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: {14fgnt, 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 $\alpha$) between two discrete distributions $p$ and $q$ with common finite support $\mathbb{S}$, given by

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

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 $\alpha$ goes to one reduces to a Shannon-type information measure (e.g., see [2] and the references therein for three different order $\alpha$ 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

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

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

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

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

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

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 \( \alpha \) 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 \( \alpha \) 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 \( \alpha \) 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 \( \alpha \) by showing that its derivative with respect to \( \alpha \) is non-positive. The same monotonicity property holds in the discrete 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

\[
\lim_{\alpha \to \infty} \frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} \tilde{c} q(x)^{\alpha - 1} = \lim_{\alpha \to \infty} \frac{1}{1 - \alpha} \ln \tilde{c} + \lim_{\alpha \to \infty} \frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} q(x)^{\alpha - 1} = \lim_{\beta \to \infty} \frac{1 - \beta}{1 - \beta} \ln \sum_{x \in \mathbb{S}} q(x)^{\beta} (\beta = \alpha - 1) = \lim_{\beta \to \infty} H_\beta(q) = - \ln q_M, \tag{4}
\]

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 \( \alpha > 1 \),

\[
\frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} p_m q(x)^{\alpha - 1} \leq \frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} p(x)q(x)^{\alpha - 1} \leq \frac{1}{1 - \alpha} \ln \sum_{x \in \mathbb{S}} p_M q(x)^{\alpha - 1},
\]

and hence by (4) we obtain

\[
\lim_{\alpha \to \infty} H_{\alpha}(p; q) = - \ln q_M. \tag{5}
\]

III. RÉNYI DIFFERENTIAL CROSS-ENTROPY FOR EXPONENTIAL FAMILY DISTRIBUTIONS

A probability distribution on \( \mathbb{R} \) or \( \mathbb{R}^n \) with parameter \( \theta \) 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

\[
f(x) = c(\theta) b(x) \exp(\eta(\theta) \cdot T(x)), \quad x \in \mathbb{S}, \tag{6}
\]

for some real-valued (measurable) functions \( c, b, \eta \) and \( T \). Here \( \eta \) 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 \( \theta \) within the parameter space

\[
\int_{\mathbb{S}} b(x) \exp(\eta(\theta) \cdot T(x)) \, dx = c(\theta)^{-1}.
\]

The pdf in (6) can also be written as

\[
f(x) = b(x) \exp(\eta \cdot T(x) + A(\eta)), \tag{7}
\]

where \( A(\eta(\theta)) = \ln c(\theta) \). 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 \( \eta_1 \) and \( \eta_2 \), 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

\[
h_{\alpha}(f_1; f_2) = \frac{A(\eta_1) - A(\eta_h) + \ln E_h}{1 - \alpha} = A(\eta_2), \tag{8}
\]

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

Proof. Using (7), we have

\[
f_1(x) f_2(x)^{\alpha - 1} = b(x) \exp(\eta_1 \cdot T(x) + A(\eta_1)) \cdot \left( b(x) \exp(\eta_2 \cdot T(x) + A(\eta_2)) \right)^{\alpha - 1} = b(x)^\alpha \exp((\eta_1 + (\alpha - 1)\eta_2) \cdot T(x)) \cdot \exp(A(\eta_1) + (\alpha - 1)A(\eta_2))
\]

Note that \( \theta \) and consequently \( T(x) \) can be vectors in cases where the distribution admits multiple parameters.
\[ b(x)^\alpha \exp (\eta_h \cdot T(x) + A(\eta_h)) \cdot \exp (A(\eta_1) + (\alpha - 1) A(\eta_2) - A(\eta_h)) \]
\[ = b(x)^{\alpha-1} f_h(x) \exp (A(\eta_1) + (\alpha - 1) A(\eta_2) - A(\eta_h)). \]

Thus,
\[ \int_S f_1(x) f_2(x)^{\alpha-1} \, dx = \int_S b(x)^{\alpha-1} f_h(x) \, dx \cdot \exp (A(\eta_1) + (\alpha - 1) A(\eta_2) - A(\eta_h)) \]
\[ = \exp (A(\eta_1) + (\alpha - 1) A(\eta_2) - A(\eta_h)) E_h, \]
and therefore,
\[ h_\alpha(f_1; f_2) = \frac{A(\eta_1) - A(\eta_h) + \ln E_h}{1 - \alpha} - A(\eta_2). \]

**Remark.** If \( b(x) = b \) is a constant for all \( x \in S \), then
\[ \ln E_h \quad \frac{1}{1 - \alpha} = -\ln b. \]

In many cases, we have that \( b(x) = 1 \) on \( S \), and thus the \( \ln E_h \) 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 \( f_i \), \( i = 1, 2 \).

### IV. Rényi differential Cross-Entropy between different distributions

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

**A. Distribution \( q \) is uniform**

Let \( q \) be uniformly distributed on \( S \). Then
\[ h_\alpha(p; q) = \frac{1}{1 - \alpha} \ln \int_S p(x) q(x)^{\alpha-1} \, dx = \ln |S|. \]

**B. Distribution \( p \) is uniform**

Now suppose \( p \) is uniformly distributed on \( S \). Then
\[ h_\alpha(p; q) = \frac{1}{1 - \alpha} \ln \int_S p(x) q(x)^{\alpha-1} \, dx = \frac{1}{1 - \alpha} \ln \frac{1}{|S|} - h_{\alpha-1}(q). \]

**C. Distribution \( q \) is exponentially distributed**

Suppose the \( S = \mathbb{R}^+ \) and \( q \) is exponential with parameter \( \lambda \). Suppose also that the moment generating function (MGF) of \( p, M_p(t) \) exists. We have
\[ h_\alpha(p; q) = \frac{1}{1 - \alpha} \ln \int_S p(x) q(x)^{\alpha-1} \, dx \]
\[ = \frac{1}{1 - \alpha} \ln \mathbb{E}_p (q(x)^{\alpha-1}) \]
\[ = \frac{1}{1 - \alpha} \ln \mathbb{E}_p \left( (\lambda \exp (-\lambda x))^{\alpha-1} \right) \]
\[ = -\ln \lambda + \frac{1}{1 - \alpha} \ln M_p (\lambda(1 - \alpha)). \]

**D. Distribution \( p \) is Gaussian**

Now assume that \( p \) 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
\[ h_\alpha(p; q) = \frac{1}{1 - \alpha} \ln \mathbb{E}_p (q(X)^{\alpha-1}) \]
\[ = \frac{1}{1 - \alpha} \ln \sigma (\sqrt{2\pi} \Gamma^{1-\alpha} (I/2) \exp (1 - \alpha) \Gamma (Y/2 \sigma^2)) \]
\[ = \ln \sigma \sqrt{2\pi} + \frac{1}{1 - \alpha} \ln M_Y (\frac{1 - \alpha}{2\sigma^2}). \]
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
\[ h_\alpha(p; q) = \ln (\sigma \sqrt{\frac{2}{\pi \sigma^2}}) + \frac{1}{1 - \alpha} \ln M_Y (\frac{1 - \alpha}{2\sigma^2}). \]

### Table I

<table>
<thead>
<tr>
<th>Name</th>
<th>( h_\alpha(f_i; f_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>( \ln B(a_2, b_2) + \frac{1}{\alpha-1} \ln B(a_1, b_1) )</td>
</tr>
<tr>
<td>( a_k := a_1 + (\alpha - 1) b_2 )</td>
<td></td>
</tr>
<tr>
<td>( b_k := b_1 + (\alpha - 1) (b_2 - 1) )</td>
<td></td>
</tr>
<tr>
<td>( x^2 )</td>
<td>( \frac{1}{1 - \beta} \left( \frac{1}{2} \ln (\alpha) - \ln \Gamma \left( \frac{1}{2} \right) + \ln \Gamma \left( \frac{1}{2} \right) \right) )</td>
</tr>
<tr>
<td>( \nu_k := \nu_1 + (\alpha - 1)(k - 2) )</td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>( \frac{1}{1 - \alpha} \ln \lambda - \ln \lambda_2 )</td>
</tr>
<tr>
<td>( \lambda := \lambda_1 + (\alpha - 1) \lambda_2 )</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>( \ln \Gamma(k_2) / k_2 \ln \theta_2 + \frac{1}{1 - \alpha} \ln \theta_2 )</td>
</tr>
<tr>
<td>( k := k_1 + (\alpha - 1) k_2 )</td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>( \frac{1}{2} \left( \ln (2\pi \sigma^2) \right) + \frac{1}{1 - \alpha} \ln \left( \frac{\sigma^2}{\sigma^2_0} \right) )</td>
</tr>
<tr>
<td>( \sigma^2_0 := \sigma^2 + (\alpha - 1) \sigma^2 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_0 = \sigma^2 )</td>
<td></td>
</tr>
<tr>
<td>Laplace</td>
<td>( \ln (2b_2) + \frac{1}{1 - \alpha} \ln \left( \frac{b_2}{b_1} \right) )</td>
</tr>
<tr>
<td>( b_k := b_2 + (1 - \alpha) b_1 )</td>
<td></td>
</tr>
</tbody>
</table>
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 $\Sigma_1$ and $\Sigma_2$, respectively, is given by

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

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

Proof. Recall that the pdf of a multivariate Gaussian with mean $0 = (0, 0, \ldots, 0)^T$ is given by:

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

for $x \in \mathbb{R}^n$. Note that this distribution is a member of the exponential family, where $T(x) = x$, $\eta = \frac{1}{2} \Sigma^{-1}$, $A(\eta) = \frac{1}{2} \ln |\Sigma| - 2\eta$ and $b(\eta) = (2\pi)^{-\frac{k}{2}}$. Hence the Rényi differential cross-entropy between two zero-mean multivariate Gaussian distributions with covariance matrices $\Sigma_1$ and $\Sigma_2$, respectively, is

$$h_\alpha(p; q) = \frac{1}{2\alpha - 2} \ln |\Sigma_1| + 2 \frac{n}{2} \ln 2\pi.$$ 

Proof. From Lemma 2, we first note that $S = \Sigma_1^{-1} + (\alpha - 1)\Sigma_2^{-1}$. With this in mind the Rényi differential cross-entropy can be rewritten using (9) as

$$1/n \left( \frac{\ln |\Sigma_{X^n}| |\Sigma_{X^n}^{-1}| B^n |\Sigma_{Y^n}^{-1}|}{2(\alpha - 1)} + \frac{1}{2} \ln |\Sigma_{Y^n}| + \frac{n}{2} \ln 2\pi \right)$$

$$= \frac{\ln 2\pi}{2} + \frac{1}{2n} \left( \ln |\Sigma_{X^n}| |\Sigma_{X^n}^{-1}| B^n |\Sigma_{Y^n}^{-1}| + \ln |\Sigma_{Y^n}| \right)$$

$$= \frac{\ln 2\pi}{2} + \frac{1}{2n} \ln \left( \frac{|\Sigma_{X^n}| |\Sigma_{X^n}^{-1}| B^n |\Sigma_{Y^n}^{-1}|}{|\Sigma_{Y^n}|} \right)$$

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

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

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

$$\lim_{n \to \infty} \frac{\ln |T^n|}{n} = \frac{1}{2\pi} \int_0^{2\pi} \left[ (2 - \alpha) \ln \tilde{g}(\lambda) - \ln \tilde{h}(\lambda) \right] d\lambda.$$ 

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 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

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

and

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

respectively, with arbitrary initial distributions, $p(i_1)$ and $q(i_1)$, $i_1 \in \mathbb{S}$. 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

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

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

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

where $1$ 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 [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
\[
\lim_{n \to \infty} \frac{1}{n} H_\alpha(X^n; Y^n) = \frac{\ln \lambda}{1 - \alpha},
\]
where $\lambda$ 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 $\lambda$ with associated positive eigenvector $b$. Let $b_m$ and $b_M$ be the minimum and maximum elements, respectively, of $b$. Then due to the non-negativity of $s$,
\[
\lambda^{n-1} s \cdot b = s R^{n-1} b \leq s R^{n-1} b_M,
\]
where $\cdot$ denotes the Euclidean inner product. Similarly, $\lambda^{n-1} s \cdot b \geq s R^{n-1} b_m$. As a result,
\[
\frac{1}{n} \ln \frac{\lambda^{n-1} s \cdot b}{b_M} \leq \frac{1}{n} \ln s R^{n-1} b \leq \frac{1}{n} \ln \frac{\lambda^{n-1} s \cdot b}{b_m}.
\]

Note that for all $n$, $\frac{sb}{b_m}$ is a constant. Thus
\[
\lim_{n \to \infty} \frac{1}{n} \ln \frac{\lambda^{n-1} s \cdot b}{b_M} = \frac{n-1}{n} \ln \lambda + \lim_{n \to \infty} \frac{1}{n} \ln \frac{s \cdot b}{b_M} = \frac{\ln \lambda}{1 - \alpha}.
\]

Similarly, we have
\[
\lim_{n \to \infty} \frac{1}{n} \ln \frac{s R^{n-1} b}{b_M} = \lim_{n \to \infty} \frac{1}{n} \ln \frac{1}{(1 - \alpha) b_m} = \frac{\ln \lambda}{1 - \alpha}.
\]

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 $\lambda_k$ be the largest positive eigenvalue of each self-communicating sub-matrix of $R$, and let $\lambda$ be the maximum of these $\lambda_k$'s. For each inessential class $C_i$, let $\lambda_j$ be the largest positive eigenvalue of the sub-matrix of each class $C_i$ that is reachable from $C_i$, and let $\lambda^i$ be the maximum of these $\lambda_j$'s. Define $\lambda = \max\{\lambda, \lambda^i\}$. Then (11) holds.

APPENDIX A: SHANNON-TYPE INFORMATION MEASURES

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon Entropy</td>
<td>$H(p) = - \sum_{x \in \mathbb{S}} p(x) \ln p(x)$</td>
</tr>
<tr>
<td>Shannon Differential</td>
<td>$h(p) = - \int_{\mathbb{X}} p(x) \ln p(x) , dx$</td>
</tr>
<tr>
<td>Shannon Cross-Entropy</td>
<td>$H(p; q) = - \sum_{x \in \mathbb{S}} p(x) \ln q(x)$</td>
</tr>
<tr>
<td>Shannon Differential</td>
<td>$h(p; q) = - \int_{\mathbb{X}} p(x) \ln q(x) dx$</td>
</tr>
<tr>
<td>KL Divergence, (Discrete)</td>
<td>$D(p</td>
</tr>
<tr>
<td>KL Divergence, (Continuous)</td>
<td>$D(p</td>
</tr>
</tbody>
</table>

APPENDIX B: DISTRIBUTIONS LISTED IN TABLE I

<table>
<thead>
<tr>
<th>Name</th>
<th>PDF $f(x)$ (Support)</th>
<th>(Parameters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>$B(a, b) x^{a-1}(1 - x)^{b-1}$</td>
<td>$(a &gt; 0, b &gt; 0)$</td>
</tr>
<tr>
<td>Normal</td>
<td>$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$</td>
<td>$(\mu, \sigma^2 &gt; 0)$</td>
</tr>
<tr>
<td>Laplace</td>
<td>$\frac{1}{\sqrt{2b}} e^{-\frac{</td>
<td>x</td>
</tr>
</tbody>
</table>

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

REFERENCES