

On Representations of Divergence Measures and Related Quantities in Exponential Families

Stefan Bedbur  and Udo Kamps * 

Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany; bedbur@isw.rwth-aachen.de

* Correspondence: kamps@isw.rwth-aachen.de

Abstract: Within exponential families, which may consist of multi-parameter and multivariate distributions, a variety of divergence measures, such as the Kullback–Leibler divergence, the Cressie–Read divergence, the Rényi divergence, and the Hellinger metric, can be explicitly expressed in terms of the respective cumulant function and mean value function. Moreover, the same applies to related entropy and affinity measures. We compile representations scattered in the literature and present a unified approach to the derivation in exponential families. As a statistical application, we highlight their use in the construction of confidence regions in a multi-sample setup.

Keywords: exponential family; cumulant function; mean value function; divergence measure; distance measure; affinity

MSC: 60E05; 62H12; 62F25



Citation: Bedbur, S.; Kamps, U. On Representations of Divergence Measures and Related Quantities in Exponential Families. *Entropy* **2021**, *23*, 726. <https://doi.org/10.3390/e23060726>

Academic Editor: Maria Longobardi

Received: 12 May 2021

Accepted: 5 June 2021

Published: 8 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a broad literature on divergence and distance measures for probability distributions, e.g., on the Kullback–Leibler divergence, the Cressie–Read divergence, the Rényi divergence, and Phi divergences as a general family, as well as on associated measures of entropy and affinity. For definitions and details, we refer to [1]. These measures have been extensively used in statistical inference. Excellent monographs on this topic were provided by Liese and Vajda [2], Vajda [3], Pardo [1], and Liese and Miescke [4].

Within an exponential family as defined in Section 2, which may consist of multi-parameter and multivariate distributions, several divergence measures and related quantities are seen to have nice explicit representations in terms of the respective cumulant function and mean value function. These representations are contained in different sources. Our focus is on a unifying presentation of main quantities, while not aiming at an exhaustive account. As an application, we derive confidence regions for the parameters of exponential distributions based on different divergences in a simple multi-sample setup.

For the use of the aforementioned measures of divergence, entropy, and affinity, we refer to the textbooks [1–4] and exemplarily to [5–10] for statistical applications, including the construction of test procedures as well as methods based on dual representations of divergences, and to [11] for a classification problem.

2. Exponential Families

Let $\Theta \neq \emptyset$ be a parameter set, μ be a σ -finite measure on the measurable space $(\mathcal{X}, \mathcal{B})$, and $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be an exponential family (EF) of distributions on $(\mathcal{X}, \mathcal{B})$ with μ -density

$$f_\vartheta(x) = C(\vartheta) \exp \left\{ \sum_{j=1}^k Z_j(\vartheta) T_j(x) \right\} h(x), \quad x \in \mathcal{X}, \quad (1)$$

of P_ϑ for $\vartheta \in \Theta$, where $C, Z_1, \dots, Z_k : \Theta \rightarrow \mathbb{R}$ are real-valued functions on Θ and $h, T_1, \dots, T_k : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}^1, \mathbb{B}^1)$ are real-valued Borel-measurable functions with $h \geq 0$.

Usually, μ is either the counting measure on the power set of \mathcal{X} (for a family of discrete distributions) or the Lebesgue measure on the Borel sets of \mathcal{X} (in the continuous case). Without loss of generality and for a simple notation, we assume that $h > 0$ (the set $\{x \in \mathcal{X} : h(x) = 0\}$ is a null set for all $P \in \mathcal{P}$). Let ν denote the σ -finite measure with μ -density h .

We assume that representation (1) is minimal in the sense that the number k of summands in the exponent cannot be reduced. This property is equivalent to Z_1, \dots, Z_k being affinely independent mappings and T_1, \dots, T_k being ν -affinely independent mappings; see, e.g., [12] (Cor. 8.1). Here, ν -affine independence means affine independence on the complement of every null set of ν .

To obtain simple formulas for divergence measures in the following section, it is convenient to use the natural parameter space

$$\Xi^* = \left\{ \zeta \in \mathbb{R}^k : \int e^{\zeta^t T} h d\mu < \infty \right\}$$

and the (minimal) canonical representation $\{P_\zeta^* : \zeta \in \mathbf{Z}(\Theta)\}$ of \mathcal{P} with μ -density

$$f_\zeta^*(x) = C^*(\zeta) e^{\zeta^t T(x)} h(x), \quad x \in \mathcal{X}, \quad (2)$$

of P_ζ^* and normalizing constant $C^*(\zeta)$ for $\zeta = (\zeta_1, \dots, \zeta_k)^t \in \mathbf{Z}(\Theta) \subset \Xi^*$, where $\mathbf{Z} = (Z_1, \dots, Z_k)^t$ denotes the (column) vector of the mappings Z_1, \dots, Z_k and $\mathbf{T} = (T_1, \dots, T_k)^t$ denotes the (column) vector of the statistics T_1, \dots, T_k . For simplicity, we assume that \mathcal{P} is regular, i.e., we have that $\mathbf{Z}(\Theta) = \Xi^*$ (\mathcal{P} is full) and that Ξ^* is open; see [13]. In particular, this guarantees that \mathbf{T} is minimal sufficient and complete for \mathcal{P} ; see, e.g., [14] (pp. 25–27).

The cumulant function

$$\kappa(\zeta) = -\ln(C^*(\zeta)), \quad \zeta \in \Xi^*,$$

associated with \mathcal{P} is strictly convex and infinitely often differentiable on the convex set Ξ^* ; see [13] (Theorem 1.13 and Theorem 2.2). It is well-known that the Hessian matrix of κ at ζ coincides with the covariance matrix of \mathbf{T} under P_ζ^* and that it is also equal to the Fisher information matrix $\mathbf{I}(\zeta)$ at ζ . Moreover, by introducing the mean value function

$$\pi(\zeta) = E_\zeta[\mathbf{T}], \quad \zeta \in \Xi^*, \quad (3)$$

we have the useful relation

$$\pi = \nabla \kappa, \quad (4)$$

where $\nabla \kappa$ denotes the gradient of κ ; see [13] (Cor. 2.3). π is a bijective mapping from Ξ^* to the interior of the convex support of ν^T , i.e., the closed convex hull of the support of ν^T ; see [13] (p. 2 and Theorem 3.6).

Finally, note that representation (2) can be rewritten as

$$f_\zeta^*(x) = e^{\zeta^t T(x) - \kappa(\zeta)} h(x), \quad x \in \mathcal{X}, \quad (5)$$

for $\zeta \in \Xi^*$.

3. Divergence Measures

Divergence measures may be applied, for instance, to quantify the “disparity” of a distribution to some reference distribution or to measure the “distance” between two distributions within some family in a certain sense. If the distributions in the family are dominated by a σ -finite measure, various divergence measures have been introduced by means of the corresponding densities. In parametric statistical inference, they serve to construct statistical tests or confidence regions for underlying parameters; see, e.g., [1].

Definition 1. Let \mathcal{F} be a set of distributions on $(\mathcal{X}, \mathcal{B})$. A mapping $D : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$ is called a divergence (or divergence measure) if:

(i) $D(P, Q) \geq 0$ for all $P, Q \in \mathcal{F}$ and $D(P, Q) = 0 \Leftrightarrow P = Q$ (positive definiteness).

If additionally

(ii) $D(P, Q) = D(Q, P)$ for all $P, Q \in \mathcal{F}$ (symmetry)

is valid, D is called a distance (or distance measure or semi-metric). If D then moreover meets

(iii) $D(P_1, P_2) \leq D(P_1, Q) + D(Q, P_2)$ for all $P_1, P_2, Q \in \mathcal{F}$ (triangle inequality), D is said to be a metric.

Some important examples are the Kullback–Leibler divergence (KL-divergence):

$$D_{KL}(P_1, P_2) = \int f_1 \ln\left(\frac{f_1}{f_2}\right) d\mu,$$

the Jeffrey distance:

$$D_J(P_1, P_2) = D_{KL}(P_1, P_2) + D_{KL}(P_2, P_1)$$

as a symmetrized version, the Rényi divergence:

$$D_{R_q}(P_1, P_2) = \frac{1}{q(q-1)} \ln\left(\int f_1^q f_2^{1-q} d\mu\right), \quad q \in \mathbb{R} \setminus \{0, 1\}, \tag{6}$$

along with the related Bhattacharyya distance $D_B(P_1, P_2) = D_{R_{1/2}}(P_1, P_2)/4$, the Cressie–Read divergence (CR-divergence):

$$D_{CR_q}(P_1, P_2) = \frac{1}{q(q-1)} \int f_1 \left[\left(\frac{f_1}{f_2}\right)^{q-1} - 1 \right] d\mu, \quad q \in \mathbb{R} \setminus \{0, 1\}, \tag{7}$$

which is the same as the Chernoff α -divergence up to a parameter transformation, the related Matusita distance $D_M(P_1, P_2) = D_{CR_{1/2}}(P_1, P_2)/2$, and the Hellinger metric:

$$D_H(P_1, P_2) = \left[\int (\sqrt{f_1} - \sqrt{f_2})^2 d\mu \right]^{1/2} \tag{8}$$

for distributions $P_1, P_2 \in \mathcal{F}$ with μ -densities f_1, f_2 , provided that the integrals are well-defined and finite.

D_{KL} , D_{R_q} , and D_{CR_q} for $q \in \mathbb{R} \setminus \{0, 1\}$ are divergences, and D_J , $D_{R_{1/2}}$, D_B , $D_{CR_{1/2}}$, and $D_M (= D_H^2)$, since they moreover satisfy symmetry, are distances on $\mathcal{F} \times \mathcal{F}$. D_H is known to be a metric on $\mathcal{F} \times \mathcal{F}$.

In parametric models, it is convenient to use the parameters as arguments and briefly write, e.g.,

$$D_{KL}(\vartheta_1, \vartheta_2) \quad \text{for} \quad D_{KL}(P_{\vartheta_1}, P_{\vartheta_2}), \quad \vartheta_1, \vartheta_2 \in \Theta,$$

if the parameter $\vartheta \in \Theta$ is identifiable, i.e., if the mapping $\vartheta \mapsto P_\vartheta$ is one-to-one on Θ . This property is met for the EF \mathcal{P} in Section 2 with minimal canonical representation (5); see, e.g., [13] (Theorem 1.13(iv)).

It is known from different sources in the literature that the EF structure admits simple formulas for the above divergence measures in terms of the corresponding cumulant function and/or mean value function. For the KL-divergence, we refer to [15] (Cor. 3.2) and [13] (pp. 174–178), and for the Jeffrey distance also to [16].

Theorem 1. Let \mathcal{P} be as in Section 2 with minimal canonical representation (5). Then, for $\zeta, \eta \in \Xi^*$, we have

$$\begin{aligned} D_{KL}(\zeta, \eta) &= \kappa(\eta) - \kappa(\zeta) + (\zeta - \eta)^t \pi(\zeta) \\ \text{and } D_J(\zeta, \eta) &= (\zeta - \eta)^t (\pi(\zeta) - \pi(\eta)). \end{aligned} \quad (9)$$

Proof. By using Formulas (3) and (5), we obtain for $\zeta, \eta \in \Xi^*$ that

$$\begin{aligned} D_{KL}(\zeta, \eta) &= \int [\ln(f_\zeta^*) - \ln(f_\eta^*)] f_\zeta^* d\mu \\ &= \int [(\zeta - \eta)^t \mathbf{T} - \kappa(\zeta) + \kappa(\eta)] f_\zeta^* d\mu \\ &= \kappa(\eta) - \kappa(\zeta) + (\zeta - \eta)^t \pi(\zeta). \end{aligned}$$

From this, the representation of D_J is obvious. \square

As a consequence of Theorem 1, D_{KL} and D_J are infinitely often differentiable on $\Xi^* \times \Xi^*$, and the derivatives are easily obtained by making use of the EF properties. For example, by using Formula (4), we find $\nabla D_{KL}(\zeta, \cdot) = \pi(\cdot) - \pi(\zeta)$ and that the Hessian matrix of $D_{KL}(\zeta, \cdot)$ at η is the Fisher information matrix $\mathbf{I}(\eta)$, where $\zeta \in \Xi^*$ is considered to be fixed.

Moreover, we obtain from Theorem 1 that the reverse KL-divergence $D_{KL}^*(\zeta, \eta) = D_{KL}(\eta, \zeta)$ for $\zeta, \eta \in \Xi^*$ is nothing but the Bregman divergence associated with the cumulant function κ ; see, e.g., [1,11,17]. As an obvious consequence of Theorem 1, other symmetrizations of the KL-divergence may be expressed in terms of κ and π as well, such as the so-called resistor-average distance (cf. [18])

$$\begin{aligned} D_{RA}(\zeta, \eta) &= 2 \left(\frac{1}{D_{KL}(\zeta, \eta)} + \frac{1}{D_{KL}(\eta, \zeta)} \right)^{-1} \\ &= \frac{2 D_{KL}(\zeta, \eta) D_{KL}(\eta, \zeta)}{D_J(\zeta, \eta)}, \quad \zeta, \eta \in \Xi^*, \zeta \neq \eta, \end{aligned} \quad (10)$$

with $D_{RA}(\zeta, \zeta) = 0$, $\zeta \in \Xi^*$, or the distance

$$D_{GA}(\zeta, \eta) = [D_{KL}(\zeta, \eta) D_{KL}(\eta, \zeta)]^{1/2}, \quad \zeta, \eta \in \Xi^*, \quad (11)$$

obtained by taking the harmonic and geometric mean of D_{KL} and D_{KL}^* ; see [19].

Remark 1. Formula (9) can be used to derive the test statistic

$$\Lambda(x) = -2 \ln \left(\frac{\sup_{\zeta \in \Xi_0} f_\zeta^*(x)}{\sup_{\zeta \in \Xi^*} f_\zeta^*(x)} \right), \quad x \in \mathcal{X},$$

of the likelihood-ratio test for the test problem

$$H_0 : \zeta \in \Xi_0 \quad \text{against} \quad H_1 : \zeta \in \Xi^* \setminus \Xi_0,$$

where $\emptyset \neq \Xi_0 \subsetneq \Xi^*$. If the maximum likelihood estimators (MLEs) $\hat{\zeta} = \hat{\zeta}(x)$ and $\hat{\zeta}_0 = \hat{\zeta}_0(x)$ of ζ in Ξ^* and Ξ_0 (based on x) both exist, we have:

$$\begin{aligned} \Lambda &= 2 \left[\ln(f_{\hat{\zeta}}^*) - \ln(f_{\hat{\zeta}_0}^*) \right] \\ &= 2 \left[\kappa(\hat{\zeta}_0) - \kappa(\hat{\zeta}) + (\hat{\zeta} - \hat{\zeta}_0)^t \mathbf{T} \right] \\ &= 2 D_{\text{KL}}(\hat{\zeta}, \hat{\zeta}_0) \end{aligned}$$

by using that the unrestricted MLE fulfils $\pi(\hat{\zeta}) = \mathbf{T}$; see, e.g., [12] (p. 190) and [13] (Theorem 5.5). In particular, when testing a simple null hypothesis with $\Xi_0 = \{\eta\}$ for some fixed $\eta \in \Xi^*$, we have $\Lambda = 2D_{\text{KL}}(\hat{\zeta}, \eta)$.

Convenient representations within EFs of the divergences in Formulas (6)–(8) can also be found in the literature; we refer to [2] (Prop. 2.22) for D_{R_q} , D_H , and D_M , to [20] for D_B , and to [9] for D_{R_q} . The formulas may all be obtained by computing the quantity

$$A_q(P_1, P_2) = \int f_1^q f_2^{1-q} d\mu, \quad q \in \mathbb{R} \setminus \{0, 1\}. \tag{12}$$

For $q \in (0, 1)$, we have the following identity (cf. [21]).

Lemma 1. Let \mathcal{P} be as in Section 2 with minimal canonical representation (5). Then, for $\zeta, \eta \in \Xi^*$ and $q \in (0, 1)$, we have:

$$A_q(\zeta, \eta) = \exp\{\kappa(q\zeta + (1 - q)\eta) - [q\kappa(\zeta) + (1 - q)\kappa(\eta)]\}.$$

Proof. Let $\zeta, \eta \in \Xi^*$ and $q \in (0, 1)$. Then,

$$\begin{aligned} A_q(\zeta, \eta) &= \int (f_{\zeta}^*)^q (f_{\eta}^*)^{1-q} d\mu \\ &= \int \exp\left\{(q\zeta + (1 - q)\eta)^t \mathbf{T} - [q\kappa(\zeta) + (1 - q)\kappa(\eta)]\right\} h d\mu \\ &= \exp\left\{\kappa(q\zeta + (1 - q)\eta) - [q\kappa(\zeta) + (1 - q)\kappa(\eta)]\right\}, \end{aligned}$$

where the convexity of Ξ^* ensures that $\kappa(q\zeta + (1 - q)\eta)$ is defined. \square

Remark 2. For arbitrary divergence measures, several transformations and skewed versions as well as symmetrization methods, such as the Jensen–Shannon symmetrization, are studied in [19]. Applied to the KL-divergence, the skew Jensen–Shannon divergence is introduced as

$$D_{\text{JS}_q}(P_1, P_2) = q D_{\text{KL}}(P_1, qP_1 + (1 - q)P_2) + (1 - q) D_{\text{KL}}(P_2, qP_1 + (1 - q)P_2)$$

for $P_1, P_2 \in \mathcal{P}$ and $q \in (0, 1)$, which includes the Jensen–Shannon distance for $q = 1/2$ (the distance $D_{\text{JS}_{1/2}}^{1/2}$ even forms a metric). Note that, for $\zeta, \eta \in \Xi^*$, the density $qf_{\zeta}^* + (1 - q)f_{\eta}^*$ of the mixture $qP_{\zeta}^* + (1 - q)P_{\eta}^*$ does not belong to \mathcal{P} , in general, such that the identity in Theorem 1 for the KL-divergence is not applicable, here.

However, from the proof of Lemma 1, it is obvious that

$$\frac{1}{A_q(\zeta, \eta)} \left(f_{\zeta}^*\right)^q \left(f_{\eta}^*\right)^{1-q} = f_{q\zeta+(1-q)\eta}^*, \quad \zeta, \eta \in \Xi^*, \quad q \in (0, 1),$$

i.e., the EF \mathcal{P} is closed when forming normalized weighted geometric means of the densities. This finding is utilized in [19] to introduce another version of the skew Jensen–Shannon divergence based on the KL-divergence, where the weighted arithmetic mean of the densities is replaced by

the normalized weighted geometric mean. The skew geometric Jensen–Shannon divergence thus obtained is given by

$$D_{GJS_q}(\zeta, \eta) = q D_{KL}(\zeta, q\zeta + (1 - q)\eta) + (1 - q) D_{KL}(\eta, q\zeta + (1 - q)\eta), \quad \zeta, \eta \in \Xi^*,$$

for $q \in (0, 1)$. By using Theorem 1, we find

$$\begin{aligned} D_{GJS_q}(\zeta, \eta) &= q [\kappa(q\zeta + (1 - q)\eta) - \kappa(\zeta) + (1 - q) (\zeta - \eta)^t \pi(\zeta)] \\ &\quad + (1 - q) [\kappa(q\zeta + (1 - q)\eta) - \kappa(\eta) + q (\eta - \zeta)^t \pi(\eta)] \\ &= \kappa(q\zeta + (1 - q)\eta) \\ &\quad - [q\kappa(\zeta) + (1 - q)\kappa(\eta)] + q(1 - q) (\zeta - \eta)^t [\pi(\zeta) - \pi(\eta)] \\ &= \ln(A_q(\zeta, \eta)) + q(1 - q) D_J(\zeta, \eta), \end{aligned} \tag{13}$$

for $\zeta, \eta \in \Xi^*$ and $q \in (0, 1)$.

In particular, setting $q = 1/2$ gives the geometric Jensen–Shannon distance:

$$D_{GJS}(\zeta, \eta) = \kappa\left(\frac{\zeta + \eta}{2}\right) - \frac{\kappa(\zeta) + \kappa(\eta)}{2} + \frac{(\zeta - \eta)^t [\pi(\zeta) - \pi(\eta)]}{4}, \quad \zeta, \eta \in \Xi^*.$$

For more details and properties as well as related divergence measures, we refer to [19,22].

Formulas for D_{R_q} , D_{CR_q} , and D_H are readily deduced from Lemma 1.

Theorem 2. Let \mathcal{P} be as in Section 2 with minimal canonical representation (5). Then, for $\zeta, \eta \in \Xi^*$ and $q \in (0, 1)$, we have

$$D_{R_q}(\zeta, \eta) = \frac{1}{q(q - 1)} [\kappa(q\zeta + (1 - q)\eta) - [q\kappa(\zeta) + (1 - q)\kappa(\eta)]],$$

$$D_{CR_q}(\zeta, \eta) = \frac{1}{q(q - 1)} [\exp \{ \kappa(q\zeta + (1 - q)\eta) - [q\kappa(\zeta) + (1 - q)\kappa(\eta)] \} - 1],$$

$$\text{and } D_H(\zeta, \eta) = \left(2 - 2 \exp \left\{ \kappa\left(\frac{\zeta + \eta}{2}\right) - \frac{\kappa(\zeta) + \kappa(\eta)}{2} \right\} \right)^{1/2}.$$

Proof. Since

$$D_{R_q} = \frac{\ln(A_q)}{q(q - 1)}, \quad D_{CR_q} = \frac{A_q - 1}{q(q - 1)}, \quad \text{and } D_H = (2 - 2A_{1/2})^{1/2},$$

the assertions are directly obtained from Lemma 1. \square

It is well-known that

$$\lim_{q \rightarrow 1} D_{R_q}(P_1, P_2) = D_{KL}(P_1, P_2) \quad \text{and} \quad \lim_{q \rightarrow 0} D_{R_q}(P_1, P_2) = D_{KL}(P_2, P_1),$$

such that Formula (9) results from the representation of the Rényi divergence in Theorem 2 by sending q to 1.

The Sharma–Mittal divergence (see [1]) is closely related to the Rényi divergence as well and, by Theorem 2, a representation in EFs is available.

Moreover, representations within EFs for so-called local divergences can be derived as, e.g., the Cressie–Read local divergence, which results from the CR-divergence by multiplying the integrand with some kernel density function; see [23].

Remark 3. Inspecting the proof of Theorem 2, D_{R_q} and D_{CR_q} are seen to be strictly decreasing functions of A_q for $q \in (0, 1)$; for $q = 1/2$, this is also true for D_H . From an inferential point of view, this finding yields that, for fixed $q \in (0, 1)$, test statistics and pivot statistics based on these divergence measures will lead to the same test and confidence region, respectively. This is not the case within some divergence families such as D_{R_q} , $q \in (0, 1)$, where different values of q correspond to different tests and confidence regions, in general.

A more general form of the Hellinger metric is given by

$$D_{H,m}(P_1, P_2) = \left(\int |f_1^{1/m} - f_2^{1/m}|^m d\mu \right)^{1/m}$$

for $m \in \mathbb{N}$, where $D_{H,2} = D_H$; see Formula (8). For $m \in 2\mathbb{N}$, i.e., if m is even, the binomial theorem then yields

$$\begin{aligned} [D_{H,m}(P_1, P_2)]^m &= \int (f_1^{1/m} - f_2^{1/m})^m d\mu \\ &= \sum_{k=0}^m (-1)^k \binom{m}{k} \int f_1^{k/m} f_2^{(m-k)/m} d\mu \\ &= \sum_{k=0}^m (-1)^k \binom{m}{k} A_{k/m}(P_1, P_2), \end{aligned}$$

and inserting for $A_{k/m}$, $k = 1, 1, \dots, m-1$, according to Lemma 1 along with $A_0 \equiv 1 \equiv A_1$ gives a formula for $D_{H,m}$ in terms of the cumulant function of the EF \mathcal{P} in Section 2. This representation is stated in [16].

Note that the representation for A_q in Lemma 1 (and thus the formulas for D_{R_q} and D_{CR_q} in Theorem 2) are also valid for $\zeta, \eta \in \Xi^*$ and $q \in \mathbb{R} \setminus [0, 1]$ as long as $q\zeta + (1-q)\eta \in \Xi^*$ is true. This can be used, e.g., to find formulas for D_{CR_2} and $D_{CR_{-1}}$, which coincide with the Pearson χ^2 -divergence

$$\begin{aligned} D_{\chi^2}(\zeta, \eta) &= \frac{1}{2} \int \frac{(f_\zeta^* - f_\eta^*)^2}{f_\eta^*} d\mu = \frac{1}{2} [A_2(\zeta, \eta) - 1] \\ &= \frac{1}{2} [\exp\{\kappa(2\zeta - \eta) - 2\kappa(\zeta) + \kappa(\eta)\} - 1] \end{aligned}$$

for $\zeta, \eta \in \Xi^*$ with $2\zeta - \eta \in \Xi^*$ and the reverse Pearson χ^2 -divergence (or Neyman χ^2 -divergence) $D_{\chi^2}^*(\zeta, \eta) = D_{\chi^2}(\eta, \zeta)$ for $\zeta, \eta \in \Xi^*$ with $2\eta - \zeta \in \Xi^*$. Here, the restrictions on the parameters are obsolete if $\Xi^* = \mathbb{R}^k$ for some $k \in \mathbb{N}$, which is the case for the EF of Poisson distributions and for any EF of discrete distributions with finite support such as binomial or multinomial distributions (with $n \in \mathbb{N}$ fixed). Moreover, quantities similar to A_q such as $\int f_\zeta^* (f_\eta^*)^\gamma d\mu$ for $\gamma > 0$ arise in the so-called γ -divergence, for which some representations can also be obtained; see [24] (Section 4).

Remark 4. If the assumption of the EF \mathcal{P} to be regular is weakened to \mathcal{P} being steep, Lemma 1 and Theorem 2 remain true; moreover, the formulas in Theorem 1 are valid for ζ lying in the interior of Ξ^* . Steep EFs are full EFs in which boundary points of Ξ^* that belong to Ξ^* satisfy a certain property. A prominent example is provided by the full EF of inverse normal distributions. For details, see, e.g., [13].

The quantity A_q in Formula (12) is the two-dimensional case of the weighted Matusita affinity

$$\rho_{w_1, \dots, w_n}(P_1, \dots, P_n) = \int \left(\prod_{i=1}^n f_i \right)^{w_i} d\mu \tag{14}$$

for distributions P_1, \dots, P_n with μ -densities f_1, \dots, f_n , weights $w_1, \dots, w_n > 0$ satisfying $\sum_{i=1}^n w_i = 1$, and $n \geq 2$; see [4] (p. 49) and [6]. ρ_{w_1, \dots, w_n} in turn, is a generalization of the Matusita affinity

$$\rho_n(P_1, \dots, P_n) = \int \left(\prod_{i=1}^n f_i \right)^{1/n} d\mu$$

introduced in [25,26]. Along the lines of the proof of Lemma 1, we find the representation

$$\rho_{w_1, \dots, w_n}(\zeta^{(1)}, \dots, \zeta^{(n)}) = \exp \left\{ \kappa \left(\sum_{i=1}^n w_i \zeta^{(i)} \right) - \sum_{i=1}^n w_i \kappa(\zeta^{(i)}) \right\}, \quad \zeta^{(1)}, \dots, \zeta^{(n)} \in \Xi^*,$$

for the EF \mathcal{P} in Section 2; cf. [27]. In [4], the quantity in Formula (14) is termed Hellinger transform, and a representation within EFs is stated in Example 1.88.

ρ_{w_1, \dots, w_n} can be used, for instance, as the basis of a homogeneity test (with null hypothesis $H_0 : \zeta^{(1)} = \dots = \zeta^{(n)}$) or in discriminant problems.

For a representation of an extension of the Jeffrey distance to more than two distributions in an EF, the so-called Toussaint divergence, along with statistical applications, we refer to [8].

4. Entropy Measures

The literature on entropy measures, their applications, and their relations to divergence measures is broad. We focus on some selected results and state several simple representations of entropy measures within EFs.

Let the EF in Section 2 be given with $h \equiv 1$, which is the case, e.g., for the one-parameter EFs of geometric distributions and exponential distributions as well as for the two-parameter EF of univariate normal distributions. Formula (5) then yields that

$$\int (f_\zeta^*)^r d\mu = \int e^{r\zeta^t T - r\kappa(\zeta)} d\mu = e^{\kappa(r\zeta) - r\kappa(\zeta)} = J_r(\zeta), \quad \text{say,}$$

for $r > 0$ and $\zeta \in \Xi^*$ with $r\zeta \in \Xi^*$. Note that the latter condition is not that restrictive, since the natural parameter space of a regular EF is usually a cartesian product of the form $A_1 \times \dots \times A_k$ with $A_i \in \{\mathbb{R}, (-\infty, 0), (0, \infty)\}$ for $1 \leq i \leq k$.

The Taneja entropy is then obtained as

$$\begin{aligned} H_T(\zeta) &= -2^{r-1} \int (f_\zeta^*)^r \ln(f_\zeta^*) d\mu = -2^{r-1} \left(\zeta^t \int T e^{r\zeta^t T - r\kappa(\zeta)} d\mu - \kappa(\zeta) J_r(\zeta) \right) \\ &= -2^{r-1} J_r(\zeta) \left(\zeta^t \int T f_{r\zeta}^* d\mu - \kappa(\zeta) \right) = -2^{r-1} e^{\kappa(r\zeta) - r\kappa(\zeta)} \left(\zeta^t \pi(r\zeta) - \kappa(\zeta) \right) \end{aligned}$$

for $r > 0$ and $\zeta \in \Xi^*$ with $r\zeta \in \Xi^*$, which includes the Shannon entropy

$$H_S(\zeta) = - \int f_\zeta^* \ln(f_\zeta^*) d\mu = \kappa(\zeta) - \zeta^t \pi(\zeta), \quad \zeta \in \Xi^*,$$

by setting $r = 1$; see [7,28].

Several other important entropy measures are functions of J_r and therefore admit respective representations in terms of the cumulant function of the EF. Two examples are

provided by the Rényi entropy and the Havrda–Charvát entropy (or Tsallis entropy), which are given by

$$H_{R_r}(\zeta) = \frac{1}{1-r} \ln(J_r(\zeta)) = \frac{\kappa(r\zeta) - r\kappa(\zeta)}{1-r}, \quad r > 0, r \neq 1,$$

and
$$H_{HC_r}(\zeta) = \frac{1}{1-r} (J_r(\zeta) - 1) = \frac{1}{1-r} (e^{\kappa(r\zeta) - r\kappa(\zeta)} - 1), \quad r > 0, r \neq 1,$$

for $\zeta \in \Xi^*$ with $r\zeta \in \Xi^*$; for the definitions, see, e.g., [1]. More generally, the Sharma–Mittal entropy is seen to be

$$\begin{aligned} H_{SM_{r,s}}(\zeta) &= \frac{1}{1-s} \left[(J_r(\zeta))^{\frac{1-s}{1-r}} - 1 \right] \\ &= \frac{1}{1-s} \left[\left(e^{\kappa(r\zeta) - r\kappa(\zeta)} \right)^{\frac{1-s}{1-r}} - 1 \right], \quad r > 0, r \neq 1, \quad s \in \mathbb{R}, s \neq 1, \end{aligned}$$

for $\zeta \in \Xi^*$ with $r\zeta \in \Xi^*$, which yields the representation for H_S as $r = s \rightarrow 1$, for H_{R_r} as $s \rightarrow 1$, and for H_{HC_r} as $s \rightarrow r$; see [29].

If the assumption $h \equiv 1$ is not met, the calculus of the entropies becomes more involved. The Shannon entropy, for instance, is then given by

$$H_S(\zeta) = \kappa(\zeta) - \zeta^t \pi(\zeta) + E_\zeta[\ln(h)], \quad \zeta \in \Xi^*,$$

where the additional additive term $E_\zeta[\ln(h)]$, as it is the mean of $\ln(h)$ under P_ζ^* , will also depend on ζ , in general; see, e.g., [17]. Since

$$\int \left(f_\zeta^* \right)^r d\mu = e^{\kappa(r\zeta) - r\kappa(\zeta)} E_{r\zeta} \left[h^{r-1} \right]$$

for $r > 0$ and $\zeta \in \Xi^*$ with $r\zeta \in \Xi^*$ (cf. [29]), more complicated expressions result for other entropies and require to compute respective moments of h . Of course, we arrive at the same expressions as for the case $h \equiv 1$ if the entropies are introduced with respect to the dominating measure ν , which is neither a counting nor a Lebesgue measure, in general; see Section 2. However, in contrast to divergence measures, entropies usually depend on the dominating measure, such that the resulting entropy values of the distributions will be different.

Representations of Rényi and Shannon entropies for various multivariate distributions including several EFs can be found in [30].

5. Application

As aforementioned, applications of divergence measures in statistical inference have been extensively discussed; see the references in the introduction. As an example, we make use of the representations of the symmetric divergences (distances) in Section 3 to construct confidence regions that are different from the standard rectangles for exponential parameters in a multi-sample situation.

Let $n_1, \dots, n_k \in \mathbb{N}$ and $X_{ij}, 1 \leq i \leq k, 1 \leq j \leq n_i$, be independent random variables, where X_{i1}, \dots, X_{in_i} follow an exponential distribution with (unknown) mean $1/\alpha_i$ for $1 \leq i \leq k$. The overall joint distribution P_α , say, has the density function

$$f_\alpha(x) = e^{\alpha^t T(x) - \kappa(\alpha)}, \tag{15}$$

with the k -dimensional statistic

$$T(x) = - (x_{1\bullet}, \dots, x_{k\bullet})^t, \quad \text{where} \quad x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij}, \quad 1 \leq i \leq k,$$

for $\mathbf{x} = (x_{11}, \dots, x_{1n_1}, \dots, x_{k1}, \dots, x_{kn_k}) \in (0, \infty)^n$, the cumulant function

$$\kappa(\boldsymbol{\alpha}) = -\sum_{i=1}^k n_i \ln(\alpha_i), \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^t \in (0, \infty)^k,$$

and $n = \sum_{i=1}^k n_i$. It is easily verified that $\mathcal{P} = \{P_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in (0, \infty)^k\}$ forms a regular EF with minimal canonical representation (15). The corresponding mean value function is given by

$$\pi(\boldsymbol{\alpha}) = -\left(\frac{n_1}{\alpha_1}, \dots, \frac{n_k}{\alpha_k}\right)^t, \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^t \in (0, \infty)^k.$$

To construct confidence regions for $\boldsymbol{\alpha}$ based on the Jeffrey distance D_J , the resistor-average distance D_{RA} , the distance D_{GA} , the Hellinger metric D_H , and the geometric Jensen–Shannon distance D_{GJS} , we first compute the KL-divergence D_{KL} and the affinity $A_{1/2}$. Note that, by Remark 3, constructing a confidence region based on D_H is equivalent to constructing a confidence region based on either $A_{1/2}$, $D_{R_{1/2}}$, or $D_{CR_{1/2}}$.

For $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^t, \boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^t \in (0, \infty)^k$, we obtain from Theorem 1 that

$$\begin{aligned} D_{KL}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= -\sum_{i=1}^k n_i \ln(\beta_i) + \sum_{i=1}^k n_i \ln(\alpha_i) - \sum_{i=1}^k \frac{n_i}{\alpha_i} (\alpha_i - \beta_i) \\ &= \sum_{i=1}^k n_i \left[\frac{\beta_i}{\alpha_i} - \ln\left(\frac{\beta_i}{\alpha_i}\right) - 1 \right], \end{aligned}$$

such that

$$\begin{aligned} D_J(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= D_{KL}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + D_{KL}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\ &= \sum_{i=1}^k n_i \left(\frac{\alpha_i}{\beta_i} + \frac{\beta_i}{\alpha_i} - 2 \right). \end{aligned}$$

D_{RA} and D_{GA} are then computed by inserting for D_{KL} and D_J in Formulas (10) and (11). Applying Lemma 1 yields

$$\begin{aligned} A_{1/2}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \left[\prod_{i=1}^k \left(\frac{\alpha_i + \beta_i}{2} \right)^{-n_i} \right] \left(\prod_{i=1}^k \alpha_i^{n_i/2} \right) \left(\prod_{i=1}^k \beta_i^{n_i/2} \right) \\ &= \prod_{i=1}^k \left[\frac{1}{2} \left(\sqrt{\frac{\alpha_i}{\beta_i}} + \sqrt{\frac{\beta_i}{\alpha_i}} \right) \right]^{-n_i}, \end{aligned}$$

which gives $D_H(\boldsymbol{\alpha}, \boldsymbol{\beta}) = [2 - 2A_{1/2}(\boldsymbol{\alpha}, \boldsymbol{\beta})]^{1/2}$ by inserting, and, by using Formula (13), also leads to

$$\begin{aligned} D_{GJS}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \ln(A_{1/2}(\boldsymbol{\alpha}, \boldsymbol{\beta})) + \frac{D_J(\boldsymbol{\alpha}, \boldsymbol{\beta})}{4} \\ &= \frac{1}{4} \sum_{i=1}^k n_i \left[\frac{\alpha_i}{\beta_i} + \frac{\beta_i}{\alpha_i} - 4 \ln \left(\frac{1}{2} \left(\sqrt{\frac{\alpha_i}{\beta_i}} + \sqrt{\frac{\beta_i}{\alpha_i}} \right) \right) - 2 \right]. \end{aligned}$$

The MLE $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_k)^t$ of $\boldsymbol{\alpha}$ based on $\mathbf{X} = (X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k})$, is given by

$$\hat{\boldsymbol{\alpha}} = \left(\frac{n_1}{X_{1\bullet}}, \dots, \frac{n_k}{X_{k\bullet}} \right)^t,$$

where $\hat{\alpha}_1, \dots, \hat{\alpha}_k$ are independent. By inserting the random distances $D_J(\hat{\alpha}, \alpha)$, $D_{RA}(\hat{\alpha}, \alpha)$, $D_{GA}(\hat{\alpha}, \alpha)$, $D_H(\hat{\alpha}, \alpha)$, and $D_{GJS}(\hat{\alpha}, \alpha)$ turn out to depend on X only through the vector $(\alpha_1/\hat{\alpha}_1, \dots, \alpha_k/\hat{\alpha}_k)^t$ of component-wise ratios, where $\alpha_i/\hat{\alpha}_i$ has a gamma distribution with shape parameter n_i , scale parameter $1/n_i$, and mean 1 for $1 \leq i \leq k$. Since these ratios are moreover independent, the above random distances form pivot statistics with distributions free of α .

Now, confidence regions for α with confidence level $p \in (0, 1)$ are given by

$$C_{\bullet} = \left\{ \alpha \in (0, \infty)^k : D_{\bullet}(\hat{\alpha}, \alpha) \leq c_{\bullet}(p) \right\},$$

where $c_{\bullet}(p)$ denotes the p -quantile of $D_{\bullet}(\hat{\alpha}, \alpha)$ for $\bullet = J, RA, GA, H, GJS$, numerical values of which can readily be obtained via Monte Carlo simulation by sampling from gamma distributions.

Confidence regions for the mean vector $m = (1/\alpha_1, \dots, 1/\alpha_k)^t$ with confidence level $p \in (0, 1)$ are then given by

$$\tilde{C}_{\bullet} = \left\{ \left(\frac{1}{\alpha_1}, \dots, \frac{1}{\alpha_k} \right)^t \in (0, \infty)^k : (\alpha_1, \dots, \alpha_k)^t \in C_{\bullet} \right\}$$

for $\bullet = J, RA, GA, H, GJS$.

In Figures 1 and 2, realizations of \tilde{C}_J , \tilde{C}_{RA} , \tilde{C}_{GA} , \tilde{C}_H , and \tilde{C}_{GJS} are depicted for the two-sample case ($k = 2$) and some sample sizes n_1, n_2 and values of $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)^t$, where the confidence level is chosen as $p = 90\%$. Additionally, realizations of the standard confidence region

$$R = \left[\frac{2n_1}{\hat{\alpha}_1 \chi_{1-q}^2(2n_1)}, \frac{2n_1}{\hat{\alpha}_1 \chi_q^2(2n_1)} \right] \times \left[\frac{2n_2}{\hat{\alpha}_2 \chi_{1-q}^2(2n_2)}, \frac{2n_2}{\hat{\alpha}_2 \chi_q^2(2n_2)} \right]$$

with a confidence level of 90% for $m = (m_1, m_2)^t$ are shown in the figures, where $q = (1 - \sqrt{0.9})/2$ and $\chi_{\gamma}^2(v)$ denotes the γ -quantile of the chi-square distribution with v degrees of freedom.

It is found that over the sample sizes and realizations of $\hat{\alpha}$ considered, the confidence regions \tilde{C}_J , \tilde{C}_{RA} , \tilde{C}_{GA} , \tilde{C}_H , and \tilde{C}_{GJS} are similarly shaped but do not coincide as the plots for different sample sizes show. In terms of (observed) area, all divergence-based confidence regions perform considerably better than the standard rectangle. This finding, however, depends on the parameter of interest, which here is the vector of exponential means; for the divergence-based confidence regions and the standard rectangle for α itself, the contrary assertion is true. Although the divergence-based confidence regions have a smaller area than the standard rectangle, this is not at the cost of large projection lengths with respect to the m_1 - and m_2 -axes, which serve as further characteristics for comparing confidence regions. Monte Carlo simulations may moreover be applied to compute the expected area and projection lengths as well as the coverage probabilities of false parameters for a more rigorous comparison of the performance of the confidence regions, which is beyond the scope of this article.

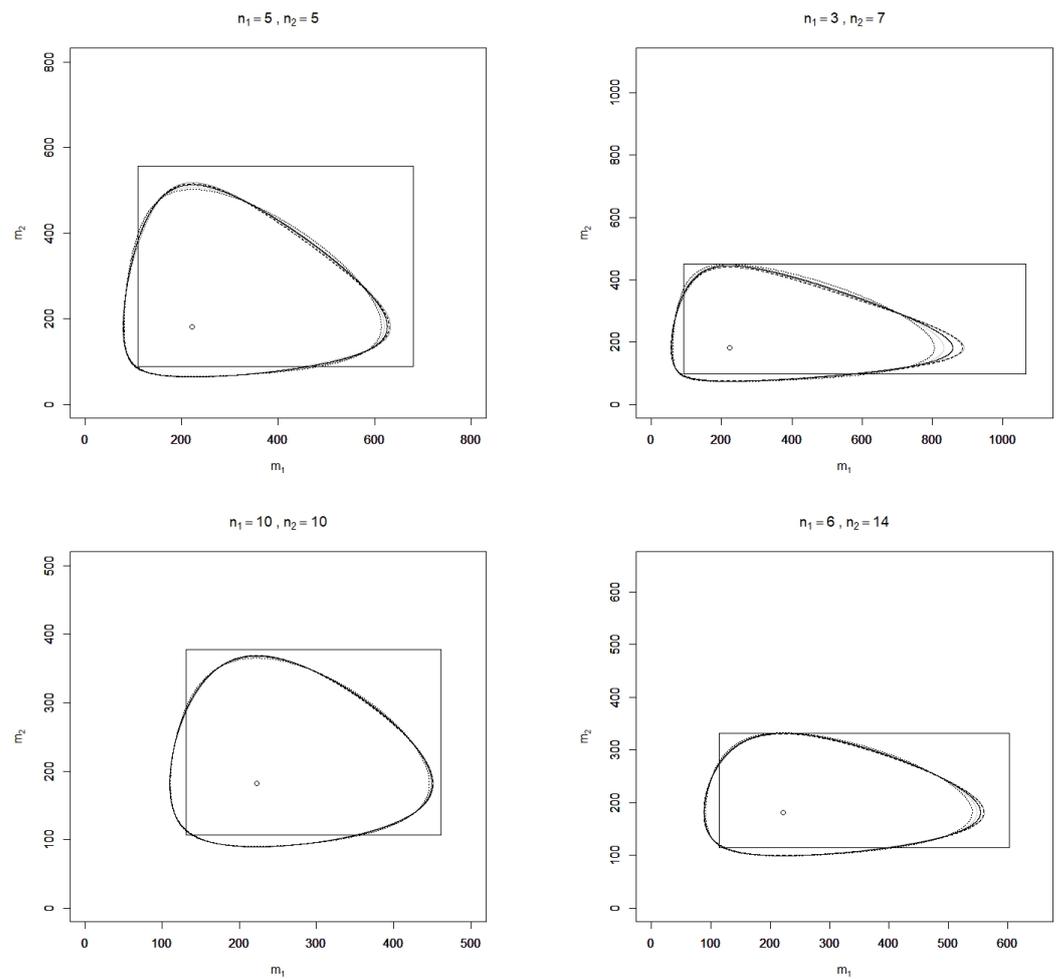


Figure 1. Illustration of the confidence regions \tilde{C}_J (solid light grey line), \tilde{C}_{RA} (solid dark grey line), \tilde{C}_{GA} (solid black line), \tilde{C}_H (dashed black line), \tilde{C}_{GJS} (dotted black line), and R (rectangle) for the mean vector $\mathbf{m} = (m_1, m_2)^t$ with level 90% and sample sizes n_1, n_2 based on a realization $\hat{\mathbf{a}} = (0.0045, 0.0055)^t$, respectively $\hat{\mathbf{m}} = (222.2, 181.8)^t$ of the MLE (circle).

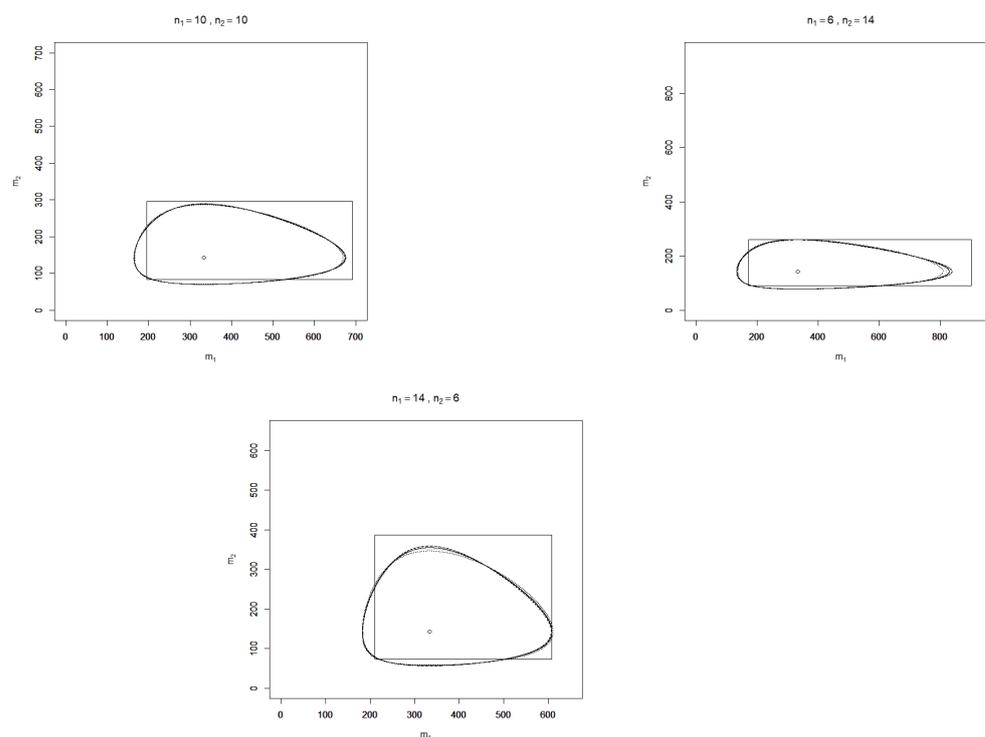


Figure 2. Illustration of the confidence regions \tilde{C}_I (solid light grey line), \tilde{C}_{RA} (solid dark grey line), \tilde{C}_{GA} (solid black line), \tilde{C}_H (dashed black line), \tilde{C}_{GJS} (dotted black line), and R (rectangle) for the mean vector $\mathbf{m} = (m_1, m_2)^t$ with level 90% and sample sizes n_1, n_2 based on a realization $\hat{\mathbf{x}} = (0.003, 0.007)^t$, respectively $\hat{\mathbf{m}} = (333.3, 142.9)^t$ of the MLE (circle).

Author Contributions: conceptualization, S.B. and U.K.; writing—original draft preparation, S.B.; writing—review and editing, U.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Abbreviations

CR	Cressie–Read
EF	exponential family
KL	Kullback–Leibler
MLE	maximum likelihood estimator

References

- Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
- Liese, F.; Vajda, I. *Convex Statistical Distances*; Teubner: Leipzig, Germany, 1987.
- Vajda, I. *Theory of Statistical Inference and Information*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
- Liese, F.; Miescke, K.J. *Statistical Decision Theory: Estimation, Testing, and Selection*; Springer: New York, NY, USA, 2008.
- Broniatowski, M.; Keziou, A. Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **2009**, *100*, 16–36. [[CrossRef](#)]
- Katzur, A.; Kamps, U. Homogeneity testing via weighted affinity in multiparameter exponential families. *Stat. Methodol.* **2016**, *32*, 77–90. [[CrossRef](#)]
- Menendez, M.L. Shannon’s entropy in exponential families: Statistical applications. *Appl. Math. Lett.* **2000**, *13*, 37–42. [[CrossRef](#)]
- Menéndez, M.; Salicrú, M.; Morales, D.; Pardo, L. Divergence measures between populations: Applications in the exponential family. *Commun. Statist. Theory Methods* **1997**, *26*, 1099–1117. [[CrossRef](#)]
- Morales, D.; Pardo, L.; Pardo, M.C.; Vajda, I. Rényi statistics for testing composite hypotheses in general exponential models. *Statistics* **2004**, *38*, 133–147. [[CrossRef](#)]

10. Toma, A.; Broniatowski, M. Dual divergence estimators and tests: Robustness results. *J. Multivar. Anal.* **2011**, *102*, 20–36. [[CrossRef](#)]
11. Katur, A.; Kamps, U. Classification into Kullback–Leibler balls in exponential families. *J. Multivar. Anal.* **2016**, *150*, 75–90. [[CrossRef](#)]
12. Barndorff-Nielsen, O. *Information and Exponential Families in Statistical Theory*; Wiley: Chichester, UK, 2014.
13. Brown, L.D. *Fundamentals of Statistical Exponential Families*; Institute of Mathematical Statistics: Hayward, CA, USA, 1986.
14. Pfanzagl, J. *Parametric Statistical Theory*; de Gruyter: Berlin, Germany, 1994.
15. Kullback, S. *Information Theory and Statistics*; Wiley: New York, NY, USA, 1959.
16. Huzurbazar, V.S. Exact forms of some invariants for distributions admitting sufficient statistics. *Biometrika* **1955**, *42*, 533–537. [[CrossRef](#)]
17. Nielsen, F.; Nock, R. Entropies and cross-entropies of exponential families. In Proceedings of the 2010 IEEE 17th International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 3621–3624.
18. Johnson, D.; Sinanovic, S. Symmetrizing the Kullback–Leibler distance. *IEEE Trans. Inf. Theory* **2001**. Available online: <https://hdl.handle.net/1911/19969> (accessed on 5 June 2021).
19. Nielsen, F. On the Jensen–Shannon symmetrization of distances relying on abstract means. *Entropy* **2019**, *21*, 485. [[CrossRef](#)]
20. Kailath, T. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* **1967**, *15*, 52–60. [[CrossRef](#)]
21. Vuong, Q.N.; Bedbur, S.; Kamps, U. Distances between models of generalized order statistics. *J. Multivar. Anal.* **2013**, *118*, 24–36. [[CrossRef](#)]
22. Nielsen, F. On a generalization of the Jensen–Shannon divergence and the Jensen–Shannon centroid. *Entropy* **2020**, *22*, 221. [[CrossRef](#)]
23. Avlogiaris, G.; Micheas, A.; Zografos, K. On local divergences between two probability measures. *Metrika* **2016**, *79*, 303–333. [[CrossRef](#)]
24. Fujisawa, H.; Eguchi, S. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* **2008**, *99*, 2053–2081. [[CrossRef](#)]
25. Matusita, K. Decision rules based on the distance, for problems of fit, two samples, and estimation. *Ann. Math. Statist.* **1955**, *26*, 631–640. [[CrossRef](#)]
26. Matusita, K. On the notion of affinity of several distributions and some of its applications. *Ann. Inst. Statist. Math.* **1967**, *19*, 181–192. [[CrossRef](#)]
27. Garren, S.T. Asymptotic distribution of estimated affinity between multiparameter exponential families. *Ann. Inst. Statist. Math.* **2000**, *52*, 426–437. [[CrossRef](#)]
28. Beitollahi, A.; Azhdari, P. Exponential family and Taneja’s entropy. *Appl. Math. Sci.* **2010**, *41*, 2013–2019.
29. Nielsen, F.; Nock, R. A closed-form expression for the Sharma–Mittal entropy of exponential families. *J. Phys. A Math. Theor.* **2012**, *45*, 032003. [[CrossRef](#)]
30. Zografos, K.; Nadarajah, S. Expressions for Rényi and Shannon entropies for multivariate distributions. *Statist. Probab. Lett.* **2005**, *71*, 71–84. [[CrossRef](#)]