Rényi Divergence Measures for Commonly Used Univariate Continuous Distributions

M. Gil[1,*], F. Alajaji, T. Linder

Department of Mathematics and Statistics, Queen's University, Kingston, ON, K7L 3N6

Abstract

Probabilistic 'distances' (also called divergences), which in some sense assess how 'close' two probability distributions are from one another, have been widely employed in probability, statistics, information theory, and related fields. Of particular importance due to their generality and applicability are the Rényi divergence measures. This paper presents closed-form expressions for the Rényi and Kullback-Leibler divergences for nineteen commonly used univariate continuous distributions as well as those for multivariate Gaussian and Dirichlet distributions. In addition, a table summarizing four of the most important information measure rates for zero-mean stationary Gaussian processes, namely Rényi entropy, differential Shannon entropy, Rényi divergence, and Kullback-Leibler divergence, is presented. Lastly, a connection between the Rényi divergence and the variance of the log-likelihood ratio of two distributions is established, thereby extending a previous result by Song [J. Stat. Plan. Infer. 93 (2001)] on the relation between Rényi entropy and the log-likelihood function. A table with the corresponding variance expressions for the univariate distributions considered here is also included.

Keywords: Rényi divergence, Rényi divergence rate, Kullback divergence, probabilistic distances, divergences, continuous distributions, log-likelihood ratio.

1. Introduction

In his 1961 work [27], Rényi introduced generalized information and divergence measures which naturally extend Shannon entropy [29] and Kullback-Leibler divergence (KLD) [19]. For a probability density f on $\mathbb{R}^n$ the Rényi entropy of order α is defined via the integral $h_\alpha(f) := (1-\alpha)^{-1} \ln\big(\int f(x)^\alpha dx\big)$

*Corresponding author

Email addresses: gilm3@mcmaster.ca (M. Gil), fady@mast.queensu.ca (F. Alajaji), linder@mast.queensu.ca (T. Linder)

[1]Present address: McMaster University, Department of Engineering Physics, 1280 Main Street West, John Hodgins Engineering Building, Room A315, Hamilton, Ontario, Canada L8S 4L7, Phone: (905) 525-9140 ext 24545, Fax: (905) 527-8409.

for $\alpha > 0$ and $\alpha \neq 1$. Under appropriate conditions, in the limit as $\alpha \to 1$, $h_\alpha(f)$ converges to the differential Shannon entropy $h(f) := -\int f(x) \ln f(x)\, dx = -E_f [\ln f]$. If g is another probability density on $\mathbb{R}^n$, the Rényi divergence of order α between f and g is given by $D_\alpha(f||g) := (\alpha-1)^{-1} \ln\big(\int f(x)^\alpha g(x)^{1-\alpha} dx\big)$ for $\alpha > 0$ and $\alpha \neq 1$. Under appropriate conditions, in the limit $\alpha \to 1$, $D_\alpha(f||g)$ converges to the KLD between f and g, $D(f||g) := \int f(x) \log(f(x)/g(x))\, dx = E_f[\log(f/g)]$.

While a significant number of other divergence measures have since been introduced [1, 3, 9], Rényi divergences are especially important because of their generality, applicability, and the fact that they posses operational definitions in the context of hypothesis testing [7, 2] as well as coding [14]. Additional applications of Rényi divergences include the derivation of a family of test statistics for the hypothesis that the coefficients of variation of k normal populations are equal [23], their use in problems of classification, indexing and retrieval, for example [15], and tests statistics for partially observed diffusion processes [28], to name a few.

The ubiquity of Rényi divergences suggests the importance of establishing their general mathematical properties as well as having a compilation of readily available analytical expressions for commonly used distributions. The mathematical properties of the Rényi information measures have been studied both directly, e.g., in [27, 32, 13], and indirectly through their relation to f-divergences [6, 21]. On the other hand, a compilation of important Rényi divergence and KLD expressions does not seem to be currently available in the literature. Some isolated results can be found in separate works, e.g.: Rényi divergence for Dirichlet distributions (via the Chernoff distance expression) [26]; Rényi divergence for two multivariate Gaussian distributions [16, 15], a result that can be traced back to [5] and [31]; Rényi divergence and KLD for a special case of univariate Pareto distributions [4]; and KLD for multivariate normal and univariate Gamma [25, 24], as well as Dirichlet and Wishart [24].

The need for an analogous compilation for Rényi and Shannon entropies has already been addressed in the literature. Closed-form expressions for differential Shannon and Rényi entropies for several univariate continuous distributions are presented in the work by Song [30]. The author also introduces an 'intrinsic loglikelihood-based distribution measure,' $\mathcal{G}_f$, derived from the Rényi entropy. Song's work was followed by [22] where the differential Shannon and Rényi entropy, as well as Song's intrinsic measure for an additional 26 continuous univariate distribution families are presented. The same authors then expanded these results for several multivariate families in [34].

The main objective of this work is to extend these compilations to include expressions for the KLD and Rényi divergence of commonly used univariate continuous distributions. Since most of the applications revolve around two distributions within the same family, this was the focus of the calculations as well. The complete derivations can be found in [11], where it is also shown that the independently derived expressions are in agreement with the aforementioned existing results in the literature. We present tables containing these expressions for nineteen univariate continuous distributions as well as those for Dirichlet and multivariate Gaussian distributions in Section 2.2. In Section 2.3 we provide a table summarizing the expressions for the Rényi entropy, Shannon entropy, Rényi divergence, and KLD rates for zero-mean stationary Gaussian sources. Finally, in Section 3 , we establish a relationship between the Rényi divergence and the variance of the log-likelihood ratio and present a table of the corresponding variance expressions for the various univariate distributions considered in this paper.

2. Closed-Form Expressions for Rényi and Kullback Divergences of Continuous Distributions

We consider commonly used families of univariate continuous distributions and present Rényi and Kullback divergence expressions between two members of a given family. While all the expressions given in this section have been independently derived in [11], we note that in the cases of distributions belonging to exponential families, the results can be computed via a closed-form formula originally obtained by Liese and Vajda [20] presented below. This result seems to be largely unknown in the literature. For example, the works [26, 4, 25, 24] do not make reference to this result, while similar work by Vajda and Darbellay [8] on differential entropy for exponential families is cited in some of the works compiling the corresponding expressions such as [22, 34], and even in the entropy section of [4].

Some expressions for Rényi divergences and/or Kullback divergences for cases where the formula by Liese and Vajda is not applicable are also derived in [11], namely the Rényi and Kullback divergence for general univariate Laplacian, general univariate Pareto, Cramér, and uniform distributions, as well as the KLD for general univariate Gumbel and Weibull densities.

The integration calculations are carried out using standard techniques such as substitution and the method of integration by parts, reparametrization of some of the integrals so as to express the integrand as a known probability distribution scaled by some factor, and applying integral representations of special functions, in particular the Gamma and related functions.

2.1. Rényi Divergence for Natural Exponential Families

In Chapter 2 of their 1987 book *Convex Statistical Distances* [20], Liese and Vajda derive a closed-form expression for the Rényi divergence between two members of a canonical exponential family. Note that their definition of Rényi divergence, here denoted by $R_\alpha(f_i||f_j)$, differs by a factor of α from the one considered in this work; that is $D_\alpha(f_i||f_j) = \alpha R_\alpha(f_i||f_j)$. Also, the authors allow the parameter α to be any real number as opposed to the more standard definitions where $\alpha > 0$. Consider a natural exponential family of probability measures $P_{\boldsymbol{\tau}}$ on $\mathbb{R}^n$ having densities $p_{\boldsymbol{\tau}}(x) = [C(\boldsymbol{\tau})]^{-1} \exp\langle \boldsymbol{\tau}, \boldsymbol{T}(x)\rangle$, where $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^k$, $\boldsymbol{\tau}$ is $k-$dimensional parameter vector, and $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^k$. Denote the natural parameter space by Θ. Then $R_\alpha(f_i||f_j)$ is given by the following cases:

- If $\alpha \notin \{0,1\}$ and $\alpha\boldsymbol{\tau_i} + (1-\alpha)\boldsymbol{\tau_j} \in \Theta$,

$$R_\alpha(f_i||f_j) = \frac{1}{\alpha(\alpha-1)} \ln \frac{C(\alpha\boldsymbol{\tau_i} + (1-\alpha)\boldsymbol{\tau_j})}{C(\boldsymbol{\tau_i})^\alpha C(\boldsymbol{\tau_j})^{1-\alpha}} \ ;$$

- If $\alpha \notin \{0,1\}$ and $\alpha\boldsymbol{\tau_i} + (1-\alpha)\boldsymbol{\tau_j} \notin \Theta$,

$$R_\alpha(f_i||f_j) = +\infty \ ;$$

- If $\alpha = 0$,

$$R_\alpha(f_i||f_j) = \Delta(\boldsymbol{\tau_i}, \boldsymbol{\tau_j});$$

- If $\alpha = 1$,

$$R_\alpha(f_i||f_j) = \Delta(\boldsymbol{\tau_j}, \boldsymbol{\tau_i}) \ ;$$

with $\Delta(\boldsymbol{\tau}, \boldsymbol{\tau_j}) := \lim_{\alpha\downarrow 0} \left(\alpha^{-1}\left[\alpha D(\boldsymbol{\tau_i}) + (1-\alpha)D(\boldsymbol{\tau_j}) - D(\alpha\boldsymbol{\tau_i} + (1-\alpha)\boldsymbol{\tau_j})\right]\right)$, and $D(\boldsymbol{\tau}) = \ln C(\boldsymbol{\tau})$.

2.2. Tables

The results presented in the following tables are derived in [11], except for the derivation of the KLD for Dirichlet distributions which can be found in [24]. A few of these results can also be found in the references mentioned in Section 1. Table 2 and Table 3 present the expressions for Rényi and Kullback divergences, respectively. We follow the convention $0 \ln 0 = 0$, which is justified by continuity. The densities associated with the distributions are given in Table 1.

The table of Rényi divergences includes a finiteness constraint for which the given expression is valid. For all other cases (and $\alpha > 0$), $D_\alpha(f_i||f_j) = \infty$. In the cases where the closed-form expression is a piecewise-defined function, the conditions for each case are presented alongside the corresponding formula, and it is implied that for all other cases $D_\alpha(f_i||f_j) = \infty$. The expressions for the Rényi divergence of Laplace and Cramer distributions are still continuous at $\alpha = \lambda_i/(\lambda_i + \lambda_j)$ (where λ is the scale parameter in the Laplace distribution) and $\alpha = 1/2$, respectively (this can be easily verified using l'Hospital's rule).

One important property of Rényi divergence is that $D_\alpha(T(X)||T(Y)) = D_\alpha(X||Y)$ for any invertible transformation T. This follows from the more general data processing inequality (see, e.g., [20, 21]). For example, the Rényi divergence between two lognormal densities is the same as that between two normal densities. Also, applying the appropriate T and reparametrizing, one can find similar relationships between other distributions, such as between Weibull and Gumbel, Gamma and Chi (or the relevant special cases), and the half-normal and Lévy under equal supports.

In the results below, $\Gamma(x)$, $\psi(x)$, and $B(\boldsymbol{x})$ denote the Gamma, Digamma, and the multivariate Beta functions, respectively. Also, $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. Lastly, note the direction of the divergence: $i \mapsto j$.

Table 1: Common Continuous Distributions

Name	**Density**	**Restrictions**
Beta	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$	$a\,,b>0\,;\ x\in(0,1)$
Chi	$\frac{2^{1-k/2}x^{k-1}e^{-x^2/2\sigma^2}}{\sigma^k\Gamma\left(\frac{k}{2}\right)}$	$\sigma>0\,,\ k\in\mathbb{N}\,;\ x>0$
$\boldsymbol{\chi^2}$	$\frac{x^{d/2-1}e^{-x/2}}{2^{d/2}\Gamma(d/2)}$	$d\in\mathbb{N}\,;\ x>0$
Cramér	$\frac{\theta}{2(1+\theta\lvert x\rvert)^2}$	$\theta>0\,;\ x\in\mathbb{R}$
Dirichlet	$\frac{1}{B(\boldsymbol{a})}\prod_{k=1}^{d}x_k^{a_k-1}$	$\boldsymbol{a}\in\mathbb{R}^d,\ a_k>0, d\geq 2\,;$ $\boldsymbol{x}\in\mathbb{R}^d, \sum x_k=1$
Exponential	$\lambda e^{-\lambda x}$	$\lambda>0\,;\ x>0$
Gamma	$\frac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)}$	$\theta>0\,,\ k>0\,;\ x>0$
Multivariate Gaussian	$\frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{n/2}\lvert\Sigma\rvert^{1/2}}$	$\boldsymbol{\mu}\in\mathbb{R}^n\,;\boldsymbol{x}\in\mathbb{R}^n$ Σ symmetric positive definite
Univariate Gaussian	$\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$	$\sigma>0\,,\mu\in\mathbb{R}\,;x\in\mathbb{R}$
Special Bivariate Gaussian	$\frac{e^{-\frac{1}{2}\boldsymbol{x}'\Phi^{-1}\boldsymbol{x}}}{2\pi(1-\rho^2)^{1/2}}$	$\rho\in(-1,1)\,,\Phi=\begin{bmatrix}1&\rho\\\rho&1\end{bmatrix};$ $\boldsymbol{x}\in\mathbb{R}^2$
Gumbel	$\frac{e^{-(x-\mu)/\beta}e^{-e^{-(x-\mu)/\beta}}}{\beta}$	$\mu\in\mathbb{R},\ \beta>0\,;x\in\mathbb{R}$
Half-Normal	$\sqrt{\frac{2}{\pi\sigma^2}}\,e^{-x^2/(2\sigma^2)}$	$\sigma>0\,;\ x>0$
Inverse Gaussian	$\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\frac{-\lambda(x-\mu)^2}{2\mu^2x}$	$\lambda>0,\ ,\mu>0\,;\ x>0$
Laplace	$\frac{1}{2\lambda}e^{-\lvert x-\theta\rvert/\lambda}$	$\lambda>0\,,\theta\in\mathbb{R}\,;\ x\in\mathbb{R}$
Lévy	$\sqrt{\frac{c}{2\pi}}\ \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}}$	$c>0\,,\mu\in\mathbb{R}\,;\ x>\mu$
Log-normal	$\frac{1}{x\sigma\sqrt{2\pi}}\,e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	$\sigma>0\,,\mu\in\mathbb{R}\,;\ x\in\mathbb{R}$
Maxwell-Boltzmann	$\sqrt{\frac{2}{\pi}}\,\frac{x^2e^{-\frac{x^2}{2\sigma^2}}}{\sigma^3}$	$\sigma>0\,;\ x>0$
Pareto	$am^ax^{-(a+1)}$	$a\,,m>0\,;x>m$
Rayleigh	$\frac{x}{\sigma^2}\,e^{-x^2/(2\sigma^2)}$	$\sigma>0\,;\ x>0$
Uniform	$\frac{1}{b-a}$	$a<x<b$
Weibull	$k\lambda^{-k}x^{k-1}e^{-(x/\lambda)^k}$	$k\,,\lambda>0\,;\ x>0$

Table 2: Rényi Divergences for Common Continuous Distributions

Name	$\boldsymbol{D_\alpha(f_i\|\|f_j)}$	**Finiteness Condition**
Beta	$\ln \frac{B(a_j,b_j)}{B(a_i,b_i)} + \frac{1}{\alpha-1}\ln\frac{B(a_\alpha,b_\alpha)}{B(a_i,b_i)}$ $a_\alpha = \alpha a_i + (1-\alpha)a_j,\ b_\alpha = \alpha b_i + (1-\alpha)b_j$	$a_\alpha, b_\alpha \geq 0$
Chi	$\ln\left(\frac{\sigma_j^{k_j}\Gamma(k_j/2)}{\sigma_i^{k_i}\Gamma(k_i/2)}\right)$ $+\frac{1}{\alpha-1}\ln\left(\frac{\Gamma(k_\alpha/2)}{\sigma_i^{k_i}\Gamma(k_i/2)}\left(\frac{\sigma_i^2\sigma_j^2}{(\sigma^2)^*_\alpha}\right)^{k_\alpha/2}\right)$ $(\sigma^2)^*_\alpha = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2,\ \ k_\alpha = \alpha k_i + (1-\alpha)k_j$	$(\sigma^2)^*_\alpha > 0,\ k_\alpha > 0$
$\boldsymbol{\chi^2}$	$\ln\left(\frac{\Gamma(d_j/2)}{\Gamma(d_i/2)}\right) + \frac{1}{\alpha-1}\ln\left(\frac{\Gamma(d_\alpha/2)}{\Gamma(d_i/2)}\right)$ $d_\alpha = \alpha d_i + (1-\alpha)d_j$	$d_\alpha > 0$
Cramér	*For* $\alpha = 1/2$ $\ln\frac{\theta_i}{\theta_j} + 2\ln\left(\frac{\theta_i}{\theta_i-\theta_j}\ln\frac{\theta_i}{\theta_j}\right)$ *For* $\alpha \neq 1/2$ $\ln\frac{\theta_i}{\theta_j} + \frac{1}{\alpha-1}\ln\left(\frac{\theta_i\left[1-(\theta_j/\theta_i)^{2\alpha-1}\right]}{(\theta_i-\theta_j)(2\alpha-1)}\right)$	
Dirichlet	$\ln\frac{B(\boldsymbol{a}_j)}{B(\boldsymbol{a}_i)} + \frac{1}{\alpha-1}\ln\left(\frac{B(\boldsymbol{a}_\alpha)}{B(\boldsymbol{a}_i)}\right)$ $\boldsymbol{a}_\alpha = \alpha\boldsymbol{a}_i + (1-a)\boldsymbol{a}_j$	$a_{\alpha_k} > 0\ \forall k$
Exponential	$\ln\frac{\lambda_i}{\lambda_j} + \frac{1}{\alpha-1}\ln\frac{\lambda_i}{\lambda_\alpha}$ $\lambda_\alpha = \alpha\lambda_i + (1-\alpha)\lambda_j$	$\lambda_\alpha > 0$
Gamma	$\ln\left(\frac{\Gamma(k_j)\theta_j^{k_j}}{\Gamma(k_i)\theta_i^{k_i}}\right)$ $+\frac{1}{\alpha-1}\ln\left(\frac{\Gamma(k_\alpha)}{\theta_i^{k_i}\Gamma(k_i)}\left(\frac{\theta_i\theta_j}{\theta^*_\alpha}\right)^{k_\alpha}\right)$ $\theta^*_\alpha = \alpha\theta_j + (1-a)\theta_i,\ \ k_\alpha = \alpha k_i + (1-\alpha)k_j$	$\theta^*_\alpha > 0$ and $k_\alpha > 0$
Multivariate Gaussian	$\frac{\alpha}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'[(\Sigma_\alpha)^*]^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ $-\frac{1}{2(\alpha-1)}\ln\frac{\|(\Sigma_\alpha)^*\|}{\|\Sigma_i\|^{1-\alpha}\|\Sigma_j\|^\alpha}$ $(\Sigma_\alpha)^* = \alpha\Sigma_j + (1-\alpha)\Sigma_i$	$\alpha\Sigma_i^{-1} + (1-\alpha)\Sigma_j^{-1}$ positive definite
Univariate Gaussian	$\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{2(\alpha-1)}\ln\left(\frac{\sigma_j^2}{(\sigma^2)^*_\alpha}\right) + \frac{1}{2}\frac{\alpha(\mu_i-\mu_j)^2}{(\sigma^2)^*_\alpha}$ $(\sigma^2)^*_\alpha = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2$	$(\sigma^2)^*_\alpha > 0$

Name	$D_\alpha(f_i\|\|f_j)$	Finiteness Condition
Special Bivariate Gaussian	$\frac{1}{2}\ln\left(\frac{1-\rho_j^2}{1-\rho_i^2}\right) - \frac{1}{2(\alpha-1)}\ln\left(\frac{1-(\rho_\alpha^*)^2}{(1-\rho_j^2)}\right)$ $\rho_\alpha^* = \alpha\rho_j + (1-\alpha)\rho_i$	$\alpha\Phi_i^{-1} + (1-\alpha)\Phi_j^{-1}$ positive definite
Gumbel Fixed Scale ($\beta_i = \beta_j$)	$\frac{\mu_i - \mu_j}{\beta} + \frac{1}{\alpha-1}\ln\frac{e^{\mu_i/\beta}}{\left(e^{\mu_i/\beta}\right)_\alpha}$ $\left(e^{\mu_i/\beta}\right)_\alpha = \alpha e^{\mu_i/\beta} + (1-\alpha)e^{\mu_j/\beta}$	$\left(e^{\mu_i/\beta}\right)_\alpha > 0$
Half-Normal	$\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{\alpha-1}\ln\left(\frac{\sigma_j^2}{(\sigma^2)_\alpha^*}\right)^{1/2}$ $(\sigma^2)_\alpha^* = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2$	$(\sigma^2)_\alpha^* > 0$
Inverse Gaussian	$\frac{1}{2}\ln\left(\frac{\lambda_i}{\lambda_j}\right) + \frac{1}{2(\alpha-1)}\ln\left(\frac{\lambda_i}{\lambda_\alpha}\right)$ $+\frac{1}{\alpha-1}\left\{\left(\frac{\lambda}{\mu}\right)_\alpha - \left[\left(\frac{\lambda}{\mu^2}\right)_\alpha \lambda_\alpha\right]^{1/2}\right\}$ $\left(\frac{\lambda}{\mu^2}\right)_\alpha = \alpha\frac{\lambda_i}{\mu_i^2} + (1-\alpha)\frac{\lambda_j}{\mu_j^2}$, $\lambda_a = \alpha\lambda_i + (1-\alpha)\lambda_j$ $\left(\frac{\lambda}{\mu}\right)_\alpha = \alpha\left(\frac{\lambda_i}{\mu_i}\right) + (1-\alpha)\left(\frac{\lambda_j}{\mu_j}\right)$.	$\left(\frac{\lambda}{\mu^2}\right)_\alpha \geq 0$ and $\lambda_\alpha > 0$
Laplace	*For* $\alpha = \lambda_i/(\lambda_i+\lambda_j)$ $\ln\frac{\lambda_j}{\lambda_i} + \frac{\|\theta_i-\theta_j\|}{\lambda_j} + \frac{\lambda_i+\lambda_j}{\lambda_j}\ln\left(\frac{2\lambda_i}{\lambda_i+\lambda_j+\|\theta_i-\theta_j\|}\right)$ *For* $\alpha \neq \lambda_i/(\lambda_i+\lambda_j)$ and $\alpha\lambda_j + (1-\alpha)\lambda_i > 0$ $\ln\frac{\lambda_j}{\lambda_i} + \frac{1}{\alpha-1}\ln\left(\frac{\lambda_i\lambda_j^2\, g(\alpha)}{\alpha^2\lambda_j^2 - (1-\alpha)^2\lambda_i^2}\right)$ where $g(\alpha) = \frac{\alpha}{\lambda_i}\exp\left(-\frac{(1-\alpha)\|\theta_i-\theta_j\|}{\lambda_j}\right) - \frac{1-\alpha}{\lambda_j}\exp\left(\frac{-\alpha\|\theta_i-\theta_j\|}{\lambda_i}\right)$	
Lévy Equal Supports ($\mu_i = \mu_j$)	$\frac{1}{2}\ln\left(\frac{c_i}{c_j}\right) + \frac{1}{2(\alpha-1)}\ln\left(\frac{c_i}{c_\alpha}\right)$ $c_\alpha = \alpha c_i + (1-\alpha)c_j$	$c_\alpha > 0$
Log-normal	$\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{2(\alpha-1)}\ln\left(\frac{\sigma_j^2}{(\sigma^2)_\alpha^*}\right) + \frac{1}{2}\frac{\alpha(\mu_i-\mu_j)^2}{(\sigma^2)_\alpha^*}$ $(\sigma^2)_\alpha^* = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2$	$(\sigma^2)_\alpha^* > 0$
Maxwell Boltzmann	$3\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{\alpha-1}\ln\left(\frac{\sigma_j^2}{(\sigma^2)_\alpha^*}\right)^{3/2}$ $(\sigma^2)_\alpha^* = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2$	$(\sigma^2)_\alpha^* > 0$
Pareto	*For* $\alpha \in (0,1)$ $\ln\frac{m_i^{a_i}}{m_j^{a_j}} + \ln\frac{a_i}{a_j} + \frac{1}{\alpha-1}\ln\frac{a_i m_i^{a_i}}{a_\alpha M^{a_\alpha}}$,	

Name	$D_\alpha(f_i\|\|f_j)$	Finiteness Condition
	$M = \max\{m_i, m_j\}$ *For* $\alpha > 1, m_i \geq m_j,$ and $a_\alpha = \alpha a_i + (1-\alpha)a_j > 0$ $\ln\left(\frac{m_i}{m_j}\right)^{a_j} + \ln\frac{a_i}{a_j} + \frac{1}{\alpha-1}\ln\frac{a_i}{a_\alpha}$	
Rayleigh	$2\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{\alpha-1}\ln\left(\frac{\sigma_j^2}{(\sigma^2)^*_\alpha}\right)$ $(\sigma^2)^*_\alpha = \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2$	$(\sigma^2)^*_\alpha > 0$
Uniform	*For* $\alpha \in (0,1)$ and $b_m = \min\{b_i, b_j\} > a_M = \max\{a_i, a_j\}$ $\ln\frac{b_j - a_j}{b_i - a_i} + \frac{1}{\alpha-1}\ln\frac{b_m - a_M}{b_i - a_i}$, *For* $\alpha > 1$, $(a_i, b_i) \subset (a_j, b_j)$ $\ln\frac{b_j - a_j}{b_i - a_i}$	
Weibull Fixed Shape ($k_i = k_j$)	$\ln\left(\frac{\lambda_j}{\lambda_i}\right)^k + \frac{1}{\alpha-1}\ln\frac{\lambda_j^k}{(\lambda^k)^*_\alpha}$ $(\lambda^k)^*_\alpha = \alpha\lambda_j^k + (1-\alpha)\lambda_i^k$	$(\lambda^k)^*_\alpha > 0$

Table 3: Kullback Divergences for Common Continuous Distributions

Name	$D(f_i\|\|f_j)$
Beta	$\ln\frac{B(a_j,b_j)}{B(a_i,b_i)} + \psi(a_i)(a_i - a_j) + \psi(b_i)(b_i - b_j)$ $+ [a_j + b_j - (a_i + b_i)]\psi(a_i + b_i)$
Chi	$\frac{1}{2}\psi(k_i/2)\,(k_i - k_j) + \ln\left[\left(\frac{\sigma_j}{\sigma_i}\right)^{k_j}\frac{\Gamma(k_j/2)}{\Gamma(k_i/2)}\right] + \frac{k_i}{2\sigma_j^2}\left(\sigma_i^2 - \sigma_j^2\right)$
χ^2	$\ln\frac{\Gamma(d_j/2)}{\Gamma(d_i/2)} + \frac{d_i - d_j}{2}\psi(d_i/2)$
Cramér	$\frac{\theta_i + \theta_j}{\theta_i - \theta_j}\ln\frac{\theta_i}{\theta_j} - 2$
Dirichlet	$\log\frac{B(\boldsymbol{a_j})}{B(\boldsymbol{a_i})} + \sum_{k=1}^{d}\left[a_{i_k} - a_{j_k}\right]\left[\psi(a_{i_k}) - \psi\left(\sum_{k=1}^{d} a_{i_k}\right)\right]$
Exponential	$\ln\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j - \lambda_i}{\lambda_i}$
Gamma	$\left(\frac{\theta_i - \theta_j}{\theta_j}\right)k_i + \ln\left(\frac{\Gamma(k_j)\theta_j^{k_j}}{\Gamma(k_i)\theta_i^{k_i}}\right) + (k_i - k_j)\,(\ln\theta_i + \psi(k_i))$
Multivariate Gaussian	$\frac{1}{2}\left(\ln\frac{\lvert\Sigma_j\rvert}{\lvert\Sigma_i\rvert} + \mathrm{tr}\left(\Sigma_j^{-1}\Sigma_i\right)\right) + \frac{1}{2}\left[(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\,\Sigma_j^{-1}\,(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) - n\right]$
Univariate Gaussian	$\frac{1}{2\sigma_j^2}\left[(\mu_i - \mu_j)^2 + \sigma_i^2 - \sigma_j^2\right] + \ln\frac{\sigma_j}{\sigma_i}$

Name	$D(f_i \| f_j)$
Special Bivariate Gaussian	$\frac{1}{2}\ln\left(\frac{1-\rho_j^2}{1-\rho_i^2}\right)+\frac{\rho_j^2-\rho_j\rho_i}{1-\rho_j^2}$
General Gumbel	$\ln\frac{\beta_j}{\beta_i}+\gamma\left(\frac{\beta_i}{\beta_j}-1\right)+e^{(\mu_j-\mu_i)/\beta_j}\Gamma\left(\frac{\beta_i}{\beta_j}+1\right)-1$
Half-Normal	$\ln\left(\frac{\sigma_j}{\sigma_i}\right)+\frac{\sigma_i^2-\sigma_j^2}{2\sigma_j^2}$
Inverse Gaussian	$\frac{1}{2}\left(\frac{\lambda_j}{\lambda_i}+\ln\left(\frac{\lambda_i}{\lambda_j}\right)+\frac{\lambda_j(\mu_i-\mu_j)^2}{\mu_i\mu_j^2}-1\right)$
Laplace	$\ln\frac{\lambda_j}{\lambda_i}+\frac{\|\theta_i-\theta_j\|}{\lambda_j}+\frac{\lambda_i}{\lambda_j}\exp\left(-\|\theta_i-\theta_j\|/\lambda_i\right)-1$
Lévy Equal Supports ($\mu_i=\mu_j$)	$\frac{1}{2}\ln\left(\frac{c_i}{c_j}\right)+\frac{c_j-c_i}{2c_i}$
Log-normal	$\frac{1}{2\sigma_j^2}\left[(\mu_i-\mu_j)^2+\sigma_i^2-\sigma_j^2\right]+\ln\frac{\sigma_j}{\sigma_i}$
Maxwell Boltzmann	$3\ln\left(\frac{\sigma_j}{\sigma_i}\right)+\frac{3(\sigma_i^2-\sigma_j^2)}{2\sigma_j^2}$
Pareto	$\ln\left(\frac{m_i}{m_j}\right)^{a_j}+\ln\frac{a_i}{a_j}+\frac{a_j-a_i}{a_i},\quad$ for $m_i\geq m_j$ and ∞ otherwise.
Rayleigh	$2\ln\left(\frac{\sigma_j}{\sigma_i}\right)+\frac{\sigma_i^2-\sigma_j^2}{\sigma_j^2}$
Uniform	$\ln\frac{b_j-a_j}{b_i-a_i}\quad$ for $(a_i,b_i)\subseteq(a_j,b_j)$ and ∞ otherwise.
General Weibull	$\ln\left(\frac{k_i}{k_j}\left[\frac{\lambda_j}{\lambda_i}\right]^{k_j}\right)+\gamma\frac{k_j-k_i}{k_i}+\left(\frac{\lambda_i}{\lambda_j}\right)^{k_j}\Gamma\left(1+\frac{k_j}{k_i}\right)-1$

2.3. Information Rates for Stationary Gaussian Processes

It is also of interest to extend divergence measures to stochastic processes, in which case one considers the *rate* of the given divergence measure. Given two real-valued processes $X=\{X_n\}_{n\in\mathbb{N}}$, $Y=\{Y_n\}_{n\in\mathbb{N}}$, the expression for the Rényi divergence rate between X and Y is $D_\alpha(X\|Y):=\lim_{n\to\infty}(1/n)D_\alpha(f_{X^n}\|f_{Y^n})$, whenever the limit exists, and where f_{X^n} and f_{Y^n} are the densities of $X^n=(X_1,...,X_n)$ and $Y^n=(Y_1,...,Y_n)$, respectively. The KLD, Rényi entropy, and differential Shannon entropy rates are similarly defined. In Table 4 we summarize these expressions for stationary zero-mean Gaussian processes with the appropriate reference. Note that $\varphi(\lambda)$ and $\psi(\lambda)$ are the power spectral densities of X and Y on $[-\pi,\pi]$, respectively.

Table 4: Information Rates for Stationary Zero-Mean Gaussian Processes

Information Measure	**Rate**	**Conditions**
Differential Entropy	$\frac{1}{2}\ln(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\varphi(\lambda)\,d\lambda$ [18, 17]	$\ln\varphi(\lambda) \in L^1[-\pi,\pi]$ Also valid for nonzero mean stationary Gaussian processes.
Rényi Entropy	$\frac{1}{2}\ \ln 2\pi\alpha^{\frac{1}{\alpha-1}} + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\varphi(\lambda)\,d\lambda$ [12]	$\ln\varphi(\lambda) \in L^1[-\pi,\pi]$ Also valid for nonzero mean stationary Gaussian processes.
KLD $D(X\|\|Y)$	$\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\varphi(\lambda)}{\psi(\lambda)} - 1 - \ln\frac{\varphi(\lambda)}{\psi(\lambda)}\right)d\lambda$ [17]	$\varphi(\lambda)/\psi(\lambda)$ is bounded *or* $\psi(\lambda) > a > 0\,,\ \forall\lambda \in [-\pi,\pi]$ and $\varphi \in L^2[-\pi,\pi]$.
Rényi Divergence $D_\alpha(X\|\|Y)$	$\frac{1}{4\pi(1-\alpha)}\int_{-\pi}^{\pi} \ln\frac{h(\lambda)}{\varphi(\lambda)^{1-\alpha}\psi(\lambda)^{\alpha}}\,d\lambda$ [31]	$h(\lambda) := \alpha\psi(\lambda) + (1-\alpha)\varphi(\lambda)$ $\psi(\lambda)$ and $\varphi(\lambda)$ are essentially bounded and essentially bounded away from zero, and $\alpha \in (0,1)$.

3. Rényi Divergence and the Log-likelihood Ratio

Song [30] pointed out a relationship between the derivative of the Rényi entropy with respect to the parameter α and the variance of the log-likelihood function of a distribution. Let f be a probability density, then

$$\lim_{\alpha\to 1} \frac{d}{d\alpha}\, h_\alpha(f) = -\frac{1}{2}\,\mathrm{Var}(\ln f(X))\ ,$$

assuming the integrals involved are well-defined and differentiation operations are legitimate. Extending Song's idea, we note that the variance of the log-likelihood ratio (LLR) between two densities can be similarly derived from an analytic formula of their Rényi divergence of order α. Let f_i and f_j be two probability densities such that $D_\alpha(f_i\,||\,f_j)$ is n times continuously differentiable with respect to α ($n \geq 2$). Assuming differentiation and integration can be interchanged, one obtains

$$\lim_{\alpha\to 1} \frac{d}{d\alpha}\, D_\alpha(f_i||f_j) = \frac{1}{2}\,\mathrm{Var}_{f_i}\left(\ln\frac{f_i(X)}{f_j(X)}\right)\ .$$

Considering the integral $G(\alpha) := \int f_i(x)^\alpha f_j^{1-\alpha}(x)\,dx$ for $\alpha > 0$, $\alpha \neq 1$, we have

$$\lim_{\alpha\to 1} \frac{d^n}{d\alpha^n}\, G(\alpha) = E_{f_i}\left[\left(\ln\frac{f_i(X)}{f_j(X)}\right)^n\right].$$

Hence, from the definition of $D_\alpha(f_i||f_j)$, we find

$$\begin{aligned}\lim_{\alpha\to 1}\frac{d}{d\alpha}D_\alpha(f_i||f_j) &= \lim_{\alpha\to 1}\frac{1}{(\alpha-1)^2}\left[(\alpha-1)G(\alpha)^{-1}\frac{dG(\alpha)}{d\alpha} - \ln G(\alpha)\right]\\ &= \frac{1}{2}\left(-E_{f_i}\left[\left(\ln\frac{f_i(X)}{f_j(X)}\right)\right]^2 + E_{f_i}\left[\left(\ln\frac{f_i(X)}{f_j(X)}\right)^2\right]\right),\end{aligned}$$

where we used l'Hospital's rule to evaluate the limit. This connection between the Rényi divergence and the LLR variance becomes practically useful in light of the Rényi divergence expressions presented in Section 2.2. Table 5 presents these quantities for the continuous univariate distributions considered in this paper. Note that $\psi^{(1)}(x)$ refers to the polygamma function of order 1.

Table 5: Var(LLR) for Common Continuous Univariate Distributions

Name	$\mathrm{Var}_{f_i}\left(\ln\frac{f_i(X)}{f_j(X)}\right)$
Beta	$(a_i-a_j)^2\psi^{(1)}(a_i)+(b_i-b_j)^2\psi^{(1)}(b_i)-(a_i-a_j+b_i-b_j)^2\psi^{(1)}(a_i+b_i)$
Chi	$\frac{2\left(\sigma_i^2-\sigma_j^2\right)\left(k_i\left(\sigma_i^2+\sigma_j^2\right)-2k_j\sigma_i^2\right)+\sigma_j^4(k_i-k_j)^2\psi^{(1)}\left(\frac{k_i}{2}\right)}{4\sigma_j^4}$
χ^2	$\frac{1}{4}(d_i-d_j)^2\psi^{(1)}\left(\frac{d_i}{2}\right)$
Cramér	$\frac{4\left[(\theta_i-\theta_j)^2-\theta_i\theta_j\log^2\left(\frac{\theta_j}{\theta_i}\right)\right]}{(\theta_i-\theta_j)^2}$
Exponential	$\frac{(\lambda_j-\lambda_i)^2}{\lambda_i^2}$
Gamma	$\frac{(\theta_i-\theta_j)(k_i(\theta_i+\theta_j)-2k_j\theta_j)+\theta_j^2(k_i-k_j)^2\psi^{(1)}(k_i)}{\theta_j^2}$
Univariate Gaussian	$\frac{2\sigma_i^2(\mu_i-\mu_j)^2+(\sigma_i^2-\sigma_j^2)^2}{2\sigma_j^4}$
Gumbel (Fixed Scale)	$e^{-\frac{2\mu_i}{\beta}}\left(e^{\frac{\mu_i}{\beta}}-e^{\frac{\mu_j}{\beta}}\right)^2$
Half-Normal	$\frac{(\sigma_j^2-\sigma_i^2)^2}{2\sigma_j^4}$
Inverse Gaussian	$\frac{\lambda_i\lambda_j^2\mu_i^4-2\lambda_i\lambda_j^2\mu_i^2\mu_j^2+\mu_j^4\left(\lambda_i\lambda_j^2+2\mu_i(\lambda_i-\lambda_j)^2\right)}{4\lambda_i^2\mu_i\mu_j^4}$
Laplace	$\frac{1}{\lambda_j^2}\left\{\lambda_i^2\left(2-e^{-\frac{2\lvert\theta_i-\theta_j\rvert}{\lambda_i}}\right)-2\lambda_i\lambda_j e^{-\frac{\lvert\theta_i-\theta_j\rvert}{\lambda_i}} -2(\lambda_i+\lambda_j)\lvert\theta_i-\theta_j\rvert e^{-\frac{\lvert\theta_i-\theta_j\rvert}{\lambda_i}}+\lambda_j^2\right\}$
Lévy Equal Supports $(\mu_i=\mu_j)$	$\frac{(c_i-c_j)^2}{2c_i^2}$
Log-normal	$\frac{2\sigma_i^2(\mu_i-\mu_j)^2+(\sigma_i^2-\sigma_j^2)^2}{2\sigma_j^4}$

Name	$\mathrm{Var}_{f_i}\left(\ln\frac{f_i(X)}{f_j(X)}\right)$
Maxwell Boltzmann	$\frac{3(\sigma_j^2-\sigma_i^2)^2}{2\sigma_j^4}$
Pareto	$\frac{(a_i-a_j)^2}{a_i^2}$
Rayleigh	$\frac{(\sigma_j^2-\sigma_i^2)^2}{\sigma_j^4}$
Weibull Fixed Scale	$\frac{\left(\lambda_i^k-\lambda_j^k\right)^2}{\lambda_j^{2k}}$

Song [30] also proposed a nonparametric estimate of $\mathrm{Var}(\ln f(X))$ based on separately estimating $\Delta_1 := E[\ln f(X)]$ and $\Delta_2 := E[(\ln f(X))^2]$ by the plug-in estimates $\hat{\Delta}_{l,n} := \int f_n(x)\big(\ln f_n(x)\big)^l\,dx$, $l=1,2$, where f_n is a kernel density estimate of f from the independent and identically distributed (i.i.d.) samples $(X_1, X_2, \ldots, X_n)$ that are drawn according to f. Theorem 3.1 in [30] establishes precise conditions under which $\hat{\Delta}_{2,n} - \hat{\Delta}_{1,n}^2$ converges with probability one to $\mathrm{Var}(\ln f(X))$ as $n\to\infty$ (the estimate is strongly consistent).

Assuming one has access to i.i.d. samples $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$ drawn according to f_i and f_j, respectively, we can propose a similar estimate of $\mathrm{Var}_{f_i}\left(\ln\frac{f_i(X)}{f_j(X)}\right)$. Let

$$\tilde{\Delta}_{l,n} := \int f_{i,n}(x)\left(\ln\frac{f_{i,n}(x)}{f_{j,n}(x)}\right)^l dx, \quad l=1,2,$$

where $f_{i,n}$ and $f_{j,n}$ are kernel density estimates of f_i and f_j based on $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$, respectively. It is an interesting problem to find exact conditions on f_i and f_j such that appropriate choices of $f_{i,n}$ and $f_{j,n}$ make $\tilde{\Delta}_{2,n}-\tilde{\Delta}_{1,n}^2$ a strongly consistent estimate of $\mathrm{Var}_{f_i}\left(\ln\frac{f_i(X)}{f_j(X)}\right)$. Investigating this problem is out of the scope of this paper, but it is likely that the consistency proof in [30] can be extended to this case under appropriate restrictions on f_i and f_j. Alternatively, the partitioning, $k(n)$ nearest-neighbor, and minimum spanning tree based methods reviewed in [33] and the references therein should also provide consistent estimates.

4. Conclusion

In this work we presented closed-form expressions for Rényi and Kullback-Leibler divergences for nineteen commonly used univariate continuous distributions, as well as those expressions for Dirichlet and multivariate Gaussian distributions. We also presented a table summarizing four of the most important information measure rates for zero-mean stationary Gaussian processes, as well as the relevant references. Lastly, we established a connection between the log-likelihood ratio between two distributions and their Rényi divergence, extending the work of Song [30], who considers the log-likelihood function and its relation to Rényi entropy. Following this connection, we have provided the corresponding expressions for the nineteen univariate distributions considered here. The present compilation is of course not complete. Particularly valuable future additions would be the Rényi divergence expressions for Student-t and Kumaraswamy distributions, as well as for the exponentiated and beta-normal distributions [10].

5. Acknowledgements

The authors are grateful to NSERC of Canada whose funds partially supported this work.

References

[1] J. Aczél, Z. Daróczy, On Measures of Information and their Characterization, Academic Press, 1975.

[2] F. Alajaji, P. N. Chen, Z. Rached, Csiszár's cutoff rates for the general hypothesis testing problem, IEEE Transactions on Information Theory 50 (4) (2004) 663 – 678.

[3] C. Arndt, Information Measures: Information and its Description in Science and Engineering, Signals and Communication Technology, Springer, 2004.

[4] M. Asadi, N. Ebrahimi, G. G. Hamedani, E. S. Soofi, Information measures for Pareto distributions and order statistics, in: Advances in Distribution Theory, Order Statistics, and Inference, Springer, 2006, pp. 207–223.

[5] J. Burbea, The convexity with respect to Gaussian distributions of divergences of order α, Utilitas Mathematica 26 (1984) 171–192.

[6] I. Csiszár, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, Magyar Tud. Akad. Mat. Kutat?? Int. Közl (8) (1963) 85–108.

[7] I. Csiszar, Generalized cutoff rates and Rényi's information measures, IEEE Transactions on Information Theory 41 (1) (1995) 26 –34.

[8] G. A. Darbellay, I. Vajda, Entropy expressions for multivariate continuous distributions, IEEE Transactions on Information Theory (2000) 709–712.

[9] P. A. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice - Hall International, 1982.

[10] N. Eugene, C. Lee, F. Famoye, Beta-normal distribution and its applications, Communications in Statistics - Theory and Methods, 31-(4) (2002) 497–512.

[11] M. Gil, On Rényi divergence measures for continuous alphabet sources, Master's thesis, Queen's University (2011).

[12] L. Golshani, E. Pasha, Rényi entropy rate for Gaussian processes, Information Sciences 180 (8) (2010) 1486–1491.

[13] L. Golshani, E. Pasha, G. Yari, Some properties of Rényi entropy and Rényi entropy rate, Information Sciences 179 (14) (2009) 2426 – 2433.

[14] P. Harremoës, Interpretations of Rényi entropies and divergences, Physica A: Statistical Mechanics and its Applications 365 (1) (2006) 57 – 62.

[15] A. O. Hero, B. Ma, O. Michel, J. Gorman, Alpha-divergence for classification, indexing and retrieval, Tech. rep., University of Michigan (2001).

[16] T. Hobza, D. Morales, L. Pardo, Rényi statistics for testing equality of autocorrelation coefficients, Statistical Methodology 6 (4) (2009) 424 – 436.

[17] S. Ihara, Information Theory for Continuous Systems, World Scientific, 1993.

[18] A. N. Kolmogorov, A new metric invariant of transient dynamical systems and automorphisms, in: in Lebesgue Spaces.??? Dokl. Akad. Nauk. SSSR 119, 1958.

[19] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[20] F. Liese, I. Vajda, Convex Statistical Distances, Treubner, 1987.

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

[22] S. Nadarajah, K. Zografos, Formulas for Rényi information and related measures for univariate distributions, Information Sciences 155 (1-2) (2003) 119 – 138.

[23] M. C. Pardo, J. A. Pardo, Use of Rényi's divergence to test for the equality of the coefficients of variation, Journal of Computational and Applied Mathematics 116 (1) (2000) 93 – 104.

[24] W. D. Penny, Kullback-Leibler divergences of normal, gamma, dirichlet and wishart densities., Tech. rep., Wellcome Department of Cognitive Neurology (2001).

[25] W. D. Penny, S. Roberts, Bayesian methods for autoregressive models, in: Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop, vol. 1, 2000.

[26] T. W. Rauber, T. Braun, K. Berns, Probabilistic distance measures of the Dirichlet and Beta distributions, Pattern Recognition 41 (2) (2008) 637 – 645.

[27] A. Rényi, On measures of entropy and information, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1961.

[28] M. J. Rivas, M. T. Santos, D. Morales, Rényi test statistics for partially observed diffusion processes, Journal of Statistical Planning and Inference 127 (1-2) (2005) 91 – 102.

[29] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27.

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

[31] I. Vajda, Theory of Statistical Inference and Information, Kluwer, Dordrecht, 1989.

[32] T. van Erven, P. Harremoës, Rényi divergence and majorization, in: Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, 2010.

[33] Q. Wang, S. Kulkarni, S. Verdú, Universal estimation of information measures for analog sources, Foundations and Trends in Communications and Information Theory, 5 (3) (2009), 265 – 353.

[34] K. Zografos, S. Nadarajah, Expressions for Rényi and Shannon entropies for multivariate distributions, Statistics & Probability Letters 71 (1) (2005) 71 – 84.