Joint Range of f-divergences

ISIT 2010

Igor Vajda (1942-2010), IEEE Fellow.
Peter Harremoës

f-divergences

f-divergences were introduced independently by

- I. Csiszár 1963
- T. Morimoto 1963
- Ali and Silvey 1966

For two probability measures P and Q on the same measurable space the f-divergence is defined by

$$D_f(P, Q) = \int_{\{q>0\}} f\left(\frac{p}{q}\right) \, dQ + f^*(0)\, P(q=0)$$

where p and q denotes the density of P and Q with respect to some domination measure.

Examples

Total variation $f(t) = |t-1|$.

χ^2-divergence $f(t) = (t-1)^2$

Information divergence $f(t) = t \log t$

Power divergence $f(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)}$

LeCam divergence $f(t) = \frac{(x-1)^2}{x+1}$

Jensen-Shannon divergence $f(t) = \frac{1}{2} \log \frac{2}{t+1} + \frac{t}{2} \log \frac{2t}{t+1}$

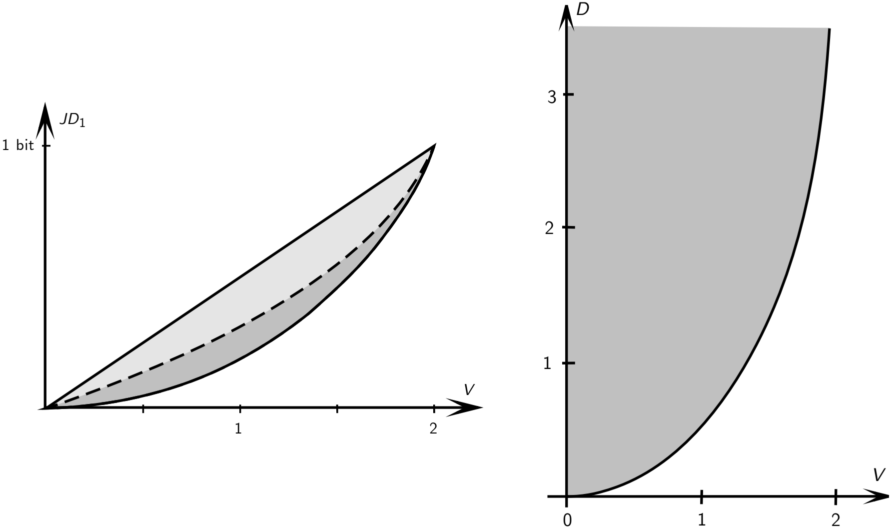


Figure:

Convexity

Definition A point $(x, y) \in \mathbb{R}^2$ is a (f, g)-*divergence pair* if there exist a Borel space $(\mathcal{X}, \mathcal{F})$ with probability measures P and Q such $(x, y) = (D_f(P, Q), D_g(P, Q))$. A (f, g)-divergence pair (x, y) is d-*achievable* if there exist probability vectors $P, Q \in \mathbb{R}^d$ such that

$$(x, y) = (D_f(P, Q), D_g(P, Q)).$$

Lemma Assume that the probability measures P_α and Q_α are singular.where $P_\alpha = (1-\alpha) P_0 + \alpha P_1$ and $Q_\alpha = (1-\alpha) Q_0 + \alpha Q_1$. Then

$$D_f(P_\alpha, Q_\alpha) = (1-\alpha) D_f(P_0, Q_0) + \alpha D_f(P_1, Q_1).$$

Theorem The set of (f, g)-divergence pairs is convex.

Main theorem

Any (f,g)-divergence pair is a convex combination of two (f,g)-divergence pairs, both of them 2-achievable. Consequently, any (f,g)-divergence pair is 4-achievable.

Proof idea Let P and Q denote probability measures on the same measurable space. Define the set $A=\{q>0\}$ and the function $X=p/q$ on A. Then Q satisfies

$$Q\left(A\right)=1,\ \int_A X\ dQ\leq 1.$$

Then

$$D_f\left(P,Q\right)=\int_A \left(f\left(X\right)\ +f^*\left(0\right)\left(1-X\right)\right)\ dQ.$$

The map $Q\rightarrow\left(D_f\left(P,Q\right),D_g\left(P,Q\right)\right)$ is linear. Extreme Q supported on (at most) three points.

Topological argument

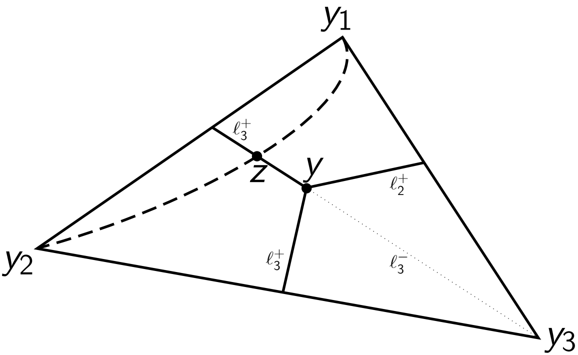


Figure: The slashed curve connects y_1 and y_2. The lines ℓ_1^- and ℓ_2^- are not illustrated.

Image of Triangle

The set of (f, g)-divergence pair that are achievable in $\mathbb{R}^2$ can be parametrized as $P = (1-p, p)$ and $Q = (1-q, q)$. If we define $\overline{(1-p, p)} = (p, 1-p)$ then $D_f (P, Q) = D_f \left(\overline{P}, \overline{Q}\right)$. Hence we may assume without loss of generality assume that $p \leq q$ and just have to determine the image of the simplex $\Delta = \{(p, q) \mid 0 \leq p \leq q \leq 1\}$. This result makes it very easy to make a numerical plot of the (f, g)-divergence pair achievable in $\mathbb{R}^2$ and the joint range is just the convex hull.

Locally 1-1 except if

$$\begin{vmatrix} \frac{\partial D_f}{\partial p} & \frac{\partial D_g}{\partial p} \\ \frac{\partial D_f}{\partial q} & \frac{\partial D_g}{\partial q} \end{vmatrix} = 0.$$

Linear inequalities

Theorem Assume that f and g are C^2 and that $f''(1) > 0$ and $g''(1) > 0$. Assume that $\lim_{t\to 0}\inf \frac{g(t)}{f(t)} > 0$, and that $\lim_{t\to\infty}\inf \frac{g(t)}{f(t)} > 0$. Then there exists $\beta > 0$ such that

$$D_g(P,Q) \geq \beta \cdot D_f(P,Q)$$

for all distributions P, Q.

Power divergences

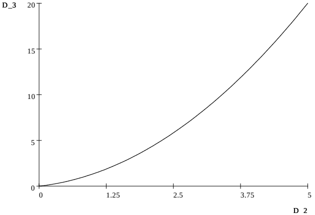
Consider the power divergences D_2 and D_3 with

$$f(t) = \phi_2(t),$$
$$g(t) = \phi_3(t).$$

Then

$$\begin{vmatrix} \frac{\partial D_f}{\partial p} & \frac{\partial D_g}{\partial p} \\ \frac{\partial D_f}{\partial q} & \frac{\partial D_g}{\partial q} \end{vmatrix} = -\frac{1}{12}\left(\frac{p-q}{q(1-q)}\right)^4 .$$

The range consists of all points on the curve $\left(x, \frac{2}{3}x(x+1)\right), x \in [0,\infty)$, and all point above this curve.



Bahadur efficiency of testing goodness-of-fit

Consider a test of
$H_0 : P = Q$
$H_a : P \neq Q$
We use $D_f\left(Emp_n(\omega), Q\right)$ as statistics.
The Bahadur efficiency is determined by

$$\inf D\left(P \| Q\right)$$

where the infimum is taken over all P with $D_f\left(P, Q\right) \geq r$. If $\inf D\left(P \| Q\right) = 0$ then information divergence is infinitely more Bahadur efficient for testing goodness of fit than $D_f\left(P, Q\right)$.

Theorem If $\lim_{t\to 0} \frac{f(t)}{t \log t} = \infty$ or $\lim_{t\to\infty} \frac{f(t)}{t \log t} = \infty$ then information divergence is infinitely more Bahadur efficient than D_f

Conclusion

- Joint range is convex.
- The extreme points are obtained on sets with two points.
- Joint range is easy to determine numerically.
- The exact range is obtained by solving differential equations.
- Consequences for Bahadur efficiency.

Announcements:

Information Geometry and its Applications 2.-6. August, Leipzig.

Prague Stochastics 30. August - 3. September, Prague,

Igor Vajda Memorial Colloquium 12.-13. November, Prague.

Links can be found on www.harremoes.dk/Peter .