Divergences

`Divergences.el`

`Divergences` is a Julia package that makes it easy to evaluate the value of divergences and their derivatives.

Definition

A divergence between $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$ is defined as

$$D(a,b) = \sum_{i=1}^{n} \gamma(a_i/b_i) b_i,$$

where $\gamma : (a_\gamma, +\infty) \to \mathbb{R}_+$, $a_\gamma \in \mathbb{R}$ is strictly convex and twice continuously differentiable on the interior of its domain. The divergence function γ is normalized as to satisfy $\gamma(1) = 0$, $\gamma'(1) = 0$, and $\gamma''(1) = 1$.

The gradient and the hessian of the divergence with respect to a are given by

$$\nabla_a D(a,b) \equiv \left.\frac{\partial D(u,v)}{\partial u}\right|_{u=a, v=b} = \begin{pmatrix} \gamma'(a_1/b_1) \\ \gamma'(a_2/b_2) \\ \vdots \\ \gamma'(a_n/b_n) \end{pmatrix}$$

and

$$\nabla_a^2 D(a,b) \equiv \left.\frac{\partial^2 D(u,v)}{\partial u \partial u}\right|_{u=a, v=b} = \begin{pmatrix} \frac{\gamma''(a_1/b_1)}{b_1} & 0 & \cdots & 0 \\ 0 & \frac{\gamma''(a_2/b_2)}{b_2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\gamma''(a_n/b_n)}{b_n} \end{pmatrix}$$

respectively. Given the normalization $\gamma'(1) = 0$, and $\gamma''(1) = 1$, we have that

$$\nabla_a D(a,a) = 0, \quad \nabla_a^2 D(a,a) = 1.$$

The divergences implemented in the packges are given in the table below together with their first and second order derivatives.

Divergence	$\gamma(u)$	Domain	$\nabla_{\gamma}(u)$	$H_{\gamma}(u)$
Kullback-Leibler	$u\log(u) - u + 1$	$(0, +\infty)$	$\log(u)$	$1/u$
Reverse Kullback-Leibler	$\log(u) + u - 1$	$(0, +\infty)$	$-\frac{1}{u} + 1$	$\frac{1}{u^2}$
Hellinger	$2u + (2 - 4\sqrt{u})$	$(0, +\infty)$	$2\left(1 - \frac{1}{\sqrt{u}}\right)$	$\frac{1}{u^{3/2}}$
Chi-Squared	$\frac{1}{2}(u-1)^2$	$(-\infty, +\infty)$	$u - 1$	1
Cressie-Read	$\frac{u^{1+\alpha}+\alpha-u(1+\alpha)}{\alpha(1+\alpha)}$	$(0, +\infty)$	$\frac{u^{\alpha}-1}{\alpha}$	$u^{\alpha-1}$

The convex conjugate conjugate of γ is defined as

$$\gamma^*(u) = \sup_{u\in\mathbb{R}} \left\{uv - \gamma(u)\right\}.$$

For continuously twice differentiable function, the convex conjugate is

$$\gamma^*(z) = (\gamma')^{-1}(z)\cdot z - \gamma\left((\gamma')^{-1}(z)\right).$$

where $(\gamma')^{-1}(z) := u : \gamma'(x) = z$. The domain of γ^* is $(-\infty, d)$, where

$$d = \lim_{u\to+\infty} \gamma(u)/u.$$

Divergences

The Cressie Read is a family of divergences. Members of this family are indexed a parameter α. This family contains the chi-squared divergence ($\alpha = 1$), the Kullback Leibler divergence ($a \to 0$), the reverse Kullback Leibler divergence ($a \to -1$), and the Hellinger distance ($a = -1/2$).

Since if $\alpha < 0$, γ in the Cressie Read family is not convex on $(-\infty 0)$ and thus we set $\gamma(u) = +\infty$.

Modified divergences

Divergence	$\gamma^*(\theta, b)$	$\lim_{u\to\infty} \frac{\gamma(u)}{u}$	$\lim_{u\to\infty} \frac{u\gamma'(u)}{\gamma(u)}$
Kullback-Leibler	$b(e^\theta - 1)$	$\log b - 1$	1
Reverse Kullback-Leibler	$b\log(1-\theta) + b$, $\theta < 1$	1	1
Hellinger	$b(1-2\sqrt{1-\theta})$, $\theta \leq 1$	2	0
Chi-Squared	$b\left(\theta + \frac{\theta^2}{2}\right)$	∞	2
Cressie-Read	Depends on α	Depends on α	Depends on α
Modified Divergence	Derived from $\gamma_0, \gamma_1, \gamma_2$	Depends on parameters	Depends on parameters
Fully Modified Divergence	Depends on $\gamma_U, \gamma_L, \rho, \phi$	Depends on ρ, ϕ	Depends on ρ, ϕ

For many of the divergences defined above the effective domain of their conjugate, γ^*, does not span $\mathbb{R}$ since $\gamma(u)/u \to l < +\infty$ as $u \to +\infty$.

For some $\vartheta > 0$, let $u_\vartheta \equiv 1 + \vartheta$. The modified divergence γ_ϑ is defined as

$$\gamma_\vartheta(u) = \begin{cases} \gamma(u_\vartheta) + \gamma'(u_\vartheta)(u - u_\vartheta) + \frac{1}{2}\gamma''(u_\vartheta)(u-u_\vartheta)^2, & u \geqslant u_\vartheta \\ \gamma(u), & u \in (0, u_\vartheta) \\ \lim_{u\to 0^+} \gamma(u), & u = 0 \\ +\infty, & u < 0 \end{cases}.$$

It is immediate to verify that this divergence still satisfies all the requirements and normalization of γ. Furthermore, it holds that

$$\lim_{u\to\infty} \frac{\gamma_\vartheta(u)}{u} = +\infty, \qquad \text{and} \qquad \lim_{u\to\infty} \frac{u\gamma_\vartheta'(u)}{\gamma_\vartheta(u)} = 2.$$

The first limit implies that the image of γ_ϑ' is the real line and thus $\overline{\operatorname{dom}\gamma_\vartheta^*} = (-\infty, +\infty)$. The expression for the conjugate is obtained by applying the Legendre-Fenchel transform to obtain

$$\gamma_\vartheta^*(u) = \begin{cases} a_\vartheta \upsilon^2 + b_\vartheta \upsilon + c_\vartheta, & \upsilon > \gamma'(u_\vartheta), \\ \gamma^*(\upsilon), & u \leqslant \gamma'(u_\vartheta) \end{cases},$$

where $a_\vartheta = 1/(2\gamma''(u_\vartheta))$, $b_\vartheta = u_\vartheta - 2a_\vartheta\gamma'(u_\vartheta)$, and $c_\vartheta = -\gamma(u_\vartheta) + a_\vartheta\gamma'(u_\vartheta) - u_\vartheta^2/a_\vartheta$. The conjugate $\gamma_\vartheta^*(u)$ will have a closed form expression when so does the original divergence function.

Fully modified divergences

For some $\vartheta > 0$ and $0 < \varphi < 1 - a_\gamma$, let $u_\vartheta \equiv 1 + \vartheta$ and $u_\varphi = a_\gamma + \varphi$. The **fully** modified divergence $\gamma_{\varphi,\vartheta}$ is defined as

$$\gamma_\vartheta(u) = \begin{cases} \gamma(u_\vartheta) + \gamma'(u_\vartheta)(u - u_\vartheta) + \frac{1}{2}\gamma''(u_\vartheta)(u - u_\vartheta)^2, & u \geqslant u_\vartheta \\ \gamma(u), & u \in (u_\varphi, u_\vartheta) \\ \gamma(u_\varphi) + \gamma'(u_\varphi)(u - u_\varphi) + \frac{1}{2}\gamma''(u_\varphi)(u - u_\varphi)^2, & u \leqslant u_\varphi \end{cases}.$$

It is immediate to verify that this divergence still satisfies all the requirements and normalization of γ, while being defined on all $\mathbb{R}$.

Example of divergences

The following divergence types are defined by `Divergences`.

Kullback-Leibler divergence

$$D^{KL}(a, b) = \sum_{i=1}^{n} \gamma^{KL}(a_i/b_i)b_i$$

$$\gamma^{KL}(u) = u\log(u) - u + 1$$

The gradient and the hessian are given by

$$\nabla_a^2 D^{KL}(a, b) = (\log(a_1/b_1), \dots, \log(a_n, b_n)), \quad \nabla_a^2 D^{KL}(a, b) = \mathrm{diag}(1/a_1, \dots, 1/a_n)$$

Reverse Kullback-Leibler divergence

$$D^{rKL}(a, b) = \sum_{i=1}^{n} \gamma^{rKL}(a_i/b_i)b_i$$

$$\gamma^{rKL}(u) = -\log(u) + u - 1$$

The gradient and the hessian are given by

$$\nabla_a^2 D^{rKL}(a, b) = (1 - b_1/a_1, \dots, 1 - b_n/a_n), \quad \nabla_a^2 D^{rKL}(a, b) = \mathrm{diag}(b_1/a_1^2, \dots, b_n/a_n^2)$$

For reverse Kullback Leibler divergence, $\gamma(u) = -\log(u) + u - 1$, we have that $\gamma(u)/u \to 0$ as $u \to \infty$. The modified reverse Kullback Leibler divergence is given by

$$\gamma_\vartheta(u) = \begin{cases} -\log(u_\vartheta) + (1 - \frac{1}{u_\vartheta})u + \frac{1}{2u_\vartheta^2}(u - u_\vartheta)^2, & u > u_\vartheta \\ -\log(u) + u - 1, & 0 < u \leqslant u_\vartheta \\ +\infty, & u \leqslant 0. \end{cases}$$

The conjugate of γ_θ is given by

$$\gamma_\vartheta(u) = \begin{cases} a_\vartheta \upsilon^2 + b_\vartheta \upsilon + c_\vartheta, & \upsilon > 1 - \frac{1}{u_\vartheta} \\ -\log(1 - \upsilon), & \upsilon \leqslant 1 - \frac{1}{u_\vartheta}, \end{cases}$$

where $a_\vartheta = u_\vartheta^2/2$, $b_\vartheta = u_\vartheta(2 - u_\vartheta)$, and $c_\vartheta = \log(u_\vartheta) - u_\vartheta - 1 + u_\vartheta(u_\vartheta - 1)/2$.

Chi-squared divergence

$$D^\chi(a, b) = \sum_{i=1}^n \gamma^\chi(a_i/b_i) b_i$$

$$\gamma^\chi(u) = u^2/2 - u + 0.5$$

The gradient and the hessian are given by

$$\nabla_a^2 D^\chi(a, b) = \left((a_1 - b_1)/b_1^2, \dots, (a_n - b_n)/b_n^2\right), \quad \nabla_a^2 D^\chi(a, b) = \operatorname{diag}\left(\frac{1}{b_1^2}, \dots, \frac{1}{b_n^2}\right)$$

Cressie-Read divergences

The type `CressieRead` is a family of divergences. Members of this family are indexed by a function γ indexed by parameter α:

$$\gamma_\alpha^{CR}(a, b) = \frac{\left(\frac{a}{b}\right)^{1+\alpha} - 1}{\alpha(\alpha + 1)} - \frac{\left(\frac{a}{b}\right) - 1}{\alpha}.$$

The gradient and the hessian are given by

$$\nabla_a^2 D_\alpha^{CR}(a, b) = \left(\frac{\left(\frac{a_1}{b_1}\right)^\alpha - 1}{\alpha b_1}, \dots, \frac{\left(\frac{a_n}{b_n}\right)^\alpha - 1}{\alpha b_n}\right), \quad \nabla_a^2 D_\alpha^{CR}(a, b) = \operatorname{diag}\left(\frac{\left(\frac{a_1}{b_1}\right)^\alpha}{a_1 b_1}, \dots, \frac{\left(\frac{a_n}{b_n}\right)^\alpha}{a_n b_n}\right)$$