

## WEIGHTED PARAMETRIC DIVERGENCE MODELS FOR DISCRETE PROBABILITY DISTRIBUTIONS

OM PARKASH<sup>1</sup> AND MUKESH SARANGAL<sup>2\*</sup>

**ABSTRACT.** In the literature of information measures, it is well acknowledged phenomenon that distance models in probability spaces discover incredible applications in a diversity of disciplines related with science and technology. The significance of these models after attaching weights to the occurring events cannot be disregarded. The present communication is a footstep in the construction of such divergence models. We have developed two new weighted parametric divergence models for the discrete probability distributions and proved their legitimacy after studying their indispensable properties.

### 1. INTRODUCTION AND PRELIMINARIES

This is to highlighting that the entropy and divergence models locate remarkable applications for the study of optimization problems associated with information theoretic measures developed for discrete as well as continuous probability distributions. Such problems include the minimization of entropy measures and maximization of divergence measures. The majority of the exhaustive study provides applications of probabilistic measures only whereas the weights of the events under study cannot be disregarded because of their importance in daily life situations. It was Shannon [17] who for the first time introduced the concept of information theoretic entropy associated with every probability distribution  $P = (p_1, p_2, \dots, p_n)$  with the help of subsequent quantitative model:

$$H(P) = - \sum_{i=1}^n p_i \log p_i. \quad (1.1)$$

Immediately after Shannon gave his measure, research workers in many fields saw the potential of the application of this expression and consequently developed a variety of entropy models for providing applications to numerous disciplines. Recently, Mekhour [9] and Haji, Moumni and Talha [4] extended the applications of this entropic model in finding the entropy solutions for nonlinear parabolic equations.

---

*Date:* Received: Oct 22, 2021; Accepted: Apr 11, 2022.

\* Corresponding author.

2010 *Mathematics Subject Classification.* 62B10; 94A15; 94A17.

*Key words and phrases.* Directed divergence, Probability distributions, Convexity, Symmetry, Weighted divergence, Continuity.

Since the entropy models have been explored as measures of amount of information for a given probability distribution, it is customary to scrutinize such measures which authorize one to appraise the amount of information shared between two probability distributions. Such measures frequently acknowledged as distance models ascertain inconceivable applications in an assortment of disciplines. In the field of information theory, the following successfully functional and most imperative divergence model is payable to Kullback and Leibler [7]:

$$D(P : Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}. \quad (1.2)$$

Kapur [5] wrought the subsequent models in probability spaces:

$$D_1(P : Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \frac{1}{a} \sum_{i=1}^n (q_i + ap_i) \ln \frac{(q_i + ap_i)}{q_i(1+a)}, a \geq -1. \quad (1.3)$$

$$D_2(P : Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \frac{1}{a} \sum_{i=1}^n (1 + ap_i) \ln \frac{(1 + ap_i)}{(1 + aq_i)}, a \geq -1. \quad (1.4)$$

Parkash and Kakkar [10, 11] made exhaustive study of divergence models for their applications towards a variety of disciplines and consequently introduced the following generalized models of directed divergence for discrete probability distributions:

$${}_{\alpha}D(P : Q) = \frac{1}{\alpha - 1} \left[ \prod_{i=1}^n \left( \frac{p_i}{q_i} \right)^{-p_i(1-\alpha)} - 1 \right], \alpha > 1, \alpha \neq 1, \quad (1.5)$$

$${}^{\alpha}D(P : Q) = \frac{1}{1 - \alpha} \left[ \prod_{i=1}^n q_i^{-p_i(1-\alpha)} - \prod_{i=1}^n p_i^{-p_i(1-\alpha)} \right], 0 \leq \alpha < 1, \alpha \neq 1, \quad (1.6)$$

$${}_{\beta}D(P : Q) = \frac{\sum_{i=1}^n p_i \beta^{\ln \frac{p_i}{q_i}} - 1}{\beta - 1}, \beta > 1. \quad (1.7)$$

Parkash and Mukesh [12, 13] also made investigations for comprehensive study of divergence models from applications point of examination and consequently introduced the subsequent generalized divergence models:

$$D^{\alpha}(P : Q) = \frac{\sum_{i=1}^n p_i \left( \alpha + \frac{1}{2} \right)^{\log \frac{p_i}{q_i}} - 1}{\alpha - \frac{1}{2}}, \alpha > 0, \alpha \neq \frac{1}{2}, \quad (1.8)$$

$$D^{\alpha, \beta}(P : Q) = \frac{\sum_{i=1}^n p_i \left( \frac{\alpha}{2\beta} + \frac{1}{2} \right)^{\log \frac{p_i}{q_i}} - 1}{\alpha - \beta}, \alpha > 0, \beta > 0, \alpha \neq \beta, \quad (1.9)$$

$$D_{\alpha, \beta}(P : Q) = \frac{\sum_{i=1}^n p_i^{\alpha - \beta + 1} q_i^{\beta - \alpha} - 1}{\alpha - \beta}, \alpha \neq \beta, \beta < \alpha + 1. \quad (1.10)$$

Kapur [5] explained certain most plausible properties of an appropriate weighted measure of distance. Motivated by this technique, Parkash, Kumar and Kakkar [12] wrought the following weighted divergence models:

$${}_αD(P, Q; W) = \frac{1}{2^{1-α} - 1} \sum_{i=1}^n w_i [(\alpha - 1)p_i - p_i^\alpha q_i^{1-\alpha} + \alpha p_i^{\frac{1}{\alpha}} q_i^{1-\frac{1}{\alpha}} + 2(1-\alpha)q_i], \alpha > 1. \quad (1.11)$$

Certain investigations of such measures are payable to Agahi [1], Pronzato, Wynn and Zhigljavsky [16], Ararat, Hamel and Rudloff [3], Kumari and Sharma [8], Pinelis [15], Ahmadzade et. al. [2], Khalaj et. al. [6], Torra, Narukawa and Sugeno [18] etc.

In the sequel, we have wrought two new weighted models.

## 2. TWO NEW PARAMETRIC WEIGHTED DIVERGENCE MODELS

### 2.1. We propose the following divergence model for the discrete probability distributions:

$$D_\alpha(P, Q; W) = \frac{1}{\alpha - 1} \sum_{i=1}^n w_i [p_i^\alpha q_i^{1-\alpha} + (\alpha - 1)q_i - \alpha p_i], \alpha > 1. \quad (2.1)$$

We observe that

$$\lim_{\alpha \rightarrow 1} D_\alpha(P, Q; W) = \sum_{i=1}^n w_i [p_i \log \frac{p_i}{q_i} - p_i + q_i],$$

which is Kapur's [5] measure of weighted divergence. Hence  $D_\alpha(P, Q; W)$  is a generalized weighted divergence model.

Properties:

- (i) At  $p_i = q_i$ ,  $D_\alpha(P, Q; W) = 0$ .
- (ii) Convexity: We have

$$\frac{\partial D_\alpha(P, Q; W)}{\partial p_i} = \frac{1}{\alpha - 1} w_i [\alpha p_i^{\alpha-1} q_i^{1-\alpha} - \alpha],$$

and

$$\frac{\partial D_\alpha(P, Q; W)}{\partial q_i} = w_i [1 - p_i^\alpha q_i^{-\alpha}],$$

Again

$$\frac{\partial^2 D_\alpha(P, Q; W)}{\partial p_i^2} = \alpha w_i p_i^{\alpha-2} q_i^{1-\alpha} > 0,$$

$$\frac{\partial^2 D_\alpha(P, Q; W)}{\partial q_i^2} = \alpha w_i p_i^\alpha q_i^{-(1+\alpha)} > 0,$$

Also

$$\frac{\partial^2 D_\alpha(P, Q; W)}{\partial p_i \partial p_j} = 0, i \neq j$$

and

$$\frac{\partial^2 D_\alpha(P, Q; W)}{\partial q_i \partial q_j} = 0, i \neq j$$

Thus the Hessian matrix of  $D_\alpha(P, Q; W)$  with respect to P is given by

$$\begin{pmatrix} \alpha w_1 p_1^{\alpha-2} q_1^{1-\alpha} & 0 & \dots & 0 \\ 0 & \alpha w_2 p_2^{\alpha-2} q_2^{1-\alpha} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha w_n p_n^{\alpha-2} q_n^{1-\alpha} \end{pmatrix}$$

The definite positivity of the above Hessian matrix proves the convexity of the divergence model  $D_\alpha(P, Q; W)$ . Similar convexity with respect to Q can be proved. Hence,  $D_\alpha(P, Q; W)$  is convex.

(iii) Non-negativity: To prove non-negativity of  $D_\alpha(P, Q; W)$ , we reflect on the following Lagrangian:

$$L = \frac{1}{\alpha - 1} \sum_{i=1}^n w_i [p_i^\alpha q_i^{1-\alpha} + (\alpha - 1)q_i - \alpha p_i] + \lambda (\sum_{i=1}^n p_i - 1)$$

Taking  $\frac{\partial L}{\partial p_i} = 0$ , we get

$$\frac{\alpha}{\alpha - 1} w_1 [p_1^{\alpha-1} q_1^{1-\alpha} - 1] = \frac{\alpha}{\alpha - 1} w_2 [p_2^{\alpha-1} q_2^{1-\alpha} - 1] = \dots = \frac{\alpha}{\alpha - 1} w_n [p_n^{\alpha-1} q_n^{1-\alpha} - 1]$$

which is possible only if  $p_i = q_i \forall i$ .

Thus, minimum value of the divergence measure exists at  $p_i = q_i$  and  $D_\alpha(P, Q; W) = 0$  which implies that  $D_\alpha(P, Q; W) \geq 0$ .

The above meticulous properties authenticate our claim that the weighted model introduced in (2.1) is a correct weighted model.

**2.2. We next propose the following weighted parametric divergence model for the discrete probability distributions:**

$$D^\alpha(P, Q; W) = \frac{1}{2^{\alpha-1} - 1} \sum_{i=1}^n w_i [p_i^\alpha q_i^{1-\alpha} - p_i^{\frac{1}{\alpha}} q_i^{1-\frac{1}{\alpha}} - (\alpha - \frac{1}{\alpha})(p_i - q_i)], \alpha > 1. \quad (2.2)$$

We observe that

$$\lim_{\alpha \rightarrow 1} D^\alpha(P, Q; W) = 2 \sum_{i=1}^n w_i [p_i \log \frac{p_i}{q_i} - p_i + q_i],$$

which is Kapur's [5] weighted divergence except a multimlicative constant. Hence  $D^\alpha(P, Q; W)$  is a generalized measure of weighted directed divergence.

Properties:

(i) At  $p_i = q_i$ ,  $D^\alpha(P, Q; W) = 0$ .

(ii) Convexity: We have the subsequent mathematical expressions:

$$\begin{aligned}\frac{\partial D^\alpha(P, Q; W)}{\partial p_i} &= \frac{w_i}{2^{\alpha-1} - 1} [\alpha p_i^{\alpha-1} q_i^{1-\alpha} - \frac{1}{\alpha} p_i^{\frac{1}{\alpha}-1} q_i^{1-\frac{1}{\alpha}} - (\alpha - \frac{1}{\alpha})] \\ \frac{\partial^2 D^\alpha(P, Q; W)}{\partial p_i^2} &= \frac{w_i(\alpha - 1)}{2^{\alpha-1} - 1} [\alpha p_i^{\alpha-2} q_i^{1-\alpha} + \frac{1}{\alpha^2} p_i^{\frac{1}{\alpha}-2} q_i^{1-\frac{1}{\alpha}}] > 0, \text{ as } \alpha > 1 \\ \frac{\partial D^\alpha(P, Q; W)}{\partial q_i} &= \frac{w_i}{2^{\alpha-1} - 1} [(1 - \alpha) p_i^\alpha q_i^{-\alpha} - (1 - \frac{1}{\alpha}) p_i^{\frac{1}{\alpha}} q_i^{-\frac{1}{\alpha}} - (\alpha - \frac{1}{\alpha})] \\ \frac{\partial^2 D^\alpha(P, Q; W)}{\partial q_i^2} &= \frac{w_i(\alpha - 1)}{2^{\alpha-1} - 1} [\alpha p_i^\alpha q_i^{-1-\alpha} + \frac{1}{\alpha^2} p_i^{\frac{1}{\alpha}} q_i^{-1-\frac{1}{\alpha}}] > 0, \text{ as } \alpha > 1\end{aligned}$$

Also

$$\frac{\partial^2 D^\alpha(P, Q; W)}{\partial p_i \partial p_j} = 0, i \neq j,$$

$$\frac{\partial^2 D^\alpha(P, Q; W)}{\partial q_i \partial q_j} = 0, i \neq j$$

Thus the Hessian matrix of  $D^\alpha(P, Q; W)$  with respect to P is given by

$$\begin{pmatrix} \frac{w_1(\alpha-1)}{2^{\alpha-1}-1} [\alpha p_1^{\alpha-2} q_1^{1-\alpha} + \frac{1}{\alpha^2} p_1^{\frac{1}{\alpha}-2} q_1^{1-\frac{1}{\alpha}}] & 0 & \dots & 0 \\ 0 & \frac{w_2(\alpha-1)}{2^{\alpha-1}-1} [\alpha p_2^{\alpha-2} q_2^{1-\alpha} + \frac{1}{\alpha^2} p_2^{\frac{1}{\alpha}-2} q_2^{1-\frac{1}{\alpha}}] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{w_n(\alpha-1)}{2^{\alpha-1}-1} [\alpha p_n^{\alpha-2} q_n^{1-\alpha} + \frac{1}{\alpha^2} p_n^{\frac{1}{\alpha}-2} q_n^{1-\frac{1}{\alpha}}] \end{pmatrix}$$

The positive definite character of the above Hessian matrix verifies the convexity of the model. Hence,  $D^\alpha(P, Q; W)$  is convex.

(iii) Non-negativity: To prove non-negativity, we reflect on the following Lagrangian:

$$L = \frac{1}{2^{\alpha-1} - 1} \sum_{i=1}^n w_i [p_i^\alpha q_i^{1-\alpha} - \frac{1}{\alpha} p_i^{\frac{1}{\alpha}} q_i^{1-\frac{1}{\alpha}} - (\alpha - \frac{1}{\alpha})(p_i - q_i)] + \lambda (\sum_{i=1}^n p_i - 1)$$

Taking  $\frac{\partial L}{\partial p_i} = 0$ , we get

$$\begin{aligned}\frac{w_1}{2^{\alpha-1} - 1} [\alpha p_1^{\alpha-1} q_1^{1-\alpha} - \frac{1}{\alpha} p_1^{\frac{1}{\alpha}-1} q_1^{1-\frac{1}{\alpha}} - \alpha + \frac{1}{\alpha}] &= \frac{w_2}{2^{\alpha-1} - 1} [\alpha p_2^{\alpha-1} q_2^{1-\alpha} - \frac{1}{\alpha} p_2^{\frac{1}{\alpha}-1} q_2^{1-\frac{1}{\alpha}} - \alpha + \frac{1}{\alpha}] \\ &= \dots = \frac{w_n}{2^{\alpha-1} - 1} [\alpha p_n^{\alpha-1} q_n^{1-\alpha} - \frac{1}{\alpha} p_n^{\frac{1}{\alpha}-1} q_n^{1-\frac{1}{\alpha}} - \alpha + \frac{1}{\alpha}]\end{aligned}$$

which is possible only if  $p_i = q_i \forall i$ .

Thus, the minimum value of the divergence measure exists at  $p_i = q_i$  and  $D^\alpha(P, Q; W) = 0$  which implies that  $D^\alpha(P, Q; W) \geq 0$ .

The above meticulous properties authenticate our claim that the weighted model introduced in (2.2) is a correct weighted model.

**Concluding Remarks:** The mathematical models of distance in the probability spaces find wonderful applications in a diversity of disciplines dealing with approximately the entire scientific systems. Since a solitary distance model cannot be tolerable for each regulation, we require a selection of generalized parametric information theoretic measures as these models induce flexibility into the system under consideration. Another perspective regarding the need of such measures is that the different attempts have been made to extend the concept of distance in various fields other than mathematical sciences. Such disciplines include economics, sociology, psychology, linguistics, genetics, biology etc. where distance measure can be successfully applied. But, distance in such cases need not to be geometrical one and hence there is necessity for its modification when it comes to the distance between two probability distributions. Here, we stress upon the modification needed when we have to consider a measure for the concept of distance between two probability distributions. Keeping this design in brain, we have created two new weighted parametric divergence models to supply their inconceivable applications in furtherance of our research findings.

## REFERENCES

1. H. Agahi, *A modified Kullback-Leibler divergence for non-additive measures based on Choquet integral*, Fuzzy Sets and Systems **367** (2019), 107-117. <https://doi.org/10.1016/j.fss.2019.01.021>
2. H. Ahmadzade, R. Gao, H. Naderi and F. Mehran, *Partial divergence measure of uncertain random variables and its application*, Soft Computing **24(1)** (2020), 501-512. <https://doi.org/10.1007/s00500-019-03929-0>
3. C. Ararat, A. H. Hamel and B. Rudloff, *Set-valued shortfall and divergence risk measures*, International Journal of Theoretical and Applied Finance **20 (5)** (2017), 1750026(1-48). <https://doi.org/10.1142/S0219024917500261>
4. B. E. Haji, E. Moumni and A. Talha, *Entropy solutions for nonlinear parabolic equations in Musielak Orlicz Spaces without 2-condition*, Gulf Journal of Mathematics **9(1)** (2020), 1-26. <https://doi.org/10.56947/gjom.v9i1.448>
5. J.N. Kapur, *Measures of Information and Their Applications*, Wiley Eastern, New York, 1994.
6. M. Khalaj, R.T. Moghaddam, F. Khalaj and A. Siadat, *New definition of the cross entropy based on the Dempster-Shafer theory and its application in a decision-making process*, Communications in Statistics - Theory and Methods **49(4)** (2020), 909-923. <https://doi.org/10.1080/03610926.2018.1554123>
7. S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics **22** (1951), 79-86. <https://doi.org/10.1214/aoms/1177729694>
8. R. Kumari and D. K. Sharma, *Generalized 'useful' non-symmetric divergence measures and inequalities*, Journal of Mathematical Inequalities **13(2)** (2019), 451-466.
9. M. Mekhour, *Entropy solutions for a nonlinear unilateral  $p(x)$  parabolic problems with data*, Gulf Journal of Mathematics **6(4)** (2018), 14-36. <https://doi.org/10.56947/gjom.v6i4.245>
10. O. Parkash and P. Kakkar, *New measures of information and their applications in coding theory*, Canadian Journal of Pure and Applied Sciences **8(2)** (2014), 2905-2912.
11. O. Parkash and P. Kakkar, *New information theoretic models, their detailed properties and new inequalities*, Canadian Journal of Pure and Applied Sciences **8(3)** (2014), 3115-3123.

12. O. Parkash, R. Kumar and P. Kakkar, *Two new weighted divergence measures in probability spaces*, International Journal of Pure and Applied Mathematics **116(19)** (2017), 643-653.
13. O. Parkash and Mukesh, *Applications of information measures for the study of Gaussian distributions*, Canadian Journal of Pure and Applied Sciences **8(1)** (2014), 2795-2799.
14. O. Parkash and Mukesh, *Portfolio optimization using measures of cross entropy*, Canadian Journal of Pure and Applied Sciences **8(2)** (2014), 2963-2967.
15. I. Pinelis, *An involution inequality for the Kullback-Leibler divergence*, Mathematical Inequalities and Applications **20(1)** (2017), 233-235.
16. L. Pronzato, H. P. Wynn and A. Zhigljavsky, *Bregman divergences based on optimal design criteria and simplicial measures of dispersion*, Statistical Papers **60(2)** (2019), 195-214.
17. C. E. Shannon, *A mathematical theory of communication*, Bell. Syst. Tech. J. **27** (1948), 379-423, 623-659. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
18. V. Torra, Y. Narukawa and M. Sugeno, *On the  $f$ -divergence for discrete non-additive measures*, Information Sciences **512** (2020), 50-63. <https://doi.org/10.1016/j.ins.2019.09.033>

<sup>1</sup> DEPARTMENT OF MATHEMATICS, GURU NANAK DEV UNIVERSITY, AMRITSAR, INDIA.  
 Email address: [omparkash777@yahoo.co.in](mailto:omparkash777@yahoo.co.in)

<sup>2</sup> DEPARTMENT OF MATHEMATICS, MAHARAJA RANJIT SINGH PUNJAB TECHNICAL UNIVERSITY, BATHINDA, INDIA.  
 Email address: [mukesh.sarangal@mrsptu.ac.in](mailto:mukesh.sarangal@mrsptu.ac.in)