On the Rényi Cross-Entropy*

Ferenc Cole Thierrin, Fady Alajaji, and Tamás Linder
Department of Mathematics and Statistics
Queen's University
Kingston, ON K7L 3N6, Canada
Emails: {14fngt, fa, tamas.linder}@queensu.ca

***Abstract*—The Rényi cross-entropy measure between two distributions, a generalization of the Shannon cross-entropy, was recently used as a loss function for the improved design of deep learning generative adversarial networks. In this work, we examine the properties of this measure and derive closed-form expressions for it when one of the distributions is fixed and when both distributions belong to the exponential family. We also analytically determine a formula for the cross-entropy rate for stationary Gaussian processes and for finite-alphabet Markov sources.**

***Index Terms*—Rényi information measures, cross-entropy, exponential family distributions, Gaussian processes, Markov sources.**

I. Introduction

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

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

for $\alpha > 0, \alpha \neq 1$, is a generalization of the Shannon entropy,[1] $H(p)$, in that $\lim_{\alpha\to 1} H_\alpha(p) = H(p)$. Similarly, the Rényi divergence (of order α) between two discrete distributions p and q with common finite support $\mathbb{S}$, given by

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

reduces to the KL divergence, $D(p\|q)$, as $\alpha \to 1$.

Since the introduction of these measures, 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 (e.g., see [2] and the references therein for three different order α extensions of Shannon's mutual information due to Sibson, Arimoto and Csiszár.)

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

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

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

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

The Rényi cross-entropy between distributions p and q is an analogous generalization of the Shannon cross-entropy $H(p;q)$. Two definitions for this measure have been suggested. In [3], mirroring the fact that Shannon's cross-entropy satisfies $H(p;q) = D(p\|q) + H(p)$, the authors define Rényi cross-entropy as

$$\tilde{H}_\alpha(p;q) := D_\alpha(p||q) + H_\alpha(p). \tag{1}$$

In contrast, prior to [3], the authors of [4] introduced the Rényi cross-entropy in their study of the so-called 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

$$H_\alpha\left(p;q\right) := \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p\left(x\right) q\left(x\right)^{\alpha-1}. \tag{2}$$

For the continuous case, the definition in (2) can be readily converted to yield the Rényi differential cross-entropy between pdfs p and q:

$$h_\alpha(p;q) := \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x) q(x)^{\alpha-1} \, dx. \tag{3}$$

As the Rényi differential divergence and entropy were already calculated for numerous distributions in [5] and [6], respectively, determining the Rényi differential cross-entropy using the definition in (1) is straightforward. As such, this paper's focus is to establish closed-form expressions of the Rényi differential cross-entropy as defined in (3) for various distributions, as well as to derive the Rényi cross-entropy rate for two important classes of sources with memory, Gaussian and Markov sources.

Motivation for determining formulae for the Rényi cross-entropy extends beyond idle curiosity. The Shannon differential cross-entropy was used as a loss function for the design of deep learning generative adversarial networks (GANs) [7]. Recently, the Rényi differential cross-entropy measures in (3) and (1), were used in [8], [9] and [3], respectively, to generalize the original GAN loss function. It is shown that in [8] and [9] that the resulting Rényi-centric generalized

*This work was supported in part by NSERC of Canada.

[1] For ease of reference, a table summarising the Shannon entropy and cross-entropy measures as well as the Kullback-Liebler (KL) divergence is provided in Appendix A.

loss function preserves the equilibrium point satisfied by the original GAN based on the Jensen-Rényi divergence [10], a natural extension of the Jensen-Shannon divergence [11]. In [3], a different Rényi-type generalized loss function is obtained and is shown to benefit from stability properties. Improved stability and system performance are shown in [8], [9] and [3] by virtue of the α parameter that can be judiciously used to fine-tune the adopted generalized loss functions which recover the original GAN loss function as $\alpha \to 1$.

The rest of this paper is organised as follows. In Section II, basic properties of the Rényi cross-entropy are examined. In Section III, the Rényi differential cross-entropy for members of the exponential family is calculated. In Section IV, the Rényi differential cross-entropy between two different distributions is obtained. In Section V, the Rényi differential cross-entropy rate is derived for stationary Gaussian sources. Finally in Section VI, the Rényi cross-entropy rate is established for finite-alphabet time-invariant Markov sources.

II. Basic Properties of the Rényi cross-entropy and differential cross-entropy

For the Rényi cross-entropy $H_\alpha(p;q)$ to deserve its name it would be preferable that it satisfies at least two key properties: it reduces to the Rényi entropy when $p = q$ and its limit as α goes to one is the Shannon cross-entropy. Similarly, it is desirable that the Rényi differential cross-entropy $h_\alpha(p;q)$ reduces to the Rényi differential entropy when $p = q$ and its limit as α tends to one yields the Shannon differential cross-entropy. In both cases, the former property is trivial, and the latter property was proven in [9] for the continuous case under some finiteness conditions (in the discrete case, the result holds directly via L'Hôpital's rule).

It is also proven in [9] that the Rényi differential cross-entropy $h_\alpha(p;q)$ is non-increasing in α by showing that its derivative with respect to α is non-positive. The same monotonicity property holds in the disrcrete case.

Like its Shannon counterpart, the Rényi cross-entropy is non-negative ($H_\alpha(p;q) \geq 0$); while the Rényi differential cross-entropy can be negative. This is easily verified when, for example, $\alpha = 2$ and p and q are both Gaussian (normal) distributions with zero mean and variance $1/(8\sqrt{\pi})$, and parallels the same lack of non-negativity of the Shannon differential cross-entropy.

We close this section by deriving the cross-entropy limit, $\lim_{\alpha\to\infty} H_\alpha(p;q)$. To begin with, for any non-zero constant $\tilde{c}$, we have

$$\begin{aligned}
&\lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} \tilde{c} q\left(x\right)^{\alpha-1} \\
&= \lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \tilde{c} + \lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} q\left(x\right)^{\alpha-1} \\
&= \lim_{\beta\to\infty} \frac{1-\beta}{-\beta} \frac{1}{1-\beta} \ln \sum_{\mathbb{S}} q\left(x\right)^{\beta} \qquad (\beta = \alpha - 1) \\
&= \lim_{\beta\to\infty} H_\beta(q) = -\ln q_M, \qquad (4)
\end{aligned}$$

where $q_M := \max_{x\in\mathbb{S}} q(x)$ and where we have used the fact that for the Rényi entropy, $\lim_{\alpha\to\infty} H_\alpha(q) = -\ln q_M$. Now, denoting the minimum and maximum values of $p(x)$ over $\mathbb{S}$ by p_m and p_M, respectively, we have that for $\alpha > 1$,

$$\frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p_m q\left(x\right)^{\alpha-1} \leq \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p(x) q\left(x\right)^{\alpha-1}$$

and

$$\frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p(x) q\left(x\right)^{\alpha-1} \leq \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p_M q\left(x\right)^{\alpha-1},$$

and hence by (4) we obtain

$$\lim_{\alpha\to\infty} H_\alpha\left(p;q\right) = -\ln q_M. \qquad (5)$$

III. Rényi Differential Cross-Entropy for Exponential Family Distributions

A probability distribution on $\mathbb{R}$ or $\mathbb{R}^n$ with parameter θ is said to belong to the exponential family (e.g., see [12]) if on its support $\mathbb{S}$ it admits a pdf of the form

$$f(x) = c(\theta) b(x) \exp\left(\eta(\theta) \cdot T(x)\right), \qquad x \in \mathbb{S}, \qquad (6)$$

for some real-valued (measurable) functions c, b, η and T.[2] Here η is known as the natural parameter of the distribution, $T(x)$ is the sufficient statistic and $c(\theta)$ is the normalization constant in the sense that for all θ within the parameter space

$$\int_{\mathbb{S}} b(x) \exp\left(\eta(\theta) \cdot T(x)\right) dx = c(\theta)^{-1}.$$

The pdf in (6) can also be written as

$$f(x) = b(x) \exp\left(\eta \cdot T(x) + A(\eta)\right), \qquad (7)$$

where $A\left(\eta(\theta)\right) = \ln c(\theta)$. Examples of distributions in the exponential family include the Gaussian, Beta, and exponential distributions.

Lemma 1. *Let $f_1(x)$ and $f_2(x)$ be pdfs of the same type in the exponential family with natural parameters η_1 and η_2, respectively. Define $f_h(x)$ as being of the same type as f_1 and f_2 but with natural parameter $\eta_h = \eta_1 + (\alpha-1)\eta_2$. Then*

$$h_\alpha\left(f_1; f_2\right) = \frac{A\left(\eta_1\right) - A\left(\eta_h\right) + \ln E_h}{1-\alpha} - A\left(\eta_2\right), \qquad (8)$$

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

Proof. Using (7), we have

$$\begin{aligned}
&f_1\left(x\right) f_2\left(x\right)^{\alpha-1} \\
&\quad = b\left(x\right) \exp\left(\eta_1 \cdot T\left(x\right) + A\left(\eta_1\right)\right) \\
&\qquad \cdot \left(b\left(x\right) \exp\left(\eta_2 \cdot T\left(x\right) + A\left(\eta_2\right)\right)\right)^{\alpha-1} \\
&\quad = b\left(x\right)^\alpha \exp\left(\left(\eta_1 + (\alpha-1)\eta_2\right) \cdot T\left(x\right)\right) \\
&\qquad \cdot \exp\left(A\left(\eta_1\right) + (\alpha-1) A\left(\eta_2\right)\right)
\end{aligned}$$

[2] Note that θ and consequently $T(x)$ can be vectors in cases where the distribution admits multiple parameters.

$$
\begin{aligned}
&= b\left(x\right)^{\alpha} \exp\left(\eta_h \cdot T\left(x\right) + A\left(\eta_h\right)\right) \\
&\quad \cdot \exp\left(A\left(\eta_1\right) + \left(\alpha - 1\right) A\left(\eta_2\right) - A\left(\eta_h\right)\right) \\
&= b\left(x\right)^{\alpha-1} f_h\left(x\right) \exp\left(A\left(\eta_1\right) + \left(\alpha - 1\right) A\left(\eta_2\right) - A\left(\eta_h\right)\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\int_{\mathbb{S}} f_1\left(x\right) f_2\left(x\right)^{\alpha-1} dx \\
&\quad = \int_{\mathbb{S}} b\left(x\right)^{\alpha-1} f_h\left(x\right) dx \\
&\qquad \cdot \exp\left(A\left(\eta_1\right) + \left(\alpha - 1\right) A\left(\eta_2\right) - A\left(\eta_h\right)\right) \\
&\quad = \exp\left(A\left(\eta_1\right) + \left(\alpha - 1\right) A\left(\eta_2\right) - A\left(\eta_h\right)\right) E_h,
\end{aligned}
$$

and therefore,

$$
h_\alpha\left(f_1; f_2\right) = \frac{A\left(\eta_1\right) - A\left(\eta_h\right) + \ln E_h}{1-\alpha} - A\left(\eta_2\right).
$$

□

Remark. *If $b(x) = b$ is a constant for all $x \in \mathbb{S}$, then*

$$
\frac{\ln E_h}{1-\alpha} = -\ln b.
$$

In many cases, we have that $b(x) = 1$ on $\mathbb{S}$, and thus the $\frac{\ln E_h}{1-\alpha}$ term disappears in (8)*.*

Table I lists Rényi differential cross-entropy expressions we derived using Lemma 1 for some common distributions in the exponential family (which we describe in Appendix B for convenience). In the table, the subscript of i is used to denote that a parameter belongs to pdf f_i, $i = 1, 2$.

TABLE I
RÉNYI DIFFERENTIAL CROSS-ENTROPIES FOR COMMON CONTINUOUS DISTRIBUTIONS

Name	$h_\alpha(f_1; f_2)$
Beta	$\ln B(a_2, b_2) + \frac{1}{\alpha-1} \ln \frac{B(a_h, b_h)}{B(a_1, b_1)}$
	$a_h := a_1 + (\alpha-1)(a_2-1)$, $b_h := b_1 + (\alpha-1)(b_2-1)$
$\boldsymbol{\chi^2}$	$\frac{1}{1-\alpha}\left(\frac{\nu_1}{2}\ln(\alpha) - \ln\Gamma\left(\frac{\nu_1}{2}\right) + \ln\Gamma\left(\frac{\nu_h}{2}\right)\right) + \frac{2-\nu_2}{2}\ln(\alpha) + \ln 2\Gamma\left(\frac{\nu_2}{2}\right)$
	$\nu_h := \nu_1 + (\alpha-1)(k-2)$
Exponential	$\frac{1}{1-\alpha}\ln\frac{\lambda_i}{\lambda_h} - \ln\lambda_2$
	$\lambda_h := \lambda_1 + (\alpha-1)\lambda_2$
Gamma	$\ln\Gamma(k_2) + k_2\ln\theta_2 + \frac{1}{1-\alpha}\left(\ln\frac{\Gamma(k_h)}{\Gamma(k_1)} - k_h\ln\theta_h - k_1\ln\theta_1\right)$
	$\theta_\alpha^* := \frac{\theta_1+(a-1)\theta_2}{(\alpha-1)\theta_1\theta_1}$, $k_h := k_i + (\alpha-1)k_2$
Gaussian	$\frac{1}{2}\left(\ln(2\pi\sigma_2^2) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h^*}\right) + \frac{(\mu_1-\mu_2^2}{(\sigma^2)_h^*}\right)$
	$(\sigma^2)_h^* := \sigma_2^2 + (\alpha-1)\sigma_1^2$
Laplace	$\ln(2b_2) + \frac{1}{1-\alpha}\ln\left(\frac{b_2}{2b_h}\right)$
$(\boldsymbol{\mu_1 = \mu_2})$	$b_h := b_2 + (1-\alpha)b_1$

IV. Rényi Differential Cross-Entropy between different distributions

Let p and q be pdfs with common support $\mathbb{S} \subseteq \mathbb{R}$. Below are some general formulae for the differential Rényi cross-entropy between one specific (common) distribution and any general distribution. If $\mathbb{S}$ is an interval below, then $|\mathbb{S}|$ denotes its length.

A. Distribution q is uniform

Let q be uniformly distributed on $\mathbb{S}$. Then

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

B. Distribution p is uniform

Now suppose p is uniformly distributed on $\mathbb{S}$. Then

$$
\begin{aligned}
h_\alpha(p;q) &= \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x) q(x)^{\alpha-1} dx \\
&= \frac{1}{1-\alpha} \ln \frac{1}{|\mathbb{S}|} - h_{\alpha-1}(q).
\end{aligned}
$$

C. Distribution q is exponentially distributed

Suppose the $\mathbb{S} = \mathbb{R}^+$ and q is exponential with parameter λ. Suppose also that the moment generating function (MGF) of p, $M_p(t)$ exists. We have

$$
\begin{aligned}
h_\alpha(p;q) &= \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x) q(x)^{\alpha-1} dx \\
&= \frac{1}{1-\alpha} \ln \mathbb{E}_p \left[q(x)^{\alpha-1}\right] \\
&= \frac{1}{1-\alpha} \ln \mathbb{E}_p \left[\left(\lambda \exp\left(-\lambda x\right)\right)^{\alpha-1}\right] \\
&= -\ln\lambda + \frac{1}{1-\alpha} \ln M_p\left(\lambda(1-\alpha)\right).
\end{aligned}
$$

D. Distribution q is Gaussian

Now assume that q is a (normal) Gaussian $\mathcal{N}(\mu, \sigma^2)$ distribution and that the MGF of $Y := (X-\mu)^2$, M_Y, exists, where X is a random variable with distribution p. Then

$$
\begin{aligned}
h_\alpha(p;q) &= \frac{1}{1-\alpha} \ln \mathbb{E}_p \left[q(X)^{\alpha-1}\right] \\
&= \frac{1}{1-\alpha} \ln \sigma(\sqrt{2\pi})^{1-\alpha} \mathbb{E}\left(\exp\left((1-\alpha)\frac{Y}{2\sigma^2}\right)\right) \\
&= \ln\sigma\sqrt{2\pi} + \frac{1}{1-\alpha} \ln M_Y\left(\frac{1-\alpha}{2\sigma^2}\right).
\end{aligned}
$$

The case where q is a half-normal distribution can be directly derived from the above. Given q is a half-normal distribution, on its support its pdf is the same as that of a normal $\mathcal{N}(0, \sigma^2)$ distribution times 2. Hence if p's support is $\mathbb{R}^+$, then $h_\alpha(p;q) = \ln\sigma\sqrt{\frac{\pi}{2}} + \frac{1}{1-\alpha}\ln M_Y\left(\frac{1-\alpha}{2\sigma^2}\right)$.

V. Rényi Differential Cross-Entropy Rate for Stationary Gaussian Processes

Lemma 2. *The Rényi differential cross-entropy between two zero-mean multivariate dimension-n Gaussian distributions with invertible covariance matrices Σ_1 and Σ_2, respectively, is given by*

$$h_\alpha(p;q) = \frac{\ln|\Sigma_1||S|}{2\alpha-2} + \frac{1}{2}\ln|\Sigma_2| + \frac{n}{2}\ln 2\pi, \qquad (9)$$

where $S := \Sigma_1^{-1} + (\alpha-1)\Sigma_2^{-1}$.

Proof. Recall that the pdf of a multivariate Gaussian with mean $\mathbf{0} = (0,0...,0)^T$ and invertible covariance matrix Σ is given by:

$$f(\mathbf{x}) = \frac{\exp(\frac{-1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x})}{(2\pi)^{k/2}|\Sigma|^{1/2}}$$

for $\mathbf{x} \in \mathbb{R}^n$. Note that this distribution is a member of the exponential family, where $T(\mathbf{x}) = \mathbf{x}$, $\eta = \frac{1}{2}\Sigma^{-1}$, $A(\eta) = \frac{1}{2}\ln|-2\eta|$ and $b(\mathbf{x}) = (2\pi)^{\frac{-n}{2}}$. Hence the Rényi differential cross-entropy between two zero-mean multivariate Gaussian distributions with covariance matrices Σ_1 and Σ_2, respectively, is

$$\begin{aligned} h_\alpha(p;q) &= \frac{1}{1-\alpha}\left(\frac{1}{2}\ln\left|2\frac{\Sigma_1^{-1}}{2}\right|\right. \\ &\qquad \left. -\frac{1}{2}\ln\left|2\frac{\Sigma_1^{-1}+(\alpha-1)\Sigma_2^{-1}}{2}\right|\right) \\ &\qquad -\frac{1}{2}\ln\left|2\frac{\Sigma_2^{-1}}{2}\right| - \ln(2\pi)^{\frac{-n}{2}} \\ &= \frac{\ln|\Sigma_1||S|}{2\alpha-2} + \frac{1}{2}\ln|\Sigma_2| + \frac{n}{2}\ln 2\pi. \end{aligned}$$

□

Let $\{X_j\}_{j=1}^\infty$ and $\{Y_j\}_{j=1}^\infty$ be stationary zero-mean Gaussian processes. For a given n, $X^n := (X_1, X_2, ..., X_n)$ and $Y^n := (Y_1, Y_2, ..., Y_n)$ are multivariate Gaussian random variables with mean $\mathbf{0}$ and covariance matrices Σ_{X^n} and Σ_{Y^n}, respectively. Since $\{X_j\}$ and $\{Y_j\}$ are stationary, their covariance matrices are Toeplitz. Furthermore, $B^n := \Sigma_{Y^n} + (\alpha-1)\Sigma_{X^n}$ is Toeplitz.

Lemma 3. *Let $\tilde{f}(\lambda)$, $\tilde{g}(\lambda)$ and $\tilde{h}(\lambda)$ be the power spectral densities of $\{X_j\}$, $\{Y_j\}$ and the zero-mean Gaussian process with covariance matrix B^n, respectively.*

Then the Rényi differential cross-entropy rate between $\{X_j\}$ and $\{Y_j\}$, $\lim_{n\to\infty}\frac{1}{n}h_\alpha(X^n;Y^n)$, is given by

$$\frac{\ln 2\pi}{2} + \frac{1}{4\pi(1-\alpha)}\int_0^{2\pi}\left[(2-\alpha)\ln\tilde{g}(\lambda) - \ln\tilde{h}(\lambda)\right]d\lambda.$$

Proof. From Lemma 2, we first note that $S = \Sigma_{X^n}^{-1}B_n\Sigma_{Y^n}^{-1}$. With this in mind the Rényi differential cross-entropy can be rewritten using (9) as

$$\begin{aligned} &\frac{1}{n}\left(\frac{\ln|\Sigma_{X^n}||\Sigma_{X^n}^{-1}B^n\Sigma_{Y^n}^{-1}|}{2(\alpha-1)} + \frac{1}{2}\ln|\Sigma_{Y^n}| + \frac{n}{2}\ln 2\pi\right) \\ &= \frac{\ln 2\pi}{2} + \frac{1}{2n}\left(\frac{\ln|\Sigma_{X^n}||\Sigma_{X^n}^{-1}||B^n||\Sigma_{Y^n}^{-1}|}{(\alpha-1)} + \ln|\Sigma_{Y^n}|\right) \\ &= \frac{\ln 2\pi}{2} + \frac{1}{2n}\left(\frac{\ln|B^n| - \ln|\Sigma_{Y^n}|}{(\alpha-1)} + \ln|\Sigma_{Y_n}|\right) \\ &= \frac{\ln 2\pi}{2} + \frac{1}{2n(1-\alpha)}\left((2-\alpha)\ln|\Sigma_{Y^n}| - \ln|B^n|\right). \end{aligned}$$

It was proven in [13] that for a sequence of Toeplitz matrices T_n with spectral density $t(\lambda)$ such that $\ln t(\lambda)$ is Reimann integrable, one has

$$\lim_{n\to\infty}\ln|T^n| = \frac{1}{2\pi}\int_0^{2\pi}\ln t(\lambda)\,d\lambda.$$

We therefore obtain that the Rényi differential cross-entropy rate is given by

$$\frac{\ln 2\pi}{2} + \frac{1}{4\pi(1-\alpha)}\int_0^{2\pi}\left[(2-\alpha)\ln\tilde{g}(\lambda) - \ln\tilde{h}(\lambda)\right]d\lambda.$$

Note that $\tilde{h}(\lambda) = \tilde{g}(\lambda) + (\alpha-1)\tilde{f}(\lambda)$. □

VI. Rényi cross-entropy rate for Markov sources

Consider two time-invariant Markov sources $\{X_j\}_{j=1}^\infty$ and $\{Y_j\}_{j=1}^\infty$ with common finite alphabet $\mathbb{S}$ and with transition distribution $P(\cdot|\cdot)$ and $Q(\cdot|\cdot)$, respectively. Then for any $i^n = (i_1,\ldots,i_n) \in \mathbb{S}^n$, their n-dimensional joint distributions are given by

$$p^{(n)}(i^n) = P(i_n|i_{n-1})P(i_{n-1}|i_{n-2})...P(i_2|i_1)q(i_1)$$

and

$$q^{(n)}(i^n) = Q(i_n|i_{n-1})Q(i_{n-1}|i_{n-2})...Q(i_2|i_1)p(i_1),$$

respectively, with arbitrary initial distributions, $p(i_1)$ and $q(i_1)$, $i_1 \in \mathbb{S}$. Define the Rényi cross-entropy rate between $\{X_j\}$ and $\{Y_j\}$ as

$$\begin{aligned} &\lim_{n\to\infty}\frac{1}{n}H_\alpha(X^n;Y^n) \\ &= \lim_{n\to\infty}\frac{1}{n}\frac{1}{1-\alpha}\ln\left(\sum_{i_n\in\mathbb{S}^n}p^{(n)}(i^n)q^{(n)}(i_n)^{\alpha-1}\right). \end{aligned}$$

Note that by defining the matrix R using the formula

$$R_{ij} = P(i|j)Q(i|j)^{\alpha-1}$$

and the row vector $\mathbf{s}$ as having components $s_i = P(i)P(i)^{\alpha-1}$, the Rényi cross-entropy rate can be written as

$$\lim_{n\to\infty}\frac{1}{n}\frac{1}{1-\alpha}\ln \mathbf{s}R^{n-1}\mathbf{1}, \qquad (10)$$

where $\mathbf{1}$ is a column vector whose dimension is the cardinatliy of the alphabet $\mathbb{S}$ and with all its entries equal to 1.

A result derived by [14] for the Rényi divergence between Markov sources can thus be used to find the Rényi cross-entropy rate for Markov sources.

Lemma 4. *Let P, Q, **s** and **R** be defined as above. If R is irreducible, then*

$$\lim_{n\to\infty}\frac{1}{n}H_\alpha(X^n;Y^n)=\frac{\ln\lambda}{1-\alpha}, \tag{11}$$

where λ is the largest positive eigenvalue of R.

Proof. Since the non-negative matrix R is irreducible, by the Frobenius theorem (e.g., cf. [15], [16]), it has a largest positive eigenvalue λ with associated positive eigenvector $\mathbf{b}$. Let b_m and b_M be the minimum and maximum elements, respectively, of $\mathbf{b}$. Then due to the non-negativity of $\mathbf{s}$,

$$\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}=\mathbf{s}R^{n-1}\mathbf{b}\le\mathbf{s}R^{n-1}\mathbf{1}b_M,$$

where $\cdot$ denotes the Euclidean inner product. Similarly, $\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}\ge\mathbf{s}R^{n-1}\mathbf{1}b_m$. As a result,

$$\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{b_M}\le\frac{1}{n}\ln\mathbf{s}R^{n-1}\mathbf{1}\le\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{b_m}.$$

Note that for all n, $\frac{\mathbf{s}\cdot\mathbf{b}}{b_M}$ is a constant. Thus

$$\lim_{n\to\infty}\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{b_M}=\lim_{n\to\infty}\frac{n-1}{n}\ln\lambda+\lim_{n\to\infty}\frac{1}{n}\ln\frac{\mathbf{s}\cdot\mathbf{b}}{b_M}=\ln\lambda.$$

Similarly, we have

$$\lim_{n\to\infty}\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{b_m}=\ln\lambda.$$

Hence,

$$\lim_{n\to\infty}\frac{1}{n}H_\alpha(X^n;Y^n)=\lim_{n\to\infty}\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{(1-\alpha)b_m}=\frac{\ln\lambda}{1-\alpha}.$$

□

Another technique can be borrowed from [14] to generalize Lemma 4 to the case where R is reducible. First R is rewritten in the canonical form detailed in Proposition 1 of [14]. Let λ_k be the largest positive eigenvalue of each self-communicating sub-matrix of R. For each inessential class C_i let λ_j be the largest positive eigenvalue of each class that is reachable from C_i. Define $\lambda=\max\{\lambda_k,\lambda_j\}$. Then (11) holds.

Appendix A: Shannon-type information measures

Name	Definition
Shannon Entropy	$H(p)=-\sum_{x\in\mathbb{S}}p(x)\ln p(x)$
Shannon Differential Entropy	$h(p)=-\int_{\mathbb{S}}p(x)\ln p(x)\,dx$
Shannon Cross-Entropy	$H(p;q)=-\sum_{x\in\mathbb{S}}p(x)\ln q(x)$
Shannon Differential Cross-Entropy	$h(p;q)=-\int_{\mathbb{S}}p(x)\ln q(x)dx$
KL Divergence, (Discrete)	$D(p\|q)=-\sum_{x\in\mathbb{S}}p(x)\ln\frac{p(x)}{q(x)}$
KL Divergence, (Continuous)	$D(p\|q)=-\int_{\mathbb{S}}p(x)\ln\frac{p(x)}{q(x)}\,dx$

Appendix B: Distributions listed in Table I

Name *(Parameters)*	**PDF** $f(x)$ (Support)
Beta $(a>0,\ b>0)$	$B(a,b)x^{a-1}(1-x)^{b-1}$ $\mathbb{S}=(0,1)$
$\boldsymbol{\chi^2}$ $(\nu\in\mathbb{Z}^+)$	$\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}$ $\mathbb{S}=\mathbb{R}^+$
Exponential $(\lambda>0)$	$\lambda e^{-\lambda x}$ $\mathbb{S}=\mathbb{R}^+$
Gamma $(k>0,\ \theta>0)$	$\frac{1}{\theta^k\Gamma(k)}x^{k-1}e^{-\frac{k}{\theta}}$ $\mathbb{S}=\mathbb{R}^+$
Gaussian $(\mu,\ \sigma^2>0)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\mathbb{S}=\mathbb{R}$
Laplace $(\mu,\ b^2>0)$	$\frac{1}{2b}e^{-\frac{\|x-\mu\|}{b}}$ $\mathbb{S}=\mathbb{R}$

Notes

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

References

[1] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 547–561.

[2] S. Verdú, "α-mutual information," in *Proceedings of the IEEE Information Theory and Applications Workshop*, 2015, pp. 1–6.

[3] A. Sarraf and Y. Nie, "RGAN: Rényi generative adversarial network," *SN Computer Science*, vol. 2, no. 1, p. 17, 2021.

[4] F. J. Valverde-Albacete and C. Peláez-Moreno, "The case for shifting the Rényi entropy," *Entropy*, vol. 21, pp. 1–21, 2019. [Online]. Available: https://doi.org/10.3390/e21010046

[5] M. Gil, F. Alajaji, and T. Linder, "Rényi divergence measures for commonly used univariate continuous distributions," *Information Sciences*, vol. 249, pp. 124–131, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025513004441

[6] K.-S. Song, "Rényi information, loglikelihood and an intrinsic distribution measure," *J. Statistical Planning and Inference*, vol. 93, 02 2001.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 2672–2680.

[8] H. Bhatia, W. Paul, F. Alajaji, B. Gharesifard, and P. Burlina, "Rényi Generative Adversarial Networks," *ArXiv:2006.02479v1*, 2020.

[9] ——, "Least kth-order and Rényi generative adversarial networks," *Neural Computation*, vol. 33, no. 9, pp. 2473–2510, Aug 2021. [Online]. Available: http://dx.doi.org/10.1162/neco_a_01416

[10] P. A. Kluza, "On Jensen-Rényi and Jeffreys-Rényi type f-divergences induced by convex functions," *Physica A: Statistical Mechanics and its Applications*, 2019.

[11] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 31, pp. 145–151, 1991.

[12] G. Casella and R. L. Berger, *Statistical Inference*. Cengage Learning, 2002.

[13] R. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, 10, 2001.

[14] Z. Rached, F. Alajaji, and L. L. Campbell, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Transactions on Information theory*, vol. 47, no. 4, pp. 1553–1561, 2001.

[15] E. Seneta, *Non-negative Matrices and Markov Chains*. Springer Science & Business Media, 2006.

[16] R. G. Gallager, *Discrete Stochastic Processes*. Springer, 1996.