The Csiszár Index and Variable Ranking

Michael Baudin[1] Cristina Butucea[2] Jean-François Delmas[3] Anne Dutfoy[1] Antoine Schoonaert[1,2,3]

[1]EDF R&D
[2]CREST, ENSAE
[3]CERMICS, École des Ponts

October 9, 2025 — ETICS

Definitions

Let $\mathbb{F}$ be the set of proper convex functions $f : \mathbb{R} \to \mathbb{R}$ such that:

- $\mathrm{dom}\,(f) \subset \mathbb{R}_+$,
- f is right-continuous at 0,
- $f(1) = 0$,
- $1 \in \mathrm{int}(\mathrm{dom}\,(f))$.

We define its convex conjugate $f^* \in \mathbb{F}$ by:

$$f^*(t) = tf\left(\frac{1}{t}\right) \quad \text{for } t \in \mathbb{R}_*^+. \tag{1}$$

Definitions

Let $\mathcal{P} = \mathcal{P}(\Omega, \mathcal{F})$ be the set of probability measures on a measurable space $(\Omega, \mathcal{F})$. We suppose $\mathcal{F}$ is not reduced to the trivial σ-field $\{\emptyset, \Omega\}$.

Definition (f-divergence [1, 4])

For a function $f \in \mathbb{F}$, we denote the f-divergence between P and Q as follows:

$$D_f(P\|Q) = \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q + f^*(0)\,P\left(\frac{\mathrm{d}P}{\mathrm{d}Q} = +\infty\right). \tag{2}$$

Definitions

Let $\mathcal{P} = \mathcal{P}(\Omega, \mathcal{F})$ be the set of probability measures on a measurable space $(\Omega, \mathcal{F})$. We suppose $\mathcal{F}$ is not reduced to the trivial σ-field $\{\emptyset, \Omega\}$.

Definition (f-divergence [1, 4])

For a function $f \in \mathbb{F}$, we denote the f-divergence between P and Q as follows:

$$D_f(P\|Q) = \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q + f^*(0)\,P\left(\frac{\mathrm{d}P}{\mathrm{d}Q} = +\infty\right). \tag{2}$$

Remark: This is not necessarily a distance.

Definitions

Let $\mathcal{P} = \mathcal{P}(\Omega, \mathcal{F})$ be the set of probability measures on a measurable space $(\Omega, \mathcal{F})$. We suppose $\mathcal{F}$ is not reduced to the trivial σ-field $\{\emptyset, \Omega\}$.

Definition (f-divergence [1, 4])

For a function $f \in \mathbb{F}$, we denote the f-divergence between P and Q as follows:

$$D_f(P\|Q) = \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q + f^*(0)\,P\left(\frac{\mathrm{d}P}{\mathrm{d}Q} = +\infty\right). \tag{2}$$

Remark: This is not necessarily a distance.

Examples:

- Kullback Leibler divergence: $f_{\mathrm{KL}}(t) = t\log(t)$.
- Conjugate Kullback Leibler divergence: $f_{\mathrm{KL}^\star}(t) = -\log(t)$.
- Total variation distance: $f_{\mathrm{TV}}(t) = |t-1|$.
- Squared Hellinger distance: $f_{\mathrm{H}}(t) = (\sqrt{t}-1)^2$.

Definitions

Let (X, Y) be a random vector, with X taking values in a measurable space $(\Omega_X, \mathcal{F}_X)$ and Y in $(\Omega_Y, \mathcal{F}_Y)$. Denote by $P_{(X,Y)}$ their joint distribution, by P_X and P_Y the marginals, and by $P_X \otimes P_Y$ the product distribution.

Definition (Csiszár Index [5])

For a function $f \in \mathbb{F}$, we denote the Csiszár index between X and Y as follows:

$$S_f(X, Y) = D_f(P_X \otimes P_Y \| P_{(X,Y)}). \tag{3}$$

Definitions

Let (X, Y) be a random vector, with X taking values in a measurable space $(\Omega_X, \mathcal{F}_X)$ and Y in $(\Omega_Y, \mathcal{F}_Y)$. Denote by $P_{(X,Y)}$ their joint distribution, by P_X and P_Y the marginals, and by $P_X \otimes P_Y$ the product distribution.

Definition (Csiszár Index [5])

For a function $f \in \mathbb{F}$, we denote the Csiszár index between X and Y as follows:

$$S_f(X, Y) = D_f(P_X \otimes P_Y \| P_{(X,Y)}). \tag{3}$$

Example: Mutual Information $\mathrm{I}(X, Y) = D_{\mathrm{KL}\star}(P_X \otimes P_Y \| P_{(X,Y)})$.

Definitions

Let (X, Y) be a random vector, with X taking values in a measurable space $(\Omega_X, \mathcal{F}_X)$ and Y in $(\Omega_Y, \mathcal{F}_Y)$. Denote by $P_{(X,Y)}$ their joint distribution, by P_X and P_Y the marginals, and by $P_X \otimes P_Y$ the product distribution.

Definition (Csiszár Index [5])

For a function $f \in \mathbb{F}$, we denote the Csiszár index between X and Y as follows:

$$S_f(X, Y) = D_f(P_X \otimes P_Y \| P_{(X,Y)}). \tag{3}$$

Example: Mutual Information $\mathrm{I}(X, Y) = D_{\mathrm{KL}\star}(P_X \otimes P_Y \| P_{(X,Y)})$.

We have the following properties for Csiszár index:

(i) **Symmetry.** $S_f(X, Y) = S_f(Y, X)$.

(ii) **Bounds.** $0 \leq S_f(X, Y) \leq f(0) + f^*(0)$.

(iii) **Independence.** If $X \perp\!\!\!\perp Y$, then $S_f(X, Y) = 0$.

(iv) **Implication of independence.** If f is strictly convex at 1, then $S_f(X, Y) = 0 \implies X \perp\!\!\!\perp Y$.

Variable Transformation in the Csiszár Index

Motivation: How does the Csiszár index change when applying a transformation to X or Y ?

Variable Transformation in the Csiszár Index

Motivation: How does the Csiszár index change when applying a transformation to X or Y ?

Proposition

Let $f \in \mathbb{F}$. For $i = 1, 2$, let X_i a random variable taking values in a measurable space $(\Omega_i, \mathcal{F}_i)$ and φ_i a measurable function defined on $(\Omega_i, \mathcal{F}_i)$ and taking values in a measurable space $(E_i, \mathcal{E}_i)$. We have:

(i) **Variable transformation reduces the Csiszár index.**

$$S_f(\varphi_1(X_1), \varphi_2(X_2)) \leq S_f(X_1, X_2). \quad (4)$$

(ii) **Marginal invariance.** *If $\mathrm{d}P_{X_1} \otimes \mathrm{d}P_{X_2} / \mathrm{d}P_{(X_1, X_2)}$ is $\sigma(\varphi_1) \otimes \sigma(\varphi_2)$-measurable, then we have:*

$$S_f(X_1, X_2) = S_f(\varphi_1(X_1), \varphi_2(X_2)). \quad (5)$$

(iii) **Invariance by injection.** *If, for $i = 1, 2$ the function φ_i is injective and bi-measurable, then we have:*

$$S_f(X_1, X_2) = S_f(\varphi_1(X_1), \varphi_2(X_2)). \quad (6)$$

Variable Transformation in the Csiszár Index

Example of invariance of the Csiszár index:

(i) **Marginal invariance.** Suppose that X and Y are real valued random variable and that (X, Y) is symmetric. Then, we have :

$$S_f(|X|, |Y|) = S_f(X, Y). \tag{7}$$

(ii) **Invariance by bijection.** Suppose that X and Y are real valued random variable and that the distributions of X and of Y are absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}$ with a positive density. Then we have:

$$S_f(F_X(X), F_Y(Y)) = S_f(X, Y). \tag{8}$$

What is a Copula ?

A d-dimensional **copula** is a cumulative distribution function (cdf) $C : [0,1]^d \to [0,1]$ with uniform marginals.

Let $Z = (Z_1, \ldots, Z_d)$ be a real random vector with cdf F_Z. We say that a d-dimensional copula C is a copula for Z if for all $z_1, \ldots, z_d \in \mathbb{R}$:

$$F_Z(z_1, \ldots, z_d) = C(F_{Z_1}(z_1), \ldots, F_{Z_d}(z_d)). \tag{9}$$

What is a Copula ?

A d-dimensional **copula** is a cumulative distribution function (cdf) $C : [0,1]^d \to [0,1]$ with uniform marginals.

Let $Z = (Z_1, \ldots, Z_d)$ be a real random vector with cdf F_Z. We say that a d-dimensional copula C is a copula for Z if for all $z_1, \ldots, z_d \in \mathbb{R}$:

$$F_Z(z_1, \ldots, z_d) = C(F_{Z_1}(z_1), \ldots, F_{Z_d}(z_d)). \tag{9}$$

Key points:

- Copulas capture the dependence structure separately from the marginals.
- For any real random vector, there exists a copula for this random vector and this copula is unique if all marginals are continuous. Sklar's Theorem [3]

Examples: Clayton copula, Gumbel copula, Gaussian copula ...

Csiszár index in Terms of the Copula

If the pair (X, Y) is a continuous random vector then we have:

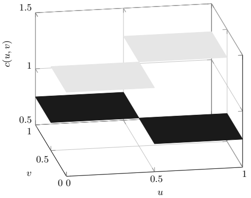
$$S_f(X, Y) = D_f(C_X \otimes C_Y \| C_{(X,Y)}). \tag{10}$$

Csiszár index in Terms of the Copula

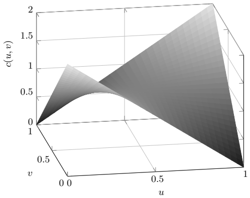
If the pair (X, Y) is a continuous random vector then we have:

$$S_f(X, Y) = D_f(C_X \otimes C_Y \| C_{(X,Y)}). \tag{10}$$

If the pair (X, Y) is not continuous, then the copula is no longer unique and Equation (10) is no longer valid in general. Let us consider X and Y as two dependent Bernoulli random variables. We consider two possible copulas for the pair (X, Y).



(a) $D_f(C_X \otimes C_Y \| C_{(X,Y)}) = 1/15$



(b) $D_f(C_X \otimes C_Y \| C_{(X,Y)}) = (\pi^2/8) - 1$

Figure: Density Plots of Two Copulas for the Pair (X, Y).

Csiszár Index in Terms of the Copula

Definition (Checkerboard copulas [2])

Let $Z = (Z_1, \ldots, Z_d)$ be a real random vector. Denote by $\Delta(Z_i) = \{z \in \mathbb{R} : \mathbb{P}(Z_i = z) > 0\}$ the set of atoms of Z_i. The (unique) checkerboard copula C_Z^{cb} is the copula where each marginal U_i is defined as:

$$U_i = F_{Z_i}(Z_i-) + T_i \sum_{x \in \Delta(Z_i)} \mathbb{P}(Z_i = x)\, \mathbf{1}_{\{Z_i = x\}}. \tag{11}$$

Csiszár Index in Terms of the Copula

Definition (Checkerboard copulas [2])

Let $Z = (Z_1, \ldots, Z_d)$ be a real random vector. Denote by $\Delta(Z_i) = \{z \in \mathbb{R} : \mathbb{P}(Z_i = z) > 0\}$ the set of atoms of Z_i. The (unique) checkerboard copula C_Z^{cb} is the copula where each marginal U_i is defined as:

$$U_i = F_{Z_i}(Z_i-) + T_i \sum_{x \in \Delta(Z_i)} \mathbb{P}(Z_i = x)\, \mathbf{1}_{\{Z_i = x\}}. \tag{11}$$

Example: Let X and Y be two Bernoulli random variables with respective parameters p and q.

The copula $C_{(X,Y)}^{\mathrm{cb}}$ has the following marginals:

$$U = F_X(X-) + T \sum_{x \in \{0,1\}} \mathbb{P}(X = x)\mathbf{1}_{\{X=x\}},$$

$$V = F_Y(Y-) + T' \sum_{y \in \{0,1\}} \mathbb{P}(Y = y)\mathbf{1}_{\{Y=y\}},$$

where $T, T' \sim \mathrm{Uniform}(0,1)$ are independent.

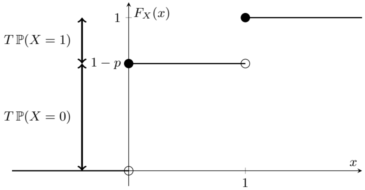


Figure: Cdf of the random variable X.

Csiszár Index in Terms of the Copula

Definition (Checkerboard copulas [2])

Let $Z = (Z_1, \dots, Z_d)$ be a real random vector. Denote by $\Delta(Z_i) = \{z \in \mathbb{R} : \mathbb{P}(Z_i = z) > 0\}$ the set of atoms of Z_i. The (unique) checkerboard copula C_Z^{cb} is the copula where each marginal U_i is defined as:

$$U_i = F_{Z_i}(Z_i-) + T_i \sum_{x \in \Delta(Z_i)} \mathbb{P}(Z_i = x)\, \mathbf{1}_{\{Z_i = x\}}. \tag{11}$$

Example: Let X and Y be two Bernoulli random variables with respective parameters p and q.

The copula $C^{\mathrm{cb}}_{(X,Y)}$ has the following marginals:

$$U = F_X(X-) + T \sum_{x \in \{0,1\}} \mathbb{P}(X = x)\mathbf{1}_{\{X=x\}},$$

$$V = F_Y(Y-) + T' \sum_{y \in \{0,1\}} \mathbb{P}(Y = y)\mathbf{1}_{\{Y=y\}},$$

where $T, T' \sim \mathrm{Uniform}(0,1)$ are independent.

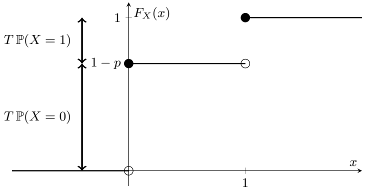


Figure: Cdf of the random variable X.

Remark: For a continuous random vector Z, the checkerboard copula equals its unique copula.

Csiszár index in Terms of the Copula

Theorem (Checkerboard copula minimises the f-divergence)

Let (X, Y) be a random vector, and $f \in \mathbb{F}$.

(i) *For any copula $C_{(X,Y)}$ with marginals C_X and C_Y:*

$$S_f(X, Y) \leq D_f(C_X \otimes C_Y \,\|\, C_{(X,Y)}). \tag{12}$$

(ii) *For the checkerboard copulas:*

$$S_f(X, Y) = D_f(C_X^{\mathrm{cb}} \otimes C_Y^{\mathrm{cb}} \,\|\, C_{(X,Y)}^{\mathrm{cb}}) = \min D_f(C_X \otimes C_Y \,\|\, C_{(X,Y)}), \tag{13}$$

where the minimum is over all copulas of (X, Y).

Remark: The checkerboard copula is not necessarily the only copula that achieves this minimum.

Gaussian Case

Motivation: A priori, the Csiszár index does not provide the same ranking as variance-based indices. Let $Y = X_1 + X_2$ with $X_1 \perp\!\!\!\perp X_2$, $X_1 \sim \mathcal{N}(0, 1.8^2)$ and $X_2 \sim$ Logistic$(0,1)$. We then have:

$$S_{f_{\mathrm{KL}\star}}(X_1, Y) \geq S_{f_{\mathrm{KL}\star}}(X_2, Y) \quad \text{whereas} \quad \mathrm{corr}(X_1, Y) \leq \mathrm{corr}(X_2, Y).$$

Are there cases where the Csiszár index behaves like variance-based indices ?

Gaussian Case

Motivation: A priori, the Csiszár index does not provide the same ranking as variance-based indices. Let $Y = X_1 + X_2$ with $X_1 \perp\!\!\!\perp X_2$, $X_1 \sim \mathcal{N}(0, 1.8^2)$ and $X_2 \sim$ Logistic$(0,1)$. We then have:

$$S_{f_{\mathrm{KL}\star}}(X_1, Y) \geq S_{f_{\mathrm{KL}\star}}(X_2, Y) \quad \text{whereas} \quad \mathrm{corr}(X_1, Y) \leq \mathrm{corr}(X_2, Y).$$

Are there cases where the Csiszár index behaves like variance-based indices ?

Theorem

Let (X, Y) be a Gaussian vector and consider the correlation matrix $C = \Sigma_X^{-1/2}\mathrm{Cov}(X, Y)\Sigma_Y^{-1/2}$ with Σ_X and Σ_Y the covariance matrices of X and Y. Let $\lambda \in [0,1]^d$ be the vector of the positive eigenvalues of $C^\top C$ (and of $CC^\top$) with $d \geq 0$ the rank of C. Then $S_f(X, Y)$ can be viewed as a function of λ, which is non-decreasing in each λ_i.

Gaussian Case

Corollary

Consider a Gaussian vector (X, X', Y)*, of dimension* $d_X \times d_{X'} \times d_Y$ *with covariance matrix* Σ_X *for* X*,* $\Sigma_{X'}$ *for* X' *and* Σ_Y *for* Y*. We have these three properties:*

(i) *If* $d_Y = d_X = d_{X'} = 1$*, then we have:*

$$\forall f \in \mathbb{F}, S_f(X,Y) \leq S_f(X',Y) \iff \mathrm{corr}(X,Y)^2 \leq \mathrm{corr}(X',Y)^2.$$

(ii) *If* $d_Y = 1$*, then we have:*

$$\forall f \in \mathbb{F}, S_f(X,Y) \leq S_f(X',Y) \iff \|\Sigma_X^{-1/2}\mathrm{Cov}(X,Y)\|_2^2 \leq \|\Sigma_{X'}^{-1/2}\mathrm{Cov}(X',Y)\|_2^2.$$

(iii) *If* $d_Y = 1$ *and both* Σ_X *and* $\Sigma_{X'}$ *are diagonal, then we have:*

$$\forall f \in \mathbb{F}, S_f(X,Y) \leq S_f(X',Y) \iff \sum_{i=1}^{d_X} \mathrm{corr}(X_i,Y)^2 \leq \sum_{i=1}^{d_{X'}} \mathrm{corr}(X'_i,Y)^2.$$

In the Gaussian vector setting, the ranking of the variables reduces to the ranking of their correlations.

The Additive Independent Model

Theorem (Comparison of the Csiszár index in the additive independent model)

Let X, X', U and W be independent random variables such that $X \overset{\mathcal{L}}{=} X'$. Denote $Z = X' + U$ and consider the additive model:

$$Y = X + Z + W. \tag{14}$$

Then for all $f \in \mathbb{F}$, we have:

$$S_f(X, Y) \leq S_f(Z, Y). \tag{15}$$

The Additive Independent Model

Theorem (Comparison of the Csiszár index in the additive independent model)

Let X, X', U and W be independent random variables such that $X \stackrel{\mathcal{L}}{=} X'$. Denote $Z = X' + U$ and consider the additive model:

$$Y = X + Z + W. \tag{14}$$

Then for all $f \in \mathbb{F}$, we have:

$$S_f(X, Y) \leq S_f(Z, Y). \tag{15}$$

Example: Consider two independent centred Gaussian variables X and Z with respective variances σ_1^2 and σ_2^2, and assume $\sigma_1^2 \leq \sigma_2^2$. Then Z can be rewritten as the sum of two independent Gaussians:

$$Z = X' + U \quad \text{with} \quad X' \sim \mathcal{N}(0, \sigma_1^2),\ U \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2),$$

and $X' \perp\!\!\!\perp U$. It follows that the ranking of the variables is determined by the order of their variances, and consequently by their correlations.

Conclusion

Our contributions:

- We introduce new transformations that preserve the Csiszár index.
- We establish a connection between the Csiszár index and the associated copula, including cases where the copula is not unique.
- We analyze the Csiszár index for Gaussian vectors and describe its main properties.
- We compare the Csiszár index in the additive independent model.

Conclusion

Our contributions:

- We introduce new transformations that preserve the Csiszár index.
- We establish a connection between the Csiszár index and the associated copula, including cases where the copula is not unique.
- We analyze the Csiszár index for Gaussian vectors and describe its main properties.
- We compare the Csiszár index in the additive independent model.

Thank you for your attention! Any questions?

References

Imre Csiszár.
Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten.
Magyer Tud. Akad. Mat. Kutato Int. Koezl., 8:85–108, 1964.

Liyuan Lin, Ruodu Wang, Ruixun Zhang, and Chaoyi Zhao.
The checkerboard copula and dependence concepts.
SIAM Journal on Financial Mathematics, 16(2):426–447, 2025.

Roger B Nelsen.
An introduction to copulas.
Springer, 2006.

Yury Polyanskiy and Yihong Wu.
Information theory: From coding to learning.
Cambridge university press, 2025.

Sharif Rahman.
The f-sensitivity index.
SIAM/ASA Journal on Uncertainty Quantification, 4(1):130–162, 2016.