Generalized Alpha-Beta Divergence, its Properties and Associated Entropy

Subhrajyoty Roy

Research Fellow, Interdisciplinary Statistical Research Unit
Indian Statistical Institute, Kolkata, India

May 19, 2025
ICORS 2025, Stresa, Italy

Co-authors

Supratik Basu

Abhik Ghosh

Ayanendranath Basu

Statistical Divergence

A **statistical divergence** $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ such that $d(f, g) = 0$ if and only if $f = g$ almost surely.

Statistical Divergence

A **statistical divergence** $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ such that $d(f, g) = 0$ if and only if $f = g$ almost surely.

Weaker than a metric.

1. $d(f, g)$ and $d(g, f)$ may be different.
2. Usually no relationship between $d(f, g)$ and $d(f, h) + d(h, g)$.

Statistical Divergence

A **statistical divergence** $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ such that $d(f, g) = 0$ if and only if $f = g$ almost surely.

Weaker than a metric.

1. $d(f, g)$ and $d(g, f)$ may be different.
2. Usually no relationship between $d(f, g)$ and $d(f, h) + d(h, g)$.

You can extend for **sub-densities**

$$\mathcal{F}^* := \left\{ f : f \geq 0, \int f d\mu \leq 1 \right\}.$$

Minimum divergence estimation

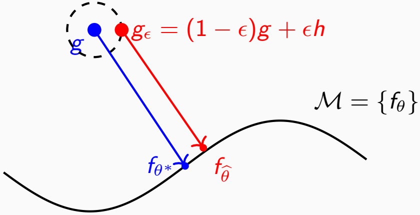
Given a statistical model $\{f_\theta : \theta \in \Theta\}$ and a divergence measure $d(\cdot, \cdot)$ between distributions, the **minimum divergence estimator** (MDE) of θ minimizes the divergence between a "proxy" data density g and the model density f_θ:

$$\theta^* = \arg\min_{\theta \in \Theta} d(g, f_\theta).$$

Minimum divergence estimation

Given a statistical model $\{f_\theta : \theta \in \Theta\}$ and a divergence measure $d(\cdot, \cdot)$ between distributions, the **minimum divergence estimator** (MDE) of θ minimizes the divergence between a "proxy" data density g and the model density f_θ:

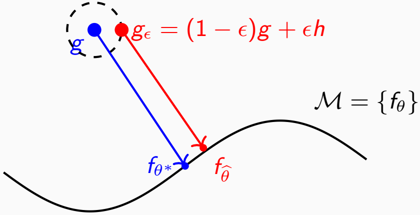
$$\theta^* = \arg\min_{\theta \in \Theta} d(g, f_\theta).$$



Minimum divergence estimation

Given a statistical model $\{f_\theta : \theta \in \Theta\}$ and a divergence measure $d(\cdot, \cdot)$ between distributions, the **minimum divergence estimator** (MDE) of θ minimizes the divergence between a "proxy" data density g and the model density f_θ:

$$\theta^* = \arg\min_{\theta \in \Theta} d(g, f_\theta).$$



The choice of divergence $d(\cdot, \cdot)$ controls the curvature of these geodesic lines.

Useful divergences for Robust Estimation

Several divergences have been proposed to balance **efficiency** and **robustness**:

- **Density Power Divergence (DPD)** [1]:

$$d_{DPD}^{\alpha}(g,f) = \int f^{1+\alpha} - \left(1+\frac{1}{\alpha}\right)\int g f^{\alpha} + \frac{1}{\alpha}\int g^{1+\alpha}.$$

- **Log-Density Power Divergence (LDPD)** [2]:

$$d_{\mathrm{LDPD}}^{\alpha}(g,f) = \log\int f^{1+\alpha} - \left(1+\frac{1}{\alpha}\right)\log\int g f^{\alpha} + \frac{1}{\alpha}\log\int g^{1+\alpha}.$$

- **Bridge Divergence** [3]

$$\begin{aligned} d_{\mathrm{BD}}^{\alpha,c_1,c_2}(g,f) = \log\left(c_1 + c_2\int f^{1+\alpha}\right) - \left(1+\frac{1}{\alpha}\right)\log\left(c_1 + c_2\int g f^{\alpha}\right) \\ + \frac{1}{\alpha}\log\left(c_1 + c_2\int g^{1+\alpha}\right). \end{aligned}$$

Useful divergences for Robust Estimation

- **S-divergence family** [4]:

$$d_{SD}^{\alpha,\lambda}(g,f) = \frac{1}{A}\int f^{1+\alpha} - \frac{1+\alpha}{AB}\int f^B g^A + \frac{1}{B}\int g^{1+\alpha},$$

where $A = 1 + \lambda(1-\alpha)$, $B = 1 + \alpha - \lambda(1-\alpha)$.

- **Logarithmic S-divergence family** [5]:

$$d_{LSD}^{\alpha,\lambda}(g,f) = \frac{1}{A}\log\left(\int f^{1+\alpha}\right) - \frac{1+\alpha}{AB}\log\left(\int f^B g^A\right) + \frac{1}{B}\log\left(\int g^{1+\alpha}\right),$$

where $A = 1 + \lambda(1-\alpha)$, $B = 1 + \alpha - \lambda(1-\alpha)$.

Generalized Alpha-Beta Divergence

We define the generalised alpha-beta divergence (GABD) as

$$d_{GAB}^{(\alpha,\beta),\psi}(f,g) = \frac{1}{\beta(\alpha+\beta)}\psi\left(\int f^{\alpha+\beta}\right) - \frac{1}{\alpha\beta}\psi\left(\int f^{\alpha}g^{\beta}\right) + \frac{1}{\alpha(\alpha+\beta)}\psi\left(\int g^{\alpha+\beta}\right),$$

where ψ is a suitable function.

If ψ is continuously differentiable, then we can define the cases $\alpha = 0, \beta = 0, (\alpha+\beta) = 0$ by taking corresponding limits.

Generalized Alpha-Beta Divergence

We define the generalised alpha-beta divergence (GABD) as

$$d_{GAB}^{(\alpha,\beta),\psi}(f,g) = \frac{1}{\beta(\alpha+\beta)}\psi\left(\int f^{\alpha+\beta}\right) - \frac{1}{\alpha\beta}\psi\left(\int f^{\alpha}g^{\beta}\right) + \frac{1}{\alpha(\alpha+\beta)}\psi\left(\int g^{\alpha+\beta}\right),$$

where ψ is a suitable function.

If ψ is continuously differentiable, then we can define the cases $\alpha = 0, \beta = 0, (\alpha+\beta) = 0$ by taking corresponding limits.

- $\psi(x) = x$, gives you super divergence.
- $\psi(x) = x, \beta = 1$ gives you density power divergence.
- $\psi(x) = \log(x)$, gives you logarithmic super divergence.
- $\psi(x) = \frac{1}{\gamma}x^{\gamma}$, gives you (ϕ, γ)-divergence.

Generalized Alpha-Beta Divergence

We define the generalised alpha-beta divergence (GABD) as

$$d_{GAB}^{(\alpha,\beta),\psi}(f,g) = \frac{1}{\beta(\alpha+\beta)}\psi\left(\int f^{\alpha+\beta}\right) - \frac{1}{\alpha\beta}\psi\left(\int f^{\alpha}g^{\beta}\right) + \frac{1}{\alpha(\alpha+\beta)}\psi\left(\int g^{\alpha+\beta}\right),$$

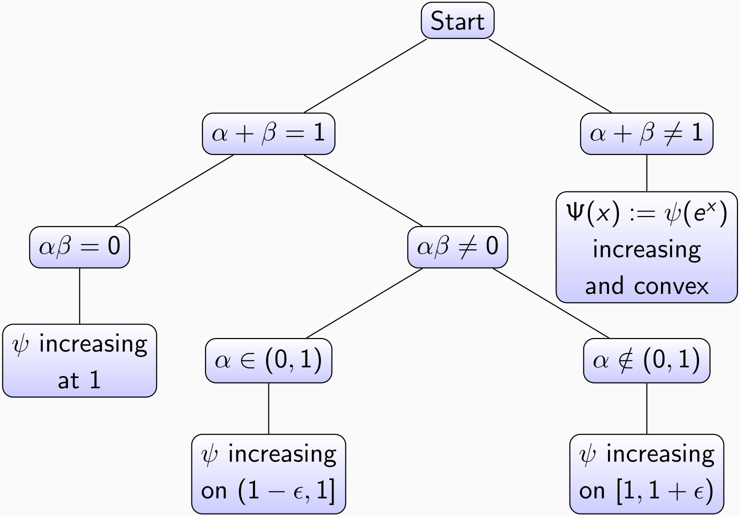
where ψ is a suitable function.

If ψ is continuously differentiable, then we can define the cases $\alpha = 0, \beta = 0, (\alpha+\beta) = 0$ by taking corresponding limits.

- $\psi(x) = x$, gives you super divergence.
- $\psi(x) = x, \beta = 1$ gives you density power divergence.
- $\psi(x) = \log(x)$, gives you logarithmic super divergence.
- $\psi(x) = \frac{1}{\gamma}x^{\gamma}$, gives you (ϕ,γ)-divergence.

Does every $\psi(\cdot)$ make it a divergence?

Key Result: Necessary and Sufficient Conditions



Recipe for new divergences

1. Start with any nonnegative function f.
2. Define, $F(x) = \int_{-\infty}^{x} f(t)dt$. If f is density, F is the cdf.
3. Define $\psi(x) = \int_{-\infty}^{\log(x)} F(u)du$.

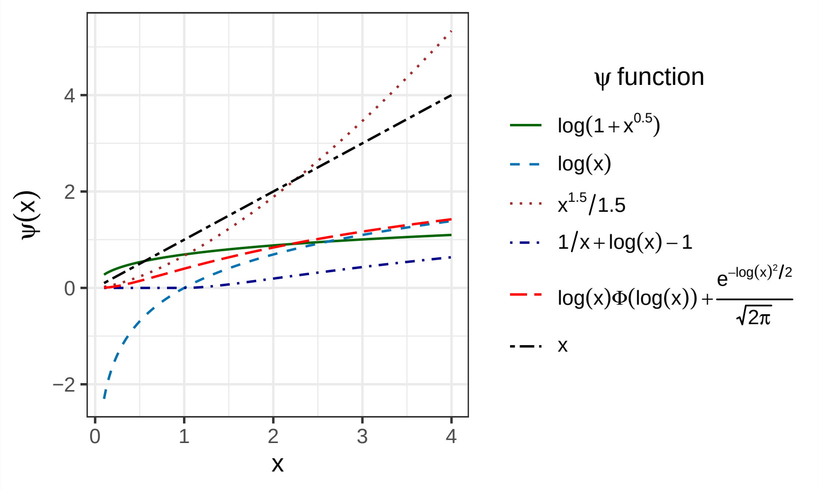
Recipe for new divergences

1. Start with any nonnegative function f.
2. Define, $F(x) = \int_{-\infty}^{x} f(t)dt$. If f is density, F is the cdf.
3. Define $\psi(x) = \int_{-\infty}^{\log(x)} F(u)du$.

Some examples are:

- LSD when $f = \delta(0)$.
- (ϕ, γ)-divergence ($x \leq 1$) when f is the density of $\exp(\gamma)$.
- Bridge divergence when f is $\text{Logistic}(\log(c_1/c_2), 1)$.

New divergences



Special cases

- In location models, all MGABDE are equivalent, i.e., $d_{GAB}^{(\alpha,\beta),\psi_1}(f,g) = h(d_{GAB}^{(\alpha,\beta),\psi_2}(f,g))$ for some monotonic function h.

Special cases

- In location models, all MGABDE are equivalent, i.e., $d_{GAB}^{(\alpha,\beta),\psi_1}(f,g) = h(d_{GAB}^{(\alpha,\beta),\psi_2}(f,g))$ for some monotonic function h.
- For a density f, let $f^{[\alpha]}(x) = f^\alpha(x)/\int f^\alpha(u)du$. When $\beta = 0, \alpha \neq 0$,

$$d_{GAB}^{(\alpha,0),\psi}(f,g) = \frac{\psi'(\int f^\alpha)\int f^\alpha}{\alpha^2}\left(d_{KL}(f^{[\alpha]}, g^{[\alpha]}) + \ln\left(\frac{\int f^\alpha}{\int g^\alpha}\right)\right) \\ - \psi\left(\int f^\alpha\right) + \psi\left(\int g^\alpha\right)$$

Special cases

- In location models, all MGABDE are equivalent, i.e., $d_{GAB}^{(\alpha,\beta),\psi_1}(f,g) = h(d_{GAB}^{(\alpha,\beta),\psi_2}(f,g))$ for some monotonic function h.
- For a density f, let $f^{[\alpha]}(x) = f^{\alpha}(x)/\int f^{\alpha}(u)du$. When $\beta = 0, \alpha \neq 0$,

$$d_{GAB}^{(\alpha,0),\psi}(f,g) = \frac{\psi'(\int f^{\alpha})\int f^{\alpha}}{\alpha^2}\left(d_{KL}(f^{[\alpha]}, g^{[\alpha]}) + \ln\left(\frac{\int f^{\alpha}}{\int g^{\alpha}}\right)\right) - \psi\left(\int f^{\alpha}\right) + \psi\left(\int g^{\alpha}\right)$$

- When $\alpha + \beta = 1, \alpha \notin \{0, 1\}$,

$$d_{GAB}^{(\alpha,1-\alpha),\psi} = \frac{1}{\alpha(1-\alpha)}\left[\psi(1) - \psi\left(d_{PD,\alpha}(f,g) + \text{constant}\right)\right]$$

Properties

- **Duality**: $d_{GAB}^{(\alpha,\beta),\psi}(f,g) = d_{GAB}^{(\beta,\alpha),\psi}(g,f)$.
- **Scaling**: $d_{GAB}^{(\alpha,\beta),\psi}(cf,cg) = d_{GAB}^{(\alpha,\beta),\psi(c^{\alpha+\beta}\cdot)}(f,g)$.
- **Zooming**: $d_{GAB}^{(\alpha,\beta),\psi}(f^{\tau},g^{\tau}) = \tau^2 d_{GAB}^{(\tau\alpha,\tau\beta),\psi}(f,g)$.

Properties

- **Duality**: $d_{GAB}^{(\alpha,\beta),\psi}(f,g) = d_{GAB}^{(\beta,\alpha),\psi}(g,f)$.
- **Scaling**: $d_{GAB}^{(\alpha,\beta),\psi}(cf,cg) = d_{GAB}^{(\alpha,\beta),\psi(c^{\alpha+\beta}\cdot)}(f,g)$.
- **Zooming**: $d_{GAB}^{(\alpha,\beta),\psi}(f^\tau,g^\tau) = \tau^2 d_{GAB}^{(\tau\alpha,\tau\beta),\psi}(f,g)$.

Theorem (Approximate Pythagorean Identity)

1. *Let ψ be continuously differentiable.*
2. $\alpha\beta(\alpha+\beta) \neq 0$.
3. $g_\epsilon^\alpha = (1-\epsilon)g^\alpha + \epsilon\delta^\alpha$.

Then,

$$d_{GAB}^{(\alpha,\beta),\psi}(g_\epsilon,f) = d_{GAB}^{(\alpha,\beta),\psi}(g_\epsilon,g) + d_{GAB}^{(\alpha,\beta),\psi}(g,f) + O_\psi(\epsilon) + O_\psi(\ln(1-\epsilon))$$

Generalized Alpha-Beta Entropy

Definition (GAB Entropy)

$$\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f) = -\frac{1}{\beta}\left[\frac{\psi(\int f^{\alpha+\beta})}{\alpha+\beta} - \frac{\psi(\int f^{\alpha})}{\alpha}\right].$$

Generalized Alpha-Beta Entropy

Definition (GAB Entropy)

$$\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f) = -\frac{1}{\beta}\left[\frac{\psi(\int f^{\alpha+\beta})}{\alpha+\beta} - \frac{\psi(\int f^{\alpha})}{\alpha}\right].$$

1. $\psi(x) = \log(x)$ yields constant $\times$ logarithmic norm entropy.
2. If $\alpha \neq 0, \beta = 0$, $\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f) = c_1 + c_2 H_\alpha(f)$ (Affine transformation of Réyni entropy).

Generalized Alpha-Beta Entropy

Definition (GAB Entropy)

$$\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f) = -\frac{1}{\beta}\left[\frac{\psi(\int f^{\alpha+\beta})}{\alpha+\beta} - \frac{\psi(\int f^{\alpha})}{\alpha}\right].$$

1. $\psi(x) = \log(x)$ yields constant $\times$ logarithmic norm entropy.
2. If $\alpha \neq 0, \beta = 0$, $\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f) = c_1 + c_2 H_\alpha(f)$ (Affine transformation of Réyni entropy).

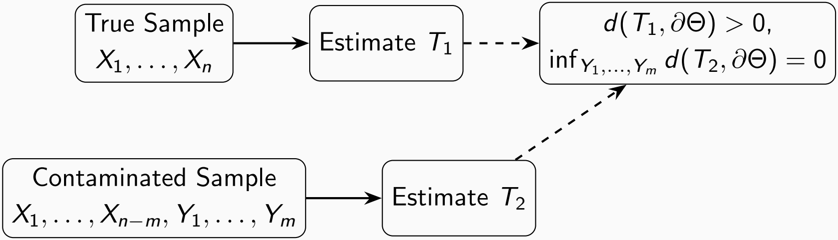
Theorem (Concavity)

$\varepsilon_{GAB}^{(\alpha,\beta),\psi}(f)$ is concave if any of the following holds:

1. *$f \mapsto \ln(\int f^\alpha)$ is convex, and either $\beta > 0, \alpha \in (-\beta, 0)$ or $\alpha > 0, \beta < -\alpha$.*
2. *$\psi(\cdot)$ is convex, and either $\alpha < 0, \beta > (1-\alpha)$ or $\alpha > 1, \beta < -\alpha$.*

Asymptotic Breakdown Point

Breakdown point of an estimator T is the maximum amount of outliers it can tolerate before giving an egregiously bad estimate.



Robustness of Generalized Alpha-Beta Divergence

1. Classical M-estimators have an asymptotic breakdown point at $1/(p+1)$. (Rousseeuw, 1985).

Robustness of Generalized Alpha-Beta Divergence

1. Classical M-estimators have an asymptotic breakdown point at $1/(p+1)$. (Rousseeuw, 1985).

Given a model family of densities $\{f_\theta\}$ and data density g, the minimum GAB divergence functional is

$$T_{MGABD}(G) = \arg\min_{\theta \in \Theta} d_{GAB}^{(\alpha,\beta),\psi}(f_\theta, g)$$

2. Asymptotic BP of MGABD functional at location model is $1/2$.

Robustness of Generalized Alpha-Beta Divergence

1. Classical M-estimators have an asymptotic breakdown point at $1/(p+1)$. (Rousseeuw, 1985).

> Given a model family of densities $\{f_\theta\}$ and data density g, the minimum GAB divergence functional is
>
> $$T_{MGABD}(G) = \arg\min_{\theta\in\Theta} d_{GAB}^{(\alpha,\beta),\psi}(f_\theta, g)$$

2. Asymptotic BP of MGABD functional at location model is $1/2$.
3. Under suitable assumptions, asymptotic breakdown point is atleast

$$\min\left\{\liminf_{d(\theta_m,\partial\Theta)\to 0}\left[\frac{\psi^{-1}\left(\frac{\alpha}{\alpha+\beta}\psi(\int f_{\theta_m}^{\alpha+\beta})\right)}{\int f_{\theta_m}^{\alpha+\beta}}\right]^{1/\beta}, 1-\left[\frac{\psi^{-1}\left(\frac{\alpha}{\alpha+\beta}\psi(\int f_{\theta g}^{\alpha+\beta})\right)}{\int f_{\theta g}^{\alpha} g^{\beta}}\right]^{1/\beta}\right\}$$

In many cases with $\psi(x) = x$, this is $\left(\frac{\alpha}{\alpha+\beta}\right)^{1/\beta}$.

Asymptotic Distribution of MGABDE

1. When $\beta = 1$, the MGABDE is an M-estimator with data-dependent $\psi_M(\cdot)$ or $\rho_M(\cdot)$ functions.
2. As a result, typical consistency and asymptotic normality holds as

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$$

$$V = \frac{N_{1+2\alpha} - M_{1+\alpha}^2}{\left| N_{1+\alpha} + M_{1+\alpha}^2 \frac{\psi''(L_{1+\alpha})}{\psi'(L_{1+\alpha})} \right|^2}$$

   where

$$L_{1+\alpha} = \int f_{\theta^g}^{1+\alpha},\ M_{1+\alpha} = \int f_{\theta^g}^{1+\alpha} u_{\theta^g},\ N_{1+\alpha} = \int f_{\theta^g}^{1+\alpha} u_{\theta^g} u_{\theta^g}^\intercal$$

3. This means, given a model family f_θ, possible to find $\psi(\cdot)$ function that achieves optimal efficiency.

References

- Basu, Ayanendranath, et al. "Robust and efficient estimation by minimising a density power divergence." Biometrika 85.3 (1998): 549-559.
- Jones, M. C., et al. "A comparison of related density-based minimum divergence estimators." Biometrika (2001): 865-873.
- Kuchibhotla, Arun Kumar, Somabha Mukherjee, and Ayanendranath Basu. "Statistical inference based on bridge divergences." Annals of the Institute of Statistical Mathematics 71 (2019): 627-656.
- Ghosh, Abhik, and Ayanendranath Basu. "The minimum S-divergence estimator under continuous models: The Basu–Lindsay approach." Statistical Papers 58 (2017): 341-372.
- Maji, Avijit, Abhik Ghosh, and Ayanendranath Basu. "The logarithmic super divergence and asymptotic inference properties." AStA Advances in Statistical Analysis 100 (2016): 99-131.

Thank you!

Questions?