Georgia Southern University

Digital Commons@Georgia Southern

2021

Variants of Mixtures: Information Properties and Applications

Omid Ardakani
Georgia Southern University

Majid Asadi
University of Isfahan

Nader Ebrahimi
Northern Illinois University

Ehsan Soofi
University of Wisconsin - Milwaukee

Variants of Mixtures: Information Properties and Applications

Omid M. Ardakani [1], Majid Asadi [2], Nader Ebrahimi [3], and Ehsan S. Soofi [4]

[1] Department of Economics, Georgia Southern University, Savannah, Georgia, USA.

[2] Department of Statistics, University of Isfahan, Isfahan, Iran.

[3] Department of Statistics, Northern Illinois University, DeKalb, Illinois, USA.

[4] Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA.

Abstract In recent years, we have studied information properties of various types of mixtures of probability distributions and introduced a new type, which includes previously known mixtures as special cases. These studies are disseminated in different fields: reliability engineering, econometrics, operations research, probability, the information theory, and data mining. This paper presents a holistic view of these studies and provides further insights and examples. We note that the insightful probabilistic formulation of the mixing parameters stipulated by Behboodian (1972) is required for a representation of the well-known information measure of the arithmetic mixture. Applications of this information measure presented in this paper include lifetime modeling, system reliability, measuring uncertainty and disagreement of forecasters, probability modeling with partial information, and information loss of kernel estimation. Probabilistic formulations of the mixing weights for various types of mixtures provide the Bayes-Fisher information and the Bayes risk of the mean residual function.

Keywords. Arithmetic Mixture, Geometric Mixture, Jensen-Shannon, Kullback-Leibler.

Omid M. Ardakani (oardakani@georgiasouthern.edu)

Majid Asadi (m.asadi@sci.ui.ac.ir)

Nader Ebrahimi (nebrahim@niu.edu)

Corresponding Author: Ehsan S. Soofi (esoofi@uwm.edu)

1 Introduction

A probability distribution has various representations. The cumulative distribution function (CDF) of random variable X gives $F(x) = P(X \leq x)$. In probability, statistics, and related fields, a mixture usually refers to the weighted arithmetic average of two or more CDFs. We will call this type of mixture the *arithmetic mixture* of CDFs and denote its CDF by F_a. To simplify the exposition, we begin with two components,

$$F_a(x) = pF_1(x) + (1 - p)F_2(x), \quad 0 < p < 1. \tag{1.1}$$

Finite, countable, and continuous mixtures will be presented in the sequel. The weights are given by a nonnegative function $\pi(p)$ that sums (integrates) to one for the finite and countable (continuous) mixtures. Behboodian (1972) provided the following probabilistic interpretation of the weight function in a mixture:

$$F_a(x) = P(D_1)F(x|D_1) + P(D_2)F(x|D_2), \tag{1.2}$$

where $D_i, i = 1, 2$ denote the event that X is drawn from F_i. Here, $p_i = P(D_i), i = 1, 2$ are assigned by the weight function $\pi(p)$, $F_i(x) = F(x|D_i), i = 1, 2$ are conditional distributions of $X|D_i$, and F_a is the marginal distribution of X. A mixture with a PDF for the weight function will be called *probabilistic mixture*. This formulation is necessary for a representation of the information measure of mixture and for the Bayes-Fisher information discussed in Section 4.

The survival function (SF) representation of F is $\bar{F}(x) = P(X > x)$ and for a continuous X, the probability density function (PDF) representation of F is $f(x) = dF(x)/dx$, when a density exists. The SF and PDF representations of F_a are:

$$\bar{F}_a(x) = p\bar{F}_1(x) + (1 - p)\bar{F}_2(x), \tag{1.3}$$

$$f_a(x) = pf_1(x) + (1 - p)f_2(x). \tag{1.4}$$

Two other well-known representations of F are the hazard (or failure) rate (HR) and the mean residual (MR) functions, which are of interest when X is a continuous nonnegative random variable with a PDF f. The HR function is defined by $r(x) = \frac{f(x)}{\bar{F}(x)}, \bar{F}(x) > 0$. The MR function (or mean excess) of a random variable with a finite mean μ is the mean of its residual or excess amount beyond a threshold,

$$m(\tau) = E_{X > \tau}(X - \tau | X > \tau), \quad \tau \geq 0, \tag{1.5}$$

where $E_{X > \tau}$ denotes the expectation with respect to the residual PDF,

$$f(x|\tau) = \frac{f(x)}{\bar{F}(\tau)}, \quad x > \tau, \ \bar{F}(\tau) > 0. \tag{1.6}$$

A distribution also can be represented by the cumulative HR function $\Lambda = -\log \bar{F}$ and the odds ratio (OR) function, $\psi(x) = \frac{F(x)}{\bar{F}(x)}$. The HR, MR, cumulative HR, and OR of F_a are not the arithmetic mixtures of r_i, m_i, Λ_i, and ψ_i corresponding to F_i of F_a.

In some application areas the arithmetic mixture of HR functions are useful. For example, for modeling the reliability of heterogeneous items, Block and Savits (1997) use F_a for the analysis of time to failure in the "burn in" problem. But Lynn and Singpurwalla (1997) argue in favor of using the Bayesian predictive HR function which is the arithmetic mixture of HRs, where the weight function gives the probability that an item's lifetime hazard rate is r_i. For two items, we have

$$r_g(x) = pr_1(x) + (1-p)r_2(x), \quad 0 < p < 1, \tag{1.7}$$

where the subscript g signifies that the corresponding SF is the following *geometric mixture* of the SFs corresponding to r_i:

$$\bar{F}_g(x) = \bar{F}_1^p(x)\bar{F}_2^{1-p}(x), \quad 0 < p < 1. \tag{1.8}$$

This model is called in Asadi et al. (2018) as mixture hazards model and the *generalized escort* of SFs. (In physics, a PDF constructed by the normalized power of another PDF is called escort distribution). The corresponding CDF, $1 - \bar{F}_g(x)$ and PDF are not the geometric mixtures of F_i and f_i.

We have studied the information properties of the arithmetic and geometric mixtures in various contexts, and introduced a more general mixture model based on the weighted power mean of SFs. We have developed applications of these information measures for comparison of coherent systems (Asadi et al., 2016), measuring the uncertainty and disagreement of economic forecasters (Shoja and Soofi, 2017), the information loss due to the kernel estimation (Beheshti et al., 2019), derivation optimal reliability models (Asadi et al., 2018), quantifying information about the mixing parameter p (Asadi et al., 2019a), and visualization plots for distinguishing components of mixture distributions (Ardakani et al., 2020). The general mixture that we have introduced is called α-mixture of SFs, where $\alpha \in \mathfrak{R}$. This is a broad family of mixtures, which includes (1.3), (1.8), and harmonic mixtures of SFs. These studies

are disseminated separately in different fields (reliability engineering, econometrics, operations research, probability, the information theory, and data mining).

The objective of this paper is to provide a holistic view of the finding of the aforementioned studies and further insights about various types of mixtures. In particular, we underscore that the insightful probabilistic representation (1.2) of Behboodian (1972) is required for a representation of the information measure of the arithmetic mixture. In addition, we present a new example which uses the estimates of the mean and variance of the arithmetic mixture of normal distributions studied by Behboodian (1970). Our previous studies considered mixtures of univariate continuous distributions. We will point out that several results are applicable to multivariate and discrete distributions.

This paper is organized as follows. Section 2 presents the information measure of f_a and its recent applications in three fields: system reliability, uncertainty and disagreement of forecasters, and kernel estimation. Section 3 gives an overview of the information properties of the geometric and α mixtures of PDFs with more interpretations and examples, and outlines the stochastic and HR properties of the α mixtures of SFs with a new result on the equivalence between the harmonic mixture of SFs and the arithmetic mixture of ORs. Section 4 gives an overview of the information properties of the arithmetic and geometric mixtures and a more general mixture of two PDFs. An example relates an information measure to Behboodian (1970). Section 5 points out the applicability of information measures to the multivariate, discrete, and categorical distributions. Section 6 concludes the paper and offers a few topics for future research.

2   Arithmetic Mixture of PDFs

In this section, we present the information measure of the finite mixture,

$$f_a(x) = \sum_{i=1}^{n} p_i f_i(x), \quad p_i > 0, \quad \sum_{i=1}^{n} p_i = 1. \tag{2.1}$$

The Jensen-Shannon (JS) divergence of the mixture PDF f_a has the following information representations:

$$JS(f_a) = JS(f_a; P) \quad = \quad H(f_a) - \sum_{i=1}^{n} p_i H(f_i), \tag{2.2}$$

$$= \quad \sum_{i=1}^{n} p_i K(f_i : f_a), \tag{2.3}$$

where P represents $p_1, \ldots, p_n$ and the information measures in (2.2) and (2.3) are as follows:

$$H(f) = - \int f(x) \log f(x) dx,$$

is the Shannon entropy, provided that the integral converges, and

$$K(f_i : f_a) = \int f_i(x) \log \frac{f_i(x)}{f_a(x)} dx,$$

is the Kullback-Leibler (KL) divergence (note that f_i is absolutely continuous with respect to f_a). Lin (1992) introduced the KL representation with $n = 2$ and $p = 1/2$ and defined (2.2) as its generalization.

Two calibrations are available for the JS divergence. The information distinguishability (ID) index is defined by the normalization of the KL divergence $ID(K) = 1 - e^{-2K}$ is proposed by Soofi et al. (1995). This index is bounded as $0 \leq ID(K) < 1$. McCulloch (1989) proposed a biased coin calibration of the KL divergence defined by $P_b(K) = .5(1 + ID(K)^{1/2})$. This index maps the difficulty of discriminating between a fair coin and a biased coin with probability $P_b > .5$.

2.1 Bounds on JS Divergence

It is known that the JS is bounded as follows:

$$JS(f_a) \leq \min \{JB(f_a), H(P)\}, \tag{2.4}$$

where $H(P) = - \sum_{i=1}^{n} p_i \log p_i$ is the discrete entropy,

$$JB(f_a) = \sum_{i<j}^{n} p_i p_j J(f_i, f_j), \tag{2.5}$$

and $J(f_i, f_j) = K(f_i : f_j) + K(f_j : f_i)$ is the Jeffreys divergence; the inequality in terms of $H(P)$ is implied by another representation of the mutual information (4.1), which was

first observed by Wang and Madiman (2014); and in terms of $JB(f_a)$ is given in Kullback (1959, p. 23) as an exercise and is shown in Asadi et al. (2016).

The entropy bound in (2.4) is distribution-free and simple. The pairwise divergence bound in (2.5) requires that all the mixture components are absolutely continuous with respect to each other. Next, we propose a bound for the mixture of maximum entropy (ME) distributions which does not require this assumption.

Let $\theta_r = E[T_r(X)]$ be the rth type of "moment" of f_a. It is easy to show that

$$\theta_r = \sum_{i=1}^{n} p_i E_i[T_r(X)] = \sum_{i=1}^{n} p_i \theta_{ri}, \quad r = 1, \ldots, b, \tag{2.6}$$

where $\theta_{ri} = E_i[T_r(X)], r = 1, \ldots, b$ denote the rth types of moments of f_i.

The ME model with these moments, if exists, is unique with PDF,

$$f_\theta^*(x) = C(\lambda) \exp \left\{ - \sum_{r=1}^{b} \lambda_r T_r(x) \right\}, \tag{2.7}$$

where $\lambda = (\lambda_1, \ldots \lambda_b)$. The entropy of this ME model is given by

$$H(f_\theta^*) = -\log C(\lambda) + \sum_{r=1}^{b} \lambda_r \theta_r. \tag{2.8}$$

Denote the ME model subject to the moment constraints $\theta_{ri} = E_i[T_r(X)], r = 1, \ldots, b$ by f_i^*. This ME model and its entropy are in the same forms as (2.7) and (2.8) with r, λ replaced by r_i, λ_i, respectively. Consequently, f_θ^* is different from the mixture

$$f_a(x) = \sum_{i=1}^{n} p_i f_i^*(x). \tag{2.9}$$

The following proposition provides a bound for $JS(f_a)$ in terms of $K(f_i^* : f_\theta^*)$.

Proposition 2.1. *Let f_i^* and f_θ^* be the ME models subject to the constraints $\theta_{ri} = E_i[T_r(X)], r = 1, \ldots, b$ and $\theta_r = E[T_r(X)]$, respectively, and f_a be the mixture (2.9). Then*

$$\sum_{i=1}^{n} p_i K(f_i^* : f_\theta^*) - \sum_{i=1}^{n} p_i K(f_i^* : f_a) = H(f_\theta^*) - H(f_a) > 0, \tag{2.10}$$

implying the following bounds for the JS and entropy of the mixture of ME models:

$$\sum_{i=1}^{n} p_i K(f_i^* : f_a) < \sum_{i=1}^{n} p_i K(f_i^* : f_\theta^*) = MEB(f_a), \tag{2.11}$$

$$H(f_a) < H(f_\theta^*). \tag{2.12}$$

The equality in (2.10) is obtained from (2.3), (2.2), and

$$\sum_{i=1}^{n} p_i K(f_i^* : f_\theta^*) = H(f_\theta^*) - \sum_{i=1}^{n} p_i H(f_i^*). \tag{2.13}$$

This is shown in Shoja and Soofi (2017) for $T_r(X) = X^r$ and holds for general $T_r(X)$. The inequality in (2.10) is due to f_θ^* being the ME model with the same moments as f_a.

The following example illustrates that the bound (2.11) can improve on the entropy bound $H(P)$ and $JB(f_a)$ given in (2.5).

Example 2.1. Let $\mu_i = E_i(X)$ and $\mu = \sum_{i=1}^{n} p_i \mu_i$. The ME model subject to $E(X) = \mu_i$ is

$$f_i^*(x) = \frac{1}{\mu_i} \exp\left(-\frac{x}{\mu_i}\right), \tag{2.14}$$

and $H(f_i^*) = \log \mu_i + 1$. The ME model subject to $E(X) = \mu$ is given by (2.14) with μ in the place μ_i and $H(f^*) = \log \mu + 1$.

Let $n = 2$ and $\mu = p\mu_1 + (1-p)\mu_2$. Then, the ME bound (2.12) for $H(f_a)$ is

$$H(f_a) < MEB(f_a) = 1 + \log[p\mu_1 + (1-p)\mu_2], \quad 0 < p < 1.$$

The left panel of Figure 1 shows plots of this bound for select values of $\rho_{21} = \frac{\mu_2}{\mu_1}$. The three bounds for $JS(f_a)$ are as follows:

$$H(P) = -p \log p - (1-p) \log(1-p),$$

$$JB(f_a) = JB(\rho_{21}) = p(1-p)(\rho_{12} + \rho_{21} - 2),$$

$$MEB(f_a) = MEB(\rho_{21}) = p \log[p + (1-p)\rho_{21}] + (1-p) \log[p\rho_{12} + (1-p)].$$

The right panel of Figure 1 shows plots of these bounds for selected values of ρ_{21}. These plots indicate the following relations: $JB(\rho_{21}) < H(P)$ for $\rho_{21} \le 4$, the inequality is reversed for $\rho_{21} \ge 6$, and the dominance for $4 < \rho_{21} < 6$ depends on p. $MEB(\rho_{21}) < JB(\rho_{21})$ for all ρ_{21}, $MEB(\rho_{21}) < H(P)$ for $\rho_{21} \le 8$, and the latter inequality is reversed for very large ρ_{21}.
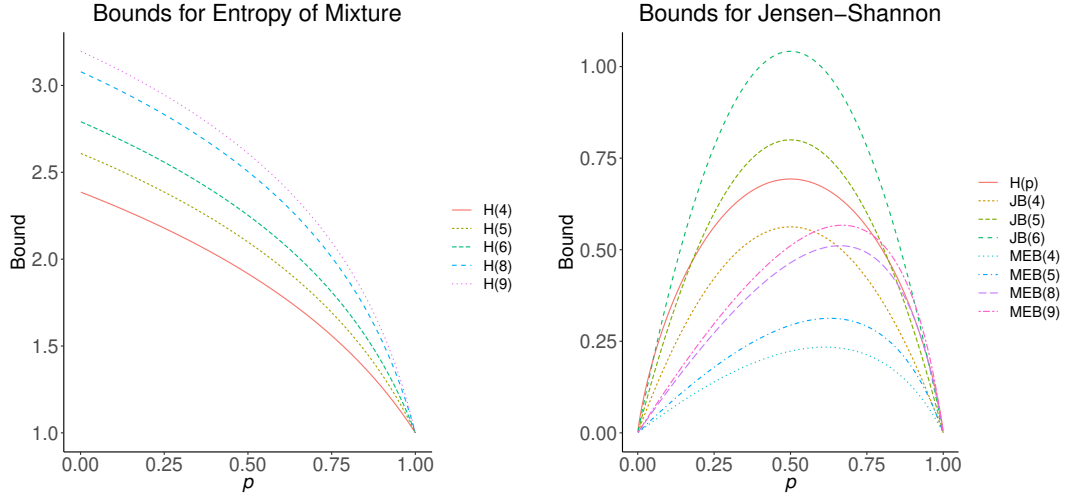
Figure 1: Three bounds for the Jensen-Shannon mixture and the maximum entropy bound for the entropy of mixtures of two exponential distributions with parameters $\mu_1 = 1$ and μ_2.

2.2 Applications of the JS Divergence

2.2.1 Comparison of Systems

Asadi et al. (2016) showed that the JS divergence provides useful information criteria for comparing coherent systems with homogeneous components. In these systems the component lifetimes $X_1, \cdots, X_n$ are independent and identically distributed (*iid*) as F_X with a continuous PDF f. The lifetime of the system T is a function of the lifetimes of its components called system's structure function, $T = \phi(X_1, \cdots, X_n)$, which is nondecreasing and every component affects the system's failure (Samaniego, 2007).

The SF of the system, $\bar{F}_T(t) = P(T > t)$, can be written as a mixture of the SFs of the order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$ of $X_1, \cdots, X_n$ as

$$\bar{F}_T(t) = \sum_{i=1}^{n} p_i \bar{F}_{i:n}(t), \quad p_i = P(T = X_{i:n}) = \frac{n_i}{n!} \geq 0, \quad \sum_{i=1}^{n} n_i = n!, \ t > 0, \tag{2.15}$$

where n_i is the number of ways that distinct $X_1, X_2, \ldots, X_n$ can be ordered such that the system lifetime $T = T(X_1, X_2, \ldots, X_n) = X_{i:n}$ (Samaniego, 1985; Kochar and Samaniego,

1999). The signature of a coherent system is defined by the vector $\boldsymbol{p} = (p_1, p_2, \cdots, p_n)$. This mixture representation is also valid when X_i's are exchangeable (Navarro and Rychlik, 2008) implying that the JS divergence of the *iid* model is applicable to coherent systems with homogeneous components.

Letting $U_i = F(X_i)$ provides $U_i, i = 1, \cdots, n$ as samples from the uniform distribution on $[0, 1]$ and order statistics $W_{i:n} = F(X_{i:n}), i = 1, \cdots, n$ with Beta$(i, n - i + 1)$ PDFs

$$g_{i:n}(w) = \frac{1}{B(i, n - i + 1)} w^{i-1}(1 - w)^{n-i}, \quad 0 \le w \le 1,$$

where $B(i, n - i + 1) = \dfrac{\Gamma(i)\Gamma(n - i + 1)}{\Gamma(n + 1)}$ is the beta function. The lifetime of the transformed system, denoted as V, is the mixture of beta distributions and its PDF is given by

$$g_V(v) = \sum_{i=1}^{n} p_i g_{i:n}(v) = \sum_{i=1}^{n} \frac{p_i}{B(i, n - i + 1)} v^{i-1}(1 - v)^{n-i}, \quad 0 \le v \le 1.$$

The invariance of the JS divergence implies that $JS(f_T) = JS(V) = JS(\boldsymbol{p})$.
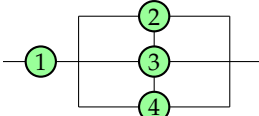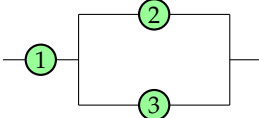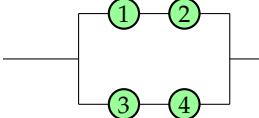
That is, $JS(\boldsymbol{p})$ is a distribution-free information criteria for comparison of coherent systems with homogeneous components (*iid* or exchangeable lifetimes) solely based on the systems' structures, enabling to compare systems with components with different lifetime distributions. In addition, $JS(\boldsymbol{p}) \ge 0$, where the equality holds if and only if $f_{k:n}(t) = f_T(t)$ almost everywhere which occurs when the kth failure is fatal to the system with probability $p_k = 1$ and $p_j = 0, j \ne k$, i.e., the case of an $(n - k + 1)$-out-of-n system.

Given the signature of a system, the second term in (2.2) can easily be computed using the entropy expression for the beta distribution. But computations of the first term in (2.2) and representation (2.3) require numerical integration. Asadi et al. (2016) illustrated applications of $JS(\boldsymbol{p})$ using two sets of systems with three and four components discussed by Shaked and Suarez-Llorens (2003). The following example gives an abridged version of their comparisons of systems with three and four components.

Example 2.2. Table 1 presents the signatures for three of the systems with three or four components presented in Asadi et al. (2016). They provide details for computations of the first term in (2.2). The systems are arranged according to the information provided by the lifetimes of the components about the lifetime of the system as measured by the JS divergence. We note that the JS divergence captures the situations where the disparity between the lifetimes of components of a larger system (four components)

and the lifetime of the system can be more or less than the disparity between the lifetimes of components of a smaller system (three components) and the lifetime of the system. The table also gives their ID and biased coin calibrations P_b of the JS divergence. The information discrepancies between the order statistics and the lifetime of these systems are like the discrepancies between loaded coins with odds of 4.9, 3.00, and 2.5 to one in favor of a "head" and a fair coin. The last two columns show the $JB(f_a)$ and $MEB(f_a)$ bounds given in (2.5) and (2.11) which are better than $H(\boldsymbol{p})$; the $MEB(f_a)$ is computed with the beta distribution which is ME subject to $E_\theta[\log W] = \theta_1$ and $E_\theta[\log(1-W)] = \theta_1$. For the first two systems, $JB(f_a) < MEB(f_a)$ and $JB(f_a) - J(f_a)$ is relatively substantial. For the last system, $JB(f_a) > MEB(f_a)$ and $MEB(f_a) - J(f_a)$ is not so substantial.

Table 1: Three coherent systems with three or four components.

System structure	System lifetime T	$\boldsymbol{p}$	$JS(\boldsymbol{p})$ $\{ID(K),\ P_b(K)\}$	$JS(\boldsymbol{p})$ Bound JB	MEB
	$\min\{X_1, \max(X_2, X_3, X_4)\}$	$(0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	.277 $\{.425,\ .83\}$	.583	.604
	$\min\{X_1, \max\{X_2, X_3\}\}$	$(\frac{1}{3}, \frac{2}{3}, 0)$	.140 $\{.244,\ .75\}$	.333	.404
	$\max\{\min(X_1, X_2), \min(X_3, X_4)\}$	$(0, \frac{2}{3}, \frac{1}{3}, 0)$	.099 $\{.180,\ .71\}$	.222	.135

2.2.2 Uncertainty and Disagreement of Forecasters

The problem of measuring the uncertainty and disagreement of a set of forecasters is addressed by numerous econometric researchers. Major part of this rich research utilizes databases such as the Survey of Economic Forecasters (SPF) of the Federal Reserve Bank of Philadelphia. The middle of each quarter, the SPF solicits subjective forecast probabilities for economic variables in the current year and the following year. The SPF offers a set of m preassigned intervals (bins) for the experts to assign

probabilities. The individual forecast distributions $f_i, i = 1, \ldots, n$ are constructed based on subjective probabilities and the arithmetic mixture of the individual probabilities with $p_i = 1/n$ is used as the consensus distribution f_a. Table 2 gives the forecast probabilities assigned (in percentage) to the intervals of the 2020 US GDP growth by three SPF respondents in the middle of the second quarter of 2020. Traditionally, uncertainty and disagreement of economic forecasters are measured using variances of f_i and f_a. The following decomposition of variance of f_a has been prominent:

$$\sigma_a^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 + \frac{1}{n} \sum_{i=1}^{n} (\mu_i - \mu_a)^2,$$

where μ_a, σ_a^2 and μ_i, σ_i^2 are the mean and variance of f_a and f_i, respectively, and the last summation is used as a measure of disagreement; we refer readers to Shoja and Soofi (2017) and Wang and Lahiri (2021) and references therein for the literature on this topic.

Shoja and Soofi (2017) proposed a general information framework for measuring uncertainty, information, and disagreement of economic forecasters. In this framework a forecaster can be a forecast distribution which can be the subjective distribution of an individual i solicited through a survey or the predictive distribution of a statistical model. They illustrated their proposed framework using the SPF data via the discrete entropy of the subjective probabilities assigned by forecasters as well as by the ME models with the first two moments of the bins midpoints. Wang and Lahiri (2021) avoided the discrete entropy and applied the information framework of Shoja and Soofi (2017) using entropies of beta and triangular distributions fitted to the histograms of the subjective probabilities.

Table 2: Subjective probabilities of three respondents of the Survey of Economic Forecaster about the 2020 US GDP growth.

	≤ 0%	0-1%	1-2%	2-3%	3-4%	4-5%	5-6%	6-7%	7-8%	> 8%
504	0	10	50	35	5	0	0	0	0	0
576	55	40	5	0	0	0	0	0	0	0
588	18	55	18	5	3	1	0	0	0	0
Average	24.33	35.00	24.33	13.33	2.67	0.33	0	0	0	0

Bajgiran et al. (2021) developed piecewise uniform ME models that only use the quantiles given by the forecasters' subjective probabilities. Quantile constraints can be

represented in terms of the following expectations:

$$E_f[\mathbf{1}(q_{k-1} < X \le q_k)] = \alpha_k - \alpha_{k-1} = p_k, \quad k = 1, \dots, m, \tag{2.16}$$

where $\mathbf{1}(A)$ is the indicator function of the event A and $\alpha_0 = 0$. The ME model with only QI constraints is in the following mixture of uniform PDFs:

$$f_\alpha^*(x) = \sum_{k=1}^{m+1} \frac{p_k}{B_k} \mathbf{1}(q_{k-1} \le x \le q_k), \quad a \le x \le b, \tag{2.17}$$

where $B_k = q_k - q_{k-1}, q_0 = a, q_{m+1} = b, \alpha_0 = 0$, and $\alpha_{m+1} = 1$. This piecewise uniform PDF is a density histogram with unequal bins B_k. Bajgiran et al. (2021) observed that, unlike the ME with moment constraints, the ME model subject to the mixtures of the average $\bar{p}_k = \sum_{i=1}^n p_i p_{ik}$ of the bin probabilities $p_{ik} = P_i[q_{k-1} \le X \le q_k]$ assigned by an individual is actually the mixture of individual forecasters ME models. Stated more formally, let $\{f_{\alpha_i}^*, i = 1, \dots, n\}$ be the set of individual ME models (2.17). Then,

$$f_a(x) = \sum_{i=1}^n p_i f_{\alpha_i}^*(x) = f_{a,\alpha}^*(x), \tag{2.18}$$

where $f_{a,\alpha}^*$ is the ME model consistent with the average bin probabilities $P[q_{k-1} \le X \le q_k] = \bar{p}_k = \sum_{i=1}^n p_i p_{ik}$. They illustrated (2.16) and (2.18) through implementing the information framework of Shoja and Soofi (2017) using the SPF data for the first and second quarters of 2020 (before and after the Covid19 outbreak).

We illustrate (2.16) and (2.18) and the ME models with the first two moment constraints using the data in Table 2 to implement the information framework for measuring the uncertainty and disagreement of the economic forecasters. These three forecasters are chosen for the purpose of illustrating Proposition 2.1 which involves distinguishing between the mixtures of the ME models with moment constraints and the ME models with the mixtures of the moments.

The upper row of Figure 2 shows the plots of ME models $f_{\alpha_i}^*$ (solid-red) and $f_{a,\alpha}^*$ (dashed-blue) for our three SPF respondents. It is common to give uniform weights to forecasters, so $p_i = 1/3, i = 1, 2, 3$. By (2.18), in this row the mixture of the ME models, $f_{\alpha_i}^*$, and the ME model with the mixture of probabilities are identical.

$$\theta_1 = E(X) = \sum_{k=1}^{m+1} p_k \frac{q_k + q_{k-1}}{2}, \quad \theta_2 = E(X^2) = \sum_{k=1}^{m+1} p_k \frac{q_k + q_k q_{k-1} + q_{k-1}}{3}.$$
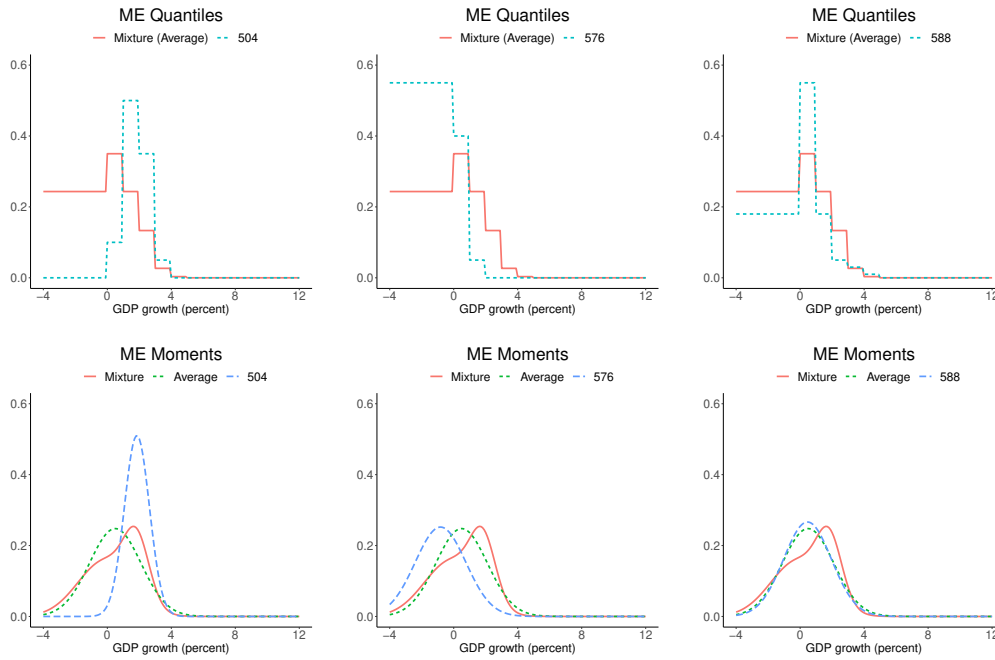
Figure 2: The ME forecast models based on the quantiles (upper row) and with the first two moments (lower row) for the data in Table 2.

Table 3: Uncertainty and disagreement of three respondents of the Survey of Economic Forecaster about the 2020 US GDP growth.

	ME with quantiles					ME with moments (θ_1, θ_2)	
	$H(f^*_{\alpha_i})$	$K(f^*_{\alpha_i} : f^*_{a,\alpha})$	θ_1	θ_2	σ^2	$H(f^*_{i,\theta})$	$K(f^*_{i,\theta} : f^*_\theta)$
504	1.094	.604	1.850	4.033	.611	1.172	.703
576	1.608	.423	-.825	3.183	2.503	1.878	.300
588	1.497	.106	.460	2.453	2.242	1.823	.019
Average ($p_i = \frac{1}{3}$)	1.399	.377	.495	3.223		1.624	.341
$\{ID(K), P_b(K)\}$		$\{.530, .86\}$					
$f^*_{a,\alpha}$	1.776	JS=.377	.495	3.223	2.591		
$f^*_\theta = N(\theta_1, \sigma^2)$			.495	3.223	2.591	1.965	
$f_{a,\theta}$			.495	3.223	2.591	$H < 1.965$	$JS < .341$

These moments are used as constraints for the moment-based ME models. The lower row of the figure shows the plots of ME models with the first two moment constraints. In this row, $f_{\theta_i}^*$ and f_θ^* are normal, but the mixture of the ME models, $f_{a,\theta}$, is bimodal.

Table 3 shows implementation of the information framework for the quantile-base ME models and moment-based ME forecast models depicted in Figure 2. The information measures of the ME models with quantile constraints are given in Bajgiran et al. (2021). Expressions for the entropy and KL divergence of normal distribution are well-known. The entropy of each distribution maps the extent of the uncertainty of the ME forecast model. The KL divergence between each individual forecast model and the pooled model maps the extent to which each forecaster disagrees with the pooled ME model. The measures for both types of ME models rank the individual forecasters similarly. The uncertainty mapped by the normal models are higher than those mapped by the quantile-based model, which can be due to their supports. The JS-divergence of the moment-based ME model can only be assessed by the bounds: $JS(f_{a,\theta}) < MEB(f_{a,\theta}) = .341 \ (ID(K) = .494, P_b(K) = .85)$, which is much smaller than $H(P) = \log 3$ and $JB(f_{a,\theta}) = 1.318 \ (ID(K) = .928, P_b(K) = .98)$.

2.2.3 Information Loss due to Kernel Estimation

Let $x_1, \ldots, x_n$ be data generated from an unknown CDF F. The kernel estimator of $F(x)$ is

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}\left(\frac{x - x_i}{h}\right), \tag{2.19}$$

where $h > 0$ is the bandwidth, $\mathcal{K}(u) = \int_{-\infty}^{u} k(z)dz$ is a proper CDF and $k(z) \geq 0$ is a symmetric kernel function such that:

$$\int_{-\infty}^{\infty} k(z)\,dz = 1, \quad \int_{-\infty}^{\infty} zk(z)\,dz = 0, \quad \int_{-\infty}^{\infty} z^2 k(z)\,dz = \kappa_2(K).$$

The kernel function $k(z)$ considered here is therefore a proper PDF.

As an estimator, (2.19) is consistent RMSE estimator with rate $n^{-1/2}$, provided that $h \to 0$ as $n \to \infty$. Raising h increases how smooth $\hat{F}$ is, while as $h \to 0$, $\hat{F}$ approaches the non-smooth empirical CDF, $\tilde{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i \leq x)$, which is a uniformly consistent estimator of $F(x)$. The corresponding kernel estimate of the PDF is given by the derivative of the smooth CDF estimator (2.19):

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{d}{dx} K\left(\frac{x - x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - x_i}{h}\right). \tag{2.20}$$

This is also RMSE consistent estimator of $f(x)$ with rate $n^{-2/5}$, provided that $h \to 0$ as $n \to \infty$, that $nh \to \infty$ as $n \to \infty$, and that $f(x)$ is twice continuously differentiable.

Beheshti et al. (2019) viewed $\hat{F}(x)$ as a smooth version of $\tilde{F}(x)$ since $\lim_{h \to 0} \hat{F}(x) = \tilde{F}(x)$ and characterized the kernel estimation of the CDF as an information transmitter with the channel consisting of the ensemble of kernel functions $\left\{ \mathcal{K}\left(\frac{x-x_i}{h}\right), i = 1, \cdots, n \right\}$. The inputs to the channel are h and a "rough" distribution $\tilde{F}$ on the support $\tilde{S} = \{x_1, \cdots, x_n\}$ and the output is the smooth kernel estimate $\hat{F}$ on a continuous support. The information measure of this system is the JS divergence where (2.2)-(4.1) are:

$$
\begin{aligned}
JS(\hat{F}) &= H(\hat{F}) - \{H(\mathcal{K}) + \log h\} \geq 0, &\text{(2.21)} \\
&= \frac{1}{n} \sum_{i=1}^{n} K\left(\mathcal{K}\left(\frac{x - x_i}{h}\right) : \hat{F}(x) \right) \geq 0, &\text{(2.22)}
\end{aligned}
$$

where the entropies and KL divergence are defined in terms of the PDFs. The use of CDFs in (2.21) and (2.22) is merely to emphasize that these measures are well-defined for $\hat{F}$ which converges to $\tilde{F}$ on $\tilde{S}$, but not for the kernel estimation of the PDF because $\hat{f}$ does not converge to $\frac{1}{n}$ on $\tilde{S}$. By the entropy representation, the output uncertainty is more than input uncertainty, which signifies $JS(\hat{F})$ as measure of information loss due to kernel estimation. This is the information cost of smoothing the $\tilde{F}$, where more smoothing (larger h) is more costly. Beheshti et al. (2019) used $JS(\hat{F})$ for evaluating various kernel functions.

Furthermore, (2.21) and (2.22) extend to the multivariate kernel estimation. The multivariate CDF kernel estimator is uniformly consistent as is the empirical CDF, so they coincide asymptotically. The bandwidth vanishes with rate $h \propto n^{-1/3}$ for all dimensions (Li and Racine, 2007). Thus as in the univariate case, the multivariate kernel estimate of CDF, $\hat{F}(x)$, is a smooth version of the multivariate empirical $\tilde{F}(x)$, and (2.21) and (2.22) give the information cost of smoothing the $\tilde{F}$, where more smoothing (larger h) is more costly.

3   Escorts and α-Mixtures

The escort distribution of order p of a probability distribution with a PDF f_1 relative to a measure ν is defined by

$$
f(x) = \frac{f_1^p(x)}{\int f_1^p(x) d\nu(x)}, \quad p > 0. \tag{3.1}
$$

These distributions are prevalent in nonextensive statistical mechanics (Tsallis, 1998), source coding (Bercher, 2011), and nanothermodynamics (Vakili, 2004). Bercher (2012) provided a ME characterization of (3.1).

Asadi et al. (2018) extended this notion to the escort of a SF $\bar{F}_{ph}(x) = \bar{F}_1^p(x)$, which is the proportional hazards model and to the escort of CDF $F_{prh}(x) = F_1^p(x)$, which is the proportional reversed hazards model, also known as the Lehman alternative.

3.1 Escorts of Two PDFs

The *generalized escort (g-escort)* distribution of order p is defined by the normalized product of two PDFs:

$$f_g(x) = \frac{f_1^p(x)f_2^{1-p}(x)}{Z_p}, \quad Z_p = \int f_1^p(x)f_2^{1-p}(x)dv(x), \ 0 \leq p \leq 1, \tag{3.2}$$

provided that f_1 is absolutely continuous relative to f_2 and the normalizing factor $Z_g < \infty$. Bercher (2012) coined the name and interpreted (3.2) as the geometric mean of two densities which as a function of $p > 0$ defines a curve, called the escort path.

Densities in the form of (3.2) have appeared in statistics literature in various contexts. The earliest version of (3.2) with $0 \leq p \leq 1$ appeared in minimization of the Bayesian probability of error of testing equally probable hypotheses (Chernoff, 1952). Kullback (1959) derived this PDF as the first example of his minimum discrimination information theorem. The power prior of Ibrahim and Chen (2000) is in this form, which was justified later by Ibrahim et al. (2003) with KL divergence-based derivation. Bercher (2012) derived (3.2) through a minimization of the KL divergence for the distribution of particles in a disequilibrium state. The three different information formulations of these authors that rendered solution in the form of (3.2) are, respectively, as follows:

$$\min_f K(f : f_2) \text{ subject to } E_f\left[\log \frac{f_1(X)}{f_2(X)}\right] = \theta, \int f(x)dx = 1, \tag{3.3}$$

$$\min_f \{wK(f : f_1) + (1 - w)K(f : f_2)\}, \int f(x)dx = 1, \quad 0 \leq w \leq 1, \tag{3.4}$$

$$\min_f K(f : f_1) \text{ subject to } K(f : f_2) = \eta, \int f(x)dx = 1, \tag{3.5}$$

where w in (3.4) is given, provided that the expectation in (3.3) is finite and $f(x) = f_1(x) = 0$ whenever $f_2(x) = 0$. Ibrahim et al. (2003) and Bercher (2012) cited Kullback

(1959), but without mentioning Kullback's derivation of (3.2). Consequently, until recently the relationships between the escort parameter p and the weight w in (3.4) and the constraint parameter η in (3.5) had remained unknown. This void was filled by Asadi et al. (2019a) via the Lagrangian dual functions which revealed the relationship between the escort parameter p and the parameter of each formulation in terms of their Lagrange multipliers as follows: $p = \lambda_K$, $p = w$, and $p = 1/(1 - \lambda_B)$, respectively.

The interpretation of Kullback's formulation (3.3) is clear in terms of expected log-likelihood ratio. The other two formulations are also insightful and have implications for various problems. We briefly describe these in our notations as follows. In (3.4), $f_1 = f_1(\beta|D_1)$ is the posterior distribution of a parameter β based on the past data D_1. Since then a new non-data opinion about β has emerged, represented by $f_2(\beta)$. The options are: (a) ignore the change of opinion and pool the data; (b) ignore the past; and (c) use a combination of $f_1(\beta)$ and $f_1 = f_1(\beta|D_1)$. Ibrahim et al. (2003) provided (3.4) for the formal justification of the power prior according to option (c). In (3.5), f_2 is the distribution of particles in the initial equilibrium state which subject to a generalized force has moved to the attractor where the particles will be distributed as f_1. The distribution at the present intermediate disequilibrium state, f, is unknown and chosen as the one that minimizes its divergence to the attractor f_1 while being hold on at the given divergence from f_2. Suppose that f_2 is the PDF of a variable in the previous stable condition. These scenarios can be applied to other problems as well. A shock has forced the distribution f_2 of a variable (a system's lifetime, an economic variable, and so on). The effect of the shock is still progressing toward a future stable condition where the distribution will be f_1. Currently, the distribution of the variable f is unknown. Formulation (3.4) applies with weights given to divergences of the current state from the two states. Formulation (3.5) applies with divergence of the current state from the distribution in future stable condition is given (η).

3.2 Escorts of Two SFs

The g-escort of two SFs is their geometric mixture $\bar{F}_g$ defined in (1.8). We recall that $d\bar{F}_g(x)/dx \neq f_g(x)$ and the geometric mixture of F_i corresponds to the arithmetic mixture of the reversed HR defined by $\tilde{r}_i(x) = \frac{f_i(x)}{F_i(x)}, F_i(x) > 0$.

Asadi et al. (2018) showed that these models are optimal solutions to various information theoretic formulations. Next, we summarize information properties of

the g-escort model (1.8).

The following is a known KL-type divergence between two SFs:

$$K_{\bar{F}}(\bar{F}_1 : \bar{F}_2) = \int \bar{F}_1(x) \log \frac{\bar{F}_1(x)}{\bar{F}_2(x)} d\nu(x) + \mu_2 - \mu_1 \geq 0,$$

where μ_i is the mean of $\bar{F}_i$. The g-escort model (1.8) is the solution to information problems analogous to (3.4) and (3.5) as follows:

$$\min_{\bar{F}} \{wK_{\bar{F}}(\bar{F} : \bar{F}_1) + (1 - w)K_{\bar{F}}(\bar{F} : \bar{F}_2)\}, \quad 0 \leq w \leq 1. \tag{3.6}$$

$$\min_{\bar{F}} K_{\bar{F}}(\bar{F} : \bar{F}_1) \text{ subject to } K_{\bar{F}}(\bar{F}_2 : \bar{F}_1) = \eta, \tag{3.7}$$

provided that all divergences exist. Interpretations of (3.4) and (3.5) apply to (3.6) and (3.7) where the distributions are represented by the SFs and $K_{\bar{F}}$ in the place of the KL divergence.

For characterizing the g-escort model (1.8) in terms of constraints analogous to (3.3)-(3.5), Asadi et al. (2018) used the expected L_1-norm between two cumulative hazard rate functions $\Lambda_i = -\log \bar{F}_i, i = 1, 2$ defined by $V_f(\Lambda_1, \Lambda_2) = \int |\Lambda_1(x) - \Lambda_2(x)| f(x) d\nu(x) \geq 0$, provided that the integral is finite. The g-escort model (1.8) is the solution to the following problems:

$$\min_{\Lambda} V_f(\Lambda, \Lambda_2) \quad \text{subject to} \quad E_f \left| \log \frac{\bar{F}_1(X)}{\bar{F}_2(X)} \right| = \theta. \tag{3.8}$$

$$\min_{\Lambda} \{wV_f(\Lambda, \Lambda_1) + (1 - w)V_f(\Lambda, \Lambda_2)\}, \quad 0 \leq w \leq 1. \tag{3.9}$$

$$\min_{\Lambda} V_f(\Lambda, \Lambda_1) \quad \text{subject to} \quad V_f(\Lambda, \Lambda_2) = \varphi, \quad 0 < p < 1, \tag{3.10}$$

provided that the expectation and all divergences exist. Interpretations of (3.3)-(3.5) apply to (3.8)-(3.10) where the distributions are represented by the cumulative hazard rate functions and V_f in the place of the KL divergence.

3.3 α-Mixture of PDFs

A more general mixture distribution appeared in information theory literature (van Erven and Harremoës, 2001), called here as α-mixture, is defined by the following PDF:

$$f_\alpha(x) = \frac{[pf_1^\alpha(x) + (1 - p)f_2^\alpha(x)]^{\frac{1}{\alpha}}}{Z_\alpha}, \quad 0 \leq p \leq 1, \alpha > 0, \tag{3.11}$$

where

$$Z_\alpha = \int [pf_1^\alpha(x) + (1-p)f_2^\alpha(x)]^{\frac{1}{\alpha}} dx \in \begin{cases} [2^{-(1-\alpha)/\alpha}, 1], & 0 < \alpha \leq 1, \\ [1, 2^{(\alpha-1)/\alpha}], & \alpha \geq 1; \end{cases}$$

The normalizing factor always exists for $\alpha > 0$. The α-mixture family (3.11) contains (1.4) as a especial case with $\alpha = 1$ and (3.2) as a limiting case for $\alpha \to 0$. Asadi et al. (2019a) derived (3.11) as extensions of (3.5) and (3.4) in terms of Tsallis divergence, defined by

$$K_q(f_1 : f_2) = \frac{1}{1-q} \int f_1^q(x)[f_1^{1-q}(x) - f_2^{1-q}(x)]dx, \quad q \in (0,1) \cup (1, \infty). \tag{3.12}$$

This measure replaces the KL divergence in the interpretations of (3.4) and (3.5).

3.4 α-Mixture of SFs

Like the $\bar{F}_a, \bar{F}_g$ extends to the geometric mixture of n SFs. Asadi et al. (2019a) introduced the finite α-mixture of SFs $\bar{F}_i, i = 1, \ldots, n$ defined by their weighted αth power mean as follows:

$$\bar{F}_\alpha(x) = \begin{cases} \left[\sum_{i=1}^n p_i \bar{F}_i^\alpha(x) \right]^{1/\alpha}, & 0 \neq \alpha \in \mathfrak{R}, \\ \bar{F}_g(x), & \alpha = 0, \end{cases} \tag{3.13}$$

where $p = (p_1, \ldots, p_n), p_i > 0, \sum_{i=1}^n p_i = 1$ and $\bar{F}_g(x) = \lim_{\alpha \to 0} \bar{F}_\alpha(x)$. We note that $d\bar{F}_\alpha(x)/dx \neq f_\alpha(x)$. This model is a broad family of mixture distributions. For example, for $\alpha = 1, \bar{F}_\alpha = \bar{F}_a$, for $\alpha = 0, \bar{F}_\alpha = \bar{F}_g$ and for $\alpha = -1$, we have the harmonic mixture (mean) of $\bar{F}_i$'s defined by

$$\bar{F}_h(x) = \left[\sum_{i=1}^n \frac{p_i}{\bar{F}_i(x)} \right]^{-1}, \quad x > 0.$$

The PDF and HR function of the α-mixture for all $\alpha \in \mathfrak{R}$ are:

$$f_\alpha(x) = \sum_{i=1}^n w_i(x, \alpha) f_i(x), \quad w_i(x, \alpha) = \left[\sum_{i=1}^n p_i \bar{F}_i^\alpha(x) \right]^{1/\alpha - 1} p_i \bar{F}_i^{\alpha-1}(x),$$

$$r_\alpha(x) = \sum_{i=1}^n p_i(x, \alpha) r_i(x), \quad p_i(x, \alpha) = p_i \left[\frac{\bar{F}_i(x)}{\bar{F}_\alpha(x)} \right]^\alpha,$$

where $w_i(x, \alpha) > 0, i = 1, \ldots, n$, but do not necessarily sum to one.

Some properties of the α-mixture found in Asadi et al. (2019a) are as follows:

(a) Stochastic order: $\bar{F}_{\alpha_1}(x) \leq \bar{F}_{\alpha_2}(x)$, $-\infty < \alpha_1 \leq \alpha_2 < \infty$. In particular, $\bar{F}_h(x) \leq \bar{F}_g(x) \leq \bar{F}_a(x)$.

(b) Stochastic distance: Let μ_a and μ_g denote the means of $\bar{F}_a$ and $\bar{F}_g$. Then,

$$SD(\bar{F}_{1/2}, \bar{F}_a) = SD(\bar{F}_g, \bar{F}_{1/2}) = \int |\bar{F}_{1/2}(x) - \bar{F}_a(x)|dx = \frac{1}{2}(\mu_a - \mu_g).$$

(c) Extensions of the well-known results of Barlow et al. (1963) on the closure of the mixture of decreasing failure rate (DFR) and DFR average (DFRA) distributions:

- If each $\bar{F}_i$ is DFR (IFR) then for $\alpha > 0$ ($\alpha < 0$) $\bar{F}_\alpha$ is DFR (IFR).
- If each $\bar{F}_i$ is DFRA (IFRA) then for $\alpha > 0$ ($\alpha < 0$) $\bar{F}_\alpha$ is DFRA (IFRA).

(d) Hazard rate order:

- Let $r_1, \ldots, r_n$ denote the HR of the components of $\bar{F}_\alpha$, $r_{\min}(x) = \min\{r_1(x), \ldots, r_n(x)\}$ for all x and $r_{\max}(x) = \max\{r_1(x), \ldots, r_n(x)\}$ for all x, respectively. Then $r_{\min}(x) \leq r_\alpha(x) \leq r_{\max}(x)$.
- If the HRs $r_i(x), i = 1, \ldots, n$ are ordered increasingly or decreasingly, then $r_\alpha(x)$ is decreasing in α for all $\alpha \in \mathcal{R}$. In particular, if HRs are ordered increasingly or decreasingly, then $r_{\min}(x) \leq r_h(x) \leq r_g(x) \leq r_a(x) \leq r_{\max}(x)$.

Countable and continuous α-mixtures are defined similarly and have analogous properties to (3.13). The following example gives an application of the closure property a countable α-mixture.

Example 3.1 (Asadi et al. (2019a)). Assume that $X_1, \ldots X_N$ are independent and identically distributed random variables with SF $\bar{F}(x)$ and N is random variable independent of X_i's. Let N have truncated Poisson distribution with parameter $\lambda > 0$. Then given $N = n$, $\min(X_1, \ldots, X_n)$ has SF $\bar{F}^n$. The harmonic mixture of the distributions of minima when the mixing distribution is the truncated Poisson given above is

$$\left[\bar{F}_h(x)\right]^{-1} = \sum_{n=1}^{\infty} \frac{g(n)}{\bar{F}^n(x)} = \frac{e^\lambda - 1}{e^{\lambda/\bar{F}(x)} - 1}, \quad x > 0.$$

Thus, if $\bar{F}$ is IFR (IFRA) then so is $\bar{F}_h$.

The following proposition established the relationship between the harmonic mixture of two SFs and the mixture of their ORs:

Proposition 3.1. *Let $\bar{F}_h(x)$ denote the survival function corresponding to harmonic mean of $\bar{F}_1$ and $\bar{F}_2$. Then*

$$\frac{F_h(x)}{\bar{F}_h(x)} = p\frac{F_1(x)}{\bar{F}_1(x)} + (1-p)\frac{F_2(x)}{\bar{F}_2(x)},$$

if and only if $\bar{F}_h(x)$ is the harmonic mean of $\bar{F}_1$ and $\bar{F}_2$.

Proof. We have

$$\frac{F_h(x)}{\bar{F}_h(x)} = \frac{1-\bar{F}_h(x)}{\bar{F}_h(x)} = \frac{1-\bar{F}_1(x)}{\bar{F}_1(x)} + (1-p)\frac{1-\bar{F}_2(x)}{\bar{F}_2(x)}.$$

This is equivalent to

$$\frac{1}{\bar{F}_h(x)} - 1 = \frac{p}{\bar{F}_1(x)} - p + \frac{1-p}{\bar{F}_2(x)} - (1-p).$$

This is easily seen that to be equivalent to

$$\bar{F}_h(x) = \frac{\bar{F}_1(x)\bar{F}_2(x)}{p\bar{F}_1(x) + (1-p)\bar{F}_2(x)}.$$

4 Probabilistic Formulations of Mixtures

The JS-divergence, in addition to (2.2) and (2.3), admits the following representation, which requires the probabilistic mixture formulation (Behboodian, 1972) given in (1.2):

$$JS(f_a) = M(X, D) = H(X) - H(X|D), \tag{4.1}$$

where $M(X, D)$ is the mutual information between the two random variables, $H(X) = H(f_a)$, and $H(X|D)$ is the conditional entropy of X given D defined by

$$H(X|D) = \sum_{i=1}^{n} p_i H(f_{X|D_i}) = \sum_{i=1}^{n} p_i H(f_i); \tag{4.2}$$

here $p_i = P(D = D_i)$, as defined in (1.2).

Lindley (1956) defined the Bayesian measure of information of an observable random variable X with a PDF $f(x|\theta)$ about the unobservable parameter Θ endowed with a prior PDF $\pi(\theta)$ in term of the following mutual information:

$$M(X, \Theta) = I(\Theta|X) - I(\Theta) = H(\Theta) - H(\Theta|X) \tag{4.3}$$

$$= \int\left[\int \pi(\theta|x)\log\frac{\pi(\theta|x)}{\pi(\theta)}d\theta\right]f(x)dx, \tag{4.4}$$

where $I(f) = -H(f)$ is the information function of f, $H(\Theta) = H(\pi(\theta))$ quantifies the prior uncertainty, the conditional entropy, $H(\Theta|X) = \int f(x)H(\Theta|x)dx$, quantifies the expected posterior uncertainty, and $f(x) = \int f(x|\theta)\pi(\theta)d\theta$ is the marginal distribution of X. The upper case symbol for the parameter signifies its association with probability distributions. That is, $M(X, \Theta)$ quantifies the information in terms of the expected uncertainty reduction. Noting that $\pi(\theta) = \int \pi(\theta|x)f(x)dx$, renders the prior distribution as a mixture, here $M(X, \Theta)$ is the continuous version of the mutual information representation of the JS-divergence (4.1).

4.1 Bayes-Fisher Information about a Parameter

This section presents an overview of Asadi et al. (2019a) with some elaborations and new examples.

Fisher information of an observable random variable X or its PDF $f(x|\theta)$ about θ that lies in an open interval $\Omega \subseteq \mathfrak{R}$ is defined by

$$\mathcal{I}(\theta) = \int \left[\frac{\partial \log f(x|\theta)}{\partial \theta} \right]^2 f(x|\theta)dx = \text{Var}_{x|\theta} \left\{ S[f(X|\theta)] \right\}, \qquad (4.5)$$

where $S[f(X|\theta)] = \partial \log f(x|\theta)/\partial \theta$ is the score function, provided that $f(x|\theta) > 0$ for all $\theta \in \Omega$ and $f(x|\theta)$ is differentiable with respect to θ. Representation (4.5) is the minimum risk of estimating the score function under the quadratic loss.

In general, $\mathcal{I}(\theta)$ is a function of θ. Under a prior $\pi(\theta)$, the expected Fisher information is defined by

$$\tilde{\mathcal{I}}(\Theta) = E_\pi[\mathcal{I}(\Theta)].$$

Based on the variance representation in (4.5), Asadi et al. (2017) called $\tilde{\mathcal{I}}(\Theta)$ the *Bayes risk of Fisher information* and in Asadi et al. (2019a) it is called *Bayes-Fisher information*, for short.

The Fisher information of a PDF f about its associated random prospect X is defined similarly to (4.5),

$$\mathcal{F}(X) = \mathcal{F}(f) = \int \left[\frac{\partial \log f(x)}{\partial x} \right]^2 f(x)dx;$$

see, for example, Cover and Thomas (2006, p. 671). Walker (2016) used this measure for the prior and posterior PDFs of Θ for developing the Lindley's information type

representation of $\tilde{I}(\Theta)$:

$$\tilde{I}(\theta) \;=\; \mathcal{F}(\Theta|X) - \mathcal{F}(\Theta) \tag{4.6}$$

$$=\; \int\int\left(\frac{\partial \log \pi(\theta|x)}{\partial \theta} - \frac{\partial \log \pi(\theta)}{\partial \theta}\right)^2 \pi(\theta|x) f(x) d\theta dx. \tag{4.7}$$

The analogy between (4.3) and (4.6) is clear. The analogy between (4.4) and (4.7) is seen by noting that the inner integral in (4.4) is $K[\pi(\theta|x) : \pi(\theta)]$ and the inner integral in (4.7) is the expected relative Fisher information distance between $\pi(\theta|x)$ and $\pi(\theta)$. Fisher distance finds applications in information geometry and Bayesian estimation (Bissiri et al., 2016; Holmes and Walker, 2017).

4.1.1 Bayes-Fisher Information of f_a, f_g, and f_α about p with Uniform Prior

The Fisher information and Bayes-Fisher information of f_a and f_g about the mixing parameter p are as follows.

(a) The Fisher information of f_a about p has the following representations:

$$\mathcal{I}_a(p) \;=\; \int \frac{[f_1(x) - f_2(x)]^2}{f_a(x)} dx$$

$$=\; \frac{1}{(1-p)^2}\chi^2(f_1 : f_a) = \frac{1}{p^2}\chi^2(f_2 : f_a)$$

$$=\; -\frac{\partial^2 JS(f_a)}{\partial p^2} = -\frac{\partial^2 H(f_a)}{\partial p^2},$$

where

$$\chi^2(f_i : f_a) = \int \frac{[f_i(x) - f_a(x)]^2}{f_a(x)} dx, \quad i = 1,2;$$

is the Chi-square divergence between f_i and f_a.

(b) The Fisher information of f_g about p has the following representations:

$$\mathcal{I}_g(p) = \int \frac{\partial f_g(x)}{\partial p} \log T(x) dx = \mathrm{Var}_{f_g}[T(X)], \quad T(X) = \log \frac{f_1(X)}{f_2(X)}.$$

(c) The Bayes-Fisher information of f_a and f_g about p with the uniform prior on $[0,1]$ are:

$$\tilde{I}_a(\mathcal{P}) = \tilde{I}_g(\mathcal{P}) = J(f_1, f_2); \tag{4.8}$$

proofs for f_a are given in Asadi et al. (2019a) and proofs for f_g are given by Bercher (2012). This result informs that under the uniform prior for p, the Bayes-Fisher information provided by f_a and f_g about the mixing parameter are equal and increasing in discrepancy between f_1 and f_2.

The following example gives a new interpretation of the well-known two sample F and T statistics as the amount of information provided by the combined sample about the mixing parameter,

Example 4.1. A product's characteristic is measured by one of two instruments with measurement variations distributed as f_i with failure rate $r_i, i = 1, 2$. The proportions of products measured by each instrument are unknown. The reliability of the product can be measured either by the probability of failure $\bar{F}_a$ or by the failure rate r_g given by (1.3) and (1.7). The instruments are tested using a sample of measurements from each instrument, $x_{i1}, \ldots, x_{in}, i = 1, 2$, assumed to be generated from $f_i = N(\mu_i, \sigma^2), i = 1, 2$. The maximum likelihood estimates of μ_i and σ^2 are (Behboodian, 1970):

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij}, \quad s^2 = \frac{s_1^2 + s_2^2}{2}, \quad i = 1, 2.$$

The empirical estimates of the distribution of a product's characteristic measured by each instrument is $N(\bar{x}_i, s^2), i = 1, 2$. By (4.8), the Bayes-Fisher information of arithmetic and geometric mixtures of this distribution about the mixing parameter the Jeffreys statistic, given by

$$\hat{I}_g(p) = \hat{I}_a(p) = J(\hat{f}_1, \hat{f}_2) = \frac{(\bar{x}_1 - \bar{x}_2)^2}{s^2},$$

That is, Jeffreys statistic is proportional to the usual T^2 and F statistics used for testing the hypothesis of $\mu_1 = \mu_2$. Jeffreys statistics $J(\hat{f}_1, \hat{f}_2)$ for discriminating between samples from two normal distributions is known. This example gives a new Bayes-Fisher interpretation for $J(\hat{f}_1, \hat{f}_2)$ as the expected Fisher information, in terms of (4.7), provided by the combined data from the two samples about the mixture parameter p.

The Fisher information of the α-mixture (3.11) has the following representation:

$$\mathcal{I}_\alpha(p) = \frac{1}{\alpha^2} \text{Var}_{f_\alpha}[S_p(X)],$$

where

$$S_p(x) = \frac{f_1^\alpha(x) - f_2^\alpha(x)}{p f_1^\alpha(x) + (1 - p) f_2^\alpha(x)} = \frac{T^\alpha(x) - 1}{p[T^\alpha(x) - 1] + 1}, \quad T(X) = \log \frac{f_1(X)}{f_2(X)}.$$

For any given α, the Bayes-Fisher information of the α-mixture (3.11) under the uniform prior $\pi(p) = 1, 0 \leq p \leq 1$ is given by

$$\tilde{\mathcal{I}}_\alpha(\mathcal{P}) = \frac{1}{\alpha} J_\alpha(f_1, f_2) + \frac{1}{\alpha(1-\alpha)} V_\alpha(h_1, h_2),$$

where

$$J_q(f_1, f_2) = K_q(f_1 : f_2) + K_q(f_2 : f_1), \quad q \in (0, 1) \cup (1, \infty), \quad (4.9)$$

is the generalized Jeffreys divergence (symmetrized Tsallis divergence), $V_\alpha(h_1, h_2) = \int_0^1 [h_1(p) - h_2(p)]^2 dp$, is the following total variation function between $h_i(p) = \frac{1}{Z_\alpha^\alpha} \mathcal{H}_\alpha(f_i : f_\alpha), i = 1, 2$, and $\mathcal{H}_q(f_i : f_\alpha) = \int f_i^q(x) f_\alpha^{1-q}(x) dx, q \in \Omega$ is the Hellinger integral.

4.1.2 Bayes-Fisher Information of f_a and f_g about p with Triangular Prior

Let $f_i, i = 1, 2$ represent the distributions of a random prospect under two conditions or two hypotheses. The uniform prior for p is reflective of being agnostic about the weights of the hypotheses. We use three triangular priors that give the highest weight to a particular combination of the two hypotheses. The triangular prior with PDF,

$$\pi_w(p) = \begin{cases} \frac{2p}{w}, & 0 < p \leq w, \\ \frac{2(1-p)}{1-w}, & w \leq p < 1. \end{cases} \quad (4.10)$$

The general triangular prior π_w is reflective of the belief that a mixture of f_1 and f_2 with a mixing weight w is the most likely case. The beta priors give higher weight for f_1 or f_2 and provide an extension of the null hypothesis belief that favors one of the models. Mcvinish et al. (2009) used the triangular prior for Bayesian robustness.

(a) Under the prior (4.10), the Bayes-Fisher information of (1.4) and (3.2) are:

$$\tilde{\mathcal{I}}_a(\mathcal{P}) = \frac{2}{w(1-w)}[wK(f_1 : f_{aw}) + (1-w)K(f_2 : f_{aw})] = \frac{2}{w(1-w)} JS(f_a), \quad (4.11)$$

$$\tilde{\mathcal{I}}_g(\mathcal{P}) = \frac{2}{w(1-w)}[wK(f_{gw} : f_1) + (1-w)K(f_{gw} : f_2)] = \frac{2}{w} R_w(f_1, f_2), \quad (4.12)$$

where f_{aw} and f_{gw} denote PDFs (1.4) and (3.2), respectively, with $p = w$ and

$$R_q(f_1 : f_2) = \frac{1}{q-1} \log \int f_1^q(x) f_2^{1-q}(x) dx, \quad q \in \Omega;$$

is Rényi divergence of order q.

(b) Under two triangular beta priors on $\pi_j(p), 0 < p < 1, j = 0, 1$, shown below, the Bayes-Fisher information of (1.4) and (3.2) are as follows:

$$
\begin{aligned}
\pi_0(p) &= 2(1 - p), & \pi_1(p) &= 2, \\
\tilde{I}_a(\mathcal{P}) &= 2K(f_1 : f_2), & \tilde{I}_a(\mathcal{P}) &= 2K(f_2 : f_1), \\
\tilde{I}_g(\mathcal{P}) &= 2K(f_2 : f_1), & \tilde{I}_g(\mathcal{P}) &= 2K(f_1 : f_2).
\end{aligned}
\tag{4.13}
$$

These results inform that also under the triangular prior the Bayes-Fisher information provided by (1.4) and (3.2) about p are increasing in the discrepancy between the components of the mixtures.

4.2 Global Mean of MR Function

At each threshold τ, the MR function is the mean of the excess distribution (1.6). This endows MR function with an intuitive and easily understood pointwise interpretation in the data analysis. At each threshold point, $m(\tau)$ is the optimal predictor of the excess under the quadratic loss function, and its risk is the variance of the excess distribution. Aggregation of the uncountably many local means, $m(\tau)$, with a continuous weight function $\pi(\tau), \tau \geq 0$ defines the following continuous mixture of the MRs:

$$
\mathcal{M}_\pi(m(\tau)) = \int \pi(\tau) m(\tau) d\tau.
$$

In the probabilistic version of this continuous mixture, $\pi(\tau)$ is a prior PDF for τ and the probabilistic mixture of $m(\tau)$ is the global mean, given by

$$
\mathcal{M}_\pi(m(\tau)) = E_\tau[m(\tau)] = E_\tau E_{x|\tau}(X - \tau|X > \tau) = E_{(x,\tau)}(X - \tau|X > \tau).
$$

This measure is used in Asadi et al. (2017), Ardakani et al. (2018), and Ardakani et al. (2020), where it is referred to as the Bayes risk of $m(\tau)$, which is a misnomer. The Bayes risk of $m(\tau)$ is given by the $E_\tau[\text{Var}(X - \tau|X > \tau)]$.

A particular prior for τ is the PDF $\pi(\tau) = f(\tau)$, where f is the PDF representation of $m(\tau)$. This prior renders the following important global mean of the MR function:

$$
\mathcal{M}_f(m(\tau)) = -\int \bar{F}(\tau) \log \bar{F}(\tau) d\tau \geq 0.
$$

Asadi and Zohrevand (2007) showed the first equality. Rao et al. (2004) introduced the integral, referred to it as the cumulative residual entropy, and showed the inequality

and it becomes equality if and only if the distribution is degenerate. That is, $\mathcal{M}_f(m(\tau))$ is a measure of concentration of f. This measure is also known by other names, including *Survival Entropy*.

We have developed a consistent estimate for $\mathcal{M}_f(m(\tau))$ and applied it for comparing forecast models (Ardakani et al., 2018). We also have shown that $\mathcal{M}_f(m(\tau)) \leq \sigma_f$, where σ_f is the standard deviation and the equality holds if and only $\bar{F}$ is exponential. With a consistent estimator of σ_f, this bound holds for the empirical version of $\mathcal{M}_f(m(\tau))$. Application to the New York City Taxi database, which includes millions of monthly observations, illustrated the following points: (a) plot of the empirical MR function, called *MR Plot*, distinguishes the distributions of taxi trip times by month and pickup time more clearly than plots of PDF, SF, and HR function; and (b) the empirical $\mathcal{M}_f(m(\tau))$ classifies the hourly distributions sensibly (Ardakani et al., 2020).

5   Discrete, Multivariate, and Categorical Models

Our previous studies considered information measures for various types of mixtures of univariate continuous distributions. In Section 2.2.3, we pointed out the applicability of Jensen-Shannon divergence, (2.21) and (2.22), for measuring information loss in the multivariate kernel estimation. Information measures for variants of mixture models are applicable to probability mass functions (PMFs) and multivariate distributions. We briefly present extensions of some of the measures in the preceding sections in the context of exponential families.

We consider the applicability of information measures of the arithmetic and geometric mixtures of PDFs to the exponential family with the following density with respect to a measure:

$$f(x) = g(x)e^{\eta T(x) - A(\eta)}, \tag{5.1}$$

where $\eta = \eta(\theta)$, called the canonical or natural parameter, is a one-to-one function of θ that appears in the "usual" form of f, $A(\eta)$ is free from x, and $g(x) > 0, T(x)$ are free from θ; η and T may be vectors. Well-known examples include discrete distributions (binomial with a common n, geometric, Poisson), multivariate distributions (normal, Pareto, Dirichlet), and categorical models such as the multinomial logit which is widely used in many fields for relating the probability of outcome j to covariates. (See Wikipedia for a table of distributions with corresponding η, θ, g, T, and A).

The geometric mixture of two members $f_i, i = 1, 2$ of (5.1) with $\eta_i = \eta(\theta_i)$ remains in

the family (Kullback, 1959, p. 40), given by the following g-escort model:

$$f_{g,\eta_1,\eta_2}(x) = \frac{g(x)}{Z_{p,\eta_1,\eta_2}} \exp\left\{ [p\eta_1 + (1-p)\eta_2]T(X) - [pA(\eta_1) + (1-p)A(\eta_2))] \right\}, \qquad (5.2)$$

where Z_{p,η_1,η_2} is the normalizing factor of f_g.

The Jensen-Shannon divergence of members of (5.1) on a finite support (binomial, multinomial logit) can be easily computed. More generally, the KL and Rényi divergences between two members of an exponential family are:

$$K(f_1 : f_2) = (\eta_1 - \eta_2)\theta_1 - A(\eta_1) + A(\eta_2), \qquad (5.3)$$

$$R_q(f_1 : f_2) = \frac{1}{q-1} \log Z_{q,\eta_1,\eta_2}, \qquad (5.4)$$

where $\theta_1 = E_1[T(X)]$ and Z_{q,η_1,η_2} is defined as in (5.2).

For all members of the family in (5.1), (5.3) provides bounds (2.4) and (2.5). It also provides bounds given in Proposition 2.1 for discrete and multivariate distributions that are ME models (geometric, normal, Pareto, Dirichlet). Divergences (5.3) and (5.4) enable computing the Bayes-Fisher information measures about the mixing parameter for f_a and f_g under the uniform prior given in (4.8) and under the triangular priors given in (4.11)-(4.13).

6 Concluding Remarks

This paper presented a holistic view of contents of ten papers which mainly or tangentially involve mixture models co-authored by at least one of us. We have discussed four types of mixtures (Arithmetic, geometric, harmonic, and α) with six representations of a probability distribution (CDF, PDF, SF, HR, MR, OR). The study of information properties of various types of mixtures involves assortments of information and divergence measures: Shannon entropy, KL, JS, Jeffreys, Chi-square, Rényi, Tsallis, and Jeffreys type symmetrized Tsallis divergences, Fisher information measure and Fisher information distance, KL type divergence between SFs, and expected L_1-norm between cumulative hazards. Areas of applications covered include reliability (comparison of systems), econometrics (uncertainty and disagreements of forecasters), statistics (kernel estimation, exponential family, comparison of two normal means), and nonextensive statistical mechanics (escort distributions).

The variants of mixtures of various representations of probability distributions, assortments of information measures, and applications presented in this paper provide ample research problems. We point out to a few specific topics for future research.

- The JS divergence of the arithmetic mixture of PDFs includes the Shannon entropy of the mixture and the KL divergence between each component and the mixture. This measure is very useful for many applications, which require numerical integration or assessed by a few bounds. Developing a user friendly general algorithm that produces accurate results for mixtures of parametric families can prove to be useful.

- The study of information properties includes only one measure for the arithmetic mixtures with multiple components. We have developed few measures for the arithmetic, geometric, and α-mixtures of two PDFs. These measures are derived under the uniform prior for the mixing parameter for all three types of mixtures and for the arithmetic and geometric mixtures under two triangular priors. Use of other types of priors (such as beta) and extensions to mixtures with multiple components provide ample research opportunities.

- Following findings in statistics and physics literature, we developed optimality of the geometric mixtures of two SFs (arithmetic mixture of two HR) and CDFs (arithmetic mixture of two reversed HR) according to formulations in terms of a KL type divergence for SFs and CDFS and according to the expected L_1-norm between two cumulative HR functions. Extensions of the univariate results for the mixtures of SFs to the multivariate and to mixtures with multiple components can lead to interesting results.

- We have not explored information properties of the α-mixture family yet. This problem offers an interesting and challenging topic for future research.

- Developing inferences about the α parameters of the α-mixtures of PDFs and SFs will be very useful. For example, if a posterior distribution of α of the α-mixture of SFs concentrates near zero (or one), then it can be concluded that the data is generated from an arithmetic mixture of the HRs (or SFs).

References

Ardakani, O. M., Ebrahimi, N., and Soofi, E. S. (2018), Ranking forecasts by stochastic error distance, information and reliability measures. *International Statistical Review*,

86(3), 442-468.

Ardakani, O. M., Asadi, M., Ebrahimi, N., and Soofi, E. S. (2020), MR plot: A big data tool for distinguishing distributions. *Statistical Analysis and Data Mining the ASA Data Science Journal*, **13**, 405-418.

Asadi, M. and Zohrevand, Y. (2007), On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference*, **137**, 1931-1941.

Asadi, M., Ebrahimi, N., and Soofi, E. S. (2017), Connections of Gini, Fisher, and Shannon, by Bayes risk under proportional hazards. *Journal of Applied Probability*, **54**, 1027-1050.

Asadi, M., Ebrahimi, N., and Soofi, E. S. (2018), Optimal hazard models based on partial information. *European Journal of Operational Research*, **270**(2), 1–11.

Asadi, M., Ebrahimi, N., and Soofi, E. S. (2019), The alpha-mixture of survival functions. *Journal of Applied Probability*, 56(4), 1151–1167.

Asadi, M., Ebrahimi, N., kharazmi, O., and Soofi, E. S. (2019), Mixture models, Bayes Fisher information, and divergence measures. *IEEE Transactions on Information Theory*, **65**, 2316–2321.

Asadi, M., Ebrahimi, N., Soofi, E. S., and Zohrevand, Y. (2016), Jensen–Shannon information of the coherent system lifetime. *Reliability Engineering and System Safety*, **156**(C), 244–255.

Bajgiran, A. H., Mardikoraem, M., and Soofi, E. S. (2021), Maximum entropy distributions with quantile information. *European Journal of Operational Research*, **290**(1), 196–209.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1963), Properties of probability distributions with monotone hazard rate. *Annals of Mathematical Statistics*, **34**, 375-389.

Behboodian, J. (1970), On a mixture of normal distributions. *Biometrika*, **57**(1), 215–217.

Behboodian, J. (1972), On the distribution of a symmetric statistics from a mixed population. *Technometrics*, **14**, 919–923.

Beheshti, N., Racine, J. S., and Soofi, E. S. (2019), Information measures of kernel estimation. *Econometric Reviews*, **38**(1), 47–68.

Bercher, J. F. (2011), Escort entropies and divergences and related canonical distribution. *Physics Letters A*, **375**, 2969-2973.

Bercher, J. F. (2012), A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians. *Physica a: Statistical Mechanics and Its Applications*, **391**(19), 4460-4469.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016), A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B*, **78**(5), 1103–1130.

Block, H. W., and Savits, T. H. (1997). Burn-in. *Statistical Science*, **12**, 1–19.

Chernoff, H. (1952), A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, **23** ,493-507.

Cover, T. M. and Thomas, J. A. (2006), *Elements of Information Theory, 2nd ed.*. New York: Wiley.

Harremoes, P. (2001), Binomial and Poisson distributions as maximum entropy distributions. *IEEE Transactions on Information Theory*, **47**, 2039–2041.

Holmes, C. C., and Walker, S. G. (2017). Assigning a value to a power likelihood in a general Bayesian model. *Biometrika*, **104**, 497-503.

Ibrahim, J. G., and Chen, M. H. (2000), Power prior distributions for regression models. *Statistical Science*, **15**, 46-60.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2003), On optimality of the power prior. *Journal of the American Statistical Association*, **98**, 204-213.

Kochar, S., Mukerjee, H., and Samaniego, F. J. (1999), The "signature" of a coherent system and its application to comparisons among systems. *Naval Research Logistics*, **46**(5), 507–523.

Kullback, S. (1959), *Information theory and statistics*. New York: Wiley (reprinted in 1968 by Dover).

Li, Q., and Racine (2007). *Nonparametric Econometrics: Theory and Practice*. New Jersey: Princeton University Press.

Lin, J. (1991), Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, **37**, 145-151.

Lindley, D. V. (1956), On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, **27**, 986–1005.

Lynn, N. J., and Singpurwalla, N. D. (1997), Comment: "Burn-in'" makes us feel good. *Statistical Science*, **12**, 13–19.

McCulloch, R. E. (1989), Local model influence. *Journal of the American Statistical Association*, **84**, 473–478.

Mcvinish, M., Rousseau, J., and Mengersen, K. (2009), Bayesian goodness of fit testing with mixtures of triangular distributions. *Scand. J Statist.*, **36**, 337-354.

Navarro, J., and Rychlik, T. (2007), Reliability and expectation bounds for coherent systems with exchangeable components. *Journal of Multivariate Analysis*, **98**(1), 102-113.

Rao, M., Chen, Y., Vemuri, B. C., and Wang, F. (2004), Cumulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory*, **50**, 1220-1228.

Samaniego, F. J. (1985), On closure of the IFR class under formation of coherent systems. *IEEE Transactions on Reliability*, **R-34**(1), 69–72.

Samaniego, F. J. (2007), *System signatures and their applications in engineering reliability*. Springer.

Shaked, M., and Shanthikumar, J. G. (2007), *Stochastic orders*. Springer.

Shaked, M., and Suarez-Llorens, A. (2003), On the comparison of reliability experiments based on the convolution order. *Journal of the American Statistical Association*, **98**(463), 693–702.

Shoja, M., and Soofi, E. S. (2017), Uncertainty, information, and disagreement of economic forecasters. *Econometric Reviews*, **36**(6-9), 796–817.

Soofi, E. S., Ebrahimi, N., and Habibullah, M. (1995), Information distinguishability with application to analysis of failure data. *Journal of the American Statistical Association*, **90**, 657–668.

Tsallis, C. (1998), Generalized entropy-based criterion for consistent testing. *Physics Review E*, **58**, 1442-1445.

Vakili-Nezhaad, G. R., and Mansoori, G. A. (2004), An application of non-extensive statistical mechanics to nanosystems. *Journal of Computational and Theortical Nanonscience*, **1**, 233-235.

van Erven, T., and Harremoës, P. (2014), Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, **60**, 3797-3820.

Walker, S. G. (2016), Bayesian information in an experiment and the Fisher information distance. *Statistics and Probability Letters*, **112**, 5–9.

Wang, W., and Lahiri, K. (2021), Estimating macroeconomic uncertainty and discord using info-metrics. *In Innovations in info-metrics A cross-disciplinary perspective on information and information processing*, 1–55.

Wang, L., and Madiman, M. (2014), Beyond the entropy power inequality, via rearrangements. *IEEE Transactions on Information Theory*, **60**(9), 5116–5137.