Rényi Cross-Entropy:
Properties and Closed-Form Expressions for Sources with and without Memory

by

Ferenc Cole Thierrin

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Master of Applied Science

Queen's University
Kingston, Ontario, Canada
January 2023

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 analyse the properties of the Rényi and Natural Rényi differential cross-entropy measures and derive their expressions in closed form for a wide class of common continuous distributions belonging to the exponential family. We also establish the Rényi-type cross-entropy rates between stationary Gaussian processes and between finite-alphabet time-invariant Markov sources.

Acknowledgments

I would like to thank my supervisors, Dr. Fady Alajaji and Dr. Tamás Linder, for their guidance and patience during my time as their student.

I would also like to thank my parents, my sister, and especially my pets for their support during these years undertaking a Master's degree. I could not have done this without them, and for that I am truly grateful.

This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada.

Contents

List of Tables

Chapter 1

Introduction

In 1948, Claude Shannon published his seminal work [16] where he introduced a core *information measure* for a distribution with probability mass function (pmf) p over a support set $\mathbb{S}$:

$$H(p) = -\sum_{x\in\mathbb{S}} p(x)\ln p(x), \tag{1.1}$$

as well as a variant for a continuous distribution with a probability density function (pdf) p over a support $\mathbb{S} \subset \mathbb{R}$:

$$h(p) = -\int_{\mathbb{S}} p(x)\ln p(x)dx. \tag{1.2}$$

The above measures are called the *entropy* and *differential entropy*, respectively, of distribution p. Much like physical entropy,[1] it represents an unavoidable cost one must pay when losslessly encoding information distributed in a given way.

Shannon's paper laid the groundwork for the modern-day field of *Information Theory*, and gave way to many information measures over one or more probability

[1] As noted by Shannon, the equation for entropy is similar to that of Gibb's entropy, derived earlier for statistical mechanics.

distributions. These include the *Kullback-Leibler divergence* between two distributions p and q with common discrete support $\mathbb{S}$:

$$D(p\|q) = \sum_{x\in\mathbb{S}} p(x) \ln \frac{p(x)}{q(x)}, \tag{1.3}$$

as well as the *Shannon cross-entropy* between p and q,

$$H(p;q) = -\sum_{x\in\mathbb{S}} p(x) \ln q(x). \tag{1.4}$$

The above two measures admit natural extensions for continuous distributions, given by

$$D(p\|q) = \int_{\mathbb{S}} p(x) \ln \frac{p(x)}{q(x)} dx, \tag{1.5}$$

and

$$h(p;q) = -\int_{\mathbb{S}} p(x) \ln q(x) dx, \tag{1.6}$$

respectively. These measures, in broad terms, characterise the cost of assuming that information with distribution p has distribution q. This characterisation makes these measures of interest to the field of machine learning (ML). This can be seen in the use of the cross-entropy in generative adversarial networks [2], an ML system where a deep neural network trained to generate data is set to compete against another deep neural network trained to differentiate real data from generated data.

Introduced in 1961 by Alfréd Rényi, [13], the *Rényi entropy* of order α for pmf p, given by

$$H_\alpha(p) = \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p(x)^\alpha \tag{1.7}$$

and the *Rényi differential entropy* for pdf p, given by

$$h_\alpha(p) = \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x)^\alpha dx \tag{1.8}$$

for $\alpha > 0, \alpha \neq 1$, are generalisations of the Shannon entropy and differential entropy, respectively, in the sense that

$$\lim_{\alpha \to 1} H_\alpha(p) = H(p)$$

and

$$\lim_{\alpha \to 1} h_\alpha(p) = h(p).$$

The same paper also introduces a similar generalisation of the Kullback-Leibler divergence, the *Rényi Divergence* of order α between two distributions p and q with common support $\mathbb{S}$:

$$D_\alpha(p\|q) = \frac{1}{\alpha-1} \ln \sum_{x \in \mathbb{S}} p(x)^\alpha q(x)^{1-\alpha}, \tag{1.9}$$

in the discrete case, and by

$$D_\alpha(p\|q) = \frac{1}{\alpha-1} \ln \int_{\mathbb{S}} p(x)^\alpha q(x)^{1-\alpha} dx. \tag{1.10}$$

in the continuous case.

The aforementioned use of the cross-entropy in ML systems lends justification towards developing a Rényi-style generalisation of the cross-entropy; the additional parameter α can allow one to fine-tune a model further than using the Shannon cross-entropy. As such, this thesis examines two such *Rényi cross-entropies* and their logical differential counterparts. These are described in Chapter 2, alongside certain properties of them. In Chapter 3 we derive closed-form expressions for the Rényi differential cross-entropy for continuous distributions belonging to an exponential family, and in Chapter 4 we consider the Rényi cross-entropy rate for two sets of sources with memory: Gaussian sources and Markov sources. Finally, we conclude the thesis in Chapter 5. Parts of this work appeared in [18] and [19].

Chapter 2

Rényi Cross-Entropy: Definitions and Properties

2.1 Two definitions for Rényi Cross-Entropy

A quick calculation will show that $H(p;q) = H(p) + D(p\|q)$, and the same is true for the differential cross-entropy. Owing to this, [14] defines the Rényi differential cross-entropy of order α between p and q as

$$\tilde{h}_\alpha(p;q) \coloneqq D_\alpha(p||q) + h_\alpha(p), \tag{2.1}$$

with a trivial discrete analogue given by

$$\tilde{H}_\alpha(p;q) \coloneqq D_\alpha(p||q) + H_\alpha(p). \tag{2.2}$$

We will refer to the measures in (2.1) and (2.2) as the *Natural Rényi differential cross-entropy* and *Natural Rényi cross-entropy*, respectively.

For an information measure to be deemed a viable candidate for a Rényi cross-entropy, it should satisfy $\lim_{\alpha\to 1} H_\alpha(p;q) = H(p;q)$. In [2], it was proven that this condition holds for a variation of the shifted Rényi differential entropy (introduced in

[20]):

$$h_\alpha(p;q) = \frac{1}{1-\alpha} \ln \int_{\mathbb{S}} p(x)q(x)^{\alpha-1}\, dx. \tag{2.3}$$

Indeed, the same can be shown for its discrete analogue:

$$H_\alpha(p;q) = \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} p(x)q(x)^{\alpha-1} \tag{2.4}$$

We will refer to the measures in (2.4) and (2.3) as the *Rényi cross-entropy* and *Rényi differential cross-entropy*, respectively. As the thorough examinations of the Rényi entropy and divergence by various authors, in particular [7] and [17], make derivations of the Rényi Natural cross-entropy and its properties much simpler, the majority of this thesis's focus will be on the Rényi cross-entropy.

2.2 Basic Properties

Lemma 1. *Given $\alpha > 0$, and pdfs p and q, $\tilde{h}_\alpha(p;q)$ and $h_\alpha(p;q)$ can each be undefined even if the other is well-defined.*

Proof. This can be illustrated by an example: as shown in Table 3.2, the Rényi differential cross-entropy between two Exponential distributions with parameters λ_1 and λ_2, respectively, is affine with respect to $\ln \lambda_h$, where $\lambda_h = \lambda_1 + (\alpha-1)\lambda_2$. Meanwhile, as shown in Table 3.3, the Natural Rényi differential cross-entropy between the same two Exponential distributions is proportional to $\ln \lambda_\alpha$, where $\lambda_\alpha = \alpha\lambda_1 + (1-\alpha)\lambda_2$.

Suppose $\lambda_1 = 1$ and $\lambda_2 = 2$. If $\alpha = 0.5$, then $\lambda_h = 0$ and $\lambda_\alpha = 1.5$. At the same time, if $\alpha = 2$, then $\lambda_h = 3$ and $\lambda_\alpha = 0$. As $\ln 0$ is undefined, the Rényi differential cross-entropy in the first scenario is undefined, likewise for the Natural Rényi differential cross-entropy in the second scenario. □

Lemma 2. *The Rényi cross-entropy is non-increasing with respect to α:*

$$\frac{dH_\alpha(p;q)}{d\alpha} = \frac{-D(\tilde{p}||p)}{(1-\alpha)^2} \leq 0,$$

where

$$\tilde{p}\left(x\right) = \frac{p\left(x\right)q\left(x\right)^{\alpha-1}}{\sum_{\mathbb{S}} p\left(x'\right)q\left(x'\right)^{\alpha-1}}, \quad x \in \mathbb{S}$$

and $D(\cdot\|\cdot)$ is the Kullback-Leibler divergence.

Proof. We have the following sequence of identities:

$$\begin{aligned}
&\frac{d}{d_\alpha}\frac{1}{1-\alpha}\ln\sum_{\mathbb{S}} p\left(x\right)q\left(x\right)^{\alpha-1} \\
&= \frac{1}{\left(1-\alpha\right)^2}\ln\sum_{\mathbb{S}} p\left(x\right)q\left(x\right)^{\alpha-1} + \frac{1}{1-\alpha}\frac{\sum_{\mathbb{S}} p\left(x\right)\ln\left(q\left(x\right)\right)q\left(x\right)^{\alpha-1}}{\sum_{\mathbb{S}} p\left(x\right)q\left(x\right)^{\alpha-1}} \\
&= \frac{1}{\left(1-\alpha\right)^2}\left(\ln\sum_{\mathbb{S}} p\left(x\right)q\left(x\right)^{\alpha-1} + (1-\alpha)\sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\sum_{\mathbb{S}} p(x)q(x)^{\alpha-1} + \sum_{\mathbb{S}}\tilde{p}(x)\ln q - \alpha\sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\frac{\sum_{\mathbb{S}} p(x')q(x')^{\alpha-1}}{q^\alpha} + \sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\frac{p(x)q(x)^{\alpha-1}}{\tilde{p}(x)q^\alpha} + \sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\frac{p(x)}{\tilde{p}(x)q} + \sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\frac{p(x)}{\tilde{p}(x)} - \sum_{\mathbb{S}}\tilde{p}(x)\ln q + \sum_{\mathbb{S}}\tilde{p}(x)\ln q\right) \\
&= \frac{1}{(1-\alpha)^2}\left(\sum_{\mathbb{S}}\tilde{p}(x)\ln\frac{p(x)}{\tilde{p}(x)}\right)
\end{aligned}$$

$$= \frac{-D(\tilde{p}||p)}{(1-\alpha)^2} \leq 0,$$

where the inequality follows from the non-negativity of the Kullback-Leibler divergence. □

Remark. *In [2], the Rényi differential cross-entropy $h_\alpha(p;q)$ is proven to be non-increasing in α by similarly showing (via a different approach) that its derivative with respect to α is non-positive.*

Lemma 3. *The limit of the Rényi cross-entropy as $\alpha \to \infty$ is given by $-\ln q_M$, where $q_M \coloneqq \max_{x\in\mathbb{S}} q(x)$.*

Proof. From [10, Section 5], we know that for finite or countably infinite supports, the Rényi entropy satisfies

$$\lim_{\beta\to\infty} H_\beta(q) = -\ln q_M.$$

In light of this result, for any positive constant $\tilde{c}$, we have

$$\begin{aligned}
&\lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} \tilde{c} q\left(x\right)^{\alpha-1} \\
&= \lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \tilde{c} + \lim_{\alpha\to\infty} \frac{1}{1-\alpha} \ln \sum_{x\in\mathbb{S}} q\left(x\right)^{\alpha-1} \\
&= 0 + \lim_{\beta\to\infty} \frac{1}{-\beta} \ln \sum_{x\in\mathbb{S}} q\left(x\right)^{\beta} \qquad (\beta = \alpha - 1) \\
&= \lim_{\beta\to\infty} \frac{1-\beta}{-\beta} \frac{1}{1-\beta} \ln \sum_{\mathbb{S}} q\left(x\right)^{\beta} \\
&= \lim_{\beta\to\infty} \frac{1}{1-\beta} \ln \sum_{\mathbb{S}} q\left(x\right)^{\beta} \\
&= \lim_{\beta\to\infty} H_\beta(q) = -\ln q_M. \qquad (2.5)
\end{aligned}$$

Let $\mathbb{S}$ be finite, and denote the minimum and maximum values of $p(x)$ over $\mathbb{S}$ by p_m and p_M, respectively. Then,

$$\frac{1}{1-\alpha}\ln\sum_{x\in\mathbb{S}} p_M q\left(x\right)^{\alpha-1} \leq \frac{1}{1-\alpha}\ln\sum_{x\in\mathbb{S}} p(x) q\left(x\right)^{\alpha-1} \tag{2.6}$$

and

$$\frac{1}{1-\alpha}\ln\sum_{x\in\mathbb{S}} p(x) q\left(x\right)^{\alpha-1} \leq \frac{1}{1-\alpha}\ln\sum_{x\in\mathbb{S}} p_m q\left(x\right)^{\alpha-1} \tag{2.7}$$

for $\alpha > 1$. Hence

$$\lim_{\alpha\to\infty} H_\alpha\left(p;q\right) = -\ln q_M. \tag{2.8}$$

If $\mathbb{S}$ is countably infinite the proof becomes harder, as p_m does not exist. Assume without loss of generality that $\mathbb{S} = \mathbb{N}$ and $q(0) = q_M$. For $n \geq 1$, we have

$$\sum_{x\in\mathbb{N}} p(x)q(x)^{\alpha-1} \geq \sum_{x=1}^{n} p(x)q(x)^{\alpha-1},$$

and hence for $\alpha > 1$,

$$H_\alpha(p;q) \leq \frac{1}{1-\alpha}\ln\sum_{x=1}^{n} p(x)q(x)^{\alpha-1}.$$

Thus for all $n \geq 1$,

$$\limsup_{\alpha\to\infty} H_\alpha(p;q) \leq \lim_{\alpha\to\infty}\frac{1}{1-\alpha}\ln\sum_{x=1}^{n} p(x)q(x)^{\alpha-1} = -\ln q_M.$$

As p_M exists, we can use the inequality (2.6) and the argument in (2.5) to conclude that $\lim_{\alpha\to\infty} H_\alpha(p;q) = -\ln q_M$. □

The previous two lemmas directly imply that $H_\alpha(p;q) \geq 0$. In addition, the Natural Rényi cross-entropy, being the sum of two non-negative measures, is non-negative. Similarly, the lack of non-negativity in the Rényi differential entropy demonstrates that the Rényi and Natural Rényi differential cross-entropies can be negative. Finally, using the expressions of $\lim_{\alpha\to\infty} H_\alpha(p)$ and $\lim_{\alpha\to\infty} D_\alpha(p\|q)$ (e.g., see [5, 10, 21]), we readily obtain the limit of the Natural Rényi cross-entropy as $\alpha \to \infty$ as follows.

Remark. *The limit of the Natural Rényi cross-entropy $\tilde{H}_\alpha(p;q)$ as $\alpha \to \infty$ is given by*

$$\begin{aligned}
\lim_{\alpha\to\infty} \tilde{H}_\alpha(p;q) &= \lim_{\alpha\to\infty} H_\alpha(p) + \lim_{\alpha\to\infty} D_\alpha(p\|q) \\
&= -\ln \max_{\{x\in\mathbb{S}\}} (p(x)) - \ln \max_{\{x\in\mathbb{S}\}} \left(\frac{p(x)}{q(x)}\right).
\end{aligned}$$

Chapter 3

Rényi Differential Cross-Entropy for Continuous Memoryless Sources

In this chapter we derive the Rényi cross-entropy measures for continuous memoryless sources belonging to an exponential family.

3.1 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 $\Theta \subseteq \mathbb{R}^j$ and functions $b : \mathbb{S} \mapsto \mathbb{R}$, $c : \Theta \mapsto \mathbb{R}_{\geq 0}$, $T : \mathbb{S} \mapsto \mathbb{R}^m$, and $\eta : \Theta \mapsto \mathbb{R}^m$ such that the pdf of distributions in a family have the form

$$f(x) = c(\theta)b(x)\exp\left(\eta(\theta)\cdot T(x)\right), \qquad x \in \mathbb{S} \tag{3.1}$$

where $\cdot$ denotes the standard inner product in $\mathbb{R}^m$. Alternatively, by using the parameter $\eta = \eta(\theta)$[1], the pdf can also be written

$$f(x) = b(x)\exp\left(\eta \cdot T(x) + A(\eta)\right), \tag{3.2}$$

[1] η is known as the natural parameter.

where $A(\eta) : \eta(\Theta) \mapsto \mathbb{R}$ with $A(\eta) = \ln c(\theta)$.[2]

Many common distributions belong to an exponential family. Many more have subsets that belong to an exponential family. The set of Pareto distributions, for example, is not an exponential family; however the set of Pareto distributions where $m = 1$ is, as is the set of Pareto distributions where $m = 2$, and so on. Table 3.1 lists the pdfs of certain distributions that form exponential families. Parameters that are required to be constant for the family to count as an exponential family are marked with the † symbol. Finally, by $\Sigma \succ 0$, we mean that the matrix Σ is positive-definite.

Table 3.1: List of PDFs from Exponential Families

Name **Parameters (Θ)**	**PDF** **Support**
Beta[3]	$\frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$
$(a > 0,\ b > 0)$	$\mathbb{S} = (0,1)$
$\boldsymbol{\chi}$ **(scaled)**	$\frac{2^{1-k/2} x^{k-1} e^{-x^2/2\sigma^2}}{\sigma^k \Gamma\left(\frac{k}{2}\right)}$
$(k > 0,\ \sigma > 0)$	$\mathbb{S} = \mathbb{R}^+$
$\boldsymbol{\chi}$ **(non-scaled)**	$\frac{2^{1-k/2} x^{k-1} e^{-x^2/2}}{\Gamma\left(\frac{k}{2}\right)}$
$(k > 0)$	$\mathbb{S} = \mathbb{R}^+$
$\boldsymbol{\chi^2}$	$\frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}$
$(\nu > 0)$	$\mathbb{S} = \mathbb{R}^+$

[2]Most literature on exponential families defines $A(\eta)$ as $-\ln c(\theta)$.
[3]$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Exponential $(\lambda > 0)$	$\lambda e^{-\lambda x}$ $\quad \mathbb{S} = \mathbb{R}^+$
Gamma $(k > 0,\, \theta > 0)$	$\frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$ $\quad \mathbb{S} = \mathbb{R}^+$
Gaussian (univariate) $(\mu,\, \sigma^2 > 0)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\quad \mathbb{S} = \mathbb{R}$
Gaussian (multivariate) $(\boldsymbol{\mu} \in \mathbb{R}^n,\, \Sigma \in \mathbb{R}^{n\times n} \succ 0)$	$\frac{1}{\sqrt{(2\pi)^n \lvert\Sigma\rvert}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$ $\quad \mathbb{S} = \mathbb{R}^n$
Half-Normal $(\sigma^2 > 0)$	$\sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ $\quad \mathbb{S} = \mathbb{R}^+$
Gumbel $(\mu^\dagger,\, \beta > 0)$	$\frac{1}{\beta} \exp -\left(\frac{x-\mu}{\beta} + e^{-\frac{x-\mu}{\beta}}\right)$ $\quad \mathbb{S} = \mathbb{R}$
Pareto $(m^\dagger > 0,\, a > 0)$	$a m^a x^{-(1+m)}$ $\quad \mathbb{S} = (m, \infty)$
Maxwell Boltzmann $(\sigma > 0)$	$\frac{2x^2}{\sqrt{\pi\sigma^6}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ $\quad \mathbb{S} = \mathbb{R}^+$
Rayleigh $(\sigma^2 > 0)$	$\frac{x}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ $\quad \mathbb{S} = \mathbb{R}^+$
Laplace $(\mu^\dagger,\, b^2 > 0)$	$\frac{1}{2b} e^{-\frac{\lvert x-\mu\rvert}{b}}$ $\quad \mathbb{S} = \mathbb{R}$

Log-Normal	$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}$
$(\mu > 0,\, \sigma^2 > 0)$	$\mathbb{S} = \mathbb{R}$

3.2 Rényi Cross-Entropies for Distributions from the Same Exponential Family

Lemma 4. *Let f_1 and f_2 be pdfs in the same exponential family with common support $\mathbb{S}$ and with natural parameters η_1 and η_2, respectively. Define f_h as being in the same family but with natural parameter $\eta_h = \eta_1 + (\alpha-1)\eta_2$.*[4] *Then the Rényi cross-entropy between f_1 and f_2 is given by*

$$h_\alpha\left(f_1; f_2\right) = \frac{A\left(\eta_1\right) - A\left(\eta_h\right) + \ln E_h}{1-\alpha} - A\left(\eta_2\right), \tag{3.3}$$

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

Proof. For $x \in \mathbb{S}$, we have

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

[4] We assume that η_h is in the natural parameter space.

Thus,

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

and therefore,

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

□

Corollary. *The Natural Rényi differential cross-entropy between f_1 and f_2 is given by*

$$\tilde{h}_\alpha(f_1; f_2) = \frac{A(\eta_\alpha) - A(\alpha\eta_1) + \ln E_\alpha}{1-\alpha} - A(\eta_2), \tag{3.4}$$

$\alpha \neq 1$, *where* $\eta_\alpha = \alpha\eta_1 + (1-\alpha)\eta_2$,[5] *and*

$$E_\alpha = \mathbb{E}_{f_{\alpha_1}}\left[b(X)^{\alpha-1}\right] = \int b(x)^{\alpha-1} f_{\alpha_1}(x)\, dx$$

where f_{α_1} refers to a distribution in the exponential family of interest with natural parameter $\alpha\eta_1$.

Proof. When $\eta_2 = \eta_1$, (3.3) reduces to

$$h_\alpha(f_1) = \frac{\alpha A(\eta_1) - A(\alpha\eta_1) + \ln E_\alpha}{1-\alpha}.$$

[5] We assume that η_α and $\alpha\eta_1$ are in the natural parameter space.

By adding the result with the formula for the Rényi divergence found in [7],

$$D_\alpha(f_1 \| f_2) = \frac{A(\eta_\alpha) - \alpha A(\eta_1)}{1-\alpha} - A(\eta_2)$$

one obtains (3.4). □

Remark. *The Natural Rényi differential cross-entropy can also be found by summing the Rényi differential entropy found in [17] with the Rényi divergence found in [7].*

3.3 Tables of Rényi and Natural Rényi Differential Cross-Entropies

Applying (3.3) to the distributions in Table 3.1 yields their Rényi Cross-Entropy, and these are listed in Table 3.2. The derivations are detailed in Appendix A. In this table, the subscript i is used to denote that a parameter belongs to pdf f_i, $i = 1, 2$.

Table 3.2: Rényi Differential Cross-Entropies

Name	$h_\alpha(f_1; f_2)$
Beta	$\ln B(a_2, b_2) + \frac{1}{\alpha - 1} \ln \frac{B(a_h, b_h)}{B(a_1, b_1)}$
	$a_h := a_1 + (\alpha - 1)(a_2 - 1), \quad a_h > 0$
	$b_h := b_1 + (\alpha - 1)(b_2 - 1), \quad b_h > 0$
χ (scaled)	$\frac{1}{2}\left(k_2 \ln \sigma_2^2 \sigma_h^2 - \ln 2\sigma_h^2\right) + \ln \Gamma\left(\frac{k_2}{2}\right) + \frac{1}{\alpha - 1}\left(\ln \Gamma\left(\frac{k_h}{2}\right) - \ln \Gamma\left(\frac{k_1}{2}\right) - \frac{k_1}{2} \ln \sigma_1^2 \sigma_h^2\right)$
	$\sigma_h^2 := \frac{1}{\sigma_1^2} + \frac{\alpha - 1}{\sigma_2^2}, \quad \sigma_h^2 > 0$
	$k_h := k_1 + (\alpha - 1)(k_2 - 1), \quad k_h > 0$

χ (non-scaled)	$\frac{1}{2}(k_2\ln\alpha - \ln 2\alpha) + \ln\Gamma\left(\frac{k_2}{2}\right) + \frac{1}{\alpha-1}\left(\ln\Gamma\left(\frac{k_h}{2}\right) - \ln\Gamma\left(\frac{k_1}{2}\right) - \frac{k_1}{2}\ln\alpha\right)$
	$k_h \coloneqq k_1 + (\alpha-1)(k_2-1), \qquad k_h > 0$
χ^2	$\frac{1}{1-\alpha}\left(\frac{\nu_1}{2}\ln(\alpha) - \ln\Gamma\left(\frac{\nu_1}{2}\right) + \ln\Gamma\left(\frac{\nu_h}{2}\right)\right) + \frac{2-\nu_2}{2}\ln(\alpha) + \ln 2\Gamma\left(\frac{\nu_2}{2}\right)$
	$\nu_h \coloneqq \nu_1 + (\alpha-1)(\nu_2-2), \qquad \nu_h > 0$
Exponential	$\frac{1}{1-\alpha}\ln\frac{\lambda_1}{\lambda_h} - \ln\lambda_2$
	$\lambda_h \coloneqq \lambda_1 + (\alpha-1)\lambda_2, \qquad \lambda_h > 0$
Gamma	$\ln\Gamma(k_2) + k_2\ln\theta_2 + \frac{1}{1-\alpha}\left(\ln\frac{\Gamma(k_h)}{\Gamma(k_1)} - k_h\ln\theta_h - k_1\ln\theta_1\right)$
	$\theta_h \coloneqq \frac{\theta_1+(a-1)\theta_2}{(\alpha-1)\theta_1\theta_1}, \qquad \theta_h > 0$ $k_h \coloneqq k_1 + (\alpha-1)k_2, \qquad k_h > 0$
Gaussian (univariate)	$\frac{1}{2}\left(\ln(2\pi\sigma_2^2) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h}\right) + \frac{(\mu_1-\mu_2)^2}{(\sigma^2)_h}\right)$
	$(\sigma^2)_h \coloneqq \sigma_2^2 + (\alpha-1)\sigma_1^2, \qquad (\sigma^2)_h > 0$
Gaussian (Multivariate)	$\frac{1}{2-2\alpha}\big(-\ln\lvert A\rvert\lvert\Sigma_1\rvert + (1-\alpha)\ln(2\pi)^n\lvert\Sigma_2\rvert - d\big)$
	$A \coloneqq \Sigma_1^{-1} + (\alpha-1)\Sigma_2^{-1}, \qquad A \succ 0$ $d \coloneqq \boldsymbol{\mu_1^T}\Sigma_1^{-1}\boldsymbol{\mu_1} + (\alpha-1)\,\boldsymbol{\mu_2^T}\Sigma_2^{-1}\boldsymbol{\mu_2}$ $-(\boldsymbol{\mu_1^T}\Sigma_1^{-1} + (\alpha-1)\,\boldsymbol{\mu_2^T}\Sigma_2^{-1})A^{-1}(\Sigma_1^{-1}\boldsymbol{\mu_1} + (\alpha-1)\,\Sigma_2^{-1}\boldsymbol{\mu_2})$
Gumbel ($\boldsymbol{\beta_1 = \beta_2 = \beta}$)	$\frac{1}{1-\alpha}\left(\ln\frac{\Gamma(2-\alpha)}{\beta} - \frac{\mu_1}{\beta} - \alpha\ln\eta_h\right) + \frac{\mu_2}{\beta}$
	$\eta_h \coloneqq e^{-\mu_1/\beta} + (\alpha-1)e^{-\mu_2/\beta}, \qquad \eta_h > 0$

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 $(\mu_1 = \mu_2 = 0)$	$\ln(2b_2) + \frac{1}{1-\alpha}\ln\left(\frac{b_2}{2b_h}\right)$
	$b_h := b_2 + (1-\alpha)b_1, \qquad b_h > 0$
Log-normal	$\frac{1}{2}\Big(\ln(2\pi\sigma_2^2) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h}\right) + \frac{(\mu_1-\mu_2)^2 + \mu_1\sigma_2^2 + (\alpha-1)\sigma_1^2(\mu_2 + 2\sigma_2^2)}{(\sigma^2)_h}\Big)$
	$(\sigma^2)_h := \sigma_2^2 + (\alpha-1)\sigma_1^2, \qquad (\sigma^2)_h > 0$
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 $(\boldsymbol{m_1 = m_2 = m})$	$-\ln m - \ln\lambda_2 + \frac{1}{1-\alpha}\ln\frac{\lambda_1}{\lambda_h}$
	$\lambda_h := \lambda_1 + (\alpha-1)(\lambda_2+1), \qquad \lambda_h > 0$
Rayleigh	$\frac{\ln\sigma_1^2 - \alpha\ln\sigma_h^2 + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln 2\sigma_2^2$
	$\sigma_h^2 := \sigma_1^{-2} + (\alpha-1)\sigma_2^{-2}, \qquad \sigma_h^2 > 0$

Similarly, applying (3.4) to the aforementioned distributions yields their Natural Rényi Cross-Entropy, summarised in Table 3.3.[6]

[6]The same notation regarding the subscripts of the parameters seen in Table 3.2 is used here.

Table 3.3: Natural Rényi Differential Cross-Entropies

Name	$\tilde{h}_\alpha(f_1; f_2)$
Beta	$\ln B(a_2, b_2) + \frac{1}{\alpha-1} \ln \frac{B(a_\alpha, b_\alpha)}{B(\alpha(a_1-1)+1, \alpha(b_1-1)+1)}$
	$a_\alpha := \alpha a_1 + (1-\alpha)a_2, \quad a_\alpha > 0$ $b_\alpha := \alpha b_1 + (1-\alpha)b_2, \quad b_\alpha > 0$
χ (scaled)	$\frac{1}{2}\left(-\ln \frac{2\sigma_1^2}{\alpha} + k_2 \ln \sigma_2^2 \sigma_\alpha^2\right) + \ln\Gamma\left(\frac{k_2}{2}\right)$ $+\frac{1}{1-\alpha}\left(\frac{\alpha k_1 \ln \frac{\sigma_\alpha^2 \sigma_1^2}{\alpha}}{2} - \ln\Gamma(\frac{k_\alpha}{2}) + \ln\Gamma\left(\frac{\alpha(k_1-1)+1}{2}\right)\right)$
	$\sigma_\alpha^2 := \frac{\alpha}{\sigma_1^2} + \frac{1-\alpha}{\sigma_2^2}, \quad \sigma_\alpha^2 > 0$ $k_\alpha := \alpha k_1 + (1-\alpha)k_2, \quad k_\alpha > 0$
χ (non-scaled)	$\frac{-\ln 2\alpha}{2} + \ln\Gamma(\frac{k_2}{2})$ $+\frac{1}{1-\alpha}\left(-\ln\Gamma(\frac{k_\alpha}{2}) - \frac{\alpha k_1 \ln \alpha}{2} + \ln\Gamma(\frac{\alpha(k_1-1)+1}{2})\right)$
	$k_\alpha := \alpha k_1 + (1-\alpha)k_2, \quad k_\alpha > 0$
χ^2	$\frac{1}{1-\alpha}\left(-\ln\Gamma\left(\frac{\nu_\alpha}{2}\right) + \alpha \ln\Gamma\left(\frac{\nu_1}{2}\right)\right) + \ln\Gamma\left(\frac{\nu_2}{2}\right)$
	$\nu_\alpha := \alpha\nu_1 + (1-\alpha)k, \quad \nu_\alpha > 0$
Exponential	$\frac{1}{1-\alpha} \ln \frac{\lambda_1}{\alpha\lambda_\alpha} - \ln \lambda_2$
	$\lambda_\alpha := \alpha\lambda_1 + (1-\alpha)\lambda_2, \quad \lambda_\alpha > 0$
Gamma	$\ln\Gamma(k_2) + k_2 \ln\theta_2$ $+\frac{1}{1-\alpha}\left(\ln \frac{\Gamma(k_1)}{\Gamma(k_\alpha)} - k_\alpha \ln\theta_\alpha - \alpha^2 k_1 \ln\theta_1\right)$
	$\theta_\alpha := \alpha\theta_1^{-1} + (1-\alpha)\theta_2^{-1}, \ k_\alpha := \alpha k_1 + (1-\alpha)k_2, \quad \theta_\alpha > 0$

Gaussian (Univariate)	$\frac{1}{2}\left(\ln(2\pi\sigma_2^2)+\frac{(\mu_1-\mu_2)^2}{(\sigma^2)_\alpha}+\frac{1}{1-\alpha}\ln\left(\frac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right)$
	$(\sigma^2)_\alpha \coloneqq \alpha\sigma_2^2+(1-\alpha)\sigma_1^2, \qquad (\sigma^2)_\alpha>0$
Gaussian (Multivariate)	$\frac{1}{2-2\alpha}\left(-\ln\lvert\alpha\rvert+\ln\lvert A\rvert\lvert\Sigma_1\rvert+d\right)+\frac{1}{2}\ln\frac{(2\pi)^n\lvert\Sigma_1\rvert^2}{\lvert\Sigma_2\rvert}$
	$A \coloneqq \alpha\Sigma_1^{-1}+(1-\alpha)\Sigma_2^{-1}, \qquad A\succ 0$ $d \coloneqq (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})^T\Sigma_1 A\Sigma_2(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})$
Gumbel ($\boldsymbol{\beta_1=\beta_2=\beta}$)	$\frac{\mu_2+\alpha\mu_1}{\beta}+\frac{1}{1-\alpha}\left(\ln\frac{\Gamma(2-\alpha)\eta_\alpha}{\alpha\beta}+\frac{\mu_1}{\beta}\right)$
	$\eta_\alpha \coloneqq \alpha e^{-\mu_1/\beta}+(1-\alpha)e^{-\mu_2/\beta}, \qquad \eta_\alpha>0$
Half-Normal	$\frac{1}{2}\left(\ln(\frac{\pi\sigma_2^2}{2})+\frac{1}{1-\alpha}\ln\left(\frac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right)$
	$(\sigma^2)_\alpha \coloneqq \alpha\sigma_2^2+(1-\alpha)\sigma_1^2\ , \qquad (\sigma^2)_\alpha>0$
Laplace ($\mu_1=\mu_2=0$)	$\frac{\ln b_\alpha+\ln\alpha b_1}{1-\alpha}+\ln 2b_2$
	$b_\alpha \coloneqq \frac{\alpha}{b_1}+\frac{1-\alpha}{b_2}, \qquad b_\alpha>0$
Log-Normal	$\frac{1}{2}\left(\frac{1-\alpha}{\alpha}\sigma_1^2+\ln(2\pi\sigma_2^2)+\frac{(\mu_1-\mu_2)^2}{(\sigma^2)_\alpha}+\frac{1}{1-\alpha}\ln\left(\frac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right)+\mu_1$
	$(\sigma^2)_\alpha \coloneqq \alpha\sigma_2^2+(1-\alpha)\sigma_1^2, \qquad (\sigma^2)_\alpha>0$
Maxwell Boltzmann	$\frac{-\ln 2+3\ln\sigma_2^2}{2}+\ln\frac{\alpha}{\sigma_1^2}$ $+\frac{1}{1-\alpha}\left(\frac{3}{2}\ln\frac{\sigma_\alpha\sigma_1^2}{\alpha}-\alpha\ln\frac{\sqrt{\pi}}{2}+\ln\Gamma\left(\alpha+\frac{1}{2}\right)\right)$
	$\sigma_\alpha^2 \coloneqq \frac{\alpha}{\sigma_1^2}+\frac{1-\alpha}{\sigma_2^2}, \qquad \sigma_\alpha^2>0$

Pareto ($m_1 = m_2 = m$)	$\frac{1}{1-\alpha}\Big(\ln\lambda_\alpha - \ln\left(1-\alpha(\lambda_1 - 1)\right)\Big) - \ln\lambda_2 m$
	$\lambda_\alpha := \alpha\lambda_1 + (1-\alpha)\,\lambda_2, \qquad \lambda_\alpha > 0$
Rayleigh	$\frac{\ln\sigma_1^2(\sigma^2)_\alpha + \ln\alpha + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln 2\sigma_1^2$
	$(\sigma^2)_\alpha := \alpha\sigma_1^{-2} + \frac{1}{2}\ln\frac{2\sigma_1^4\sigma_2^4}{\alpha}, \qquad (\sigma^2)_\alpha > 0$

Remark. *While Tables 3.2 and 3.3 focused exclusively on continuous distributions, it should be noted that* (3.3) *and* (3.4) *also hold for discrete exponential family distributions. In this case, we note that*

$$\mathbb{E}_{f_h}\left[b(X)^{\alpha-1}\right] = \sum_{x\in\mathbb{S}} b(x)^{\alpha-1} f_h(x)$$

and

$$\mathbb{E}_{f_{\alpha_1}}\left[b(X)^{\alpha-1}\right] = \sum_{x\in\mathbb{S}} b(x)^{\alpha-1} f_{\alpha_1}(x).$$

Chapter 4

Rényi Cross-Entropy Rate for Sources with Memory

4.1 Rényi Differential Cross-Entropy Rate for Stationary Gaussian Processes

In this section, we consider the Rényi differential cross-entropy rate, $\lim_{n\to\infty} \frac{1}{n} H_\alpha(p;q)$, between two stationary zero-mean Gaussian processes, $\{X_j\}_{j=1}^{\infty}$ and $\{Y_j\}_{j=1}^{\infty}$, respectively. That is to say, given $n \in \mathbb{Z}^+$, $X^n := (X_1, X_2, ..., X_n)$ and $Y^n := (Y_1, Y_2, ..., Y_n)$ are multivariate Gaussian random vectors with mean $\mathbf{0}$ and invertible covariance matrices Σ_{X^n} and Σ_{Y^n}, respectively. These covariance matrices are obtained via the inverse Fourier transforms of the power spectral densities associated with $\{X_j\}$ and $\{Y_j\}$, denoted by $\tilde{f}(\lambda)$ and $\tilde{g}(\lambda)$ respectively: the element in the r^{th} row and c^{th} column of Σ_{X^n} is ([9])

$$\frac{1}{2\pi}\int_0^{2\pi} \tilde{f}(\lambda) e^{-i2\pi(r-c)}\, d\lambda. \tag{4.1}$$

One will note that this formula for the covariance matrix depends entirely on the difference between indices r and c, due to the stationarity of the Gaussian processes.

As a result of this, the covariance matrices of $\{X_j\}$ and $\{Y_j\}$ are Toeplitz. One can also remark that $B^n := \Sigma_{Y^n} + (\alpha - 1)\Sigma_{X^n}$ will likewise be Toeplitz.

Lemma 5. *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. For the range of α such that B^n is positive-definite, 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*

$$\lim_{n\to\infty} \frac{1}{n} h_\alpha(X^n; Y^n) = \frac{\ln 2\pi}{2} + \frac{1}{4\pi(1-\alpha)} \int_0^{2\pi} \left[(2-\alpha) \ln \tilde{g}(\lambda) - \ln \tilde{h}(\lambda) \right] d\lambda.$$

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

Proof. As shown in Table 3.2, the Rényi cross-entropy between two zero-mean multivariate Gaussian random variables is given by

$$\frac{\ln |\Sigma_{X^n}||A|}{2(\alpha - 1)} + \frac{1}{2} \ln |\Sigma_{Y^n}| + \frac{n}{2} \ln 2\pi. \tag{4.2}$$

Noting that $A = \Sigma_{X^n}^{-1} B^n \Sigma_{Y^n}^{-1}$, we rewrite (4.2) and normalise by n 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 [8, Lemma 4.4] that for the sequence of $n \times n$ Hermitian Toeplitz matrices $\{T_n\}$ with spectral density $t(\lambda)$ such that $\ln t(\lambda)$ is Riemann integrable, one has

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

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} h_\alpha(X^n; Y^n) = \frac{\ln 2\pi}{2} + \frac{1}{4\pi(1-\alpha)} \int_0^{2\pi} \left[(2-\alpha)\ln \tilde{g}(\lambda) - \ln \tilde{h}(\lambda)\right] d\lambda,$$

where $\tilde{h}(\lambda)$ is the spectral density associated with the sequence of matrices generated by B_n. □

The above Rényi differential cross-entropy rate, alongside with the Shannon and Natural Rényi differential cross-entropy rates,[1] are summarised in Table 4.1. Here,

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

4.2 Rényi Cross-Entropy Rate for Markov Sources

Markov sources are sources that exhibit the Markov property; the value of the source at the next time unit is dependent solely on its value at the present. In more formal terms, for a Markov source $\{X_n\}_{n=1}^{\infty}$ with alphabet $\mathbb{S}$, we have

$$P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_1 = x_1) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

[1]These latter two measures are calculated via manipulating the Shannon and Rényi Divergence rates between two stationary zero-mean Gaussian processes as detailed in [7].

Table 4.1: Differential Cross-Entropy Rates for Stationary Zero-Mean Gaussian Sources

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

for all $x_1, x_2 ... x_n \in \mathbb{S}$. The Markov source $\{X_n\}$ may be time-invariant:

$$P(X_n = i | X_{n-1} = j) = P(X_{n+k} = i | X_{n+k-1} = j)$$

for all $i, j \in \mathbb{S}, k \geq 2$ and $n \geq 1$. In this case, the Markov source may be characterised by an $|\mathbb{S}| \times |\mathbb{S}|$ matrix $\mathcal{P}$, where the element in the i^{th} row and j^{th} column equals $P(X_n = x_i | X_{n-1} = x_j), x_i, x_j \in \mathbb{S}$. In practice, many information sources are modelled as Markov chains.

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 kernels $P(\cdot|\cdot)$ and $Q(\cdot|\cdot)$, respectively. Then for any vector $i^n = (i_1, \dots, i_n) \in \mathbb{S}^n$, its n-dimensional joint distributions are given by

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

and

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

respectively, with arbitrary initial distributions, $p(i_1)$ and $q(i_1)$, $i_1 \in \mathbb{S}$. For simplicity, we assume that $p(i)$, $q(i), Q(j|i) > 0$ for all $i, j \in \mathbb{S}$. Define the Rényi cross-entropy rate between $\{X_j\}$ and $\{Y_j\}$ as

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

Note that by defining the matrix R using the formula

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

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

$$\lim_{n\to\infty} \frac{1}{n} \frac{1}{1-\alpha} \ln \mathbf{s} R^{n-1} \mathbf{1}, \tag{4.3}$$

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

A result derived by [12] for the Rényi divergence between Markov sources can thus be used to find the Rényi cross-entropy rate for Markov sources.

Lemma 6. *Define P, Q, $\mathbf{s}$ and R as above. If R is irreducible, then*

$$\lim_{n\to\infty} \frac{1}{n} H_\alpha(X^n; Y^n) = \frac{\ln \lambda}{1-\alpha}, \tag{4.4}$$

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, 6]), 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,

$$\lim_{n\to\infty}\frac{1}{n}H_\alpha(X^n;Y^n) = \lim_{n\to\infty}\frac{1}{n}\ln\frac{\lambda^{n-1}\mathbf{s}\cdot\mathbf{b}}{(1-\alpha)b_m} = \frac{\ln\lambda}{1-\alpha}. \tag{4.5}$$

□

Another technique can be borrowed from [12] to generalize Lemma 6 to the case where R is reducible. First R is rewritten in the canonical form detailed in Proposition 1 of [12]. Let λ_k be the largest positive eigenvalue of each self-communicating sub-matrix of R, and let $\tilde{\lambda}$ be the maximum of these λ_k's. For each inessential class C_i, let λ_j be the largest positive eigenvalue of the sub-matrix of each class C_j that is reachable from C_i, and let $\lambda^\dagger$ be the maximum of these λ_j's. Define $\lambda(R) = \max\{\tilde{\lambda}, \lambda^\dagger\}$. Then (4.4) holds.

The Rényi cross-entropy rate, alongside the Shannon and Natural Rényi cross-entropy rates, is summarised in Table 4.2. For simplicity, we assume that the initial distribution vectors, p and q, of both Markov chains also have positive entries ($p > 0$ and $q > 0$).[2] 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 where 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.

Table 4.2: Cross-Entropy Rates for Time-Invariant Markov Sources

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

[2] This condition can be relaxed via the approach used to prove [12, Theorem 1].

Chapter 5

Conclusion

In this thesis, a gap in the literature on information measures was filled by the derivation of properties of the Rényi and Natural Rényi cross-entropies. Having been given the equations for the Rényi and Natural Rényi cross-entropies, we chose to explore how these formulae would evaluate when applied to commonly used distributions. We were able to derive general expressions for the Rényi and Natural Rényi differential cross-entropies for an exponential family. We used these to derive closed-form formulae for commonly used exponential families. In addition, we built upon the work of [7] and [17] to derive cross-entropy rates for stochastic processes with memory: Gaussian sources and Markov sources. Further work upon this topic includes investigating cross-entropy measures based on the f-divergences [1, 3, 4, 11].

Appendix A

Derivations of Rényi Differential and Natural Rényi Differential Cross-Entropies

A.1 Univariate Gaussian

A.1.1 Rényi Differential Cross-Entropy

For $x \in \mathbb{R}$, we have

$$f_1(x)f_2(x)^{\alpha-1} = \frac{1}{\sigma_1\sqrt{2\pi}}(\frac{1}{\sigma_2\sqrt{2\pi}})^{\alpha-1}\exp(\frac{-(x-\mu_1)^2}{2\sigma_1^2} + \frac{-(x-\mu_2)^2}{2\sigma_2^2}(\alpha-1))$$
$$= \frac{\sigma_2}{(\sigma_2\sqrt{2\pi})^{\alpha}\sigma_1}\exp(\frac{-x^2+2x\mu_1-\mu_1^2}{2\sigma_1^2} + \frac{-x^2(1-\alpha)+2x\mu_2(\alpha-1)-\mu_2^2(\alpha-1)}{2\sigma_2^2}).$$

Define $a = \frac{1}{2\sigma_1^2} + \frac{\alpha-1}{2\sigma_2^2}$, $b = \frac{\mu_1}{\sigma_1^2} + \frac{(\alpha-1)\mu_2}{\sigma_2^2}$, $c = \frac{-\mu_1^2}{2\sigma_1^2} - \frac{(\alpha-1)\mu_2^2}{2\sigma_2^2}$.

Note that a will be positive as long as $\alpha \geq 1$, since $\sigma_1, \sigma_2 > 0$.

Therefore,

$$\begin{aligned}\int_R f_1(x)f_2(x)^{\alpha-1}dx &= \frac{\sigma_2}{(\sigma_2\sqrt{2\pi})^\alpha\sigma_1}\sqrt{\frac{2\pi\sigma_1^2\sigma_2^2}{\sigma_2^2+\sigma_1^2(\alpha-1)}}\\ &\times\exp\left(\frac{(\frac{\mu_1}{\sigma_1^2}+\frac{(\alpha-1)\mu_2}{\sigma_2^2})^2}{4(\frac{1}{2\sigma_1^2}+\frac{\alpha-1}{2\sigma_2^2})}-\frac{\mu_1^2}{2\sigma_1^2}-\frac{(\alpha-1)\mu_2^2}{2\sigma_2^2}\right)\\ &=\frac{\sigma_2}{(\sigma_2\sqrt{2\pi})^{\alpha-1}}\sqrt{\frac{1}{\sigma_2^2+\sigma_1^2(\alpha-1)}}\\ &\times\exp\left(\frac{1}{2\sigma_1^2\sigma_2^2}\Big(\frac{(\mu_1\sigma_2^2+(\alpha-1)\mu_2\sigma_1^2)^2}{\sigma_2^2+(\alpha-1)\sigma_1^2}-\mu_1^2\sigma_2^2-(\alpha-1)\mu_2^2\sigma_1^2\Big)\right).\end{aligned}$$

Thus,

$$\begin{aligned}(1-\alpha)h_\alpha(f_1;f_2) &= \ln\sigma_2-(\alpha-1)\ln\sigma_2\sqrt{2\pi}-\frac{\ln\left(\sigma_2^2+\sigma_1^2\left(\alpha-1\right)\right)}{2}\\ &\quad+\frac{1}{2\sigma_1^2\sigma_2^2}\left(\frac{\left(\mu_1\sigma_2^2+\left(\alpha-1\right)\mu_2\sigma_1^2\right)^2}{\sigma_2^2+\left(\alpha-1\right)\sigma_1^2}-\mu_1^2\sigma_2^2-\left(\alpha-1\right)\mu_2^2\sigma_1^2\right)\\ &=\ln\sigma_2+(1-\alpha)\ln\sigma_2\sqrt{2\pi}-\frac{\ln\left(\sigma_2^2+\sigma_1^2\left(\alpha-1\right)\right)}{2}\\ &\quad+\frac{1}{2\sigma_1^2\sigma_2^2}\left(\frac{2\mu_1\sigma_2^2\left(\alpha-1\right)\mu_2\sigma_1^2-\left(\alpha-1\right)\sigma_1^2\sigma_2^2\left(\mu_2^2+\mu_1^2\right)}{\sigma_2^2+\left(\alpha-1\right)\sigma_1^2}\right).\end{aligned}$$

Hence,

$$\begin{aligned}h_\alpha(f_1;f_2) &= \frac{1}{1-\alpha}\left(\ln\sigma_2-\frac{\ln\left(\sigma_2^2+\sigma_1^2\left(\alpha-1\right)\right)}{2}\right)+\ln\sigma_2\sqrt{2\pi}\\ &\quad+\frac{\left(\mu_1-\mu_2\right)^2}{2\sigma_2^2+2\left(\alpha-1\right)\sigma_1^2}\\ &=-\frac{\ln\left(\sigma_2^2+\sigma_1^2\left(\alpha-1\right)\right)}{2-2\alpha}+\ln\sigma_2\sqrt{2\pi}+\frac{\ln\sigma_2}{1-\alpha}+\frac{\left(\mu_1-\mu_2\right)^2}{2\sigma_2^2+2\left(\alpha-1\right)\sigma_1^2}.\end{aligned}$$

A.1.2 Natural Rényi Differential Cross-Entropy

We have

$$\begin{aligned} h_\alpha(f_1) &= \frac{1}{1-\alpha}(\ln\sigma_1 + \frac{\ln 2\pi}{2} - \frac{\ln\alpha}{2} - \alpha\ln\sigma_1 - \alpha\frac{\ln 2\pi}{2}) \\ &= \ln\sigma_1 + \frac{\ln 2\pi}{2} - \frac{\ln\alpha}{2-2\alpha} \end{aligned}$$

and

$$D_\alpha(f_1\|f_2) = \ln\frac{\sigma_2}{\sigma_1} + \frac{\ln\frac{\sigma_2^2}{\sigma_\alpha^2}}{2\alpha-2} + \frac{\alpha(\mu_1-\mu_2)^2}{2\sigma_\alpha^2}$$

where $\sigma_\alpha^2 = \alpha\sigma_1^2 + (1-\alpha)\sigma_2^2$. Therefore,

$$\tilde{h}_\alpha(f_1;f_2) = \ln\sigma_2 + \frac{\ln 2\pi}{2} - \frac{\ln\alpha}{2-2\alpha} + \frac{\ln\frac{\sigma_2^2}{\sigma_\alpha^2}}{2\alpha-2} + \frac{\alpha(\mu_1-\mu_2)^2}{2\sigma_\alpha^2}.$$

A.2 Multivariate Gaussian

A.2.1 Rényi Differential Cross-Entropy

The following depends heavily on material in [7], specifically the derivation of the Rényi divergence. For $x \in \mathbb{R}^n$,

$$\begin{aligned} f_1(x)f_2(x)^{\alpha-1} &= \left((2\pi)^n\left|\Sigma_1\right|\right)^{\frac{-1}{2}}\left((2\pi)^n\left|\Sigma_2\right|\right)^{\frac{1-\alpha}{2}} \\ &\quad\times\exp\left(-.5\left(\left(x-\boldsymbol{\mu_1}\right)^T\Sigma_1^{-1}\left(x-\boldsymbol{\mu_1}\right)+\left(\alpha-1\right)\left(x-\boldsymbol{\mu_2}\right)^T\Sigma_2^{-1}\left(x-\boldsymbol{\mu_2}\right)\right)\right) \end{aligned}$$

By being careful to note where there is an α that appears due to the fact that f_1 is raised to the power of α, we can rewrite the term in the exponent (ignoring the -0.5

term for the moment, though taking care to remember to add it in later) as

$$
\begin{aligned}
&x^T\left(\Sigma_1^{-1}+(\alpha-1)\,\Sigma_2^{-1}\right)x - x^T\left(\Sigma_1^{-1}\boldsymbol{\mu_1}+(\alpha-1)\,\Sigma_2^{-1}\boldsymbol{\mu_2}\right)\\
&-\left(\boldsymbol{\mu_1}\Sigma_1^{-1}+(\alpha-1)\,\mu_2^T\Sigma_2^{-1}\right)x+\mu_1^T\Sigma_1^{-1}\boldsymbol{\mu_1}+(\alpha-1)\,\mu_2^T\Sigma_2^{-1}\boldsymbol{\mu_2}\\
&=x^TAx-2x^Tb+c,
\end{aligned}
$$

where

$$
\begin{aligned}
A&=\Sigma_1^{-1}+(\alpha-1)\,\Sigma_2^{-1},\\
b&=\Sigma_1^{-1}\boldsymbol{\mu_1}+(\alpha-1)\,\Sigma_2^{-1}\boldsymbol{\mu_2},
\end{aligned}
$$

and

$$
c=\mu_1^T\Sigma_1^{-1}\boldsymbol{\mu_1}+(\alpha-1)\,\mu_2^T\Sigma_2^{-1}\boldsymbol{\mu_2}.
$$

Note that if A is symmetric and positive-definite we can define ν and d in the same way [7] does, i.e., $\nu=A^{-1}b$ and $d=c-b^T\nu$.

This allows us to rewrite $f_1(x)f_2(x)^{\alpha-1}$ as

$$
\begin{aligned}
&\left((2\pi)^n\left|\Sigma_1\right|\right)^{\frac{-1}{2}}\left((2\pi)^n\left|\Sigma_2\right|\right)^{\frac{1-\alpha}{2}}e^{\frac{-d}{2}}\\
&\qquad\times\left((2\pi)^n\left|\Sigma_1\right|\right)^{\frac{-1}{2}}\left((2\pi)^n\left|\Sigma_2\right|\right)^{\frac{1-\alpha}{2}}e^{\frac{-d}{2}}\exp\left(-0.5\,(x-\nu)^T A\,(x-\nu)\right)\\
&=\left(|A||\Sigma_1|\right)^{\frac{-1}{2}}\left((2\pi)^n\left|\Sigma_2\right|\right)^{\frac{1-\alpha}{2}}e^{\frac{-d}{2}}\left(|A|^{-1}\,(2\pi)^n\right)^{\frac{-1}{2}}\exp\left(-0.5\,(x-\nu)^T A\,(x-\nu)\right)\\
&=\left(|A||\Sigma_1|\right)^{\frac{-1}{2}}\left((2\pi)^n\left|\Sigma_2\right|\right)^{\frac{1-\alpha}{2}}e^{\frac{-d}{2}}f\,(x)),
\end{aligned}
$$

where $f\,(x)$ is the pdf of a multivariate Gaussian with mean ν and covariance matrix A^{-1}. As the integral of $f(x)$ over $\mathbb{R}$ will be 1, and none of the other terms depend on x, we conclude that

$$\int_{R^n} f_1(x) f_2(x)^{\alpha-1} dx = (|A||\Sigma_1|)^{\frac{-1}{2}} \left((2\pi)^n |\Sigma_2|\right)^{\frac{1-\alpha}{2}} e^{\frac{-d}{2}}.$$

Hence,

$$\begin{aligned} h_\alpha(f_1; f_2) &= \frac{1}{1-\alpha}\left(\frac{-\ln|A||\Sigma_1|}{2} + \frac{(1-\alpha)\ln(2\pi)^n|\Sigma_2|}{2} - \frac{d}{2}\right) \\ &= \frac{1}{2-2\alpha}\left(-\ln|A||\Sigma_1| + (1-\alpha)\ln(2\pi)^n|\Sigma_2| - d\right). \end{aligned}$$

A.2.2 Natural Rényi Differential Cross-Entropy

Let $f_2 = f_1$ in the above equation, i.e., $\Sigma_1 = \Sigma_2 \coloneqq \Sigma$, and $\boldsymbol{\mu_1} = \boldsymbol{\mu_2} \coloneqq \mu$. Thus, $A = \Sigma^{-1} + (\alpha-1)\Sigma^{-1} = \alpha\Sigma^{-1}$, $b = \Sigma^{-1}\mu + (\alpha-1)\Sigma^{-1}\mu = \alpha\Sigma^{-1}\mu$, and $c = \mu^T\Sigma^{-1}\mu + (\alpha-1)\mu^T\Sigma^{-1} = \alpha\mu^T\Sigma^{-1}\mu$.

Additionally,

$$d = \alpha\mu^T\Sigma^{-1}\mu - \left(\alpha\Sigma^{-1}\mu\right)^T\left(\alpha\Sigma^{-1}\right)^{-1}\alpha\Sigma^{-1}\mu = c - c = 0.$$

Hence,

$$\begin{aligned} \tilde{h}_\alpha(f_1) &= \frac{1}{2-2\alpha}\left(-\ln|\alpha\Sigma^{-1}||\Sigma| + (1-\alpha)\ln(2\pi)^n|\Sigma| - 0\right) \\ &= \frac{1}{2-2\alpha}\left(-\ln|\alpha| + (1-\alpha)\ln(2\pi)^n|\Sigma|\right), \end{aligned}$$

and thus

$$\tilde{h}_\alpha(f_1; f_2) = \frac{1}{2-2\alpha}\left(-\ln|\alpha| + (1-\alpha)\ln(2\pi)^n|\Sigma_1| + (1-\alpha)\ln\frac{|\Sigma_1|}{|\Sigma_2|} + \ln|A||\Sigma_1| + d\right)$$

$$= \frac{1}{2-2\alpha}\left(-\ln|\alpha| + \ln|A||\Sigma_1| + d\right) + \frac{1}{2}\ln\frac{(2\pi)^n|\Sigma_1|^2}{|\Sigma_2|}.$$

A.3 Exponential distribution

A.3.1 Rényi Differential Cross-Entropy

For $x \in \mathbb{R}^+$, we have

$$f_1(x)f_2(x)^{1-\alpha} = \lambda_1\lambda_2^{\alpha-1}e^{-\lambda_1 x}e^{-\lambda_2 x(\alpha-1)} = \lambda_1\lambda_2^{\alpha-1}e^{-(\lambda_1+\lambda_2(\alpha-1))x}.$$

Thus,

$$\int_0^\infty f_1(x)f_2(x)^{1-\alpha}dx = \frac{\lambda_1\lambda_2^{\alpha-1}}{(\lambda_1+\lambda_2(\alpha-1))}$$

and

$$h_\alpha(f_1;f_2) = \frac{\ln\lambda_1 - \ln(\lambda_1+\lambda_2(\alpha-1))}{1-\alpha} - \ln\lambda_2.$$

A.3.2 Natural Rényi Differential Cross-Entropy

As shown in [17] , the Rényi differential entropy of an exponential distribution, f_1, is given by

$$\begin{aligned} h_\alpha(f_1) = -\ln(\alpha\lambda_1) - \frac{\alpha}{1-\alpha}\ln(\alpha) &= \frac{(\alpha-1)(\ln\alpha+\ln\lambda_1) - \alpha\ln\alpha}{1-\alpha} \\ &= \frac{(\alpha-1)(\ln\alpha+\ln\lambda_1) - \alpha\ln\alpha}{1-\alpha} \\ &= \frac{-\ln\alpha}{1-\alpha} - \ln\lambda_1. \end{aligned}$$

Similarily, as shown in [7],

$$D_\alpha(f_1\|f_2) = \ln\frac{\lambda_1}{\lambda_2} + \frac{1}{1-\alpha}\ln\frac{\lambda_1}{\lambda_\alpha}.$$

Thus,

$$\tilde{h}_\alpha(f_1; f_2) = \frac{-\ln \alpha}{1-\alpha} - \ln \lambda_1 + \ln \frac{\lambda_1}{\lambda_2} + \frac{1}{1-\alpha} \ln \frac{\lambda_1}{\lambda_\alpha}$$
$$= -\ln \lambda_2 + \frac{1}{1-\alpha} \ln \frac{\lambda_1}{\alpha \lambda_\alpha}.$$

A.4 Beta Distribution

Rényi Differential Cross-Entropy

For $x \in \mathbb{S}$,

$$f_1(x) f_2(x)^{\alpha-1} = \frac{x^{a_1-1}(1-x)^{b_1-1}}{B(a_1,b_1)} \left(\frac{x^{a_2-1}(1-x)^{b_2-1}}{B(a_2,b_2)} \right)^{\alpha-1}$$
$$= \frac{B(a_2,b_2)}{B(a_1,b_1)} \frac{x^{a_h-1}(1-x)^{b_h-1}}{B(a_2,b_2)^\alpha},$$

where

$$a_h = a_1 + (a_2-1)(\alpha-1)$$

and

$$b_h = b_1 + (b_2-1)(\alpha-1) = \frac{B(a_2,b_2)}{B(a_1,b_1)} \frac{B(a_h,b_h)}{B(a_2,b_2)^\alpha} \frac{x^{a_h-1}(1-x)^{b_h-1}}{B(a_h,b_h)}.$$

Therefore,

$$\int_R f_1(x) f_2(x)^{\alpha-1} dx = \frac{B(a_2,b_2)}{B(a_1,b_1)} \frac{B(a_h,b_h)}{B(a_2,b_2)^\alpha} \int_0^1 \frac{x^{a_h-1}(1-x)^{b_h-1}}{B(a_h,b_h)} dx$$
$$= \frac{B(a_2,b_2)}{B(a_1,b_1)} \frac{B(a_h,b_h)}{B(a_2,b_2)^\alpha}.$$

Hence,

$$h_\alpha(f_1; f_2) = \ln B(a_2, b_2) + \frac{1}{1-\alpha} \ln \frac{B(a_h, b_h)}{B(a_1, b_1)}.$$

Natural Rényi Differential Cross-Entropy

By summing the expressions for $h_\alpha(f_1)$ and $D_\alpha(f_1 \| f_2)$ found in [17] and [7] respectively, we obtain

$$\begin{aligned}
\tilde{h}_\alpha (f_1; f_2) &= \frac{1}{1-\alpha} \ln \frac{B(\alpha(a_1-1)+1, \alpha(b_1-1)+1)}{B(a_1, b_1)^\alpha} \\
&\qquad + \ln \frac{B(a_2, b_2)}{B(a_1, b_1)} + \frac{1}{1-\alpha} \ln \frac{B(a_1, b_1)}{B(a_\alpha, b_\alpha)} \\
&= \frac{1}{1-\alpha} \ln \frac{B(\alpha(a_1-1)+1, \alpha(b_1-1)+1)}{B(a_\alpha, b_\alpha)} \\
&\qquad + \ln \frac{B(a_2, b_2)}{B(a_1, b_1)} + \frac{1}{1-\alpha} \ln \frac{B(a_1, b_1)}{B(a_1, b_1)^\alpha} \\
&= \frac{1}{1-\alpha} \ln \frac{B(\alpha(a_1-1)+1, \alpha(b_1-1)+1)}{B(a_\alpha, b_\alpha)} + \ln B(a_2, b_2)
\end{aligned}$$

where $a_\alpha = \alpha a_1 + (1-\alpha) a_2$ and $b_\alpha = \alpha b_1 + (1-\alpha) b_2$.

A.5 Gamma

For a *Gamma* (k, θ) distribution,

$$\eta[1] = k - 1,$$

$$\eta[2] = -\theta^{-1},$$

and

$$\begin{aligned}
A(\eta) &= (\eta[1]+1) \ln(-\eta[2]) - \ln \Gamma(\eta[1]+1) \\
&= k \ln\left(\theta^{-1}\right) - \ln \Gamma(k).
\end{aligned}$$

A.5.1 Rényi Differential Cross-Entropy

We have

$$\begin{aligned} h_\alpha(f_1; f_2) &= \frac{k_1 \ln\left(\theta_1^{-1}\right) - \ln\Gamma(k_1) - k_h \ln\left(\theta^{-1*}\right) + \ln\Gamma(k_h)}{1-\alpha} - k_2 \ln\left(\theta_2^{-1}\right) + \ln\Gamma(k_2) \\ &= \frac{-k_1 \ln(\theta_1) - \ln\Gamma(k_1) - k_h \ln\left(\theta^{-1*}\right) + \ln\Gamma(k_h)}{1-\alpha} + k_2 \ln(\theta_2) + \ln\Gamma(k_2) \end{aligned}$$

where $\theta^{-1*} = \theta_1^{-1} + (\alpha - 1)\,\theta_2^{-1}$ and $k_h = k_1 + (\alpha - 1)(k_2 - 1)$.

A.5.2 Natural Rényi Differential Cross-Entropy

We have

$$\begin{aligned} \tilde{h}_\alpha(f_1; f_2) &= \frac{k_\alpha \ln\left(\theta_\alpha^{-1}\right) - \ln\Gamma(k_\alpha) - \alpha^2 k_1 \ln\left(\theta_1^{-1}\right) + \ln\Gamma(\alpha k_1)}{1-\alpha} - k_2 \ln\left(\theta_2^{-1}\right) + \ln\Gamma(k_2) \\ &= \frac{-k_\alpha \ln\left(\theta_\alpha^{-1}\right) - \ln\Gamma(k_\alpha) + \alpha^2 k_1 \ln(\theta_1) + \ln\Gamma(k_1)}{1-\alpha} + k_2 \ln(\theta_2) + \ln\Gamma(k_2) \end{aligned}$$

where $\theta_\alpha^{-1} = \alpha\theta_1^{-1} + (1-\alpha)\,\theta_2^{-1}$ and $k_\alpha = \alpha k_1 + (1-\alpha)\,k_2$.

A.5.3 Chi-Squared Distribution

The chi-squared distribution is a special form of the Gamma distribution, in that $\chi^2(\nu) \sim \Gamma(\nu/2, 2)$.

Rényi Differential Cross-Entropy

We have

$$h_\alpha(f_1; f_2) = \frac{-\frac{\nu_1}{2}\ln(2) - \ln\Gamma\left(\frac{\nu_1}{2}\right) + \left(\frac{\nu_1}{2} + \frac{(\alpha-1)(\nu_2-2)}{2}\right)\ln(2\alpha) + \ln\Gamma\left(\frac{\nu_h}{2}\right)}{1-\alpha}$$

$$
\begin{aligned}
&+ \frac{\nu_2}{2} \ln(2) + \ln\Gamma\left(\frac{\nu_2}{2}\right) \\
&= \frac{\frac{\nu_1}{2}\ln(\alpha) - \ln\Gamma\left(\frac{\nu_1}{2}\right) + \ln\Gamma\left(\frac{\nu_h}{2}\right)}{1-\alpha} - \frac{\nu_2 - 2}{2}\ln(2\alpha) + \frac{\nu_2}{2}\ln(2) + \ln\Gamma\left(\frac{\nu_2}{2}\right) \\
&= \frac{\frac{\nu_1}{2}\ln(\alpha) - \ln\Gamma\left(\frac{\nu_1}{2}\right) + \ln\Gamma\left(\frac{\nu_h}{2}\right)}{1-\alpha} + \frac{2-\nu_2}{2}\ln(\alpha) + \ln 2\Gamma\left(\frac{\nu_2}{2}\right)
\end{aligned}
$$

where $\nu_h = \nu_1 + (\alpha - 1)(\nu_2 - 2)$.

Natural Rényi Differential Cross-Entropy

We have

$$
\tilde{h}_\alpha(f_1; f_2) = \frac{-\ln\Gamma\left(\frac{\nu_\alpha}{2}\right) + \alpha\ln\Gamma\left(\frac{\nu_1}{2}\right)}{1-\alpha}) + \ln\Gamma\left(\frac{\nu_2}{2}\right).
$$

A.6 Pareto, Constant Minimum

A.6.1 Rényi Differential Cross-Entropy

For the Pareto distribution:

$$
\eta = -\lambda - 1,
$$

and

$$
A(\eta) = \ln(-1-\eta) - (1+\eta)\ln m.
$$

Hence,

$$
\begin{aligned}
h_\alpha(f_1; f_2) = &\frac{\ln(-1-\eta_1) - (1+\eta_1)\ln m - \ln(-1-\eta_h) + (1+\eta_h)\ln m}{1-\alpha} \\
&- \ln(-1-\eta_2) + (1+\eta_2)\ln m
\end{aligned}
$$

$$= \frac{1}{1-\alpha}\Bigg(\ln\lambda_1 + \lambda_1 \ln m - \ln\left(-1 - \lambda_2 + \lambda_1 + \alpha\left(\lambda_2 + 1\right)\right)$$
$$+ \left(1 + \lambda_2 - \lambda_1 - \alpha\left(\lambda_2 + 1\right)\right)\ln m\Bigg) - \ln\lambda_2 - \lambda_2 \ln m$$
$$= \frac{\ln\lambda_1 - \ln\lambda_h}{1-\alpha} - \ln m\lambda_2,$$

where $\lambda_h = \lambda_1 + (\alpha - 1)(\lambda_2 + 1)$.

A.6.2 Natural Rényi Differential Cross-Entropy

Define $\lambda_\alpha = \alpha\lambda_1 + (1-\alpha)\lambda_2$. Then,

$$\tilde{h}_\alpha(f_1; f_2) = \frac{\ln(-1-\eta_\alpha) - (1+\eta_\alpha)\ln m - \ln(-1-\alpha\eta_1) + (1+\alpha\eta_1)\ln m}{1-\alpha}$$
$$- \ln(-1-\eta_2) + (1+\eta_2)\ln m$$
$$= \frac{\ln(\lambda_\alpha) - (\lambda_\alpha)\ln m - \ln(1-\alpha(\lambda_1 - 1)) - (1-\alpha(\lambda_1-1))\ln m}{1-\alpha}$$
$$- \ln(\lambda_2) - \lambda_2 \ln m$$
$$= \frac{\ln(\lambda_\alpha) - \ln(1-\alpha(\lambda_1-1))}{1-\alpha} - \ln(\lambda_2) - \ln m.$$

A.7 Laplace, Constant Mean

Here,

$$\eta = -\frac{1}{b},$$

and

$$A(\eta) = \ln\frac{-\eta}{2}.$$

A.7.1 Rényi Differential Cross-Entropy

We have

$$
\begin{aligned}
h_\alpha\left(f_1; f_2\right) &= \frac{\ln \frac{-\eta_1}{2} - \ln \frac{-\eta_h}{2}}{1-\alpha} - \ln \frac{-\eta_2}{2} \\
&= \frac{-\ln 2b_1 - \ln b_h + \ln b_1 + \ln b_2}{1-\alpha} + \ln 2b_2 \\
&= \frac{-\ln 2 - \ln b_h + \ln b_2}{1-\alpha} + \ln 2b_2
\end{aligned}
$$

where $b_h = b_2 + (1-\alpha)b_1$.

A.7.2 Natural Rényi Differential Cross-Entropy

We have

$$
\begin{aligned}
\tilde{h}_\alpha\left(f_1; f_2\right) &= \frac{\ln \frac{-\eta_\alpha}{2} - \ln \frac{-\alpha\eta_1}{2}}{1-\alpha} - \ln \frac{-\eta_2}{2} \\
&= \frac{\ln b_\alpha + \ln \alpha b_1}{1-\alpha} + \ln 2b_2
\end{aligned}
$$

where $b_\alpha = \frac{\alpha}{b_1} + \frac{1-\alpha}{b_2}$.

A.8 Gumbel, Fixed Scale Parameter

Here, $\eta = e^{-\mu/\beta}$ and $A(\eta) = \ln \eta$. Moreover, $b(x) = \beta^{-1}e^{-x/\beta}$.

A.8.1 Rényi Cross-Entropy

$$
E_h = \beta^{-1}\mathbb{E}_{f_h}(e^{-x/\beta}).
$$

One can therefore use the Moment-Generating Function of f_h to derive E_h:

$$\begin{aligned}E_h &= \beta^{-1}\Gamma\left(1-\frac{\beta(\alpha-1)}{\beta}\right)e^{-\beta\ln\eta_h(\alpha-1)/\beta}\\ &= \beta^{-1}\Gamma(2-\alpha)\eta_h^{1-\alpha}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}h_\alpha(f_1;f_2) &= \frac{\ln\eta_1-\ln\eta_h+\ln\beta^{-1}\Gamma(2-\alpha)\eta_h^{1-\alpha}}{1-\alpha}-\ln\eta_2\\ &= \frac{-\mu_1/\beta-\alpha\ln\eta_h+\ln\beta^{-1}\Gamma(2-\alpha)}{1-\alpha}+\mu_2/\beta,\end{aligned}$$

where $\eta_h = e^{-\mu_1/\beta}+(\alpha-1)e^{-\mu_2/\beta}$.

A.8.2 Natural Rényi Cross-Entropy

We have

$$\begin{aligned}E_\alpha &= \beta^{-1}\mathbb{E}_{f_\alpha}(e^{-x/\beta})\\ &= \beta^{-1}\Gamma(1-\frac{\beta(\alpha-1)}{\beta})e^{\alpha\mu_1(\alpha-1)/\beta}.\end{aligned}$$

Therefore,

$$\begin{aligned}\tilde{h}_\alpha(f_1;f_2) &= \frac{\ln\eta_\alpha-\ln\alpha\eta_1+\ln\beta^{-1}\Gamma(2-\alpha)e^{\alpha\mu_1(\alpha-1)/\beta}}{1-\alpha}-\ln\eta_2\\ &= \frac{\ln\frac{\Gamma(2-\alpha)\eta_\alpha}{\alpha\beta}+\frac{\mu_1}{\beta}}{1-\alpha}+\frac{\boldsymbol{\mu_2}+\alpha\mu_1}{\beta}.\end{aligned}$$

A.9 Rayleigh

Here, $b(x) = x$, $\eta = \frac{-1}{(2\sigma^2)}$, and $A(\eta) \ln -2\eta$. Thus, $E_h = \mathbb{E}_{f_h}(x^{\alpha-1})$.

A.9.1 Rényi Differential Cross-Entropy

Using the formula for the raw moments of the Rayleigh distribution, we obtain that $E_h = (-\eta_h)^{1-\alpha}\Gamma(1+\frac{\alpha-1}{2})$. Therefore,

$$\begin{aligned}
h_\alpha(f_1; f_2) &= \frac{\ln -2\eta_1 - \ln -2\eta_h + \ln(-\eta_h)^{1-\alpha}\Gamma(1+\frac{\alpha-1}{2})}{1-\alpha} - \ln -2\eta_2 \\
&= \frac{-\ln\sigma_1^2 - \ln(\frac{1}{\sigma_1^2} + \frac{\alpha-1}{\sigma_2^2}) + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln\sigma_2^2 + \ln\eta_h \\
&= \frac{\ln\sigma_1^2 - \ln\sigma_h^2 + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln 2\sigma_2^2 + \ln\sigma_h^2 \\
&= \frac{\ln\sigma_1^2 - \alpha\ln\sigma_h^2 + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \ln 2\sigma_2^2,
\end{aligned}$$

where $\sigma_h^2 = \sigma_1^{-2} + (\alpha-1)\sigma_2^{-2}$.

A.9.2 Natural Rényi Differential Cross-Entropy

Here,

$$E_\alpha = \sigma_1^{1-\alpha}\left(\frac{2}{\alpha}\right)^{\frac{1-\alpha}{2}}\Gamma\left(\frac{1-\alpha}{2}\right).$$

Hence,

$$\begin{aligned}
\tilde{h}_\alpha(f_1; f_2) &= \frac{\ln -2\eta_\alpha - \ln -2\alpha\eta_1 + \ln\sigma_1^{1-\alpha}(\frac{2}{\alpha})^{\frac{1-\alpha}{2}}\Gamma(\frac{1-\alpha}{2})}{1-\alpha} - \ln -2\eta_2 \\
&= \frac{\ln\sigma_1^2\sigma_\alpha - \ln\alpha + \ln\Gamma(\frac{1-\alpha}{2})}{1-\alpha} + \frac{1}{2}\ln\frac{2\sigma_1^4\sigma_2^4}{\alpha}.
\end{aligned}$$

A.10 Half-Normal

The Half-Normal distribution's pdf is similar to that of the normal distribution, albeit with a zero mean and $b(x)$ multiplied by 2. This means we can use what we found above, set $\mu_1 = \mu_2 = 0$, and add $0.5 \ln 2\pi$:

$$\begin{aligned} h_\alpha(f_1; f_2) &= -\frac{\ln\left(\sigma_2^2 + \sigma_1^2(\alpha-1)\right)}{2-2\alpha} + \ln \sigma_2 \sqrt{2\pi} - \frac{\ln \sigma_2}{1-\alpha} \\ &\quad + \frac{(0-0)^2}{2\sigma_2^2 + 2(\alpha-1)\sigma_1^2} + 0.5 \ln 2\pi \\ &= -\frac{\ln\left(\sigma_2^2 + \sigma_1^2(\alpha-1)\right)}{2-2\alpha} + \ln \sigma_2 - \frac{\ln \sigma_2}{1-\alpha} + \ln 2\pi \\ &= -\frac{\ln\left(\sigma_2^2 + \sigma_1^2(\alpha-1)\right) + \alpha \ln \sigma_2^2}{2-2\alpha} + \ln 2\pi. \end{aligned}$$

A.11 Chi Distribution

Here $\eta = \left[\frac{-1}{2\sigma^2}, k-1\right]$, and $A(\eta) = \frac{(\eta[2]+1)\ln -2\eta[1] - (\eta[2]-1)\ln 2}{2} - \ln\Gamma(\frac{\eta[2]+1}{2})$.

A.11.1 Rényi Differential Cross-Entropy

We have

$$\begin{aligned} h_\alpha(f_1; f_2) &= \frac{1}{1-\alpha}\left(\frac{(\eta_1[2]+1)\ln -2\eta_1[1] - (\eta_1[2]-1)\ln 2}{2} - \ln\Gamma\left(\frac{\eta_1[2]+1}{2}\right)\right. \\ &\quad \left. + \frac{(\eta_h[2]-1)\ln 2 - (\eta_h[2]+1)\ln -2\eta_h[1]}{2} + \ln\Gamma\left(\frac{\eta_h[2]+1}{2}\right)\right) \\ &\quad + \frac{(\eta_2[2]-1)\ln 2 - (\eta_2[2]+1)\ln -2\eta_2[1]}{2} + \ln\Gamma\left(\frac{\eta_2[2]+1}{2}\right) \\ &= \frac{1}{1-\alpha}\left(\frac{-k_1 \ln \sigma_1^2 - (k_1-2)\ln 2}{2} - \ln\Gamma\left(\frac{k_1}{2}\right)\right. \\ &\quad + \frac{(k_1 + (\alpha-1)k_2 - \alpha - 1)\ln 2 - (k_1 + (\alpha-1)(k_2-1))\ln \sigma_h^2}{2} \\ &\quad \left. + \ln\Gamma\left(\frac{k_1 + (\alpha-1)(k_2-1)}{2}\right)\right) \end{aligned}$$

$$+\frac{(k_2-2)\ln 2+k_2\ln\sigma_2^2}{2}+\ln\Gamma\left(\frac{k_2}{2}\right)$$
$$=\frac{1}{1-\alpha}\frac{-k_1\ln\sigma_1^2\sigma_h^2}{2}-\ln\Gamma\left(\frac{k_1}{2}\right)+\ln\Gamma\left(\frac{k_1+(\alpha-1)(k_2-1)}{2}\right)$$
$$-\frac{\ln 2\sigma_h^2+k_2\ln\sigma_2^2\sigma_h^2}{2}+\ln\Gamma\left(\frac{k_2}{2}\right)$$

where $\sigma_h^2=\frac{1}{\sigma_1^2}+\frac{\alpha-1}{\sigma_2^2}$.

A.11.2 Natural Rényi Differential Cross-Entropy

We have

$$\tilde{h}_\alpha\left(f_1;f_2\right)=\frac{\frac{k_\alpha\ln\sigma_\alpha-(\alpha(k_1-1)+1)\ln\frac{\alpha}{\sigma_1^2}}{2}-\ln\Gamma(\frac{k_\alpha}{2})+\ln\Gamma\left(\frac{\alpha(k_1-1)+1}{2}\right)}{1-\alpha}$$
$$+\frac{-\ln 2+k_2\ln\sigma_2^2}{2}+\ln\Gamma\left(\frac{k_2}{2}\right)$$

where $\sigma_\alpha=\frac{\alpha}{\sigma_1^2}+\frac{1-\alpha}{\sigma_2^2}$ and $k_\alpha=\alpha k_1+(1-\alpha)k_2$.

A.11.3 Non-scaled case

In the non-scaled case, $\sigma_1^2=\sigma_2^2=1$, and thus $\sigma_h^2=\alpha$. Therefore,

$$h_\alpha\left(f_1;f_2\right)=\frac{\frac{-k_1\ln\alpha}{2}-\ln\Gamma\left(\frac{k_1}{2}\right)+\ln\Gamma(\frac{k_1+(\alpha-1)(k_2-1)}{2})}{1-\alpha}$$
$$+\frac{-\ln 2+(k_2-1)\ln\alpha}{2}+\ln\Gamma\left(\frac{k_2}{2}\right).$$

Furthermore,

$$\tilde{h}_\alpha\left(f_1;f_2\right)=\frac{-\ln\Gamma(\frac{k_\alpha}{2})-\frac{\alpha k_1\ln\alpha}{2}+\ln\Gamma(\frac{\alpha(k_1-1)+1}{2})}{1-\alpha}+\frac{-\ln 2\alpha}{2}+\ln\Gamma\left(\frac{k_2}{2}\right).$$

A.11.4 Maxwell-Boltzmann distribution

The Maxwell-Boltzmann distribution is a special case of the chi distribution where $k_1 = k_2 = 3$. Therefore,

$$
\begin{aligned}
h_\alpha(f_1; f_2) &= \frac{\frac{-3\ln\sigma_1^2\sigma_h^2}{2} - \ln\Gamma(\frac{3}{2}) + \ln\Gamma\left(\frac{3+(\alpha-1)(3-1)}{2}\right)}{1-\alpha} \\
&\quad + \frac{-\ln 2 + 3\ln\sigma_2^2\sigma_h^2 - \ln\sigma_h^2}{2} + \ln\Gamma\left(\frac{3}{2}\right) \\
&= \frac{\frac{-3\ln\sigma_1^2\sigma_h^2}{2} - \alpha(\ln\sqrt{\pi} - \ln 2) + \ln\Gamma(\alpha + \frac{1}{2})}{1-\alpha} \\
&\quad + \frac{-\ln 2 + 3\ln\sigma_2^2\sigma_h^2 - \ln\sigma_h^2}{2} \\
&= \frac{\frac{-3\ln\sigma_1^2\sigma_h^2}{2} - \alpha(\ln\sqrt{\pi} - \ln 2) + \ln 2^{1-2\alpha}\sqrt{\pi}\Gamma(2\alpha) - \ln\Gamma(\alpha)}{1-\alpha} \\
&\quad + \frac{-\ln 2 + 3\ln\sigma_2^2\sigma_h^2 - \ln\sigma_h^2}{2} \\
&= \frac{\frac{-3\ln\sigma_1^2\sigma_h^2}{2} + \ln\Gamma(2\alpha) - \ln\Gamma(\alpha)}{1-\alpha} + \frac{\ln 2\pi + 3\ln\sigma_2^2\sigma_h^2 - \ln\sigma_h^2}{2}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\tilde{h}_\alpha(f_1; f_2) = \frac{1}{1-\alpha}&\left(\frac{3\ln\sigma_\alpha - (2\alpha+1)\ln\frac{\alpha}{\sigma_1^2}}{2} - \alpha\ln\frac{\sqrt{\pi}}{2} + \ln\Gamma\left(\alpha + \frac{1}{2}\right)\right) \\
&+ \frac{3\ln\sigma_2^2 - \ln 2}{2}.
\end{aligned}
$$

A.12 Uniform

Let $a_M = \max(a_1, a_2)$, and $b_m = \min(b_1, b_2)$. Then,

$$
\begin{aligned}
\int_R f_1(x) f_2(x)^{\alpha-1} dx &= \int_{a_M}^{b_m} \frac{(b_2 - a_2)^{1-\alpha}}{b_1 - a_1} dx \\
&= \frac{(b_m - a_M)(b_2 - a_2)^{1-\alpha}}{b_1 - a_1}.
\end{aligned}
$$

Therefore,
$$h_\alpha(f_1; f_2) = \frac{1}{1-\alpha} \ln \frac{(b_m - a_M)}{b_1 - a_1} + \ln(b_2 - a_2).$$

A.13 Log-Normal Distribution

The only difference between the pdf of a Normal distribution and a log-normal distribution is $b(x)$ and T(x). Since T(x) is not a factor of the cross-entropy the only difference between the cross-entropy of the normal distribution and the log-normal distribution will be the E_h term.

A.13.1 Rényi Differential Cross-Entropy

The n^{th} moment of $X \sim Log-normal(\mu, \sigma^2)$ is $\exp(n\mu + \frac{1}{2}n^2\sigma^2)$. Thus,

$$\begin{aligned}\frac{1}{1-\alpha} \ln E_h &= \frac{1}{1-\alpha} \ln \frac{1}{\sqrt{2\pi}^{\alpha-1}} + \mathbb{E}_{f_h}(X^{1-\alpha}) \\ &= \frac{\mu_1\sigma_2^2 + (\alpha-1)\mu_2\sigma_1^2 + (2\alpha-2)\sigma_1^2\sigma_2^2}{2(\sigma^2)_h} + \frac{1}{2}\ln 2\pi,\end{aligned}$$

where $\sigma_h^2 = \sigma_2^2 + (\alpha-1)\sigma_1^2$ As we know $\frac{\ln E_h}{1-\alpha} = \frac{1}{2}\ln 2\pi$ for the Gaussian distribution,

$$\begin{aligned}h_\alpha(f_1; f_2) = \frac{1}{2}&\left(\ln(2\pi\sigma_2^2) + \frac{1}{1-\alpha}\ln\left(\frac{\sigma_2^2}{(\sigma^2)_h}\right) + \frac{(\mu_1-\mu_2)^2}{(\sigma^2)_h}\right) \\ &+ \frac{\mu_1\sigma_2^2 + (\alpha-1)\mu_2\sigma_1^2 + (2\alpha-2)\sigma_1^2\sigma_2^2}{2(\sigma^2)_h}\end{aligned}$$

A.13.2 Natural Rényi Differential Cross-Entropy

$$\frac{1}{1-\alpha}\ln E_h = \mu_1 + \frac{1-\alpha}{2\alpha}\sigma_1^2 + \frac{1}{2}\ln 2\pi$$

Thus,
$$\tilde{h}_\alpha(f_1; f_2) = \frac{1}{2}\left(\ln(2\pi\sigma_2^2) + \frac{(\mu_1-\mu_2)^2}{(\sigma^2)_\alpha} + \frac{1}{1-\alpha}\ln\left(\frac{\alpha\sigma_2^2}{(\sigma^2)_\alpha}\right)\right) + \mu_1 + \frac{1-\alpha}{2\alpha}\sigma_1^2$$

Bibliography

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142, 1966.

[2] H. Bhatia, W. Paul, F. Alajaji, B. Gharesifard, and P. Burlina. Least kth-order and Rényi generative adversarial networks. *Neural Computation*, 33(9):2473–2510, Aug 2021.

[3] I. Csiszar. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizitat on Markhoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Series A*, 8, 01 1963.

[4] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungarica*, 2:299–318, 1967.

[5] I. Csiszár. Generalized cutoff rates and Rényi's information measures. *IEEE Transactions on information theory*, 41(1):26–34, 1995.

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

[7] M. Gil, F. Alajaji, and T. Linder. Rényi divergence measures for commonly used univariate continuous distributions. *Information Sciences*, 249:124–131, 2013.

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

[9] A. Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109(1):604–615, 1934.

[10] M. Kovačević, I. Stanojević, and V. Šenk. Some properties of Rényi entropy over countably infinite alphabets. *Problems of Information Transmission*, 49(2):99–110, 2013.

[11] F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006.

[12] Z. Rached, F. Alajaji, and L. L. Campbell. Rényi's divergence and entropy rates for finite alphabet Markov sources. *IEEE Transactions on Information Theory*, 47(4):1553–1561, 2001.

[13] A. Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, 1961.

[14] A. Sarraf and Y. Nie. RGAN: Rényi generative adversarial network. *SN Computer Science*, 2(1):17, 2021.

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

[16] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[17] K.-S. Song. Rényi information, loglikelihood and an intrinsic distribution measure. *Journal of Statistical Planning and Inference*, 93(1-2):51–69, 2001.

[18] F. C. Thierrin, F. Alajaji, and T. Linder. On the Rényi cross-entropy. In *Proceedings of the 17th Canadian Workshop on Information Theory* (see also arXiv e-prints, arXiv:2206.14329), pages 1–5, 2022.

[19] F. C. Thierrin, F. Alajaji, and T. Linder. Rényi cross-entropy measures for common distributions and processes with memory. *Entropy*, 24(10):1417, 2022.

[20] F. J. Valverde-Albacete and C. Peláez-Moreno. The case for shifting the Rényi entropy. *Entropy*, 21:1–21, 2019.

[21] T. Van Erven and P. Harremos. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.