

Article

Discrete Versions of Jensen–Fisher, Fisher and Bayes–Fisher Information Measures of Finite Mixture Distributions

Omid Kharazmi ¹ and Narayanaswamy Balakrishnan ^{2,*} 

¹ Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, P.O. Box 518, Rafsanjan, Iran; Omid.kharazmi@vru.ac.ir

² Department of Mathematics and Statistics, McMaster University, Hamilton, ON L8S 4K1, Canada

* Correspondence: bala@mcmaster.ca

Abstract: In this work, we first consider the discrete version of Fisher information measure and then propose Jensen–Fisher information, to develop some associated results. Next, we consider Fisher information and Bayes–Fisher information measures for mixing parameter vector of a finite mixture probability mass function and establish some results. We provide some connections between these measures with some known informational measures such as chi-square divergence, Shannon entropy, Kullback–Leibler, Jeffreys and Jensen–Shannon divergences.

Keywords: Fisher information; Bayes–Fisher information; chi-square divergence; Kullback–Leibler divergence; Jensen–Shannon entropy



Citation: Kharazmi, O.; Balakrishnan, N. Discrete Versions of Jensen–Fisher, Fisher and Bayes–Fisher Information Measures of Finite Mixture Distributions. *Entropy* **2021**, *23*, 363. <https://doi.org/10.3390/e23030363>

Academic Editors: Joaquín Abellán and Yong Deng

Received: 2 February 2021

Accepted: 16 March 2021

Published: 18 March 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

Over the last seven decades, several different criteria have been introduced in the literature for measuring uncertainty in a probabilistic model. Shannon entropy and Fisher information are the most important information measures that have been used rather extensively. Information theory started with Shannon entropy, introduced in the pioneering work of Shannon [1], based on a study of systems described by probability density (or mass) functions. About two decades earlier, Fisher [2] had proposed another information measure, describing the interior properties of a probabilistic model, that plays an important role in likelihood-based inferential methods. Fisher information and Shannon entropy are fundamental criteria in statistical inference, physics, thermodynamics and information theory. Complex systems can be described by means of their behavior (Shannon) and their architecture (Fisher) information. For more discussions, see Zegers [3] and Balakrishnan and Stepanov [4].

Let X be a discrete random variable with probability mass function (PMF) $\mathbf{P} = (p_1, \dots, p_n)$. Then, the Shannon entropy of random variable X is defined as

$$H(X) = H(\mathbf{P}) = - \sum_{i=1}^n p_i \log p_i,$$

where “log” denotes the natural logarithm. For more details, see Shannon [1]. Following the work of Shannon [1], considerable attention has been paid to providing some extensions of Shannon entropy. Jensen–Shannon (JS) divergence is one such important extension of Shannon entropy that has been widely used; see Lin [5]. The Jensen–Shannon divergence between two probability mass functions $\mathbf{P} = (p_1, p_2, \dots, p_n)$ and $\mathbf{Q} = (q_1, q_2, \dots, q_n)$, for $0 \leq \alpha \leq 1$, is defined as

$$JS(\mathbf{P}, \mathbf{Q}; \alpha) = H(\alpha\mathbf{P} + (1 - \alpha)\mathbf{Q}) - \alpha H(\mathbf{P}) - (1 - \alpha)H(\mathbf{Q}).$$

The JS divergence is a smoothed and symmetric version of the most important divergence measure of information theory, namely, Kullback–Leibler divergence. Recently, Jensen–Fisher (JF) and Jensen–Gini (JG) divergence measures have been introduced by Sánchez-Moreno et al. [6] and Mehrli et al. [7], respectively.

In the present paper, motivated by the idea of JS divergence, we consider discrete versions of Fisher information (DFI) and Fisher information distance (DFID), and then develop a new information measure associated with DFI measure. In addition, we provide some results for the Fisher information of a finite mixture probability mass function through a Bayesian perspective. The discrete Fisher information of a random variable X with PMF $\mathbf{P} = (p_1, p_2, \dots, p_n)$ is defined as

$$\mathcal{I}(\mathbf{P}) = \sum_{i=1}^n \frac{(p_{i+1} - p_i)^2}{p_i}, \tag{1}$$

with $p_{n+1} = 0$.

The Fisher information in (1) has been made use of in the processing of complex and stationary signals. For example, the discrete version of Fisher information has been used in detecting epileptic seizures in EEG signals recorded in humans and turtles, in detecting dynamical changes in many non-linear models such as logistic map and Lorenz model, and also in the analysis of geoelectrical signals; see Martin et al. [8], Ramírez-Pacheco et al. [9] and Ramírez-Pacheco et al. [10] for pertinent details.

The discrete Fisher information distance (DFID) between two probability mass functions $\mathbf{P} = (p_1, p_2, \dots, p_n)$ and $\mathbf{Q} = (q_1, q_2, \dots, q_n)$ is defined as

$$\mathcal{D}(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^n \left(\frac{p_{i+1}}{p_i} - \frac{q_{i+1}}{q_i} \right)^2 p_i, \tag{2}$$

where, as above, $p_{n+1} = q_{n+1} = 0$. For some of its properties, one may refer to Ramírez-Pacheco et al. [10] and Johnson [11].

With regard to informational properties of finite mixture models, one may refer to Contreras-Reyes and Cortés [12] and Abid et al. [13]. These authors have provided upper and lower bounds for Shannon and Rényi entropies of non-gaussian finite mixtures, skew-normal and skew-t distributions, respectively. Kolchinsky and Tracey [14] have studied the upper and lower bounds for the entropy of Gaussian mixture distributions using the Bhattacharyya and Kullback–Leibler divergences.

The first purpose of this paper is to propose Jensen–Fisher information for discrete random variables X_1, \dots, X_n , with probability mass functions $\mathbf{P}_1, \dots, \mathbf{P}_n$, respectively. For this purpose, we first define discrete version of Jensen–Fisher information for two PMFs \mathbf{P} and \mathbf{Q} , and then provide some results concerning this new information measure. Then, this idea is extended to the general case of PMFs $\mathbf{P}_1, \dots, \mathbf{P}_n$.

The second purpose of this work is to study Fisher and Bayes–Fisher information measures for the mixing parameter of a finite mixture probability mass function. Let $\mathbf{P}_1, \dots, \mathbf{P}_n$ be n probability mass functions, where $\mathbf{P}_j = (p_{j1}, \dots, p_{jk})$. Then, a finite mixture probability mass function with mixing parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{n-1})$, for $n \geq 2$, is given by $\mathbf{P}_{\boldsymbol{\theta}} = (p_{\boldsymbol{\theta}}^1, \dots, p_{\boldsymbol{\theta}}^k)$, where

$$p_{\boldsymbol{\theta}}^j = \frac{1}{n-1} \sum_{i=1}^{n-1} \theta_i p_{ij} + \left(1 - \frac{\sum_{i=1}^{n-1} \theta_i}{n-1} \right) p_{nj}, \quad j = 1, \dots, k, \tag{3}$$

$0 \leq \theta_i \leq 1, i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} \theta_i \leq 1$.

Let X and Y be two discrete random variables with PMFs $\mathbf{P} = (p_1, \dots, p_n)$ and $\mathbf{Q} = (q_1, \dots, q_n)$, respectively. Then, the Kullback–Leibler (KL) distance between X and Y (or \mathbf{P} and \mathbf{Q}) is defined as

$$KL(X||Y) = KL(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right).$$

The Kullback–Leibler discrimination between Y and X can be defined similarly. For more details, see Kullback and Leibler [15]. The chi-square divergence between PMFs \mathbf{P} and \mathbf{Q} is defined by

$$\chi^2(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{p_i}.$$

For pertinent details, see Broniatowski [16] and Cover and Thomas [17].

The rest of this paper is organized as follows. In Section 2, we first consider discrete version of Fisher information and then propose the discrete Jensen–Fisher information (DJFI) measure. We show that DJFI measure can be represented based on the mixture of discrete Fisher information distance measures. In Section 3, we consider a finite mixture probability mass function and establish some results for the Fisher information measure of the mixing parameter vector. We show that the Fisher information of the mixing parameter vector is connected to chi-square divergence. Next, in Section 4, we discuss the Bayes–Fisher information for the mixing parameter vector of probability mass functions under some prior distributions for the mixing parameter. We then show that this measure is connected to Shannon entropy, Jensen–Shannon entropy, Kullback–Leibler and Jeffreys divergence measures. Finally, we present some concluding remarks in Section 5.

2. Discrete Version of Jensen–Fisher Information

In this section, we first give a result for the DFI measure based on the log-convex and log-concave property of the probability mass function. Then, we define the discrete Jensen–Fisher information measure, and establish some interesting properties of it.

Theorem 1. Let $\mathbf{P} = (p_1, p_2, \dots, p_n)$ be a probability mass function.

- (i) If \mathbf{P} is log-concave, then $\mathcal{I}(\mathbf{P}) \leq p_1$;
- (ii) If \mathbf{P} is log-convex, then $\mathcal{I}(\mathbf{P}) \geq p_1$.

Proof. \mathbf{P} is log-convex (log-concave) if $p_i^2 \geq (\leq) p_{i-1} p_{i+1} \forall i$. So, from the definition of DFI in (1), we have

$$\mathcal{I}(\mathbf{P}) = \sum_{i=1}^n \frac{(p_{i+1} - p_i)^2}{p_i} \geq (\leq) p_1.$$

□

2.1. Discrete Jensen–Fisher Information Based on Two Probability Mass Functions \mathbf{P} and \mathbf{Q}

We first define a symmetric version of DFID measure in (2), and then propose the discrete Jensen–Fisher information measure involving two probability mass functions.

Definition 1. Let \mathbf{P} and \mathbf{Q} be two probability mass functions given by $\mathbf{P} = (p_1, p_2, \dots, p_n)$ and $\mathbf{Q} = (q_1, q_2, \dots, q_n)$. Then, a symmetric version of discrete Fisher information distance in (2) is defined as

$$\mathcal{SD}(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \mathcal{D} \left(\mathbf{P}, \frac{\mathbf{P} + \mathbf{Q}}{2} \right) + \frac{1}{2} \mathcal{D} \left(\mathbf{Q}, \frac{\mathbf{P} + \mathbf{Q}}{2} \right).$$

Definition 2. Let P and Q be two probability mass functions given by $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$. Then, the discrete Jensen–Fisher information is defined as

$$\mathcal{JFI}(P, Q) = \frac{\mathcal{I}(P) + \mathcal{I}(Q)}{2} - \mathcal{I}\left(\frac{P + Q}{2}\right). \tag{4}$$

In the following theorem, we show that the discrete Jensen–Fisher information measure can be obtained based on mixtures of Fisher information distances.

Theorem 2. Let P and Q be two probability mass functions given by $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$. Then,

$$\begin{aligned} \mathcal{JFI}(P, Q) &= \frac{1}{2}\mathcal{D}\left(P, \frac{P + Q}{2}\right) + \frac{1}{2}\mathcal{D}\left(Q, \frac{P + Q}{2}\right) \\ &= \mathcal{SD}(P, Q). \end{aligned}$$

Proof. From the definition of DFID in (2), we get

$$\begin{aligned} \mathcal{D}\left(P, \frac{P + Q}{2}\right) &= \sum_{i=1}^n \left(\frac{p_{i+1}}{p_i} - \frac{p_{i+1} + q_{i+1}}{p_i + q_i}\right)^2 p_i \\ &= \sum_{i=1}^n \left\{ \left(\frac{p_{i+1}}{p_i} - 1\right) - \left(\frac{p_{i+1} + q_{i+1}}{p_i + q_i} - 1\right) \right\}^2 p_i \\ &= \sum_{i=1}^n \left(\frac{p_{i+1}}{p_i} - 1\right)^2 p_i - 2 \sum_{i=1}^n \left(\frac{p_{i+1}}{p_i} - 1\right) \left(\frac{p_{i+1} + q_{i+1}}{p_i + q_i} - 1\right) p_i \\ &\quad + \sum_{i=1}^n \left(\frac{p_{i+1} + q_{i+1}}{p_i + q_i} - 1\right)^2 p_i \\ &= \sum_{i=1}^n \frac{(p_{i+1} - p_i)^2}{p_i} - 2 \sum_{i=1}^n (p_{i+1} - p_i) \frac{(p_{i+1} + q_{i+1} - (p_i + q_i))}{p_i + q_i} \\ &\quad + \sum_{i=1}^n \frac{(p_{i+1} + q_{i+1} - (p_i + q_i))^2}{(p_i + q_i)^2} p_i. \end{aligned}$$

In a similar way, we get

$$\begin{aligned} \mathcal{D}\left(Q, \frac{P + Q}{2}\right) &= \sum_{i=1}^n \frac{(q_{i+1} - q_i)^2}{q_i} - 2 \sum_{i=1}^n (q_{i+1} - q_i) \frac{(p_{i+1} + q_{i+1} - (p_i + q_i))}{p_i + q_i} \\ &\quad + \sum_{i=1}^n \frac{(p_{i+1} + q_{i+1} - (p_i + q_i))^2}{(p_i + q_i)^2} q_i. \end{aligned}$$

Upon adding the above two expressions, we obtain

$$\begin{aligned} \mathcal{D}\left(P, \frac{P + Q}{2}\right) + \mathcal{D}\left(Q, \frac{P + Q}{2}\right) &= \sum_{i=1}^n \frac{(p_{i+1} - p_i)^2}{p_i} + \sum_{i=1}^n \frac{(q_{i+1} - q_i)^2}{q_i} \\ &\quad - \sum_{i=1}^n \frac{(p_{i+1} + q_{i+1} - (p_i + q_i))^2}{p_i + q_i} \\ &= \mathcal{I}(P) + \mathcal{I}(Q) - 2\mathcal{I}\left(\frac{P + Q}{2}\right) \\ &= 2\mathcal{JFI}(P, Q), \end{aligned}$$

as required. \square

Example 1. Let

$$X = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1 - p, \end{cases}$$

and

$$Y = \begin{cases} 1, & \text{with probability } q, \\ 0, & \text{with probability } 1 - q. \end{cases}$$

The corresponding PMFs of variables X and Y are given by $\mathbf{P} = (p, 1 - p)$ and $\mathbf{Q} = (q, 1 - q)$, respectively. From Theorem 2, we then have

$$\mathcal{JFI}(\mathbf{P}, \mathbf{Q}) = \frac{p + q}{2} \left(\frac{1 - p}{p} - \frac{1 - q}{q} \right)^2.$$

A 3D-plot of this $\mathcal{JFI}(\mathbf{P}, \mathbf{Q})$ is presented in Figure 1.

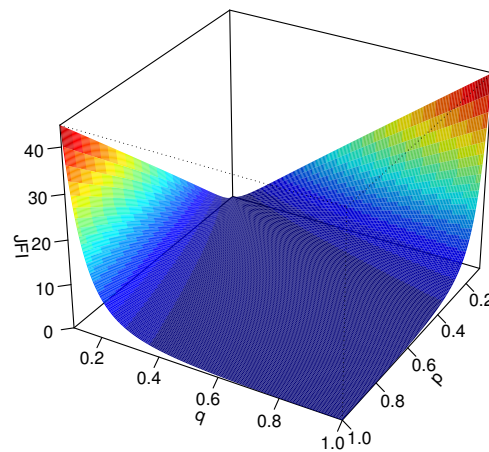


Figure 1. 3D-plot of the DJFI divergence between the PMFs $\mathbf{P} = (p, 1 - p)$ and $\mathbf{Q} = (q, 1 - q)$.

2.2. Discrete Jensen–Fisher Information Based on n Probability Mass Functions $\mathbf{P}_1, \dots, \mathbf{P}_n$

Let $\mathbf{P}_1, \dots, \mathbf{P}_n$ be n probability mass functions, where $\mathbf{P}_i = (p_{i1}, \dots, p_{ik})$. In the following definition, we extend the discrete Jensen–Fisher information measure in (4) to the case of n probability mass functions.

Definition 3. Let $\mathbf{P}_1, \dots, \mathbf{P}_n$ be n probability mass functions given by $\mathbf{P}_i = (p_{i1}, p_{i2}, \dots, p_{ik})$, $i = 1, 2, \dots, n$, with $\sum_{j=1}^k p_{ij} = 1$, and $\alpha_1, \dots, \alpha_n$ be non-negative real numbers such that $\sum_{i=1}^n \alpha_i = 1$. Then, the discrete Jensen–Fisher information (DJFI) based on the n probability mass functions is defined as

$$\begin{aligned} \mathcal{JFI}(\mathbf{P}_1, \dots, \mathbf{P}_n; \underline{\alpha}) &= \sum_{i=1}^n \alpha_i \mathcal{I}(\mathbf{P}_i) - \mathcal{I} \left(\sum_{i=1}^n \alpha_i \mathbf{P}_i \right) \\ &= \sum_{i=1}^n \alpha_i \sum_{j=1}^k \frac{(p_{ij+1} - p_{ij})^2}{p_{ij}} - \sum_{j=1}^k \frac{(\sum_{i=1}^n \alpha_i p_{ij+1} - \sum_{i=1}^n \alpha_i p_{ij})^2}{\sum_{i=1}^n \alpha_i p_{ij}}, \end{aligned} \tag{5}$$

where $\underline{\alpha} = (\alpha_1, \dots, \alpha_n)$.

Theorem 3. Let $\mathbf{P}_1, \dots, \mathbf{P}_n$ be n probability mass functions given by $\mathbf{P}_i = (p_{i1}, p_{i2}, \dots, p_{ik})$, $i = 1, 2, \dots, n$, and $\sum_{j=1}^k p_{ij} = 1$. Then, the DJFI measure can be expressed as a mixture of DFID measures in (2) as follows:

$$\mathcal{JFI}(\mathbf{P}_1, \dots, \mathbf{P}_n, \underline{\alpha}) = \sum_{i=1}^n \alpha_i \mathcal{D}(\mathbf{P}_i, \mathbf{P}_T),$$

where $\mathbf{P}_T = \sum_{i=1}^n \alpha_i \mathbf{P}_i$ is the weighted PMF.

Proof. From the definition in (5), we get

$$\begin{aligned}
 \sum_{i=1}^n \alpha_i \mathcal{D}(\mathbf{P}_i, \mathbf{P}_T) &= \sum_{i=1}^n \alpha_i \sum_{j=1}^k \left(\frac{p_{ij+1}}{p_{ij}} - \frac{\sum_{i=1}^n \alpha_i p_{ij+1}}{\sum_{i=1}^n \alpha_i p_{ij}} \right)^2 p_{ij} \\
 &= \sum_{i=1}^n \alpha_i \sum_{j=1}^k \left\{ \frac{p_{ij+1}}{p_{ij}} - 1 - \left(\frac{\sum_{i=1}^n \alpha_i p_{ij+1}}{\sum_{i=1}^n \alpha_i p_{ij}} - 1 \right) \right\}^2 p_{ij} \\
 &= \sum_{i=1}^n \alpha_i \sum_{j=1}^k \frac{(p_{ij+1} - p_{ij})^2}{p_{ij}} - 2 \sum_{j=1}^k \frac{(\sum_{i=1}^n \alpha_i p_{ij+1} - \sum_{i=1}^n \alpha_i p_{ij})^2}{\sum_{i=1}^n \alpha_i p_{ij}} \\
 &\quad + \sum_{j=1}^k \frac{(\sum_{i=1}^n \alpha_i p_{ij+1} - \sum_{i=1}^n \alpha_i p_{ij})^2}{\sum_{i=1}^n \alpha_i p_{ij}} \\
 &= \sum_{i=1}^n \alpha_i \sum_{j=1}^k \frac{(p_{ij+1} - p_{ij})^2}{p_{ij}} - \sum_{j=1}^k \frac{(\sum_{i=1}^n \alpha_i p_{ij+1} - \sum_{i=1}^n \alpha_i p_{ij})^2}{\sum_{i=1}^n \alpha_i p_{ij}} \\
 &= \sum_{i=1}^n \alpha_i \mathcal{I}(\mathbf{P}_i) - \mathcal{I} \left(\sum_{i=1}^n \alpha_i \mathbf{P}_i \right) \\
 &= \mathcal{JFI}(\mathbf{P}_1, \dots, \mathbf{P}_n, \underline{\alpha}),
 \end{aligned}$$

as required. \square

3. Fisher Information of a Finite Mixture Probability Mass Function

In this section, we discuss Fisher information for parameter θ of a finite mixture probability mass function.

Theorem 4. The Fisher information of PMF in (3) about parameter $\theta_i, i = 1, \dots, n - 1$, is given by

$$\mathcal{I}(\theta_i) = \frac{1}{(\theta_i - (n - 1))^2} \chi^2(\mathbf{P}_{\theta_{-i}}, \mathbf{P}_{\theta}), \quad i = 1, \dots, n - 1, \tag{6}$$

where $\mathbf{P}_{\theta_{-i}} = (p_{\theta_{-i}}^1, \dots, p_{\theta_{-i}}^k)$,

$$p_{\theta_{-i}}^j = \frac{n - 2}{n - 1} p_{ij} + \frac{1}{n - 1} \sum_{t=1, t \neq i}^{n-1} \theta_t p_{tj} + \frac{1}{n - 1} \left(1 - \sum_{t=1, t \neq i}^{n-1} \theta_t \right) p_{nj}, \quad j = 1, \dots, k,$$

and $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_{n-1})$.

Proof. From the definition of Fisher information in (1) and for $i = 1, \dots, n - 1$, we have

$$\begin{aligned}
 \mathcal{I}(\theta_i) &= \sum_{j=1}^k \left[\frac{\partial \log \mathbf{P}_{\theta}}{\partial \theta_i} \right]^2 p_{\theta}^j \\
 &= \frac{1}{(n - 1)^2} \sum_{j=1}^k \frac{(p_{ij} - p_{nj})^2}{(p_{\theta}^j)^2} p_{\theta}^j \\
 &= \frac{1}{(\theta_i - (n - 2))^2} \sum_{j=1}^k \frac{(p_{\theta_{-i}}^j - p_{\theta}^j)^2}{p_{\theta}^j} \\
 &= \frac{1}{(\theta_i - (n - 1))^2} \chi^2(\mathbf{P}_{\theta_{-i}}, \mathbf{P}_{\theta}), \quad i = 1, \dots, n - 1,
 \end{aligned} \tag{7}$$

where the third equation follows from the fact that, for $i = 1, \dots, n - 1$,

$$p_{ij} - p_{nj} = \frac{n-1}{\theta_i - (n-2)} (p_{\theta}^j - p_{\theta_{-i}}^j).$$

□

4. Bayes–Fisher Information of a Finite Mixture Probability Mass Function

In this section, we discuss Bayes–Fisher information for the mixing parameter vector θ of the finite mixture probability mass function in (3) under some prior distributions for the mixing parameter vector. We now introduce two notations that will be used in the sequel. Consider the parameter vector $\theta = (\theta_1, \dots, \theta_{n-1})$, and then define $(0_i, \theta) = (\theta_1, \dots, \theta_{i-1}, 0, \theta_{i+1}, \dots, \theta_{n-1})$ and $(1_i, \theta) = (\theta_1, \dots, \theta_{i-1}, 1, \theta_{i+1}, \dots, \theta_{n-1})$.

Theorem 5. *The Bayes–Fisher information for parameter θ_i , $i = 1, \dots, n - 1$, of the finite mixture PMF in (3), under the uniform prior on $[0, 1]$, is given by*

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= KL(\mathbf{P}_{(1_i, \theta)}, \mathbf{P}_{(0_i, \theta)}) + KL(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}) \\ &= J(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}), \end{aligned}$$

where $\mathbf{P}_{(1_i, \theta)} = (p_{(1_i, \theta)}^1, \dots, p_{(1_i, \theta)}^n)$, with

$$p_{(1_i, \theta)}^j = \frac{1}{n-1} p_{ij} + \frac{1}{n-1} \sum_{t=1, t \neq i}^{n-1} \theta_t + \left(1 - \frac{1}{n-1} \left(1 + \sum_{t=1, t \neq i}^{n-1} \theta_t\right)\right) p_{nj}, \tag{8}$$

and $\mathbf{P}_{(0_i, \theta)} = (p_{(0_i, \theta)}^1, \dots, p_{(0_i, \theta)}^n)$, with

$$p_{(0_i, \theta)}^j = \frac{1}{n-1} \sum_{t=1, t \neq i}^{n-1} \theta_t p_{tj} + \left(1 - \frac{1}{n-1} \sum_{t=1, t \neq i}^{n-1} \theta_t\right) p_{nj}, \tag{9}$$

and J corresponds to Jeffreys’ divergence.

Proof. By definition and from (7), for $i = 1, \dots, n - 1$, we have

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= E[\mathcal{I}(\Theta_i)] = \frac{1}{(n-1)^2} \int_0^1 \left\{ \sum_{j=1}^k \frac{(p_{ij} - p_{nj})^2}{p_{\theta}^j} \right\} d\theta_i \\ &= \frac{1}{n-1} \sum_{j=1}^k (p_{ij} - p_{nj}) \left\{ \int_0^1 \frac{1}{n-1} \frac{p_{ij} - p_{nj}}{p_{\theta}^j} d\theta_i \right\} \\ &= \frac{1}{n-1} \sum_{j=1}^k (p_{ij} - p_{nj}) \left\{ \log(p_{\theta}^j) \Big|_0^1 \right\}. \end{aligned} \tag{10}$$

On the other hand, we have

$$p_{(1_i, \theta)} - p_{(0_i, \theta)} = \frac{1}{n-1} (p_{ij} - p_{nj}). \tag{11}$$

Hence, upon substituting (11) into (10), we obtain

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= \frac{1}{n-1} \sum_{j=1}^k (p_{ij} - p_{nj}) \log \left\{ \frac{p_{(1_i, \theta)}^j}{p_{(0_i, \theta)}^j} \right\} \\ &= \sum_{j=1}^k (p_{(1_i, \theta)} - p_{(0_i, \theta)}) \log \left\{ \frac{p_{(1_i, \theta)}^j}{p_{(0_i, \theta)}^j} \right\} \\ &= KL(\mathbf{P}_{(1_i, \theta)}, \mathbf{P}_{(0_i, \theta)}) + KL(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}) \\ &= J(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}), \end{aligned}$$

as required. \square

Theorem 6. For the mixture model with PMF in (3), we have the following:

- (i) The Bayes–Fisher information for $\theta_i, i = 1, \dots, n - 1$, under Beta(2, 1) prior with PMF $\pi(\theta_i) = 2\theta_i, \theta_i \in [0, 1]$, is

$$\tilde{\mathcal{I}}(\theta_i) = 2KL(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}), i = 1, \dots, n - 1;$$

- (ii) The Bayes–Fisher information for parameter $\theta_i, i = 1, \dots, n - 1$, under Beta(1, 2) prior with PMF $\pi(\theta_i) = 2(1 - \theta_i), \theta_i \in [0, 1]$, is

$$\tilde{\mathcal{I}}(\theta_i) = 2KL(\mathbf{P}_{(1_i, \theta)}, \mathbf{P}_{(0_i, \theta)}), i = 1, \dots, n - 1.$$

Proof. By definition, and from (7), for $i = 1, \dots, n - 1$, we have

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= E[\mathcal{I}(\Theta_i)] = \frac{1}{(n - 1)^2} \int_0^1 \left\{ \sum_{j=1}^k \frac{(p_{ij} - p_{nj})^2}{p_{\theta}^j} \right\} \pi(\theta_i) d\theta_i \\ &= \frac{2}{n - 1} \sum_{j=1}^k (p_{ij} - p_{nj}) \left\{ \int_0^1 \frac{\theta_i}{n - 1} \frac{p_{ij} - p_{nj}}{p_{\theta}^j} d\theta_i \right\} \\ &= \frac{2}{n - 1} \sum_{j=1}^k (p_{ij} - p_{nj}) \left\{ \int_0^1 \frac{p_{\theta}^j - p_{(0_i, \theta)}^j}{p_{\theta}(x)} d\theta_i \right\} \\ &= \frac{2}{n - 1} \sum_{j=1}^k (p_{ij} - p_{nj}) \left\{ 1 - \frac{(n - 1)p_{(0_i, \theta)}^j}{p_{ij} - p_{nj}} \log(p_{\theta}^j) \Big|_0^1 \right\} \\ &= -2 \sum_{j=1}^k p_{(0_i, \theta)}^j \log \left\{ \frac{p_{(1_i, \theta)}^j}{p_{(0_i, \theta)}^j} \right\} \\ &= 2 \sum_{j=1}^k p_{(0_i, \theta)}^j \log \left\{ \frac{p_{(0_i, \theta)}^j}{p_{(1_i, \theta)}^j} \right\} \\ &= 2KL(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}), \end{aligned}$$

as required for Part (i). Part (ii) can be proved in an analogous manner.

Let us now consider the following general triangular prior for the parameter $\theta_i, i = 1, \dots, n - 1$:

$$\pi_{\alpha}(\theta_i) = \begin{cases} \frac{2\theta_i}{\alpha}, & 0 < \theta_i \leq \alpha, \\ \frac{2(1 - \theta_i)}{1 - \alpha}, & \alpha \leq \theta_i < 1, \end{cases} \tag{12}$$

for some $\alpha \in (0, 1)$. \square

Theorem 7. The Bayes–Fisher information for parameter $\theta_i, i = 1, \dots, n - 1$, with the general triangular prior with density $\pi_{\alpha}(\theta_i)$ in (12), is given by

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= \frac{2}{\alpha(1 - \alpha)} \left[\alpha KL(\mathbf{P}_{(1_i, \theta)}, \mathbf{P}_{\alpha}) + (1 - \alpha) KL(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{\alpha}) \right] \\ &= \frac{2}{\alpha(1 - \alpha)} JS(\mathbf{P}_{(0_i, \theta)}, \mathbf{P}_{(1_i, \theta)}; \alpha), \end{aligned}$$

where $\mathbf{P}_\alpha = (p_\alpha^1, \dots, p_\alpha^k)$ is a finite mixture PMF, with

$$p_\alpha^j = \frac{\alpha}{n-1} p_{ij} + \frac{1}{n-1} \sum_{l=1, l \neq i}^{n-1} \theta_l p_{lj} + \left(1 - \frac{1}{n-1} \left(\alpha + \sum_{l=1, l \neq i}^{n-1} \theta_l\right)\right) p_{nj}$$

and $\mathbf{P}_{(1,i,\theta)}$ and $\mathbf{P}_{(0,i,\theta)}$ are as defined in (8) and (9), respectively.

Proof. From the assumptions made, for $i = 1, \dots, n-1$, we have

$$\begin{aligned} \tilde{\mathcal{I}}(\theta_i) &= E[\mathcal{I}(\Theta_i)] = \int_0^\alpha \mathcal{I}(\theta_i) \pi_\alpha d\theta_i + \int_\alpha^1 \mathcal{I}(\theta_i) \pi_\alpha d\theta_i \\ &= \frac{2}{(n-1)\alpha} \sum_{j=1}^k (p_{ij} - p_{nj}) \left[\int_0^\alpha \frac{\theta_i}{n-1} \frac{p_{ij} - p_{nj}}{p_\theta^j} d\theta_i \right] \\ &\quad + \frac{2}{(n-1)(1-\alpha)} \sum_{j=1}^k (p_{ij} - p_{nj}) \left[\int_\alpha^1 \frac{1-\theta_i}{n-1} \frac{(p_{ij} - p_{nj})}{p_\theta^j} d\theta_i \right] \\ &= \frac{2}{(n-1)\alpha} \sum_{j=1}^k (p_{ij} - p_{nj}) \left[\int_0^\alpha \left(1 - \frac{p_{(0,i,\theta)}^j}{p_\theta^j}\right) d\theta_i \right] \\ &\quad - \frac{2}{(n-1)(1-\alpha)} \sum_{j=1}^k (p_{ij} - p_{nj}) \left[\int_0^\alpha \left(1 - \frac{p_{(1,i,\theta)}^j}{p_\theta^j}\right) d\theta_i \right] \\ &= \frac{2}{\alpha} \sum_{j=1}^k p_{(0,i,\theta)}^j \log \left\{ \frac{p_\alpha^j}{p_{(0,i,\theta)}^j} \right\} + \frac{2}{1-\alpha} \sum_{j=1}^k p_{(1,i,\theta)}^j \log \left\{ \frac{p_{(1,i,\theta)}^j}{p_\alpha^j} \right\} \\ &= \frac{2}{\alpha(1-\alpha)} \left[\alpha KL(\mathbf{P}_{(1,i,\theta)}, \mathbf{P}_\alpha) + (1-\alpha) KL(\mathbf{P}_{(0,i,\theta)}, \mathbf{P}_\alpha) \right] \\ &= \frac{2}{\alpha(1-\alpha)} JS(\mathbf{P}_{(0,i,\theta)}, \mathbf{P}_{(1,i,\theta)}; \alpha), \end{aligned}$$

as required. \square

5. Concluding Remarks

In this paper, we have introduced the discrete version of Jensen–Fisher information measure, and have shown that this information measure can be expressed as a mixture of discrete Fisher information distance measures. Further, we have considered a finite mixture probability mass function and have derived Fisher information and Bayes–Fisher information for the mixing parameter vector. We have shown that the Fisher information for the mixing parameter is connected to chi-square divergence. We have also studied the Bayes–Fisher information for the mixing parameter of a finite mixture model under some prior distributions. These results have provided connections between the Bayes–Fisher information and some known informational measures such as Shannon entropy, Kullback–Leibler, Jeffreys and Jensen–Shannon divergence measures.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

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

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
2. Fisher, R.A. Tests of significance in harmonic analysis. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1929**, *125*, 54–59.
3. Zegers, P. Fisher information properties. *Entropy* **2015**, *17*, 4918–4939. [[CrossRef](#)]
4. Balakrishnan, N.; Stepanov, A. On the Fisher information in record data. *Stat. Probab. Lett.* **2006**, *76*, 537–545. [[CrossRef](#)]
5. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [[CrossRef](#)]
6. Sánchez-Moreno, P.; Zarzo, A.; Dehesa, J.S. Jensen divergence based on Fisher's information. *J. Phys. A Math. Theor.* **2012**, *45*, 125305. [[CrossRef](#)]
7. Mehrali, Y.; Asadi, M.; Kharazmi, O. A Jensen-Gini measure of divergence with application in parameter estimation. *Metron* **2018**, *76*, 115–131. [[CrossRef](#)]
8. Martin, M.T.; Pennini, F.; Plastino, A. Fisher's information and the analysis of complex signals. *Phys. Lett. A* **1999**, *256*, 173–180. [[CrossRef](#)]
9. Ramírez-Pacheco, J.; Torres-Román, D.; Rizo-Dominguez, L.; Trejo-Sanchez, J.; Manzano-Pinzón, F. Wavelet Fisher's information measure of $1/f^\alpha$ signals. *Entropy* **2011**, *13*, 1648–1663. [[CrossRef](#)]
10. Ramírez-Pacheco, J.; Torres-Román, D.; Argaez-Xool, J.; Rizo-Dominguez, L.; Trejo-Sanchez, J.; Manzano-Pinzón, F. Wavelet q -Fisher information for scaling signal analysis. *Entropy* **2012**, *14*, 1478–1500. [[CrossRef](#)]
11. Johnson, O. *Information Theory and the Central Limit Theorem*; World Scientific Publishers: Singapore, 2004.
12. Contreras-Reyes, J.E.; Cortés, D.D. Bounds on Rényi and Shannon entropies for finite mixtures of multivariate skew-normal distributions: Application to swordfish (*Xiphias gladius linnaeus*). *Entropy* **2017**, *18*, 382. [[CrossRef](#)]
13. Abid, S.H.; Quaez, U.J.; Contreras-Reyes, J.E. An information-theoretic approach for multivariate skew-t distributions and applications. *Mathematics* **2021**, *9*, 146. [[CrossRef](#)]
14. Kolchinsky, A.; Tracey, B.D. Estimating mixture entropy with pairwise distances. *Entropy* **2017**, *19*, 361. [[CrossRef](#)]
15. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
16. Broniatowski, M. Minimum divergence estimators, Maximum likelihood and the generalized bootstrap. *Entropy* **2021**, *23*, 185. [[CrossRef](#)] [[PubMed](#)]
17. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2006.