Rényi Cross-Entropy Measures for Common Distributions and Processes with Memory

Ferenc Cole Thierrin, Fady Alajaji * and **Tamás Linder**

Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada; 14fngt@queensu.ca (F.C.T.); tamas.linder@queensu.ca (T.L.)

* Correspondence: fa@queensu.ca

Abstract: Two Rényi-type generalizations of the Shannon cross-entropy, the Rényi cross-entropy and the Natural Rényi cross-entropy, were recently used as loss functions for the improved design of deep learning generative adversarial networks. In this work, we derive the Rényi and Natural Rényi differential cross-entropy measures in closed form for a wide class of common continuous distributions belonging to the exponential family, and we tabulate the results for ease of reference. We also summarise the Rényi-type cross-entropy rates between stationary Gaussian processes and between finite-alphabet time-invariant Markov sources.

Keywords: Rényi information measures; cross-entropy; divergence measures; exponential family distributions; Gaussian processes; Markov sources

1. Introduction

The Rényi entropy [1] of order α of a probability mass function p with finite support $\mathbb{S}$ is defined as

$$H_\alpha(p) = \frac{1}{1-\alpha} \ln \sum_{x \in \mathbb{S}} p(x)^\alpha \tag{1}$$

for $\alpha > 0, \alpha \neq 1$. The Rényi entropy generalizes the Shannon entropy,

$$H(p) = -\sum_{x \in \mathbb{S}} p(x) \ln p(x), \tag{2}$$

in the sense that as $\alpha \to 1$, $H_\alpha(p) \to H(p)$. Several other Rényi-type information measures have been put forward, each obeying the condition that their limit as α goes to one reduces to a Shannon-type information measure. This includes the Rényi divergence (of order α) between two discrete distributions p and q with common finite support $\mathbb{S}$, given by

$$D_\alpha(p||q) = \frac{1}{\alpha - 1} \ln \sum_{x \in \mathbb{S}} p(x)^\alpha q(x)^{1-\alpha} \tag{3}$$

which reduces to the familiar Kullback–Leibler divergence,

$$D_{\text{KL}}(p||q) = \sum_{x \in \mathbb{S}} p(x) \ln \frac{p(x)}{q(x)}, \tag{4}$$

as $\alpha \to 1$. Note that in some cases [2], there may exist multiple Rényi-type generalisations for the same information measure (particularly for mutual information).

Many of these definitions admit natural counterparts in the case when the involved distributions have a probability density function (pdf). This gives rise to information measures such as the Rényi differential entropy for pdf p with support $\mathbb{S}$,

$$h_\alpha(p) = \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x)^\alpha \, dx, \tag{5}$$

and the Rényi differential divergence between pdfs p and q with common support $\mathbb{S}$,

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \ln \int_{\mathbb{S}} p(x)^\alpha q(x)^{1-\alpha} \, dx. \tag{6}$$

The Rényi cross-entropy between distributions p and q is an analogous generalization of the Shannon cross-entropy

$$H(p;q) = -\sum_{x \in \mathbb{S}} p(x) \ln q(x). \tag{7}$$

Two definitions for this measure have been recently suggested. In light of the fact that Shannon's cross-entropy satisfies $H(p;q) = D(p\|q) + H(p)$, a natural definition of the Rényi cross-entropy is:

$$\tilde{H}_\alpha(p;q) := D_\alpha(p\|q) + H_\alpha(p). \tag{8}$$

This definition was indeed proposed in [3] in the continuous case, with the differential cross-entropy measure given by

$$\tilde{h}_\alpha(p;q) := D_\alpha(p\|q) + h_\alpha(p). \tag{9}$$

In contrast, prior to [3], the authors of [4] introduced the Rényi cross-entropy in their study of shifted Rényi measures expressed as the logarithm of weighted generalized power means. Specifically, upon simplifying Definition 6 in [4], their expression for the Rényi cross-entropy between distributions p and q is given by

$$H_\alpha(p;q) := \frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} p(x) q(x)^{\alpha-1}. \tag{10}$$

For the continuous case, (10) can be readily converted to yield the Rényi differential cross-entropy between pdfs p and q:

$$h_\alpha(p;q) := \frac{1}{1 - \alpha} \ln \int_{\mathbb{S}} p(x) q(x)^{\alpha-1} \, dx. \tag{11}$$

Note that both (8) and (10) reduce to the Shannon cross-entropy $H(p;q)$ as $\alpha \to 1$ [5]. A similar result holds for (9) and (11), where the Shannon differential cross-entropy,

$$h(p;q) = -\int_{\mathbb{S}} p(x) \ln q(x) \, dx, \tag{12}$$

is obtained. Further, the Rényi (differential) entropy is recovered in all equations when $p = q$ (almost everywhere). These properties alone make these definitions viable extensions of the Shannon (differential) cross-entropy.

Finding closed-form expressions for the cross-entropy measure in (9) for continuous distributions is direct, since the Rényi divergence and the Rényi differential entropy were already calculated for numerous distributions in [6,7], respectively. However, deriving the measure in (11) is more involved. We hereafter refer to the measures $\tilde{H}_\alpha(p;q)$ in (8) and $\tilde{h}_\alpha(p;q)$ in (9) as the *Natural Rényi cross-entropy* and the *Natural Rényi differential cross-entropy*, respectively; while we plainly call the measures $H_\alpha(p;q)$ in (10) and $h_\alpha(p;q)$ (11) as the *Rényi cross-entropy* and the *Rényi differential cross-entropy*, respectively.

In a recent conference paper [8], we showed how to calculate the Rényi differential cross-entropy $h_\alpha(p;q)$ between distributions of the same type from the exponential family. Building upon the results shown there, the purpose of this paper is to derive in closed form the expression of $h_\alpha(p;q)$ for thirteen commonly used univariate distributions from the exponential family, as well as for multivariate Gaussians, and tabulate the results for ease of reference. We also analytically derive the Natural Rényi differential cross-entropy $\tilde{h}_\alpha(p;q)$ for the same set of distributions. Finally, we present tables summarising

the Rényi and Natural Rényi (differential) cross-entropy rate measures, along with their Shannon counterparts, for two important classes of sources with memory, namely stationary Gaussian sources and finite-state time-invariant Markov sources.

Motivation for determining formulae for the Rényi cross-entropy originates from the use of the Shannon differential cross-entropy as a loss function for the design of deep learning generative adversarial networks (GANs) in [9]. The parameter α, ubiquitous to all Rényi information measures, allows one to fine-tune the loss function to improve the quality of the GAN-generated output. This can be seen in [3,5,10], which used the Rényi differential cross-entropy, and the Natural Rényi differential cross-entropy measures, respectively, to generalize the original GAN loss function (which is recovered as $\alpha \to 1$), resulting in both improved GAN system stability and performance for multiple image datasets. It is also shown in [5,10] that the introduced Rényi-centric generalized loss function preserves the equilibrium point satisfied by the original GAN via the so-called Jensen-Rényi divergence [11], a natural extension of the Jensen–Shannon divergence [12] upon which the equilibrium result of [9] is established. Other GAN systems that utilize different generalized loss functions were recently developed and analysed in [13–15] (see also the references therein for prior work).

The rest of this paper is organised as follows. In Section 2, the formulae for the Rényi differential cross-entropy and Natural Rényi differential cross-entropy for distributions from the exponential family are given. In Section 3, these calculations are systematically carried out for fourteen pairs of distributions of the same type within the exponential family, and the results are presented in two tables. The Rényi and Natural Rényi differential cross-entropy rates are presented in Section 4 for stationary Gaussian sources; furthermore, the Rényi and Natural Rényi cross-entropy rates are provided in Section 5 for finite-state time-invariant Markov sources. Finally, the paper is concluded in Section 6.

2. Rényi and Natural Rényi Differential Cross-Entropies for Distributions from the Exponential Family

An exponential family is a class of probability distributions over a support $\mathbb{S} \subseteq \mathbb{R}^n$ defined by a parameter space $\Theta \subseteq \mathbb{R}^m$ and functions $b : \mathbb{S} \mapsto \mathbb{R}$, $c : \Theta \mapsto \mathbb{R}$, $T : \mathbb{S} \mapsto \mathbb{R}^m$, and $\eta : \Theta \mapsto \mathbb{R}^m$ such that the pdf in this family has the form

$$f(x) = c(\theta)b(x)\exp(\langle \eta(\theta), T(x)\rangle), \qquad x \in \mathbb{S} \tag{13}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^m$. Alternatively, by using the (natural) parameter $\eta = \eta(\theta)$, the pdf can also be written

$$f(x) = b(x)\exp(\langle \eta, T(x)\rangle + A(\eta)), \tag{14}$$

where $A(\eta) : \eta(\Theta) \mapsto \mathbb{R}$ with $A(\eta) = \ln c(\theta)$. Examples of important pdfs we consider from the exponential family are included in Appendix A.

In [8], the cross-entropy between pdfs f_1 and f_2 of the same type from the exponential family was proven to be

$$h_\alpha(f_1; f_2) = \frac{A(\eta_1) - A(\eta_h) + \ln E_h}{1 - \alpha} - A(\eta_2), \tag{15}$$

where

$$E_h = \mathbb{E}_{f_h}\left[b(X)^{\alpha-1}\right] = \int b(x)^{\alpha-1} f_h(x)\, dx. \tag{16}$$

Here, f_h refers to a distribution of the same type as f_1 and f_2 within the exponential family with natural parameter

$$\eta_h := \eta_1 + (\alpha - 1)\eta_2. \tag{17}$$

It can also be shown that the Natural Rényi differential cross-entropy between f_1 and f_2 is given by

$$\tilde{h}_\alpha(f_1; f_2) = \frac{A(\eta_\alpha) - A(\alpha\eta_1) + \ln E_\alpha}{1 - \alpha} - A(\eta_2), \tag{18}$$

where

$$\eta_\alpha = \alpha\eta_1 + (1 - \alpha)\eta_2, \tag{19}$$

and

$$E_\alpha = \mathbb{E}_{f_{\alpha 1}}\left[b(X)^{\alpha - 1}\right] = \int b(x)^{\alpha - 1} f_{\alpha 1}(x)\, dx \tag{20}$$

where $f_{\alpha 1}$ refers to a distribution of the same type as f_1 and f_2 within the exponential family with natural parameter $\alpha\eta_1$.

3. Tables of Rényi and Natural Rényi Differential Cross-Entropies

Tables 1 and 2 list Rényi and Natural differential cross-entropy expressions, respectively, between common distributions of the same type from the exponential family (which we describe in Appendix A for convenience). The closed-form expressions were derived using (15) and (18), respectively. In the tables, the subscript i is used to denote that a parameter belongs to pdf f_i, $i = 1, 2$.

Table 1. Rényi differential cross-entropies.

Name	$h_\alpha(f_1; f_2)$
Beta	$\ln B(a_2, b_2) + \dfrac{1}{\alpha - 1} \ln \dfrac{B(a_h, b_h)}{B(a_1, b_1)}$
	$a_h := a_1 + (\alpha - 1)(a_2 - 1), \qquad a_h > 0$ $b_h := b_1 + (\alpha - 1)(b_2 - 1), \qquad b_h > 0$
χ **(scaled)**	$\dfrac{1}{2}\left(k_2 \ln \sigma_2^2 \sigma_h^2 - \ln 2\sigma_h^2\right) + \ln \Gamma\left(\dfrac{k_2}{2}\right)$ $+ \dfrac{1}{\alpha - 1}\left(\ln \Gamma\left(\dfrac{k_h}{2}\right) - \ln \Gamma\left(\dfrac{k_1}{2}\right) - \dfrac{k_1}{2} \ln \sigma_1^2 \sigma_h^2\right)$
	$\sigma_h^2 := \frac{1}{\sigma_1^2} + \frac{\alpha - 1}{\sigma_2^2}, \qquad \sigma_h^2 > 0$ $k_h := k_1 + (\alpha - 1)(k_2 - 1), \qquad k_h > 0$
χ **(non-scaled)**	$\dfrac{1}{2}(k_2 \ln \alpha - \ln 2\alpha) + \ln \Gamma\left(\dfrac{k_2}{2}\right)$ $+ \dfrac{1}{\alpha - 1}\left(\ln \Gamma\left(\dfrac{k_h}{2}\right) - \ln \Gamma\left(\dfrac{k_1}{2}\right) - \dfrac{k_1}{2} \ln \alpha\right)$
	$k_h := k_1 + (\alpha - 1)(k_2 - 1), \qquad k_h > 0$
χ^2	$\dfrac{1}{1 - \alpha}\left(\dfrac{\nu_1}{2} \ln(\alpha) - \ln \Gamma\left(\dfrac{\nu_1}{2}\right) + \ln \Gamma\left(\dfrac{\nu_h}{2}\right)\right)$ $+ \dfrac{2 - \nu_2}{2} \ln(\alpha) + \ln 2\Gamma\left(\dfrac{\nu_2}{2}\right)$
	$\nu_h := \nu_1 + (\alpha - 1)(\nu_2 - 2), \qquad \nu_h > 0$
Exponential	$\dfrac{1}{1 - \alpha} \ln \dfrac{\lambda_1}{\lambda_h} - \ln \lambda_2$
	$\lambda_h := \lambda_1 + (\alpha - 1)\lambda_2, \qquad \lambda_h > 0$
Gamma	$\ln \Gamma(k_2) + k_2 \ln \theta_2$ $+ \dfrac{1}{1 - \alpha}\left(\ln \dfrac{\Gamma(k_h)}{\Gamma(k_1)} - k_h \ln \theta_h - k_1 \ln \theta_1\right)$
	$\theta_h := \frac{\theta_1 + (\alpha - 1)\theta_2}{(\alpha - 1)\theta_1 \theta_1}, \qquad \theta_h > 0$ $k_h := k_1 + (\alpha - 1)k_2, \qquad k_h > 0$

Table 1. *Cont.*

Name	$h_\alpha(f_1; f_2)$
Gaussian (univariate)	$\frac{1}{2}\left(\ln(2\pi\sigma_2^2) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h}\right) + \frac{(\mu_1 - \mu_2)^2}{(\sigma^2)_h}\right)$
	$(\sigma^2)_h := \sigma_2^2 + (\alpha - 1)\sigma_1^2, \qquad (\sigma^2)_h > 0$
Gaussian (multivariate)	$\frac{1}{2-2\alpha}\left(-\ln|A||\Sigma_1| + (1-\alpha)\ln(2\pi)^n|\Sigma_2| - d\right)$
	$A := \Sigma_1^{-1} + (\alpha - 1)\Sigma_2^{-1}, \qquad A \succ 0$ $d := \boldsymbol{\mu_1}^T\Sigma_1^{-1}\boldsymbol{\mu_1} + (\alpha - 1)\boldsymbol{\mu_2}^T\Sigma_2^{-1}\boldsymbol{\mu_2}$ $-(\boldsymbol{\mu_1}^T\Sigma_1^{-1} + (\alpha - 1)\boldsymbol{\mu_2}^T\Sigma_2^{-1})A^{-1}(\Sigma_1^{-1}\boldsymbol{\mu_1} + (\alpha - 1)\Sigma_2^{-1}\boldsymbol{\mu_2})$
Gumbel $(\beta_1 = \beta_2 = \beta)$	$\frac{1}{1-\alpha}\left(\ln\frac{\Gamma(2-\alpha)}{\beta} - \frac{\mu_1}{\beta} - \alpha\ln\eta_h\right) + \frac{\mu_2}{\beta}$
	$\eta_h := e^{-\mu_1/\beta} + (\alpha - 1)e^{-\mu_2/\beta}, \qquad \eta_h > 0$
Half-Normal	$\frac{1}{2}\left(\ln(\frac{\pi\sigma_2^2}{2}) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h}\right)\right)$
	$(\sigma^2)_h := \sigma_2^2 + (\alpha - 1)\sigma_1^2, \qquad (\sigma^2)_h > 0$
Laplace $(\mu_1 = \mu_2 = 0)$	$\ln(2b_2) + \frac{1}{1-\alpha}\ln\left(\frac{b_2}{2b_h}\right)$
	$b_h := b_2 + (1-\alpha)b_1, \qquad b_h > 0$
Maxwell Boltzmann	$\frac{1}{2}\left(\ln 2\pi + 3\ln\sigma_2^2\right) + \ln\sigma_h^2$ $+\frac{1}{1-\alpha}\left(\ln\frac{\Gamma(2\alpha)}{\Gamma(\alpha)} - \frac{3}{2}\ln\sigma_1^2\sigma_h^2\right)$
	$\sigma_h^2 := \sigma_1^{-2} + (\alpha - 1)\sigma_2^{-2}, \qquad \sigma_h^2 > 0$
Pareto $(m_1 = m_2 = m)$	$-\ln m - \ln\lambda_2 + \frac{1}{1-\alpha}\ln\frac{\lambda_1}{\lambda_h}$
	$\lambda_h := \lambda_1 + (\alpha - 1)(\lambda_2 + 1), \qquad \lambda_h > 0$
Rayleigh	$\frac{\ln\sigma_1^2 - \alpha\ln\sigma_h^2 + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln 2\sigma_2^2$
	$\sigma_h^2 := \sigma_1^{-2} + (\alpha - 1)\sigma_2^{-2}, \qquad \sigma_h^2 > 0$

Table 2. Natural Rényi differential cross-entropies.

Name	$\tilde{h}_\alpha(f_1; f_2)$
Beta	$\ln B(a_2, b_2) + \frac{1}{\alpha - 1}\ln\frac{B(a_\alpha, b_\alpha)}{B(\alpha(a_1 - 1) + 1, \alpha(b_1 - 1) + 1)}$
	$a_\alpha := \alpha a_1 + (1-\alpha)a_2, \qquad a_\alpha > 0$ $b_\alpha := \alpha b_1 + (1-\alpha)b_2, \qquad b_\alpha > 0$
χ (scaled)	$\frac{1}{2}\left(-\ln\frac{2\sigma_1^2}{\alpha} + k_2\ln\sigma_2^2\sigma_\alpha^2\right) + \ln\Gamma\left(\frac{k_2}{2}\right)$ $+\frac{1}{1-\alpha}\left(\frac{\alpha k_1\ln\frac{\sigma_\alpha^2\sigma_1^2}{\alpha}}{2} - \ln\Gamma(\frac{k_\alpha}{2}) + \ln\Gamma\left(\frac{\alpha(k_1 - 1) + 1}{2}\right)\right)$
	$\sigma_\alpha^2 := \frac{\alpha}{\sigma_1^2} + \frac{1-\alpha}{\sigma_2^2}, \qquad \sigma_\alpha^2 > 0$ $k_\alpha := \alpha k_1 + (1-\alpha)k_2, \qquad k_\alpha > 0$
χ (non-scaled)	$\frac{-\ln 2\alpha}{2} + \ln\Gamma(\frac{k_2}{2})$ $+\frac{1}{1-\alpha}\left(-\ln\Gamma(\frac{k_\alpha}{2}) - \frac{\alpha k_1\ln\alpha}{2} + \ln\Gamma(\frac{\alpha(k_1 - 1) + 1}{2})\right)$
	$k_\alpha := \alpha k_1 + (1-\alpha)k_2, \qquad k_\alpha > 0$

Table 2. *Cont.*

Name	$\tilde{h}_\alpha(f_1; f_2)$
χ^2	$\dfrac{1}{1-\alpha}\left(-\ln\Gamma\left(\dfrac{\nu_\alpha}{2}\right)+\alpha\ln\Gamma\left(\dfrac{\nu_1}{2}\right)\right)+\ln\Gamma\left(\dfrac{\nu_2}{2}\right)$
	$\nu_\alpha := \alpha\nu_1 + (1-\alpha)k, \qquad \nu_\alpha > 0$
Exponential	$\dfrac{1}{1-\alpha}\ln\dfrac{\lambda_1}{\alpha\lambda_\alpha}-\ln\lambda_2$
	$\lambda_\alpha := \alpha\lambda_1 + (1-\alpha)\lambda_2, \qquad \lambda_\alpha > 0$
Gamma	$\ln\Gamma(k_2)+k_2\ln\theta_2$ $+\dfrac{1}{1-\alpha}\left(\ln\dfrac{\Gamma(k_1)}{\Gamma(k_\alpha)}-k_\alpha\ln\theta_\alpha-\alpha^2k_1\ln\theta_1\right)$
	$\theta_\alpha := \alpha\theta_1^{-1}+(1-\alpha)\theta_2^{-1}, \; k_\alpha := \alpha k_1+(1-\alpha)k_2, \qquad \theta_\alpha > 0$
Gaussian (univariate)	$\dfrac{1}{2}\left(\ln(2\pi\sigma_2^2)+\dfrac{(\mu_1-\mu_2)^2}{(\sigma^2)_\alpha}+\dfrac{1}{1-\alpha}\ln\left(\dfrac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right)$
	$(\sigma^2)_\alpha := \alpha\sigma_2^2+(1-\alpha)\sigma_1^2, \qquad (\sigma^2)_\alpha > 0$
Gaussian (multivariate)	$\dfrac{1}{2-2\alpha}(-\ln|\alpha|+\ln|A||\Sigma_1|+d)+\dfrac{1}{2}\ln\dfrac{(2\pi)^n|\Sigma_1|^2}{|\Sigma_2|}$
	$A := \alpha\Sigma_1^{-1}+(1-\alpha)\Sigma_2^{-1}, \qquad A \succ 0$ $d := (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})^T\Sigma_1 A\Sigma_2(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})$
Gumbel $(\beta_1 = \beta_2 = \beta)$	$\dfrac{\mu_2+\alpha\mu_1}{\beta}+\dfrac{1}{1-\alpha}\left(\ln\dfrac{\Gamma(2-\alpha)\eta_\alpha}{\alpha\beta}+\dfrac{\mu_1}{\beta}\right)$
	$\eta_\alpha := \alpha e^{-\mu_1/\beta}+(1-\alpha)e^{-\mu_2/\beta}, \qquad \eta_\alpha > 0$
Half-Normal	$\dfrac{1}{2}\left(\ln(\dfrac{\pi\sigma_2^2}{2})+\dfrac{1}{1-\alpha}\ln\left(\dfrac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right)$
	$(\sigma^2)_\alpha := \alpha\sigma_2^2+(1-\alpha)\sigma_1^2, \qquad (\sigma^2)_\alpha > 0$
Laplace $(\mu_1 = \mu_2 = 0)$	$\dfrac{\ln b_\alpha+\ln\alpha b_1}{1-\alpha}+\ln 2b_2$
	$b_\alpha := \dfrac{\alpha}{b_1}+\dfrac{1-\alpha}{b_2}, \qquad b_\alpha > 0$
Maxwell Boltzmann	$\dfrac{-\ln 2+3\ln\sigma_2^2}{2}+\ln\dfrac{\alpha}{\sigma_1^2}$ $+\dfrac{1}{1-\alpha}\left(\dfrac{3}{2}\ln\dfrac{\sigma_\alpha\sigma_1^2}{\alpha}-\alpha\ln\dfrac{\sqrt{\pi}}{2}+\ln\Gamma\left(\alpha+\dfrac{1}{2}\right)\right)$
	$\sigma_\alpha^2 := \dfrac{\alpha}{\sigma_1^2}+\dfrac{1-\alpha}{\sigma_2^2}, \qquad \sigma_\alpha^2 > 0$
Pareto $(m_1 = m_2 = m)$	$\dfrac{1}{1-\alpha}\left(\ln\lambda_\alpha-\ln(1-\alpha(\lambda_1-1))\right)-\ln\lambda_2 m$
	$\lambda_\alpha := \alpha\lambda_1+(1-\alpha)\lambda_2, \qquad \lambda_\alpha > 0$
Rayleigh	$\dfrac{\ln\sigma_1^2(\sigma^2)_\alpha+\ln\alpha+\ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha}+\ln 2\sigma_1^2$
	$(\sigma^2)_\alpha := \alpha\sigma_1^{-2}+\dfrac{1}{2}\ln\dfrac{2\sigma_1^4\sigma_2^4}{\alpha}, \qquad (\sigma^2)_\alpha > 0$

4. Rényi and Natural Rényi Differential Cross-Entropy Rates for Stationary Gaussian Processes

In [8], the Rényi differential cross-entropy rate between two stationary zero-mean Gaussian processes is derived. We present its formula alongside the expressions of the Shannon and Natural Rényi differential cross-entropy rates in Table 3. Here, $f(\lambda)$ is the

spectral density of the first zero-mean Gaussian process, $g(\lambda)$ is the spectral density of the second Gaussian process,

$$h(\lambda) = f(\lambda) + (\alpha - 1)g(\lambda), \tag{21}$$

and

$$j(\lambda) = \alpha f(\lambda) + (1 - \alpha)g(\lambda). \tag{22}$$

Table 3. Differential cross-entropy rates for stationary zero-mean Gaussian sources.

Information Measure	Rate	Constraint
Shannon Differential Cross-Entropy	$\dfrac{1}{2}\ln 2\pi + \dfrac{1}{4\pi}\displaystyle\int_0^{2\pi}\left[\ln g(\lambda) + \dfrac{f(\lambda)}{g(\lambda)}\right]d\lambda$	$g(\lambda) > 0$
Natural Rényi Differential Cross-Entropy	$\dfrac{1}{2}\ln 4\pi^2\alpha^{\frac{1}{\alpha-1}} + \dfrac{1}{4\pi(1-\alpha)}\displaystyle\int_0^{2\pi}\ln\dfrac{j(\lambda)}{g(\lambda)^\alpha}\,d\lambda$	$\dfrac{j(\lambda)}{g(\lambda)} > 0$
Rényi Differential Cross-Entropy	$\dfrac{\ln 2\pi}{2} + \dfrac{1}{4\pi(1-\alpha)}\displaystyle\int_0^{2\pi}[(2-\alpha)\ln g(\lambda) - \ln h(\lambda)]d\lambda$	$\dfrac{g(\lambda)}{h(\lambda)} > 0$

5. Rényi and Natural Rényi Cross-Entropy Rates for Markov Sources

In [8], the Rényi cross-entropy rate between finite-state time-invariant Markov sources was established, using, as in [16], tools from the theory of non-negative matrices and Perron–Frobenius theory (e.g., cf. [17,18]). This measure, as well as the Shannon and Natural Rényi differential cross-entropy rates, are derived and summarised in Table 4. Here, P and Q are the $m \times m$ (stochastic) transition matrices associated with the first and second Markov sources, respectively, where both sources have a common alphabet of size m. To allow any value of the Rényi parameter α in $(0,1) \cup (1,\infty)$, we assume that the transition matrix Q of the second Markov chain has positive entries ($Q > 0$); however, the transition matrix P of the first Markov chain is taken to be an *arbitrary* stochastic matrix. For simplicity, we assume that the initial distribution vectors, p and q, of both Markov chains also have positive entries ($p > 0$ and $q > 0$). This condition can be relaxed via the approach used to prove Theorem 1 in [16]. Moreover, π_p^T denotes the stationary probability row vector associated with the first Markov chain, and $\mathbf{1}$ is an m-dimensional column vector in which each element equals one. Furthermore, $\odot$ denotes element-wise multiplication (i.e., the Hadamard product operation), and $\dot{\ln}$ is the element-wise natural logarithm.

Finally, the definition of $\lambda(R) : \mathbb{R}^{m \times m} \mapsto \mathbb{R}$ for a matrix R is more involved. If R is irreducible, $\lambda(R)$ is its largest positive eigenvalue. Otherwise, rewriting R in its canonical form as detailed in Proposition 1 in [16], we have that $\lambda(R) = \max(\lambda^*, \lambda_*)$, where λ^* is the maximum of all largest positive eigenvalues of (irreducible) sub-matrices of R corresponding to self-communicating classes, and λ_* is the maximum of all largest positive eigenvalues of sub-matrices of R corresponding to classes reachable from an inessential class.

Table 4. Cross-entropy rates for time-invariant Markov sources.

Information Measure	Rate
Shannon Cross-Entropy	$-\pi_p^T(P \odot \dot{\ln}Q)\mathbf{1}$
Natural Rényi Cross-Entropy	$\dfrac{1}{\alpha - 1}\ln\dfrac{\lambda(P^\alpha \odot Q^{1-\alpha})}{\lambda(P^\alpha)}$
Rényi Cross-Entropy	$\dfrac{1}{1 - \alpha}\ln\lambda(P \odot Q^{\alpha-1})$

6. Conclusions

We have derived closed-form formulae for the Rényi and Natural Rényi differential cross-entropies of commonly used distributions from the exponential family. This is of potential use to further studies in information theory and machine learning, particularly problems where deep neural networks, trained according to a Shannon cross-entropy loss function, can be improved via generalized Rényi-type loss functions in virtue of the extra degree of freedom provided by the Rényi (α) parameter. In addition, we have provided formulae for the Rényi and Natural Rényi differential cross-entropy rates for stationary zero-mean Gaussian processes and expressions for the cross-entropy rates for Markov sources. Further work includes expanding the present collection by considering distributions such as Levy or Weibull and investigating cross-entropy measures based on the f-divergence [19–21], starting with Arimoto's divergence [22].

Author Contributions: Conceptualization, investigation, and manuscript preparation, all authors; derivation, F.C.T.; validation and supervision, F.A. and T.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In Table A1, we describe the distributions of Tables 1 and 2 (for the multivariate Gaussian, $\boldsymbol{\mu} \in \mathbb{R}^n$ is a mean vector, and $\Sigma^2 \in \mathbb{R}^n \times \mathbb{R}^n$ is a positive definite covariance matrix).

Table A1. Distributions listed in Tables 1 and 2.

Name	Parameters (Θ)	PDF $f(x)$	Support
Beta	$(a > 0, b > 0)$	$B(a,b)x^{a-1}(1-x)^{b-1}$	$\mathbb{S} = (0,1)$
χ **(scaled)**	$(k > 0, \sigma > 0)$	$\dfrac{2^{1-k/2}x^{k-1}e^{-x^2/2\sigma^2}}{\sigma^k \Gamma\left(\frac{k}{2}\right)}$	$\mathbb{S} = \mathbb{R}^+$
χ **(non-scaled)**	$(k > 0)$	$\dfrac{2^{1-k/2}x^{k-1}e^{-x^2/2}}{\Gamma\left(\frac{k}{2}\right)}$	$\mathbb{S} = \mathbb{R}^+$
χ^2	$(\nu > 0)$	$\dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}$	$\mathbb{S} = \mathbb{R}^+$
Exponential	$(\lambda > 0)$	$\lambda e^{-\lambda x}$	$\mathbb{S} = \mathbb{R}^+$
Gamma	$(k > 0, \theta > 0)$	$\dfrac{1}{\theta^k \Gamma(k)}x^{k-1}e^{-\frac{k}{\theta}}$	$\mathbb{S} = \mathbb{R}^+$
Gaussian (univariate)	$(\mu, \sigma^2 > 0)$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\mathbb{S} = \mathbb{R}$
Gaussian (multivariate)	$(\boldsymbol{\mu}, \Sigma^2)$	$\dfrac{1}{\sqrt{(2\pi)^n|\Sigma|}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$	$\mathbb{S} = \mathbb{R}^n$
Half-Normal	$(\sigma^2 > 0)$	$\sqrt{\dfrac{2}{\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$	$\mathbb{S} = \mathbb{R}^+$
Gumbel	$(\mu, \beta > 0)$	$\dfrac{1}{\beta}\exp-\left(\dfrac{x-\mu}{\beta} + e^{-\frac{x-\mu}{\beta}}\right)$	$\mathbb{S} = \mathbb{R}$

Table A1. *Cont.*

Name	Parameters (Θ)	PDF $f(x)$	Support
Pareto	$(m > 0, a > 0)$	$am^a x^{-(1+m)}$	$\mathbb{S} = (m, \infty)$
Maxwell Boltzmann	$(\sigma > 0)$	$\dfrac{2x^2}{\sqrt{\pi\sigma^6}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$	$\mathbb{S} = \mathbb{R}^+$
Rayleigh	$(\sigma^2 > 0)$	$\dfrac{x}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$	$\mathbb{S} = \mathbb{R}^+$
Laplace	$(\mu, b^2 > 0)$	$\dfrac{1}{2b} e^{-\frac{|x-\mu|}{b}}$	$\mathbb{S} = \mathbb{R}$

Notes

- $\mathrm{B}(a, b) = \displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}\, dt$ is the Beta function.
- $\Gamma(z) = \displaystyle\int_0^\infty x^{z-1} e^{-x}\, dx$ is the Gamma function.

References

1. Rényi, A. On measures of entropy and information. *Fourth Berkeley Symp. Math. Stat. Probab.* **1961**, *1*, 547–561.
2. Verdú, S. α-Mutual Information. In Proceedings of the 2015 Information Theory and Applications Workshop (ITA), 2015, San Diego, CA, USA, 1–6 February 2015; pp. 1–6.
3. Sarraf, A.; Nie, Y. RGAN: Rényi generative adversarial network. *SN Comput. Sci.* **2021**, *2*, 17. [CrossRef]
4. Valverde-Albacete, F.J.; Peláez-Moreno, C. The Case for shifting the Rényi entropy. *Entropy* **2019**, *21*, 46. [CrossRef] [PubMed]
5. Bhatia, H.; Paul, W.; Alajaji, F.; Gharesifard, B.; Burlina, P. Least kth-Order and Rényi generative adversarial networks. *Neural Comput.* **2021**, *33*, 2473–2510. [CrossRef] [PubMed]
6. Gil, M.; Alajaji, F.; Linder, T. Rényi divergence measures for commonly used univariate continuous distributions. *Inf. Sci.* **2013**, *249*, 124–131. [CrossRef]
7. Song, K.S. Rényi information, loglikelihood and an intrinsic distribution measure. *J. Statist. Plann. Inference* **2001**, *93*, 51–69. [CrossRef]
8. Thierrin, F.C.; Alajaji, F.; Linder, T. On the Rényi cross-entropy. In Proceedings of the 17th Canadian Workshop on Information Theory, Ottawa, ON, Canada, 5–8 June 2022; pp. 1–5. [CrossRef]
9. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
10. Bhatia, H.; Paul, W.; Alajaji, F.; Gharesifard, B.; Burlina, P. Rényi generative adversarial Networks. *arXiv* **2020**, arXiv:2006.02479v1.
11. Kluza, P.A. On Jensen-Rényi and Jeffreys-Rényi type f-divergences induced by convex functions. *Phys. Stat. Mech. Its Appl.* **2019**, *548*, 122527. [CrossRef]
12. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *31*, 145–151. [CrossRef]
13. Pantazis, Y.; Paul, D.; Fasoulakis, M.; Stylianou, Y.; Katsoulakis, M. Cumulant GAN. *arXiv* **2020**, arXiv:2006.06625.
14. Kurri, G.R.; Sypherd, T.; Sankar, L. Realizing GANs via a tunable loss function. In Proceedings of the IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021; pp. 1–6.
15. Kurri, G.R.; Welfert, M.; Sypherd, T.; Sankar, L. α-GAN: Convergence and estimation guarantees. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 312–317.
16. Rached, Z.; Alajaji, F.; Campbell, L.L. Rényi's divergence and entropy rates for finite alphabet Markov sources. *IEEE Trans. Inf. Theory* **2001**, *47*, 1553–1561. [CrossRef]
17. Seneta, E. *Non-Negative Matrices and Markov Chains*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
18. Gallager, R.G. *Discrete Stochastic Processes*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 31–55.
19. Csiszar, I. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizitat on Markhoffschen Ketten. *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **2006**, *8*, 85–108.
20. Csiszár, I. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.* **1967**, *2*, 299–318.
21. Ali, S.M.; Silvey, S.D. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Society. Ser. B (Methodol.)* **1966**, *28*, 131–142. [CrossRef]
22. Liese, F.; Vajda, I. On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412. [CrossRef]