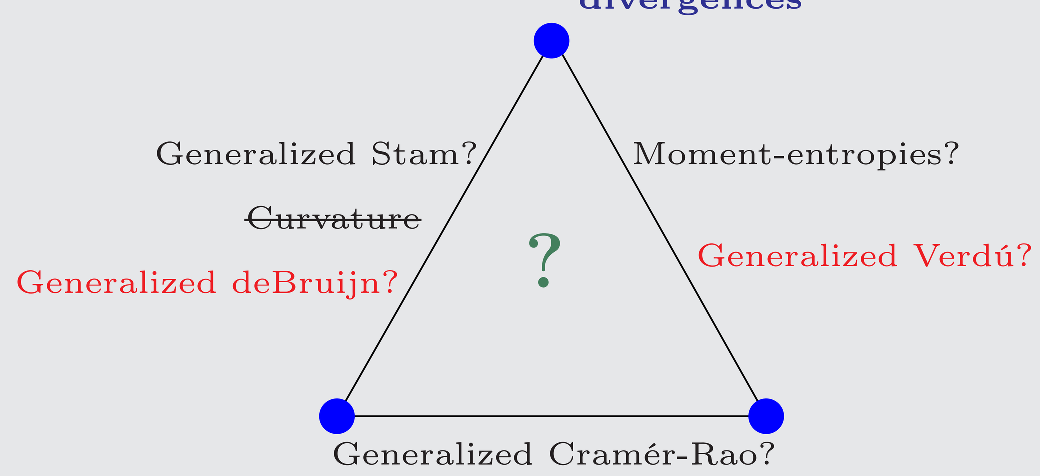


$J^{(\theta)}(p) = \int_\Omega \left(\frac{\partial}{\partial\theta} \log p(x)\right)^2 p(x)\, dx$

$J^{(\theta)}(p\|q) = \int_\Omega \left(\frac{\partial}{\partial\theta} \log\left(\frac{p(x)}{q(x)}\right)\right)^2 p(x)\, dx$

θ position: $\frac{\partial}{\partial\theta} \equiv -\frac{\partial}{\partial x}$ & notation J

$\rightsquigarrow$

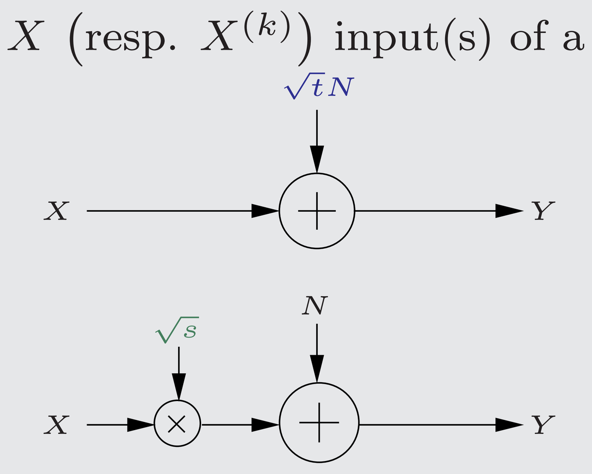


e.g., Bercher'12 or Lutwak et al.'12

- Stretched Gaussian $\left(1-(\lambda-1)|x|^\alpha\right)_+^{\frac{1}{1-\lambda}}$
- Rényi–Tsallis entropies $\mathbb{E}\left[p(X)^{\lambda-1}\right]$
- Generalized Fisher $\mathbb{E}\left[(\nabla_x \log p(X))^\beta\, p^{\beta(\lambda-1)}\right]$
- Generalized moments $\mathbb{E}\left[|X|^\alpha\right]$

The usual deBruijn identity for the Gaussian channel

X (resp. $X^{(k)}$) input(s) of a noisy channel, independent on the zero-mean standard Gaussian N and on parameter t and s (preamplification); output(s) $Y \sim p_Y$ (resp. $Y^{(k)}$)



deBruijn (Stam'59): $\frac{d}{dt} H(p_Y) = \frac{1}{2} J(p_Y)$

Barron'86: $\frac{d}{dt} D_{\mathrm{kl}}(p_{Y^{(1)}}\|p_{Y^{(0)}}) = -\frac{1}{2} J(p_{Y^{(1)}}\|p_{Y^{(0)}})$

Guo–Shamai–Verdú'05: $\frac{d}{ds} I(X;Y) = \frac{1}{2}\,\mathrm{MMSE}(X|Y) \equiv \frac{1}{2}\mathbb{E}\left[(X-\mathbb{E}[X|Y])^2\right]$

Consequences:

- Stam's inequality
- Entropy Power Inequality

ϕ−deBruijn identity for the Gaussian channel

$\phi : \mathbb{R}_+ \mapsto \mathbb{R}$, convex

ϕ−entropy of Salicrú, $H_\phi(p) = -\int_\Omega \phi(p(x))\, dx$

e.g.,
- $\phi(p) = p \log p$: Shannon
- $\phi(p) = p^\lambda$: Rényi–Tsallis...

$$\frac{d}{dt} H_\phi(p_Y) = \frac{1}{2} J_\phi(p_Y)$$

$\Downarrow$

ϕ−Fisher information: $J_\phi(p) = \int_\Omega \left(\frac{\partial}{\partial x} \log p(x)\right)^2 \left(\phi''(p(x)) p^2(x)\right) dx$

ϕ−divergence of Csizàr: $D_\phi(p\|q) = \int_\Omega \phi\left(\frac{p(x)}{q(x)}\right) q(x)\, dx$

e.g.,
- $\phi(l) = l \log l$: Kullback–Leibler
- $\phi(l) = \frac{l}{2}\log l - \frac{l+1}{2}\log\left(\frac{l+1}{2}\right)$: Jensen–Shannon...

$$\frac{d}{dt} D_\phi(p_{Y^{(1)}}\|p_{Y^{(0)}}) = -\frac{1}{2} J_\phi(p_{Y^{(1)}}\|p_{Y^{(0)}})$$

$\Downarrow$

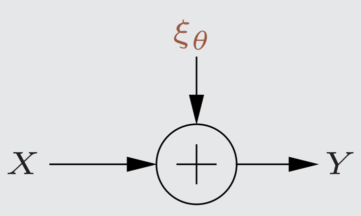
ϕ−Fisher divergence: $J_\phi(p\|q) = \int_\Omega \left(\frac{\partial}{\partial x} \log\left(\frac{p(x)}{q(x)}\right)\right)^2 \left(\frac{\phi''\left(\frac{p(x)}{q(x)}\right) p^2(x)}{q(x)}\right) dx$

Proof– The pdf of $\sqrt{t}N$ satisfies the Heat equation $\frac{\partial}{\partial t} p = \frac{1}{2}\frac{\partial^2}{\partial x^2} p \Rightarrow p_Y$ ($p_{Y^{(k)}}$) satisfies the same equation. Derivation of H_ϕ (or D_ϕ) vs t, heat equation and integration by parts.

The ϕ−Fisher information/divergence charaterizes a gain of entropy (loss of information)/convergence of the ouput vs the variation of the noise level.

Variation à la Verdú: $$\frac{d}{ds} D_\phi(p_{X,Y}\|p_X p_Y) = \frac{1}{2}\,\mathrm{MMSE}_\phi(X|Y) \equiv \frac{1}{2}\int_\Omega \left(x - \mathbb{E}[X|Y=y]\right)^2 \frac{\phi''\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) p_{X,Y}^2(x,y)}{p_X(x)p_Y(y)}\, dx\, dy$$

To more general channels...



pdf of ξ_θ satisfying the PDE: $\alpha_2(\theta)\frac{\partial^2 p_{\xi_\theta}(x)}{\partial\theta^2} + \alpha_1(\theta)\frac{\partial p_{\xi_\theta}(x)}{\partial\theta} = \beta_2(\theta)\frac{\partial^2 p_{\xi_\theta}(x)}{\partial x^2} + \beta_1(\theta)\frac{\partial p_{\xi_\theta}(x)}{\partial x}$

Entropic version: $$\alpha_2(\theta)\frac{d^2}{d\theta^2} H_\phi(p_Y) + \alpha_1(\theta)\frac{d}{d\theta} H_\phi(p_Y) = \beta_2(\theta) J_\phi(p_Y) - \alpha_2(\theta) J_\phi^{(\theta)}(p_Y)$$

Divergence version: $$\alpha_2(\theta)\frac{d^2}{d\theta^2} D_\phi(p_{Y^{(1)}}\|p_{Y^{(0)}}) + \alpha_1(\theta)\frac{d}{d\theta} D_\phi(p_{Y^{(1)}}\|p_{Y^{(0)}}) = \alpha_2(\theta) J_\phi^{(\theta)}(p_{Y^{(1)}}\|p_{Y^{(0)}}) - \beta_2(\theta) J_\phi(p_{Y^{(1)}}\|p_{Y^{(0)}})$$

- Fokker–Planck equation (state-independent drift and diffusion, $\alpha_2 = 0$): the parametric & nonparametric ϕ−Fisher information characterize the channel.
- Lévy: $\xi_\theta = \theta^2 L$, $\alpha_2 = 1, \beta_1 = 2$; the parametric ϕ−Fisher information characterizes the curvature of the ϕ−entropy/divergence (see Johnson'04 for D_{kl}).
- Cauchy: $\xi_\theta = \theta C$, $\alpha_2 = -\beta_2 = 1$; sum parametric-nonparametric ϕ−Fisher information characterizes the curvature of the ϕ−entropy/divergence (see Johnson'04 for D_{kl}).

Discussion

Generalizations of the usual deBruijn identity, in terms of entropies, or in terms of divergences:

- For Salicrú's ϕ−entropies & Csizàr's ϕ− divergences,
- Beyond the Gaussian channel, (noise pdf satisfying a well-suited second order partial differential equation)

Without many efforts, they extend to the multivariate laws and for a multivariate parameter θ.

- General interpretation?
- Potential implications (e.g., generalized Entropy Power Inequality & Central Limit Theorem)?

References

A. R. Barron, *The Annals of Proba.*, **14**: 336 (1986)
J.-F. Bercher, *J. Math. Phys.*, **53**: 063303 (2012)
J. of Phys. A, **45**: 255303 (2012)
I. Csizàr, *Studia Sci. Math. Hungarica*, **2**: 299 (1967)
O. Johnson, "Info. Th. and The CLT" *Imp. Col. Press* (2004)
E. Lutwak et al., *IEEE Trans. on IT*, **58**:1319 (2012)
M. Salicrú et al., *Com. in Stat.*, **22**: 2015 (1993)
A. J. Stam, *Information and Control*, **2**:101 (1959)