Connections between cumulative information measures and activation functions

ENBIS Spring Meeting 2026 – Next-Gen Session

Speaker: **Giulia Pisano**
email: *gpisano@unisa.it*

Dipartimento di Matematica,
Università degli Studi di Salerno

enbis European Network for Business and Industrial Statistics

Table of contents

The cumulative information ψ-measure

A generalized framework for cumulative information measures

Formal definitions

Definition 1

Let X be a random variable with SF $\overline{F}$. Let $\psi : [0,1] \to \mathbb{R}_0^+$ be a Riemann-integrable function. The cumulative information ψ-measure of X is defined as

$$\mathcal{CI}_\psi(X) = \int_{\mathbb{R}} \psi(\overline{F}(t))\mathrm{d}t, \tag{1}$$

provided that the latter integral is finite.

Related measures

Let $\Psi = \{\psi : [0,1] \to \mathbb{R}_0^+;\ \psi$ is continuous, such that $\psi(0) = \psi(1) = 0\}$.

	$\psi \in \Psi$	$\mathcal{CI}_\psi(X)$
(i)	$-u \log u$	$\mathrm{CRE}(X) = -\int_{\mathbb{R}} \overline{F}(t) \log \overline{F}(t) \mathrm{d}t$, cumulative residual entropy (Rao et al. [17])
(ii)	$-(1-u)\log(1-u)$	$\mathrm{CE}(X) = -\int_{\mathbb{R}} F(t) \log F(t) \mathrm{d}t$, cumulative entropy (Di Crescenzo and Longobardi [10])
(iii)	$-u \log u - (1-u)\log(1-u)$	$\mathrm{CPE}(X) = \mathrm{CRE}(X) + \mathrm{CE}(X)$, cumulative paired Shannon entropy (Klein et al. [13])
(iv)	$\frac{1}{n!}\, u\,[-\log u]^n,\ n \in \mathbb{N}_0$	$\mathrm{CRE}_n(X) = \frac{1}{n!} \int_{\mathbb{R}} \overline{F}(t) \left[-\log \overline{F}(t)\right]^n \mathrm{d}t$, generalized cumulative residual entropy (Psarrakos and Navarro [14])
(v)	$\frac{1}{n!}\,(1-u)\,[-\log(1-u)]^n,\ n \in \mathbb{N}$	$\mathrm{CE}_n(X) = \frac{1}{n!} \int_{\mathbb{R}} F(t) \left[-\log F(t)\right]^n \mathrm{d}t$, generalized cumulative entropy (Kayal [12])

Table: Some special cases of $\mathcal{CI}_\psi(X)$ defined in Eq. (1) for different choices of $\psi \in \Psi$.

Related measures

	$\psi \in \Psi$	$\mathcal{CI}_\psi(X)$
(vi)	$\frac{1}{\Gamma(\nu+1)}\, u\,[-\log u]^\nu,\ \nu \in \mathbb{R}_0^+$	$\mathrm{CRE}_\nu(X) = \frac{1}{\Gamma(\nu+1)} \int_{\mathbb{R}} \overline{F}(t)\,\left[-\log \overline{F}(t)\right]^\nu \mathrm{d}t$, fractional generalized cumulative residual entropy (Di Crescenzo et al. [9])
(vii)	$\frac{1}{\Gamma(\nu+1)}\,(1-u)\,[-\log(1-u)]^\nu,\ \nu \in \mathbb{R}^+$	$\mathrm{CE}_\nu(X) = \frac{1}{\Gamma(\nu+1)} \int_{\mathbb{R}} F(t)\,[-\log F(t)]^\nu \mathrm{d}t$, fractional generalized cumulative entropy (Di Crescenzo et al. [9])
(viii)	$\frac{1}{s-1}\,(u-u^s),\ s \in \mathbb{R}^+\backslash\{1\}$	$\mathrm{CRTE}_s(X) = \frac{1}{s-1} \int_{\mathbb{R}} \left\{\overline{F}(t) - \left[\overline{F}(t)\right]^s\right\} \mathrm{d}t$, cumulative residual Tsallis entropy (Rajesh and Sunoj [16])
(ix)	$\frac{1}{s-1}\,[(1-u)-(1-u)^s],\ s \in \mathbb{R}^+\backslash\{1\}$	$\mathrm{CTE}_s(X) = \frac{1}{s-1} \int_{\mathbb{R}} \left\{F(t) - [F(t)]^s\right\} \mathrm{d}t$, cumulative Tsallis entropy (Simon and Dulac [18])
(x)	$2(1-u)u$	$\mathrm{GMD}(X) = 2 \int_{\mathbb{R}} F(t)\overline{F}(t)\mathrm{d}t$, Gini mean difference (Arnold and Sarabia [2])
(xi)	$(1-u)^\alpha u^\beta,\ \alpha, \beta \in \mathbb{R}\backslash\{0\}$	$\mathrm{G}_X(\alpha,\beta) = \int_{\mathbb{R}} (F(t))^\alpha (\overline{F}(t))^\beta \mathrm{d}t$, cumulative information generating function (Capaldo et al. [5])
(xii)	$d_1(1-u)\,d_2(u)$	$\widehat{\mathrm{G}}_X(\boldsymbol{d}) = \int_{\mathbb{R}} d_1\,(F(t))\,d_2\,(\overline{F}(t))\,\mathrm{d}t$, distorted Gini function (Capaldo et al. [5])

Table: Some special cases of $\mathcal{CI}_\psi(X)$ defined in Eq. (1) for different choices of $\psi \in \Psi$. In (xii) $\boldsymbol{d} = (d_1, d_2)$, where d_1 and d_2 are suitable distortions.

New generalized entropies

Some cases of the previous Table leads to the following new generalized entropies:

- Let ψ be the sum of the ψ-functions given in cases (iv) and (v). Then, we define the *generalized cumulative paired entropy* as

$$\mathrm{CPE}_n(X) = \mathrm{CRE}_n(X) + \mathrm{CE}_n(X), \qquad n \in \mathbb{N},$$

  which is a suitable generalized version of the cumulative paired entropy given in (iii).

- Let ψ be the sum of the ψ-functions given in (vi) and (vii). Then, we define the *fractional generalized cumulative paired entropy* as

$$\mathrm{CPE}_\nu(X) = \mathrm{CRE}_\nu(X) + \mathrm{CE}_\nu(X), \qquad \nu \in \mathbb{R}^+,$$

  which represents a new fractional version of the cumulative paired entropy given in (iii).

- Let ψ be the sum of the ψ-functions given in (viii) and (ix). Then, the *cumulative paired Tsallis entropy* is given by

$$\mathrm{CPTE}_s(X) = \mathrm{CRTE}_s(X) + \mathrm{CTE}_s(X), \qquad s \in \mathbb{R}^+ \setminus \{1\}.$$

Alternative representations

Let X be a random variable with SF $\overline{F}$, CDF F and mean $\mathsf{E}(X)$. Let $\psi : [0,1] \to \mathbb{R}_0^+$ be a Riemann-integrable function.

(i) **Covariance representation:**
If ψ is differentiable and such that $\lim_{t\to\pm\infty} t\,\psi(\overline{F}(t)) = 0$, then

$$\mathcal{CI}_\psi(X) = \mathrm{Cov}(X, \psi'(\overline{F}(X))) + (\psi(1) - \psi(0))\,\mathsf{E}(X). \tag{2}$$

(ii) **Quantile-based representation:**
If X has absolutely continuous SF and PDF f, then

$$\mathcal{CI}_\psi(X) = \int_0^1 \psi(u)\widetilde{q}(u)\mathrm{d}u,$$

where $\widetilde{q}(u) = 1/f\left(\overline{F}^{-1}(u)\right)$, for $u \in (0,1)$, is the dual quantile density function (DQDF) of X, see Capaldo et al. [7].

Informational properties

The cumulative information ψ-measure in Eq. (1) is a variability measure in the sense of Bickel and Lehmann [3].

Proposition 2

Let f_X (f_Y) and $\overline{F}_X$ ($\overline{F}_Y$) be the PDF and the SF of X (Y), respectively.
For any Riemann-integrable function $\psi : [0,1] \to \mathbb{R}_0^+$ such that $\mathcal{CI}_\psi$ is well-defined, it holds:

(i) *if $X =_{st} Y$ (i.e., $\overline{F}_X \equiv \overline{F}_Y$), then $\mathcal{CI}_\psi(X) = \mathcal{CI}_\psi(Y)$ (law invariance);*

(ii) *$\mathcal{CI}_\psi(X + \delta) = \mathcal{CI}_\psi(X)$, for all $\delta \in \mathbb{R}$ (translation invariance);*

(iii) *$\mathcal{CI}_\psi(\upsilon X) = \upsilon\, \mathcal{CI}_\psi(X)$, for all $\upsilon \in \mathbb{R}^+$ (positive homogeneity);*

(iv) *$\mathcal{CI}_\psi(X) \geq 0$ for any random variable X. One also has $\mathcal{CI}_\psi(X) = 0$ for any degenerate random variable X and provided that $\psi(0) = \psi(1) = 0$ (non-negativity);*

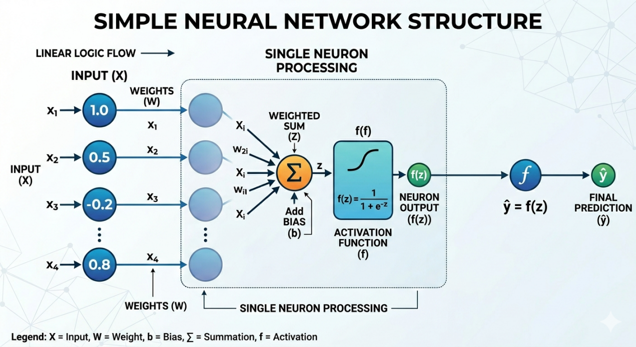
(v) *If $X \leq_{disp} Y$, then $\mathcal{CI}_\psi(X) \leq \mathcal{CI}_\psi(Y)$, provided that X and Y have absolutely continuous distribution (consistency with the dispersive order).*

Note that $X \leq_{disp} Y$ if and only if $f_X(\overline{F}_X^{-1}(u)) \geq f_Y(\overline{F}_Y^{-1}(u))$, for any $u \in [0,1]$.

Connections with neural networks

A bridge to machine learning

What is an activation function?



Example:

- $\mathrm{ReLU}(x) = \max\{0, x\}$, $x \in \mathbb{R}$
- $\text{Leaky ReLU}(x) = \max\{0.01x, x\}$, $x \in \mathbb{R}$
- $\mathrm{Sigmoid}(x) = \frac{e^x}{1+e^x}$, $x \in \mathbb{R}$
- $\mathrm{Tanh}(x) = \frac{e^x - ^{-x}}{e^x+e^{-x}}$, $x \in \mathbb{R}$

A new activation function

Definition 3

For $\nu, \omega \in \mathbb{R}^+$, the generalized logistic-linear activation function is defined as $\gamma_{\nu,\omega} : \mathbb{R} \to [0, 1]$, where

$$\gamma_{\nu,\omega}(x) = \begin{cases} 0, & x \leq -\frac{\nu}{\omega}, \\ \dfrac{1}{1 + \left(1 - \frac{2\omega x}{\omega x + \nu}\right)^{\nu}}, & |x| < \frac{\nu}{\omega}, \\ 1, & x \geq \frac{\nu}{\omega}. \end{cases} \tag{3}$$

(i) $\gamma_{\nu,\omega}(x)$ is a strictly increasing and absolutely continuous CDF for $x \in \left(-\frac{\nu}{\omega}, \frac{\nu}{\omega}\right)$;

(ii) $\gamma_{\nu,\omega}(0) = 1/2$;

(iii) $\gamma_{1,\omega}(x)$ is a linear activation function, i.e., $\gamma_{1,\omega}(x) = \dfrac{\omega}{2} x + \dfrac{1}{2}$, for $x \in \left(-\frac{1}{\omega}, \frac{1}{\omega}\right)$;

(iv) $\lim\limits_{\nu \to +\infty} \gamma_{\nu,\omega}(x) = \left(1 + e^{-2\omega x}\right)^{-1}$, for any $x \in \mathbb{R}$, that is a sigmoid activation function.

Let us define the ψ-function related to the generalized logistic-linear activation function given in (3), i.e.,

$$\psi_{\nu,\omega}(u) = -\int_0^u \gamma_{\nu,\omega}^{-1}(t)\,\mathrm{d}t, \qquad u \in [0,1]\,, \;\; \nu,\omega \in \mathbb{R}^+, \tag{4}$$

Theorem 4

Let $\psi_1 = \psi_{\nu_1,\omega_1}$ and $\psi_2 = \psi_{\nu_2,\omega_2}$, with $\nu_1, \nu_2, \omega_1, \omega_2 \in \mathbb{R}^+$. If one of the following conditions holds:

(i) $\nu_1 < \nu_2$ *and* $\omega_1 = \omega_2$,

(ii) $\nu_1 = \nu_2$ *and* $\omega_1 > \omega_2$,

then

$$\mathcal{CI}_{\psi_1}(X) < \mathcal{CI}_{\psi_2}(X),$$

for any random variable X.

For instance, $\mathcal{CI}_{\psi_{1,1}}(X) = \mathrm{GMD}(X) < CPE(X) = \mathcal{CI}_{\psi_{\nu\to+\infty,1}}(X)$.

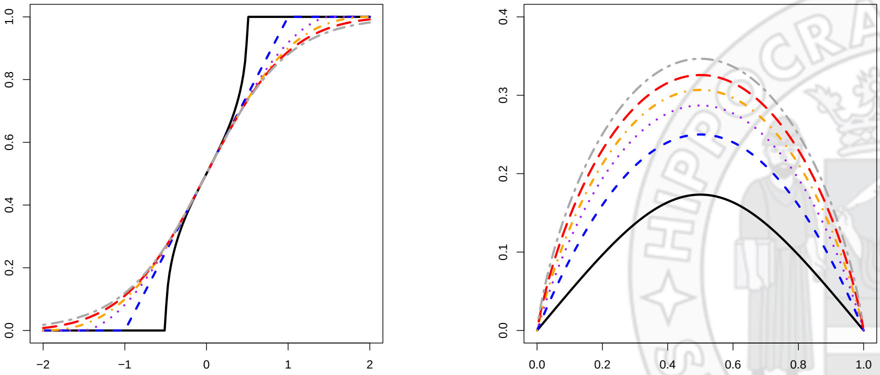
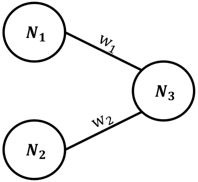


Figure: Plots (left) of the activation functions $\gamma_{\nu,\omega}$ obtained from Eq. (3) for different choices of ν and $\omega = 1$, and (right) of the corresponding $\psi_{\nu,\omega}$. Specifically, solid line refer to case $\nu = 0.5$, dashed to $\nu = 1$, dotted to $\nu = 1.5$, dotdash to $\nu = 2$, longdash to $\nu = 3$ and twodash to $\nu \to +\infty$.

Application

Application



Let us consider a neural network made up of three nodes. We assume that the weights are given by

$$(\mathrm{w}_1, \mathrm{w}_2) = \left(\frac{\nu}{\omega} - \epsilon, -\frac{\nu}{\omega} - \epsilon\right), \qquad \nu, \omega \in \mathbb{R}^+, \quad \epsilon \in \left[0, \frac{\nu}{\omega}\right), \tag{5}$$

where ϵ denotes the bias.

Let X_i be the random variable describing the information provided by the node N_i, for $i = 1, 2, 3$. Assume that X_1 and X_2 are independent and uniformly distributed in $(0, 1)$, and

$$X_3 = \gamma_{\nu,\omega}\left(\left(\frac{\nu}{\omega} - \epsilon\right) X_1 - \left(\frac{\nu}{\omega} + \epsilon\right) X_2 + \epsilon\right), \qquad \nu, \omega \in \mathbb{R}^+, \quad \epsilon \in \left[0, \frac{\nu}{\omega}\right),$$

where $\gamma_{\nu,\omega}$ is the activation function defined in Eq. (3). Note that X_3 has median $m = \mathrm{E}(X_3) = 1/2$ and $X_3 - \frac{1}{2} =_{st} \frac{1}{2} - X_3$.

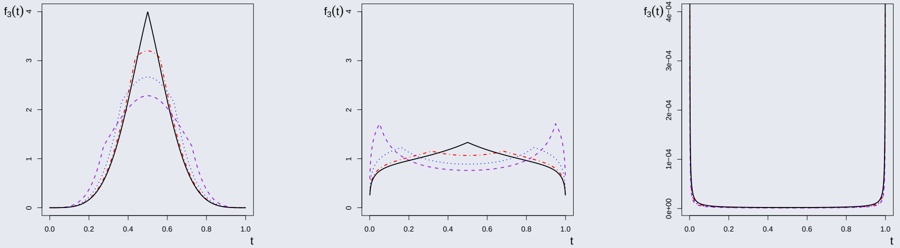


Figure: Plots of the PDF of X_3 for $\omega = 1$ and (left) $\nu = 0.5$, (center) $\nu = 1.5$ and (right) $\nu = 10^6$. In each case, the solid (dotdash, dotted, dashed) line corresponds to $\epsilon = 0$ ($\nu/4$, $\nu/2$, $3\nu/4$).

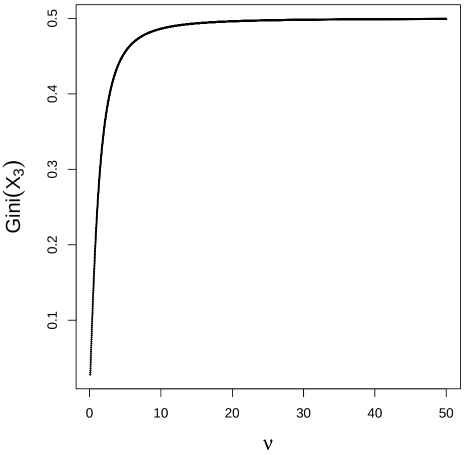


Figure: Plot of the Gini index of X_3 as a function of ν, by assuming $\epsilon = 0$.

- The level of non-linearity in the present neural network depends on the choice of $\nu \in \mathbb{R}^+$.
- The information (dispersion) quantified by the Gini index of X_3 also depends of $\nu \in \mathbb{R}^+$, so that it is connected with the complexity of the related activation function.

Conclusions

What about the future?

- **Optimal Activation Selection**: Developing a systematic method to select the best activation function for a neural network based on the underlying model.
- **Solving the Inverse Problem**: Using known ψ-functions to derive entirely new information-based activation functions for machine learning.
- **Exploring New Measures**: Investigating and interpreting the open practical applications of the newly introduced symmetric and generalized entropy tables.

Main references I

[1] A. Apicella, F. Donnarumma, F. Isgrò, and R. Prevete, "A survey on modern trainable activation functions," *Neural Networks*, vol. 138, pp. 14–32, 2021. DOI: 10.1016/j.neunet.2021.01.026.

[2] B.C. Arnold and J.M. Sarabia, *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*. Cham, Switzerland: Springer, 2018. DOI: 10.1007/978-3-319-93773-1.

[3] P.J. Bickel and E.L. Lehmann, "Descriptive Statistics for Nonparametric Models IV. Spread," in *Selected Works of E. L. Lehmann*, J. Rojo, Ed. Boston, MA, USA: Springer, 2012, pp. 519–526. DOI: 10.1007/978-1-4614-1412-4_45.

[4] M. Capaldo, A. Di Crescenzo, and A. Meoli, "Cumulative information generating function and generalized Gini functions," *Metrika*, vol. 87, pp. 775–803, 2024. DOI: 10.1007/s00184-023-00931-3.

[5] M. Capaldo, A. Di Crescenzo, and F. Pellerey, "Generalized Gini's mean difference through distortions and copulas, and related minimizing problems," *Statistics & Probability Letters*, vol. 206, pp. 109981, 2024. DOI: 10.1016/j.spl.2023.109981.

[6] M. Capaldo, M., A. Di Crescenzo and **G. Pisano** (2026+). "Information measures and activation functions," *submitted*.

[7] A. Di Crescenzo, S. Kayal, and A. Meoli, "Fractional generalized cumulative entropy and its dynamic version," *Communications in Nonlinear Science and Numerical Simulation*, vol. 102, pp. 105899, 2021. DOI: 10.1016/j.cnsns.2021.105899.

[8] A. Di Crescenzo and M. Longobardi, "On cumulative entropies," *Journal of Statistical Planning and Inference*, vol. 139, no. 12, pp. 4072–4087, 2009. DOI: 10.1016/j.jspi.2009.05.038.

Main references II

[9] S. Kayal, “On generalized cumulative entropies,” *Probability in the Engineering and Informational Sciences*, vol. 30, no. 4, pp. 640–662, 2016. DOI: 10.1017/S0269964816000218.

[10] I. Klein, B. Mangold, and M. Doll, “Cumulative paired ϕ-entropy,” *Entropy*, vol. 18, no. 7, Art. no. 248, 2016. DOI: 10.3390/e18070248.

[11] G. Psarrakos and J. Navarro, “Generalized cumulative residual entropy and record values,” *Metrika*, vol. 76, pp. 623–640, 2013. DOI: 10.1007/s00184-012-0408-6.

[12] G. Rajesh and S.M. Sunoj, “Some properties of cumulative Tsallis entropy of order α,” *Statistical Papers*, vol. 60, pp. 933–943, 2019. DOI: 10.1007/s00362-016-0855-7.

[13] M. Rao, Y. Chen, B.C. Vemuri, and F. Wang, “Cumulative residual entropy: A new measure of information,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1220–1228, 2004. DOI: 10.1109/TIT.2004.828057.

[14] T. Simon and G. Dulac, “On cumulative Tsallis entropies,” *Acta Applicandae Mathematicae*, vol. 188, no. 1, Art. no. 9, 2023. DOI: 10.1007/s10440-023-00620-3.

Thank you for your kind attention!