A NEW DISCRETE ENTROPIC MODEL AND ITS APPLICATION TO CONTINGENCY TABLE INFERENCE

Poonam Kumari[1], Nishi Kumari[2]

•

[1] Department of Mathematics, Magadh Mahila College, Patna University, Patna, India
poonamkumari1865@gmail.com
[2]Department of Mathematics, Patna University, Patna, India
nishiraj1510@gmail.com

Abstract

A variety of parametric and non-parametric information models, both discrete and continuous, are well-established. However, there remains a continued need to develop new parametric models that offer greater flexibility in analyzing complex systems. Information entropy is closely linked to the field of statistics, providing valuable insights. This paper introduces a novel discrete information entropic model and explores its applications in statistical analysis. We have demonstrated the efficacy of this new model by applying it to the inference of contingency tables, a cornerstone of statistical analysis. Using the maximum entropy principle and Lagrange method of multiplier, we have derived an important relation between the chi-square and the entropy measure. By incorporating this novel entropy measure, we have expanded the scope and applicability of the maximum entropy principle in this domain, enabling more robust and informative statistical inferences.

Keywords: Discrete probability distribution, entropy, entropic model, maximum entropy principle, Lagrange method of multiplier, contingency table.

I. Introduction

Entropic models have significantly advanced fields like Physics and Engineering, and now play a fundamental role across all areas of Mathematical Sciences. This concept underwent a remarkable transformation in the 20th century with the development of Shannon's [1] mathematical theory of communication. He proposed entropy as a measure of information content, which also facilitates decision-making in problems related to message transmission over noisy or noiseless channels. The non-parametric nature of this entropic model has limited its applicability in various domains. This necessitates the development of more flexible parametric entropic models. Recognizing the inherent strengths and limitations of this entropy, numerous novel generalized entropic models have been formulated, finding successful applications across diverse disciplines.

One particular area where entropic models are of paramount importance is in the analysis of contingency tables, which are frequently used in statistics to analyze the relationship between two or more categorical variables. Learning from contingency tables involves understanding the dependencies and associations between different variables, often relying on methods such as the maximum entropy principle to make inferences about the underlying data distribution. By imposing constraints based on observed data, the maximum entropy principle provides a way to estimate distributions in situations where limited information is available.

This paper proposes a novel discrete entropic model designed to enhance the flexibility and applicability of the maximum entropy principle in the context of contingency table inference. By extending the range of possible entropy distributions, our model allows for more accurate estimations, even in cases of limited or incomplete data. This new model is based on an innovative formulation that better reflects the underlying structure of the data and can adapt to various types of relationships between variables. The paper is organized as follows: Section II outlines the existing literature review, while Section III describes the proposed entropic model. Section IV demonstrates the application of this entropic model to contingency table inference. Finally, Section V discusses the conclusion and suggests directions for future research.

II. Literature Review

In 1948, Shannon [1] established the following expression for the influential and fundamental additive discrete entropic model:

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

where $P = (p_1, p_2, \dots, p_n)$ is the probability measure on a finite set $\{x_1, x_2, \dots, x_n\}$ such that $p_i \geq 0$ for all $i = 1,2,\dots,n$ and $p_1+p_2+\dots+p_n = 1$. The logarithm in equation (1) is taken to the base 2, and by convention, $0\log 0 = 0$.

Arimoto [2] proposed the subsequent generalized entropic model represented by the expression

$$H_t(P) = \frac{t}{t-1}\left[1-\left(\sum_{i=1}^{n} p_i^t\right)^{1/t}\right]; \quad t>0, t \neq 1. \tag{2}$$

Behara and Chawla [3] derived the generalized entropic model recognized as τ - entropy and defined as

$$H_\tau(P) = \frac{1-\left(\sum_{i=1}^{n} p_i^{1/\tau}\right)^{\tau}}{1-2^{\tau-1}}; \ \tau>0, \tau\neq 1. \tag{3}$$

Sharma and Mittal [4] introduced a two-parameter entropy measure in the following form:

$$H_{\alpha,\beta}(P) = \frac{1}{1-\beta}\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{1-\beta}{1-\alpha}} - 1\right]; \ \alpha,\beta>0, \alpha\neq 1, \beta\neq 1, \alpha\neq\beta. \tag{4}$$

Kapur [5] developed several probabilistic models of entropy, one of which is

$$H_b^a(P) = \frac{1}{a+b-2}\left[\sum_{i=1}^{n} p_i^a + \sum_{i=1}^{n} p_i^b - 2\right]; \ a\geq 1, b\leq 1 \text{ or } a\leq 1, b\geq 1. \tag{5}$$

Recognizing the broad scope of entropic models in probability spaces, Parkash and Kakkar [6] examined the following model:

$$H^a(P) = \frac{\sum_{i=1}^{n} p_i^a \log\left(\dfrac{p_i^a}{\sum_{i=1}^{n} p_i^a}\right)}{\sum_{i=1}^{n} p_i^a}; \ a>0. \tag{6}$$

The applications of the above model were extended to the field of coding theory for the development of innovative codeword lengths. To advance the literature in the area of coding theory, Parkash and Kakkar [6], [7] extended existing information theoretic entropic models and derived the following expressions:

$$S_a(P)=\frac{1}{1-a}\left(\prod_{i=1}^{n} p_i^{\,p_i(a-1)}-1\right);\ a>0, a\neq 1, \tag{7}$$

$$_aS(P)=\frac{\sum_{i=1}^{n} p_i a^{\log p_i}-1}{1-a};\ a>1, \tag{8}$$

$$^aS(P)=\frac{a^{\sum_{i=1}^{n} p_i \ln p_i}-1}{1-a};\ a>1, \tag{9}$$

$$S_a^b(P)=\frac{a^{\frac{1}{b\text{-}1}\ln\left(\sum_{i=1}^{n} p_i^b\right)}-1}{1-a};\ a>1, 0<b<1. \tag{10}$$

Bulinski and Kozhevin [8] investigated the applications of entropy functions to obtain features that can be effectively used for feature selection problems. Lu et al. [9] observed that measures of predictability in physiological indicators, based on entropy metrics, have been widely utilized in the field of medical diagnosis and clinical documentation. The authors proposed a novel entropy-based pattern recognition approach by integrating singular spectrum analysis with entropy measures for the assessment of physiological signals.

Sholehkerdar et al. [10] pointed out that entropy-based metrics are frequently proposed for evaluating the effectiveness of image fusion, owing to their single-layer execution of entropy and minimal parameter set. In their study, the authors presented a theoretical exploration of an image fusion quality measure based on Tsallis [11] entropy. The aim of their research was to assess whether the proposed quality measure could meet the expected performance standards of an ideal information-based image fusion distinction metric. To evaluate the Tsallis [11] quality measure, the authors employed an image realization model to derive a closed-form expression for distinction, while weighted averaging was used as the fusion technique. Their findings showed that the Tsallis-based quality measure affected the anticipated performance, particularly in relation to variations in signal-to-noise ratio and the influence of entropy order on the measured distinction. Additionally, the authors conducted experiments with real images, and the results demonstrated consistency with their theoretical predictions.

Lenormand et al. [12] investigated the applications of entropy-based models in urban environments. They observed that characterizing and quantifying spatial inequalities across urban landscapes remains a complex and challenging task, exacerbated by the growing availability of vast geo-located datasets. Their research findings demonstrated that the attractiveness of a location, as quantified by entropy, is a significant predictor of the socioeconomic status of the surrounding area and can thus be utilized as an indicator for various socioeconomic factors.

Manzoor et al. [13] presented entropic models in the field of chemistry, noting that, inspired by Shannon's [1] entropy, graph entropies with topological indices have effectively quantified the information content of chemical graphs and complex systems.

Recently, Parkash and Mukesh [14] developed the following parametric information model to support the continuation of their findings:

$$H_{a,b}(P)=\frac{\sum_{i=1}^{n} p_i^a-\sum_{i=1}^{n} p_i^b-n^{1-a}+n^{1-b}}{(b-a)(b+a-1)};\ \ a\geq 1, b\leq 1 \text{ or } a\leq 1, b\geq 1. \tag{11}$$

Siddiqui et al. [15] computed and analyzed several distinct entropy measures in conjunction with diverse crystallographic structures. They worked on the Zagreb entropies, hyper and augmented Zagreb entropies, and forgotten and Balaban entropies for the crystallographic structures of the cuprite and titanium difluoride.

Teixeira et al. [16] remarked that there is no universally established definition for conditional Tsallis [11] entropy. The traditional definition of Tsallis entropy relies on a parameter α, which converges to Shannon's [1] entropy as α approaches 1. In their research, the authors presented three proposed classifications of conditional Tsallis entropy as outlined in the literature. Furthermore, they introduced an alternative natural approach for conditional Tsallis entropy and compared it with the existing ones. Finally, the authors provided insightful suggestions, emphasizing that the application of entropies across various fields, particularly in information theory and its connection to communication, holds significant potential for the development of numerous valuable information measures, such as mutual information, information symmetry, and information distances.

Segura et al. [17] emphasized that entropy plays a crucial role as a physical concept in any configuration or environment, particularly within the business context. They pointed out that within organizations; a common issue is inefficiency, disruption of relationships, and widespread misunderstandings, which ultimately lead to dysfunction. In this context, a system is defined as a group of individuals characterized by specific attributes, which interact with one another and exist within a particular environment to achieve certain objectives. Entropy, in this sense, refers to a measure of the constraints present within a system that influence its operations. The authors further noted that, in the face of market changes and uncertainty, entropy remains a persistent factor in companies. However, it can also be part of a progression, as new business decisions are made. In this regard, entropy allows us to understand the underlying causes of disorder within organizations and provides tools to monitor changes or disruptions in information.

Saraiva [18] provided a concise and unexpected perspective on Shannon's entropy, encompassing specific properties and demonstrating the model's applicability in two distinct domains from its original context: biological diversity and an innovative study on student migration.

Zhang and Shi [19] emphasized that Shannon's entropy is a fundamental building block of information theory and an essential component of Machine Learning (ML) algorithms. The authors highlighted asymptotic properties that require no assumptions about the underlying data distribution. These properties enable interval estimation and statistical testing with general Shannon's entropy. Friesner et al. [20] investigated the applications of entropy in the field of survey research. Stoyanov et al. [21] applied the maximum entropy approach to derive a novel principle for the *M*-indeterminacy of distributions on the positive half-line (Stieltjes case). Additionally, the authors illustrated how maximum entropy enhances the understanding of symmetry properties and *M*-indeterminacy. Some noteworthy related work on entropy measures and their applications has been carried out by Jizba and Korbel [22], Gao and Deng [23], Elgawad et al. [24], Bajic et al. [25], Fowler and Heckman [26], Li et al. [27], Bisikalo et al. [28], Gregorio et al. [29], among others.

III. A New Discrete Entropic Model

We introduce a new parametric entropic model of order a and type b defined by the expression

$$H_{a,b}(P) = -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b-a}\sum_{i=1}^{n}\left[1 - p_i^{(b-a)p_i}\right];\ a \neq b, b-a > 0. \tag{12}$$

Obviously, for $b = 1$, we have

$$H_{a,1}(P) = H_a(P) = -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{1-a}\sum_{i=1}^{n}\left[1 - p_i^{(1-a)\,p_i}\right];\ a > 0. \tag{13}$$

It can be proved that $\lim_{a\to 0} \frac{1}{1-a}\sum_{i=1}^{n}\left(1-p_i^{(1-a)\,p_i}\right)=-\sum_{i=1}^{n} p_i \log p_i$. Therefore the entropy measure (13) reduces to $\lim_{a\to 0} H_a(P) = -2\sum_{i=1}^{n} p_i \log p_i$, which is similar to Shannon's entropy. Thus, equation (12) represents a generalized measure of entropy.

To prove the validity of the proposed entropy measure, we verify its fundamental properties.

(i) The entropy is non-negative.

For all $b-a>0$, we have $p_i^{(b-a)p_i} \leq 1$. Therefore, $\frac{1}{b-a}\sum_{i=1}^{n}\left[1-p_i^{(b-a)p_i}\right] \geq 0$ for all $b-a>0$.

Also, we have $-\sum_{i=1}^{n} p_i \log p_i \geq 0.$

Thus, we get $$-\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b-a}\sum_{i=1}^{n}\left[1-p_i^{(b-a)p_i}\right] \geq 0$$

i.e., $$H_{a,b}(P) \geq 0.$$

(ii) A degenerate distribution concentrated at a single element has entropy zero.

If we take $(p_1, p_2, p_3, \ldots, p_n) = \delta_1 = (1,0,0,\ldots,0)$ or $(p_1, p_2, p_3, \ldots, p_n) = \delta_2 = (0,1,0,\ldots,0)$ or $(p_1, p_2, p_3, \ldots, p_n) = \delta_n = (0,0,0,\ldots,1)$, then, obviously, $H_{a,b}(P) = 0$. Therefore the entropy defined by equation (12) vanishes for a degenerate distribution.

(iii) The entropy does not change on inserting an impossible event.

We have

$$H_{a,b}(p_1, p_2, \ldots, p_n, 0) = -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b-a}\sum_{i=1}^{n}\left[1-p_i^{(b-a)\,p_i}\right] + \frac{1}{b-a}\left[1-(0)^{(b-a)0}\right]$$

$$= -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b-a}\sum_{i=1}^{n}\left[1-p_i^{(b-a)\,p_i}\right]$$

$$= H_{a,b}(p_1, p_2, \ldots, p_n).$$

Thus, the entropy is unaffected by inserting an impossible event.

(iv) The entropy is a symmetric function of $p_1, p_2, \ldots, p_n$.

Obviously, if $p_1, p_2, \ldots, p_n$ are reordered, then $H_{a,b}(P)$ remains unchanged. Hence, $H_{a,b}(P)$ is a symmetric function of $p_1, p_2, \ldots, p_n$.

(v) The entropy is a concave function of $p_1, p_2, \ldots, p_n$.

Differentiating $H_{a,b}(P)$ w.r.to p_i, we get

$$\frac{dH_{a,b}(P)}{dp_i} = -1 - \log p_i - p_i^{(b-a)p_i}(1+\log p_i).$$

$$\therefore \quad \frac{d^2 H_{a,b}(P)}{dp_i^2} = -\frac{1}{p_i} - p_i^{(b-a)p_i}\left\{(b-a)(1+\log p_i)^2 + \frac{1}{p_i}\right\} < 0.$$

Hence $H_{a,b}(P)$ is a concave function of $p_1, p_2, \ldots, p_n$.

The concavity of $H_{a,b}(P)$ can also be seen from the graph of $H_{a,b}(P)$ plotted against p_i for $n=2$ and $a=2, b=4$.

Table 1: $H_{a,b}(P)$ *against* p_i *for* $n=2$ *and* $a=2, b=4$

p_i	$H_{a,b}(P)$
0	0
0.1	0.739892
0.2	1.109399
0.3	1.335030
0.4	1.459862
0.5	1.50000
0.6	1.459862
0.7	1.335030
0.8	1.109399
0.9	0.739892
1	0

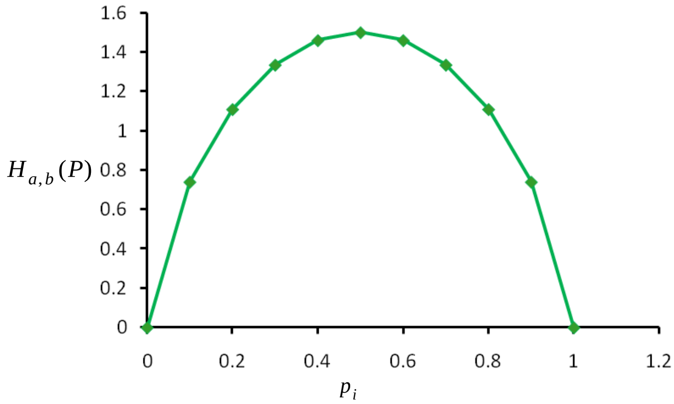

Figure 1: Concavity of $H_{a,b}(P)$ with respect to p_i

(vi) The entropy is a continuous function of $p_1, p_2, \ldots, p_n$.

The continuity of $H_{a,b}(P)$ is obvious from the graph of $H_{a,b}(P)$ plotted against p_i.

(vii) The entropy attains its maximum value at the uniform distribution.

Consider the Lagrange's function

$$L = -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b-a}\sum_{i=1}^{n}\left(1 - p_i^{(b-a)\,p_i}\right) - \lambda\left(\sum_{i=1}^{n} p_i - 1\right).$$

Differentiating the above equation w.r.to p_i and equating the derivative to zero, we get

$$\log p_i + p_i^{(b-a)\,p_i}(1+\log p_i) = -(1+\lambda).$$

The above relation is true only if $p_1 = p_2 = \ldots = p_n$. Further, since $\sum_{i=1}^{n} p_i = 1$, we have $p_i = \frac{1}{n}$ for all $i = 1,2,\ldots,n$. Thus, the maximum value of $H_{a,b}(P)$ is attained at the uniform distribution.

(viii) When entropy is at its maximum, it becomes an increasing function of *n*.

If $\phi(n)$ denotes the maximum value of the entropy, then

$$\phi(n) = \log n + \frac{n}{b-a}\left[1 - \left(\frac{1}{n}\right)^{\frac{b-a}{n}}\right].$$

Therefore $$\phi'(n) = \frac{1}{n} + \left(\frac{1}{n}\right)^{\frac{b-a}{n}+1}\left[1 + \log\frac{1}{n}\right] + \frac{1}{b-a}\left[1 - \left(\frac{1}{n}\right)^{\frac{b-a}{n}}\right] > 0.$$

This shows that $\phi(n)$ is an increasing function of n.

IV. Application of the Proposed Entropy Measure to Contingency Table Inference

Let ω_{ij} denote the observed frequency in the i^{th} row and j^{th} column of a $p \times q$ contingency table. Let $r_1, r_2, \ldots, r_p$ denote the sum of all the elements in $1^{\text{st}}, 2^{\text{nd}}, \ldots, p^{\text{th}}$ rows and let $c_1, c_2, \ldots, c_q$ denote the sum of all the elements in $1^{\text{st}}, 2^{\text{nd}}, \ldots, q^{\text{th}}$ columns respectively. Then, we have

$$\sum_{j=1}^{q} \omega_{ij} = r_i\ ;\ \ i = 1, 2, \ldots, p, \tag{14}$$

$$\sum_{i=1}^{p} \omega_{ij} = c_j\ ;\ \ j = 1, 2, \ldots, q, \tag{15}$$

$$\sum_{j=1}^{q}\sum_{i=1}^{p} \omega_{ij} = \sum_{i=1}^{p} r_i = \sum_{j=1}^{q} c_j = X. \tag{16}$$

Now, we suppose that the frequencies are unknown and we estimate these values from the available information of row totals and column totals. Since row totals and column totals are equal, there are $p+q-1$ independent constraints. With $p+q-1$ constraints and pq cells, our problem is to find the frequency in each of the pq cells.

Let ν_{ij} denote the estimated frequency in the i^{th} row and j^{th} column. To find ν_{ij}, we use the entropy function introduced in equation (12) and apply the Maximum Entropy Principle. Since the probability of the $(i, j)^{\text{th}}$ cell frequency is $\frac{\nu_{ij}}{X}$, the corresponding entropy function is

$${}_1H^{a,b} = -\sum_{j=1}^{q}\sum_{i=1}^{p} \frac{\nu_{ij}}{X}\log\frac{\nu_{ij}}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1 - \left(\frac{\nu_{ij}}{X}\right)^{(b-a)\frac{\nu_{ij}}{X}}\right]. \tag{17}$$

Next, we maximize (17) subject to the constraints

$$\sum_{j=1}^{q} \nu_{ij} = r_i\ ;\ \ i = 1, 2, \ldots, p, \tag{18}$$

$$\sum_{i=1}^{p} \nu_{ij} = c_j\ ;\ \ j = 1, 2, \ldots, q, \tag{19}$$

$$\sum_{j=1}^{q}\sum_{i=1}^{p} \nu_{ij} = \sum_{i=1}^{p} r_i = \sum_{j=1}^{q} c_j = X. \tag{20}$$

Now, we apply Lagrange's method to find the maximum value of ${}_1H^{a,b}$ subject to the constraints (18), (19) and (20). The corresponding Lagrangian is

$$L \cong -\sum_{j=1}^{q}\sum_{i=1}^{p} \frac{\nu_{ij}}{X}\log\frac{\nu_{ij}}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1 - \left(\frac{\nu_{ij}}{X}\right)^{(b-a)\frac{\nu_{ij}}{X}}\right]$$
$$-\lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right)$$

or $$L \cong -\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{\nu_{ij}}{X}\log\frac{\nu_{ij}}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1 - e^{(b-a)\frac{\nu_{ij}}{X}\log\left(\frac{\nu_{ij}}{X}\right)}\right] - \lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right)$$

or $$L \cong -\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{\nu_{ij}}{X}\log\frac{\nu_{ij}}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1 - \left(1 + (b-a)\frac{\nu_{ij}}{X}\log\left(\frac{\nu_{ij}}{X}\right)\right)\right] - \lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right)$$

or $$L \cong -2\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{\nu_{ij}}{X}\log\frac{\nu_{ij}}{X} - \lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right)$$

or $$L \cong -\frac{2}{X}\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij}(\log\nu_{ij} - \log X) - \lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right)$$

or $$L \cong -\frac{2}{X}\left[\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij}\log\nu_{ij} - \log X\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij}\right] - \lambda_i\left(\sum_{j=1}^{q}\nu_{ij} - r_i\right) - \mu_j\left(\sum_{i=1}^{p}\nu_{ij} - c_j\right) - \xi\left(\sum_{j=1}^{q}\sum_{i=1}^{p}\nu_{ij} - X\right).$$

Differentiating the above equation with respect to ν_{ij} and equating the derivative to zero, we get

$$-\frac{2}{X}[(1+\log\nu_{ij}) - \log X] - \lambda_i - \mu_j - \xi = 0$$

or $$\log\nu_{ij} = -\frac{1}{2}(X\lambda_i - 2\log X + 2) - \frac{X}{2}(\mu_j + \xi)$$

or $$\nu_{ij} = e^{-\frac{1}{2}(X\lambda_i - 2\log X + 2)}e^{-\frac{X}{2}(\mu_j+\xi)}$$

or $$\nu_{ij} = \theta_i\psi_j, \tag{21}$$

where $$\theta_i = e^{-\frac{1}{2}(X\lambda_i - 2\log X + 2)} \text{ and } \psi_j = e^{-\frac{X}{2}(\mu_j+\xi)}.$$

Substituting the value of ν_{ij} from equation (21) in equations (18), (19) and (20), we get

$$\sum_{j=1}^{q}\theta_i\psi_j = r_i;\quad i = 1, 2, \ldots, p,$$

$$\sum_{i=1}^{p}\theta_i\psi_j = c_j;\quad j = 1, 2, \ldots, q,$$

$$\sum_{j=1}^{q}\sum_{i=1}^{p}\theta_i\psi_j = \sum_{i=1}^{p}r_i = \sum_{j=1}^{q}c_j = X$$

or $$\theta_i\sum_{j=1}^{q}\psi_j = r_i;\quad i = 1, 2, \ldots, p, \tag{22}$$

$$\psi_j\sum_{i=1}^{p}\theta_i = c_j;\quad j = 1, 2, \ldots, q, \tag{23}$$

$$\left(\sum_{j=1}^{q}\psi_j\right)\left(\sum_{i=1}^{p}\theta_i\right) = X. \tag{24}$$

Multiplying equations (22) and (23), we get

$$\theta_i\psi_j\left(\sum_{j=1}^{q}\psi_j\right)\left(\sum_{i=1}^{p}\theta_i\right) = r_i c_j$$

or $$\theta_i\psi_j X = r_i c_j$$ [Using equation (24)] or $$\theta_i \psi_j = \frac{r_i c_j}{X}$$

or $$\nu_{ij} = \frac{r_i c_j}{X}. \tag{25}$$

This ν_{ij} gives the maximum frequency estimate of $(i, j)^{\text{th}}$ cell. We shall represent this maximum frequency estimate by ρ_{ij}. Then, from equation (20), we have

$$\sum_{j=1}^{q}\sum_{i=1}^{p}\rho_{ij} = X. \tag{26}$$

The corresponding maximum entropy $H_{\max}^{a,b}$ is given by

$$H_{\max}^{a,b} = -\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{\rho_{ij}}{X}\log\frac{\rho_{ij}}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1-\left(\frac{\rho_{ij}}{X}\right)^{(b-a)\frac{\rho_{ij}}{X}}\right]$$

$$= -\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{r_i c_j}{X^2}\log\frac{r_i c_j}{X^2} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1-\left(\frac{r_i c_j}{X^2}\right)^{(b-a)\frac{r_i c_j}{X^2}}\right]$$

$$= -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} - \sum_{j=1}^{q}\frac{c_j}{X}\log\frac{c_j}{X} + \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}\left[1-e^{(b-a)\frac{r_i c_j}{X^2}\log\left(\frac{r_i c_j}{X^2}\right)}\right]$$

$$= -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} - \sum_{j=1}^{q}\frac{c_j}{X}\log\frac{c_j}{X} - \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}(b-a)\frac{r_i c_j}{X^2}\log\left(\frac{r_i c_j}{X^2}\right)$$

$$= -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} - \sum_{j=1}^{q}\frac{c_j}{X}\log\frac{c_j}{X} - \frac{1}{b-a}\sum_{j=1}^{q}\sum_{i=1}^{p}(b-a)\frac{r_i c_j}{X^2}\left(\log\frac{r_i}{X} + \log\frac{c_j}{X}\right)$$

$$= -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} - \sum_{j=1}^{q}\frac{c_j}{X}\log\frac{c_j}{X} - \frac{1}{b-a}\left(\sum_{j=1}^{q}\frac{c_j}{X}\right)\left[\sum_{i=1}^{p}(b-a)\frac{r_i}{X}\log\frac{r_i}{X}\right] - \frac{1}{b-a}\left(\sum_{i=1}^{p}\frac{r_i}{X}\right)\left[\sum_{j=1}^{q}(b-a)\frac{c_j}{X}\log\frac{c_j}{X}\right]$$

$$= -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} - \sum_{j=1}^{q}\frac{c_j}{X}\log\frac{c_j}{X} - \frac{1}{b-a}\sum_{i=1}^{p}\log\left(\frac{r_i}{X}\right)^{(b-a)\frac{r_i}{X}} - \frac{1}{b-a}\sum_{j=1}^{q}\log\left(\frac{c_j}{X}\right)^{(b-a)\frac{c_j}{X}}$$

i.e., $$H_{\max}^{a,b} = H_1^{a,b} + H_2^{a,b}, \tag{27}$$

where $H_1^{a,b}$ represents the entropy of the row totals, i.e.,

$$H_1^{a,b} = -\sum_{i=1}^{p}\frac{r_i}{X}\log\frac{r_i}{X} + \frac{1}{b-a}\sum_{i=1}^{p}\left[1-\left(\frac{r_i}{X}\right)^{(b-a)\frac{r_i}{X}}\right]$$

and $H_2^{a,b}$ represents the entropy of the column totals, i.e.,

$$H_2^{a,b} = -\sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X} + \frac{1}{b-a} \sum_{j=1}^{q} \left[1 - \left(\frac{c_j}{X} \right)^{(b-a)\frac{c_j}{X}} \right].$$

Let $H^{a,b}$ denote the entropy of the observed cell frequencies ω_{ij}. Then, we have

$$H^{a,b} = -\sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X} + \frac{1}{b-a} \sum_{j=1}^{q} \sum_{i=1}^{p} \left[1 - \left(\frac{\omega_{ij}}{X} \right)^{(b-a)\frac{\omega_{ij}}{X}} \right].$$

Obviously, $\qquad H_{\max}^{a,b} \geq H^{a,b}$

i.e., $\qquad H_1^{a,b} + H_2^{a,b} \geq H^{a,b}.$

Thus the gap between the two frequencies is

$$G = H_1^{a,b} + H_2^{a,b} - H^{a,b} \geq 0.$$

Now, $\qquad G = H_1^{a,b} + H_2^{a,b} - H^{a,b}$

$$= -\sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} + \frac{1}{b-a} \sum_{i=1}^{p} \left[1 - \left(\frac{r_i}{X} \right)^{(b-a)\frac{r_i}{X}} \right] - \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X}$$

$$+ \frac{1}{b-a} \sum_{j=1}^{q} \left[1 - \left(\frac{c_j}{X} \right)^{(b-a)\frac{c_j}{X}} \right] + \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X} - \frac{1}{b-a} \sum_{j=1}^{q} \sum_{i=1}^{p} \left[1 - \left(\frac{\omega_{ij}}{X} \right)^{(b-a)\frac{\omega_{ij}}{X}} \right]$$

$$= -\sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} + \frac{1}{b-a} \sum_{i=1}^{p} \left[1 - e^{(b-a)\frac{r_i}{X} \log \frac{r_i}{X}} \right] - \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X}$$

$$+ \frac{1}{b-a} \sum_{j=1}^{q} \left[1 - e^{(b-a)\frac{c_j}{X} \log \frac{c_j}{X}} \right] + \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X} - \frac{1}{b-a} \sum_{j=1}^{q} \sum_{i=1}^{p} \left[1 - e^{(b-a)\frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X}} \right]$$

$$= -\sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} - \sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} - \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X} - \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X}$$

$$+ \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X} + \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X}$$

i.e., $$G = 2 \left[-\sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} - \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X} + \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{X} \right]. \tag{28}$$

From equation (14), we have

$$\sum_{j=1}^{q} \omega_{ij} = r_i$$

$$\Rightarrow \qquad \frac{r_i}{X} = \sum_{j=1}^{q} \frac{\omega_{ij}}{X}$$

$$\Rightarrow \qquad \frac{r_i}{X} \log \frac{r_i}{X} = \sum_{j=1}^{q} \frac{\omega_{ij}}{X} \log \frac{r_i}{X}$$

$$\Rightarrow \qquad \sum_{i=1}^{p} \frac{r_i}{X} \log \frac{r_i}{X} = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\omega_{ij}}{X} \log \frac{r_i}{X}. \tag{29}$$

Similarly, from equation (15), we have

$$\sum_{i=1}^{p} \omega_{ij} = c_j$$

$$\Rightarrow \qquad \frac{c_j}{X} = \sum_{i=1}^{p} \frac{\omega_{ij}}{X}$$

$$\Rightarrow \qquad \frac{c_j}{X} \log \frac{c_j}{X} = \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{c_j}{X}$$

$$\Rightarrow \qquad \sum_{j=1}^{q} \frac{c_j}{X} \log \frac{c_j}{X} = \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{c_j}{X}. \tag{30}$$

Using equations (29) and (30), equation (28) reduces to

$$G = 2 \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\frac{\omega_{ij}}{X}}{\frac{r_i}{X} \frac{c_j}{X}}$$

or

$$G = 2 \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\omega_{ij}}{X} \log \frac{\omega_{ij}}{\rho_{ij}}. \tag{31}$$

Obviously, $G = 0$ if and only if $\frac{\omega_{ij}}{\rho_{ij}} = 1$, i.e., $\omega_{ij} = \rho_{ij}$, where ω_{ij} and ρ_{ij} are the observed frequencies and the maximum frequency estimates respectively..

Thus we find that frequency estimates determined from Maximum Entropy Principle give the same values as provided by the hypothesis of independence.

Now, we consider the case when $G > 0$. In this case, let

$$\omega_{ij} = \rho_{ij} + \varepsilon_{ij}, \tag{32}$$

where $-1 < \varepsilon_{ij} < 1$ and ε_{ij} are very small quantities. Then

$$\sum_{j=1}^{q} \sum_{i=1}^{p} \omega_{ij} = \sum_{j=1}^{q} \sum_{i=1}^{p} \rho_{ij} + \sum_{j=1}^{q} \sum_{i=1}^{p} \varepsilon_{ij}.$$

But, from equations (16) and (26), $\sum_{j=1}^{q} \sum_{i=1}^{p} \omega_{ij} = \sum_{j=1}^{q} \sum_{i=1}^{p} \rho_{ij} = X$. Therefore

$$\sum_{j=1}^{q} \sum_{i=1}^{p} \varepsilon_{ij} = 0. \tag{33}$$

Now,

$$G = 2 \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\rho_{ij} + \varepsilon_{ij}}{X} \log \frac{\rho_{ij} + \varepsilon_{ij}}{\rho_{ij}}$$

$$= \frac{2}{X} \sum_{j=1}^{q} \sum_{i=1}^{p} (\rho_{ij} + \varepsilon_{ij}) \log \left(1 + \frac{\varepsilon_{ij}}{\rho_{ij}} \right)$$

$$= \frac{2}{X} \sum_{j=1}^{q} \sum_{i=1}^{p} (\rho_{ij} + \varepsilon_{ij}) \left[\frac{\varepsilon_{ij}}{\rho_{ij}} - \frac{(\varepsilon_{ij} / \rho_{ij})^2}{2} + \dots \right]$$

$$= \frac{2}{X} \sum_{j=1}^{q} \sum_{i=1}^{p} \varepsilon_{ij} + \frac{1}{X} \sum_{j=1}^{q} \sum_{i=1}^{p} \frac{\varepsilon_{ij}^{\,2}}{\rho_{ij}}.$$

But, from equation (33), we have $\sum_{j=1}^{q}\sum_{i=1}^{p}\varepsilon_{ij} = 0$. Therefore

$$G = \frac{1}{X}\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{\varepsilon_{ij}^{2}}{\rho_{ij}} = \frac{1}{X}\sum_{j=1}^{q}\sum_{i=1}^{p}\frac{(\omega_{ij}-\rho_{ij})^{2}}{\rho_{ij}} = \frac{1}{X}\chi^{2},$$

where χ^2 denotes the chi-square. Thus, we have

$$\chi^2 = XG \tag{34}$$

i.e., chi-square is X times the additional information provided by the observed frequencies over the information previously obtained by the row totals and column totals. This interpretation aligns with the fundamental purpose of the chi-square test in contingency tables: assessing how well the observed data fits the expected distribution.

V. Conclusion

Through the present work, we have tried to contribute to the ongoing development of more flexible and powerful tools for statistical analysis, offering a new perspective on contingency table inference and providing a foundation for future advancements in entropy-based methods. The proposed model leverages entropy as a central concept to capture the underlying dependencies and associations between variables in a more flexible and computationally efficient manner than traditional methods. The significance of the work lies not only in the development of a new entropy model but also in its practical applications. By expanding the applicability of the maximum entropy principle, the proposed model provides a powerful tool for statistical inference in a wide range of domains, including social sciences, biology, economics, and any field where contingency tables are commonly used. Additionally, the enhanced flexibility of the model opens up new possibilities for exploring complex dependencies and interactions within datasets, ultimately leading to more accurate and meaningful insights.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. Bell. System Technical Journal, 27: 379–423, 623–659.

[2] Arimoto, S. (1971). Information theoretical consideration on estimation problems. *Information and Control,* 19: 181–194.

[3] Behara, M. and Chawla, J. S. (1974). Generalized gamma-entropy. *Selecta statistica canadiana,* 2: 15–38.

[4] Sharma, B. D. and Mittal, D. P. (1975). New non-additive measures of entropy for discrete probability distributions. Journal of Mathematical Sciences, 10: 28–40.

[5] Kapur, J.N. (1986). Four families of measures of entropy. *Indian Journal of Pure and Applied Mathematics,* 17: 429–449.

[6] Parkash, O. and Kakkar, P. (2012). Development of two new mean codeword lengths. Information Sciences, 207: 90–97.

[7] Parkash, O. and Kakkar, P. (2014). New information theoretic models, their detailed properties and new inequalities. Canadian Journal of Pure and Applied Sciences, 8(3): 3115-3123.

[8] Bulinski, A. and Kozhevin, A. (2019). Statistical estimation of conditional Shannon entropy. *ESAIM Probabability and Statistics,* 23: 350–386.

[9] Lu,Y., Wang, M., Wu, W., Zhang, Q., Han, Y., Kausar, T., Chen, S., Liu, M., and Wang, B. (2020). Entropy-Based Pattern Learning Based on Singular Spectrum Analysis Components for Assessment of Physiological Signals. Complexity, Article ID 4625218.

[10] Sholehkerdar, A., Tavakoli, J. and Liu, Z. (2020). Theoretical analysis of Tsallis Entropy-Based quality measure for weighted averaging Image Fusion. Information Fusion, 58: 69-81.

[11] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52: 479–487.

[12] Lenormand, M., Samaniego, H., Chaves, J.C., Vieira, V.D.F., Barbosa da Silva, M.A.H., Evsukoff, A.G. (2020). Entropy as a measure of attractiveness and socioeconomic complexity in Rio de Janeiro metropolitan area. *Entropy,* 22: 368.

[13] Manzoor, S., Siddiqui, M. K. and Ahmad, S. (2020). On entropy measures of molecular graphs using topological indices. Arabian Journal of Chemistry, 13(8):6285–6298.

[14] Parkash, O. and Mukesh (2021). Two new parametric entropic models for discrete probability distributions. Turkish Journal of Computer and Mathematics Education, 12(6): 2949–2954.

[15] Siddiqui, M. K., Manzoor, S., Ahmad, S. and K. A. Kaabar, M. K.A. (2021). On Computation and Analysis of Entropy Measures for Crystal Structures. Graph Invariants and Their Applications.

[16] Teixeira, A. Souto, A. and Antunes, L. (2021). On Conditional Tsallis Entropy. *Entropy*, *23*(11): 1427.

[17] Segura, E. A., Zamar, M. D. G. and Squillante, M. (2021). Examining the Research on Business Information-Entropy Correlation in the Accounting Process of Organizations. *Entropy, 23*(11):1493.

[18] Saraiva, P. (2023). On Shannon entropy and its applications. Kuwait Journal of Science, 50(3): 194–199.

[19] Zhang, J. and Shi, J. (2022). Asymptotic normality for plug-in estimators of generalized Shannon's entropy. *Entropy*, 24: 683. 62.

[20] Friesner, D., Bozman, CS, McPherson, M., Valente, F. and Zhang, A. (2023). Information entropy as a quality control tool in survey research. *Journal of Marketing Development and Competitiveness, 17(1).*

[21] Stoyanov, J.M., Tagliani, A. and Inverardi, P.L.N. Maximum (2024). Entropy Criterion for Moment Indeterminacy of Probability Densities. *Entropy, 26*(2), 121.

[22] Jizba, P. and Korbel, J. (2019). Maximum entropy principle in statistical inference: Case for Non-Shannonian entropies. *Physical Review Letters,* 122, 120601.

[23] Gao, X. and Deng, Y. (2020). The Pseudo-Pascal Triangle of Maximum Deng Entropy. *International Journal of Computers, Communications & Control*, 15: 1–10.

[24] Elgawad M. A. A., Barakat H. M., Xiong S. and Alyami S. A. (2021). Information Measures for Generalized Order Statistics and Their Concomitants under General Framework from Huang-Kotz FGM Bivariate Distribution. Entropy 23(3): 335(1–17).

[25] Bajic D., Đaji´c V. and Milovanovi´c B. (2021). Entropy analysis of COVID-19 cardiovascular signals. *Entropy*, 23(1): 87 (1–15).

[26] Fowler, R. and Heckman, J.J. (2022). Misanthropic entropy and renormalization as a communication channel. *International Journal of Modern Physics, A* 37(16): 2250109, 18.

[27] Li, Y., Geng, B. and Jiao, S. (2022). Dispersion entropy-based Lempel-Ziv complexity: A new metric for signal analysis. Chaos Solitons & Fractals 161: 112400.

[28] Bisikalo, O., Kharchenko,V. Kovtun, V., Krak, L. and Pavlov, S. (2023). Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. *Entropy, 25* (2), 184.

[29] Gregorio, J. D., Sánchez, D. and Toral, R. (2024). Entropy Estimators for Markovian Sequences: A Comparative Analysis, *Entropy*, 26(1), 79.