entropy

MDPI

Review

Shannon, Rényi, Tsallis, and Beyond: A Unified Sharma–Mittal Framework for Parameter-Integrated Feature Selection

Nida Oruç Ünal [1,*], Muzaffer Göztaş [2] and Doğan Yıldız [2]

[1] Department of Management Information Systems, Faculty of Economics, Administrative and Social Sciences, Istanbul Topkapi University, Istanbul 34662, Türkiye
[2] Department of Statistics, Faculty of Arts and Sciences, Yıldız Technical University, Istanbul 34220, Türkiye; muzaffer.goztas@yildiz.edu.tr (M.G.); dyildiz@yildiz.edu.tr (D.Y.)
* Correspondence: nidaoruc@topkapi.edu.tr

Abstract

Entropy originated in thermodynamics and has since become a central tool for quantifying uncertainty in probabilistic and information-theoretic systems. This study provides a unified mathematical synthesis of the Shannon, Rényi, Tsallis, and Sharma–Mittal entropies, focusing on their axiomatic foundations, functional structures, and limiting relationships. We show that Rényi entropy reduces to Shannon entropy as $\alpha \to 1$, Tsallis entropy reduces to Shannon entropy as $q \to 1$, and the two-parameter Sharma–Mittal entropy encompasses both Rényi (as $\beta \to 1$) and Tsallis (when $\beta = \alpha$), converging to Shannon entropy in the joint limit $(\alpha, \beta) \to (1, 1)$. Additivity, pseudo-additivity, concavity, and stability are compared across the four families, consolidating the unifying role of Sharma–Mittal entropy noted in earlier work rather than treating it merely as a more flexible formulation. Building on this synthesis, we connect it to a previously developed parameter-integrated framework that addresses the practical difficulty introduced by the Sharma–Mittal entropy's two-parameter structure: fixing (α, β) at a single, arbitrary pair makes any derived quantity—and, in feature selection, the resulting variable ranking—dependent on that choice. The framework integrates the conditional Sharma–Mittal entropy and information-gain surfaces volumetrically over a bounded region of the α-β plane, yielding three parameter-integrated indices—PICSME, PIGSME, and NIGSME—that are exact monotone transformations of one another and therefore induce an identical variable ranking for a fixed dataset and integration region. This illustrates how the unifying position of Sharma–Mittal entropy can be operationalized into a practical feature-ranking tool that does not require the selection of a single predefined (α, β) pair.

Keywords: Shannon entropy; Rényi entropy; Tsallis entropy; Sharma–Mittal entropy; generalized entropy; feature selection; kernel density estimation

Check for updates

Academic Editor: Andrea Murari

Received: 9 October 2026
Revised: 30 October 2026
Accepted: 3 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The concept of entropy first emerged in the context of thermodynamics to explain the direction of heat flow and the degree of disorder in physical systems. Initially treated by Clausius [1] as a state function related to the second law of thermodynamics, entropy later acquired a probabilistic and statistical meaning through the work of Boltzmann [2] and Gibbs [3]. In this process, entropy was linked to the number of microscopic arrangements corresponding to a macroscopic state and to the statistical structure of the system. Thus, entropy evolved from being merely a concept expressing disorder in physical systems to becoming a fundamental mathematical tool for measuring uncertainty in probabilistic systems.

In the classical thermodynamic formulation, Clausius [1] introduced entropy as a state function associated with reversible heat transfer, expressed by $dS = \frac{\delta Q_{\text{rev}}}{T}$, where T denotes absolute temperature, and δQ_{rev} is the infinitesimal heat transferred reversibly. This formulation established entropy within the second law of thermodynamics, independently of an explicit probability-based representation.

The statistical interpretation was subsequently developed through Boltzmann's principle [2], commonly written as $S = k_B \ln W$, where k_B is the Boltzmann constant, and W denotes the number of equally probable microstates compatible with a macroscopic state. Gibbs [3] extended this probabilistic description through $S = -k_B \sum_{i=1}^{n} p_i \ln p_i$, where p_i represents the probability of the i-th microstate. For an equiprobable distribution, $p_i = 1/W$, the Gibbs expression reduces to Boltzmann's principle. Ben-Naim [4] argues that thermodynamic entropy should be distinguished from a general information measure defined over arbitrary probability distributions.

The modern interpretation of entropy in information theory was shaped by Shannon's 1948 paper, "A Mathematical Theory of Communication" [5]. According to Shannon, the fundamental problem of communication is the ability to reproduce a message selected at one point, either exactly or approximately, at another point. In this approach, what matters is not so much the semantic content of the message as the fact that it was selected from a set of possible messages. Shannon [5] noted that a logarithmic measure is natural for quantifying the amount of information and discussed the suitability of this measure from both practical and mathematical perspectives. This framework treated entropy as a measure of average uncertainty associated with a probability distribution and established a fundamental mathematical structure for communication theory, coding theory, statistical inference, and subsequent data science applications.

A key strength of Shannon entropy is that it can be uniquely determined under certain axiomatic properties. These axioms, systematically addressed by Khinchin [6], include the principles of continuity, maximality, extensibility, and additivity. In particular, the property of additivity is one of the fundamental principles that defines Shannon entropy's central role in classical information theory. According to this principle, the joint entropy of two statistically independent systems is equal to the sum of their individual entropies. Rényi [7] also emphasized this additivity property for independent experiments when discussing Shannon entropy and noted that it is one of the most important properties of entropy.

However, the additive nature of Shannon entropy may not be sufficient for every system. In particular, there has been a need for more flexible measures of entropy in systems involving long-range interactions, multifractal structures, strong correlations, or those that go beyond the classical Boltzmann–Gibbs statistics. In this line of development, Rényi entropy generalizes Shannon entropy through the concept of entropy order, thereby allowing for the control of the effect of different probability weights on the measurement of uncertainty [7]. In Rényi's work [7], Shannon entropy is derived as a special limiting case of Rényi entropy.

Tsallis entropy, on the other hand, was proposed to generalize the Boltzmann–Gibbs statistic and has gained significant prominence, particularly in the context of multifractal structures and generalized statistical ensembles [8]. Tsallis [8] proposed a new form of entropy defined by the parameter q and demonstrated that it reduces to the classical Boltzmann–Gibbs statistic in the limit as $q \to 1$. In this respect, Tsallis entropy can be regarded as an important single-parameter generalization that diverges from the Shannon/Boltzmann–Gibbs line but preserves the classical structure in the limit.

Sharma–Mittal entropy, on the other hand, extends this line of development to a broader mathematical framework. This two-parameter structure, proposed by Sharma and Mittal [9], holds a significant position among non-additive entropy measures due to the limit

and special-case relationships it establishes with Shannon, Rényi, and Tsallis entropies. As demonstrated in the review by Esteban and Morales [10] on entropy statistics, many entropy measures—such as Shannon, Rényi, Havrda–Charvát, Arimoto, and Sharma–Mittal—can be treated as special or limiting cases under a more general functional framework. In this context, the Sharma–Mittal entropy occupies a unifying position in the literature on generalized entropies because it goes beyond single-parameter entropy families to offer a two-parameter structure.

The aim of this study is to provide a unified mathematical synthesis of the Shannon, Rényi, Tsallis, and Sharma–Mittal entropies by examining their axiomatic foundations, functional structures, compositional properties, and limiting relationships. Beyond this theoretical synthesis, the study connects the two-parameter structure of the Sharma–Mittal entropy to a previously developed parameter-integrated volumetric framework for feature selection, thereby illustrating how generalized entropy theory can be operationalized for continuous-variable analysis.

This study does not claim new axiomatic or limiting results for the individual entropy families. Sections 2–5 consolidate established definitions and properties, and the unifying role of the Sharma–Mittal entropy has been discussed previously by Masi (see Sections 5.1 and 6.3). The contribution of this work lies in presenting the four families within a common notation, placing their axiomatic foundations, composition rules, and limiting transitions side by side (Section 6), and connecting this structure, within a single treatment, to the parameter-integrated indices PICSME, PIGSME, and NIGSME (Section 7), so that the entropy families and the indices that operationalize their parameter dependence are presented as one coherent framework. The indices themselves were introduced in the authors' earlier work (see Section 7.1), where their derivations, estimation protocol, and empirical validation are reported; Section 7 does not introduce a new method.

The remainder of the study is organized as follows. Sections 2–5 present the mathematical foundations of the Shannon, Rényi, Tsallis, and Sharma–Mittal entropies, respectively, including their axiomatic structure, functional form, and limiting relationships. Section 6 provides a comparative synthesis of their parameter structures, compositional properties, and limiting relationships. Section 7 connects this theoretical structure to a parameter-integrated volumetric feature-ranking framework.

Section 8 summarizes the main findings, acknowledges the limitations, and outlines directions for future research. Generative AI tools were used as assistive tools only. ChatGPT (OpenAI; GPT-6, high) was used to assist in creating the figures, which were checked by the authors for consistency with the text, and Claude (Anthropic; Claude Sonnet 5.5) was used to help refine English phrasing. All mathematical content, definitions, equations, and citations were prepared and verified by the authors, and no AI tool was used to generate data or results.

2. Shannon Entropy

2.1. Definition of Shannon Entropy

Shannon entropy is a fundamental functional in information theory that measures uncertainty based on a discrete probability distribution. Let the possible outcomes of a random variable defined on a finite sample space be $x_1, x_2, x_3, \ldots, x_n$ and let the corresponding probabilities be $p_1, p_2, p_3, \ldots, p_n$. Here, the conditions $P_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$ hold. Shannon entropy is defined as follows [5]:

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

Since $-\log p_i = \log(1/p_i)$, Shannon entropy can be interpreted as the probability-weighted average of the information content of each event. Here, $\log \frac{1}{p_i}$ represents the amount of information gained when event x_i occurs, while $p_i \log \frac{1}{p_i}$ represents the expected contribution of this information. Thus, Shannon entropy measures the average uncertainty that exists prior to the outcome of a random experiment.

The base of the logarithm determines the unit of measurement for entropy. When the base of the logarithm is 2, entropy is expressed in "bits"; when the natural logarithm is used, it is expressed in "nat."

For a discrete probability distribution over physical microstates, the Gibbs entropy is mathematically proportional to the Shannon entropy defined in Equation (1). When the natural logarithm is used, the relationship is $S = k_B H(x)$, whereas for a base-2 logarithm, the corresponding relationship is $S = k_B(\ln 2)H(x)$. Thus, the two expressions share the same functional structure, differing by a dimensional scaling factor and, where applicable, a logarithmic-base conversion. Nevertheless, this mathematical correspondence does not imply that their physical interpretations are identical. Ben-Naim distinguishes Shannon's measure of information, which is defined on any probability distribution, from thermodynamic entropy, which is defined on a much more restricted set of distributions, while also arguing that the former provides a quantitative basis for interpreting the latter. In discussing Boltzmann's H-theorem, he further argues that the H-function is a Shannon measure of information but not thermodynamic entropy [4].

The intuitive meaning of Shannon entropy is that it measures how "uncertain" or "spread out" a distribution is. If the distribution is concentrated on a single outcome—that is, if one outcome occurs with probability 1—there is no uncertainty, and the entropy is equal to zero. Conversely, uncertainty reaches its highest level when all outcomes have equal probability. In the case of a uniform distribution with n outcomes, where $p_i = \frac{1}{n}$, Shannon entropy takes the value

$$H(x) = \log n \tag{2}$$

This result shows that uncertainty increases as the number of possible outcomes increases.

2.2. Axiomatic Foundations

The importance of Shannon entropy stems not only from the fact that it is a suitable functional for measuring uncertainty, but also from the fact that it can be derived uniquely under certain natural axioms.

This axiomatic framework was later systematically addressed by Khinchin [6] and has been referred to in the literature as the Shannon–Khinchin axioms. In general, these axioms can be summarized around four fundamental principles. The first is continuity: the entropy function must behave continuously in response to small changes in probabilities. The second is the maximality principle: for a fixed number of possible outcomes, entropy must take its maximum value in a uniform distribution where all outcomes have equal probability. The third is extensibility: adding an impossible event with zero probability to the distribution should not change the entropy value. The fourth is the principle of additivity or separability: in a multi-stage selection process, the total uncertainty must be expressed as the probability-weighted sum of the uncertainty in the first stage and the conditional uncertainties of the subsequent stages.

Rényi [7] summarizes Faddeev's postulates, stating that the properties of symmetry, continuity, normalization, and separability characterize Shannon entropy [11]. In this context, Faddeev's postulates show that Shannon entropy is not merely an intuitive measure of uncertainty, but also a functional that arises necessarily under certain mathematical conditions that are taken for granted.

The additivity axiom is one of the most distinctive features of Shannon entropy. If, for two statistically independent systems or experiments X and Y, the joint distribution can be written as $p(x,y) = p(x)p(y)$, then the Shannon entropy can be expressed as

$$H(X,Y) = H(X) + H(Y) \tag{3}$$

Rényi [7] also considered this property to be one of the most important characteristics of Shannon entropy. This property explains why Shannon entropy plays such a central role in classical information theory. At the same time, however, a significant portion of the later generalized entropy measures were developed precisely by modifying or relaxing this additivity axiom.

2.3. Main Mathematical Properties

Shannon entropy has many fundamental mathematical properties. First, entropy is never negative. Since $\log(1 / p_i) \geq 0$ for $0 < p_i \leq 1$, each term $p_i \log(1 / p_i)$ contributes a non-negative value. Therefore,

$$H(x) \geq 0 \tag{4}$$

holds. Entropy being zero corresponds to the case where the distribution is completely certain. In other words, if the probability of one event is 1, and the probability of all other events is 0, uncertainty disappears.

Second, Shannon entropy reaches its maximum value under a uniform distribution. If all n possible outcomes have equal probability, that is, $P_i = 1 / n$, then

$$H(x) = log(n) \tag{5}$$

is obtained.

Third, Shannon entropy is a concave function. This property states that the entropy of a mixture of two probability distributions will not be less than the weighted average of the entropies of the distributions. In other words, an increase in mixing or uncertainty within a probability distribution does not reduce entropy. Concavity is one of the fundamental properties that enables the use of Shannon entropy in the context of statistical uncertainty and the maximum entropy principle.

A geometric interpretation of Shannon entropy can be obtained by considering probability distributions on the simplex $\Delta_{n-1} = \{(p_1, \ldots, p_n) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}$. The uniform distribution $(1/n,\ldots,1/n)$ corresponds to the barycenter of this simplex and, as noted above, maximizes Shannon entropy. Entropy values therefore define level sets over the probability simplex, providing a geometric representation of uncertainty. A related but distinct geometric application was presented by Bormashenko et al. [12], who introduced Shannon-type informational measures of symmetry based on the occurrence of symmetry operations and symmetry elements in patterns composed of one- and two-dimensional shapes.

Fourth, Shannon entropy is invariant under permutations. When the order of the probabilities is changed, the entropy value remains the same because entropy depends only on the set of probability values. Therefore,

$$H(p_1,p_2,\ldots,p_n) = H(p_{\pi(1)},p_{\pi(2)},\ldots,p_{\pi(n)}) \tag{6}$$

where π denotes any permutation of the probabilities. This property indicates that entropy is influenced not by the names or order of events, but solely by their probability structure.

Fifth, Shannon entropy forms the basis for derived measures of information such as conditional entropy, joint entropy, and mutual information. Shannon [5] discussed the relationship between the uncertainty of joint events and conditional uncertainty, noting that when one variable is known, the uncertainty regarding the other variable does not

increase. This framework enables the definition of mutual information through the concepts of joint entropy and conditional entropy. Therefore, Shannon entropy not only measures the uncertainty of a single distribution but also plays a fundamental role in quantifying information sharing and the structure of dependencies among variables.

In conclusion, Shannon entropy is a classical measure of uncertainty possessing strong mathematical properties such as continuity, maximality, extensibility, additivity, concavity, and invariance under permutations. Due to these properties, it is regarded as the cornerstone of information theory. However, the additivity property, in particular, can be limiting in complex systems because it relies on the classical independence assumption. This limitation constitutes one of the primary motivations for the development of the Rényi, Tsallis, and Sharma–Mittal entropies, which will be discussed in subsequent sections.

3. Rényi Entropy

3.1. Definition of Rényi Entropy

Rényi entropy is one of the most fundamental single-parameter generalizations of Shannon entropy. Rényi [7] proposed examining the amount of uncertainty for a finite discrete probability distribution not only through the classical expected information content but also by weighting the probabilities in the distribution to varying degrees. This approach differs from Shannon entropy in that it offers a family of entropies capable of controlling sensitivity to different regions of the probability distribution.

Let P be a finite probability distribution

$$P = (p_1, p_2, \ldots, p_n) \tag{7}$$

such that $p_i \geq 0$ and

$$\sum_{i=1}^{n} p_i = 1 \tag{8}$$

For $\alpha > 0$ and $\alpha \neq 1$, the Rényi entropy is defined as follows:

$$H_\alpha^R(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right) \tag{9}$$

Here, α is the parameter that determines the order or degree of the entropy. This parameter controls which regions of the probability distribution are given greater weight. When $\alpha < 1$, the influence of low-probability events increases relatively, whereas when $\alpha > 1$, the influence of high-probability events becomes more pronounced. For this reason, Rényi entropy is considered not merely a single measure of uncertainty, but a family of parametric entropies that examines different aspects of the distribution through different values of α [7,13].

The term

$$\sum_{i=1}^{n} p_i^\alpha \tag{10}$$

in the structure of Rényi entropy provides an α-order moment-like summary of the probability distribution. The logarithmic transformation of this term ensures that Rényi entropy preserves its additivity property under independent systems. In this respect, Rényi entropy can be viewed as a generalization that modifies the way probabilities are weighted, without completely departing from the axiomatic framework of Shannon entropy.

In the literature on generalized entropy, Rényi entropy is recognized, along with Tsallis entropy, as one of the most widely used single-parameter entropy families. Amigó and colleagues [13] evaluated Rényi and Tsallis entropies as typical generalizations of the Boltzmann–Gibbs–Shannon entropy and note that both are linked to classical entropy via a limit. Fuentes and Gonçalves [14], on the other hand, emphasize that Rényi entropy

emerged in information theory as a parametric relaxation of Shannon entropy and was later addressed in the context of statistical mechanics.

3.2. Relationship with Shannon Entropy

The most fundamental relationship between Rényi entropy and Shannon entropy arises in the limit as $\alpha \to 1$. Rényi entropy is not directly defined for $\alpha = 1$, because both the numerator and denominator in the formula approach zero. However, this is a removable singularity, and taking the limit yields Shannon entropy:

$$\lim_{\alpha \to 1} H_\alpha^R(P) = -\sum_{i=1}^{n} p_i log p_i \tag{11}$$

This result shows that Rényi entropy includes Shannon entropy as a special limit case [7,13,14]. Therefore, Rényi entropy can be interpreted not as an independent measure replacing Shannon entropy, but rather as a broader parametric family that encompasses Shannon entropy.

The limit can be verified with L'Hôpital's rule. As $\alpha \to 1$, both $\log \sum_{i=1}^{n} p_i^\alpha$ and $1 - \alpha$ tend to zero, and differentiating both with respect to α gives

$$\lim_{\alpha \to 1} \frac{\log \sum_{i=1}^{n} p_i^\alpha}{1-\alpha} = \lim_{\alpha \to 1} \frac{\sum_{i=1}^{n} p_i^\alpha \log p_i / \sum_{i=1}^{n} p_i^\alpha}{-1} = -\sum_{i=1}^{n} p_i \log p_i \tag{12}$$

Thus, Shannon entropy emerges as the limit of Rényi entropy as α approaches 1.

This limit relationship is one of the fundamental elements that determine the position of Rényi entropy among generalized entropies. While Shannon entropy measures the expected logarithmic information content of probabilities, Rényi entropy can modify the distribution's sensitivity to different probability levels via the α parameter. Nevertheless, when $\alpha \to 1$, this parametric sensitivity disappears, and the classical Shannon form is restored [7,13].

3.3. Main Mathematical Properties

(i) Additivity: Rényi entropy preserves the additivity property for independent systems. For two statistically independent random variables X and Y, if their joint probabilities

$$p_{ij} = p_i q_j \tag{13}$$

then the Rényi entropy satisfies the following equation:

$$H_\alpha^R(X, Y) = H_\alpha^R(X) + H_\alpha^R(Y) \tag{14}$$

This follows from the factorization $\sum_{i,j} \left(p_i q_j\right)^\alpha = \left(\sum_i p_i^\alpha\right)\left(\sum_j q_j^\alpha\right)$, since the logarithm converts the product into a sum.

In this respect, Rényi entropy differs from Tsallis entropy because Tsallis entropy exhibits pseudo-additivity rather than classical additivity for independent systems. Therefore, Rényi entropy can be regarded as a generalization that preserves the additivity property of Shannon entropy but parameterizes the probability weighting [7,13].

(ii) Monotonicity: Rényi entropy is monotonic with respect to the parameter α. In general, for a fixed probability distribution, the Rényi entropy does not increase as α increases; in other words, since larger α values assign greater weight to the higher-probability components of the distribution, the entropy value may decrease or remain the same [13].

(iii) Special cases: Certain parameter values of Rényi entropy correspond to classical measures of information. In the limit as $\alpha \to 0$, Rényi entropy corresponds to Hartley entropy, which depends on the support size of the distribution:

$$H_0^R(P) = \log|\mathrm{supp}(p)| \tag{15}$$

Here, $|\mathrm{sup}p(P)|$ denotes the number of outcomes with positive probability. The Rényi entropy obtained for $\alpha = 2$ is referred to as collision entropy in most contexts:

$$H_2^R(P) = -\log\left(\sum_{i=1}^{n} p_i^2\right) \tag{16}$$

As α approaches ∞, the Rényi entropy approaches the min-entropy:

$$H_\infty^R(P) = -\log\left(\max_i p_i\right) \tag{17}$$

(iv) Permutation invariance: The entropy depends only on $\sum_i p_i^\alpha$ and is therefore unchanged under any permutation of the probabilities.

(v) Concavity and stability: These properties are not as straightforward as for Shannon entropy. While Rényi entropy exhibits more regular concavity properties in the range $0 < \alpha < 1$, it is generally not concave for $\alpha > 1$. Furthermore, the literature indicates that Rényi entropy occupies a more delicate position regarding Lesche stability compared to Shannon and Tsallis entropies [13,14].

4. Tsallis Entropy

4.1. Definition of Tsallis Entropy

Tsallis entropy has been proposed as a single-parameter, non-additive generalization of the classical Boltzmann–Gibbs–Shannon entropy. Tsallis [8] introduced a new form of entropy defined by the parameter q, particularly for multifractal structures and systems where the classical statistical mechanics framework may prove inadequate. The aim of this approach is not to completely abandon the fundamental structure of Shannon/Boltzmann–Gibbs entropy, but rather to generalize it through a more flexible parameter.

For a probability distribution P as in Equations (7) and (8), and for $q > 0$ and $q \neq 1$, Tsallis entropy is defined as follows:

$$S_q^T(P) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{n} p_i^q\right) \tag{18}$$

Here, q is the parameter that determines the non-additive nature of the entropy and its sensitivity to different regions of the probability distribution. The q parameter controls the weighting of probabilities in a manner similar to the a parameter in Rényi entropy; however, in Tsallis entropy, this parametric structure is based on a deformed logarithm rather than a logarithmic transformation [13].

Tsallis entropy can also be expressed using the q-logarithm. For $x > 0$, the q-logarithm

$$\ln_q(x) = \frac{x^{1-q} - 1}{1-q} \tag{19}$$

is defined, and it reduces to the classical natural logarithm as q to 1.

In this case, the Tsallis entropy can be written as in [13]:

$$S_q^T(P) = \sum_{i=1}^{n} p_i \ln_q\left(\frac{1}{p_i}\right) \tag{20}$$

This representation allows Tsallis entropy to be interpreted as a q-deformed version of the classical logarithmic information content in Shannon entropy. Thus, Tsallis entropy is a generalized version of Shannon entropy not only at the level of the formula but also at the level of how information content is measured.

The primary motivation for Tsallis entropy stems from the fact that classical Boltzmann–Gibbs entropy is particularly suitable for statistically independent or weakly interacting systems, but may prove inadequate for systems with strong correlations or long-range interactions. For this reason, Tsallis entropy has gained a significant place in the context of non-extensive statistical mechanics and has contributed to a broader understanding of Boltzmann–Gibbs theory [8,15]. This connection to non-extensive statistical mechanics provides a physical motivation for the q-deformation, complementing the additive, order-dependent generalization represented by Rényi entropy.

4.2. Relationship with Shannon Entropy

The most fundamental relationship between Tsallis entropy and Shannon entropy arises in the limit as $q \to 1$. Tsallis entropy is not directly defined for $q = 1$, because both the numerator and denominator in the formula approach zero. However, this is a removable singularity, and taking the limit yields Shannon entropy:

$$\lim_{q \to 1} S_q^T(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{21}$$

This result shows that Tsallis entropy includes Shannon entropy as a special limit case [8,13]. In this respect, Tsallis entropy is not a completely independent measure that replaces Shannon entropy; rather, it is a family of generalized entropies that extends classical entropy via the parameter q.

This limit relationship can be demonstrated using L'Hôpital's rule applied to Equation (18).

Since the probabilities sum to one, both the numerator and the denominator of Equation (18) tend to zero as $q \to 1$ (a 0/0 form). The derivative of the numerator with respect to q is

$$-\sum_{i=1}^{n} p_i^q \log p_i \tag{22}$$

and the derivative of the denominator is 1. Thus,

$$\lim_{q \to 1} S_q^T(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{23}$$

is obtained. This result shows that Tsallis entropy is consistent with classical Shannon entropy and that the q-deformation vanishes at $q = 1$.

This connection is one of the fundamental elements defining the position of Tsallis entropy in the literature on generalized entropies. While Shannon entropy uses the classical logarithmic information content, Tsallis entropy measures information content via the q-deformed logarithm. Therefore, the parameter q allows for the uncertainty structure of a distribution to be evaluated in different ways. However, in the limit as $q \to 1$, this parametric flexibility disappears, and the classical Shannon form is restored [8,13].

4.3. Pseudo-Additivity and Main Mathematical Properties

The most fundamental way in which Tsallis entropy differs from Shannon and Rényi entropies is that it exhibits pseudo-additivity rather than classical additivity for independent systems. For two statistically independent systems A and B, if the joint probabilities

$$p_{ij} = p_i^A p_j^B \tag{24}$$

can be written in this form, then Tsallis entropy satisfies the following composition rule:

$$S_q^T(A, B) = S_q^T(A) + S_q^T(B) + (1 - q) S_q^T(A) S_q^T(B) \tag{25}$$

This property clearly demonstrates the non-additive nature of the Tsallis entropy. When $q = 1$, the third term vanishes, and classical Shannon additivity is restored:

$$S(A,B) = S(A) + S(B) \tag{26}$$

Thus, Tsallis entropy does not completely reject classical additivity; rather, it incorporates it as a special case within a more general composition rule via the parameter q [8,13].

This so-called additivity property explains the importance of Tsallis entropy in the context of non-extensive statistical mechanics. Classical Boltzmann–Gibbs entropy exhibits an additive and, in most cases, extensive structure in statistically independent or weakly interacting subsystems. However, in systems where correlations cannot be neglected at all scales, this classical structure may not be sufficient. Anastasiadis [15] emphasizes that, in this regard, Tsallis entropy has become a fundamental entropy measure for non-extensive statistical mechanics, generalizing the Boltzmann–Gibbs theory.

Another important property of Tsallis entropy is that it attains its maximum value in a uniform distribution. If all n possible outcomes have equal probability, that is, if p_i=1/n, then

$$S_q^T(P) = \frac{1}{q-1}\left(1 - n^{1-q}\right) \tag{27}$$

is obtained.

This expression approaches the value of log n as q approaches 1 and yields the maximum value of Shannon entropy under a uniform distribution. This result shows that Tsallis entropy generalizes the principle of maximum uncertainty in a manner consistent with classical entropy [8].

Tsallis entropy exhibits concavity for $q > 0$. This property ensures that the entropy is sensitive to increases in mixing and uncertainty within the probability distribution. Furthermore, under the condition $q > 0$, Tsallis entropy exhibits a structure that is acceptable in terms of generalized entropy axioms; it satisfies fundamental properties such as continuity, maximality, and extensibility [13]. Therefore, the $q > 0$ range can be considered the fundamental parameter region in terms of the mathematical and physical interpretations of Tsallis entropy.

Tsallis entropy also holds an important position in terms of Lesche stability. Lesche stability refers to the property that small changes in the probability distribution lead to controlled and small changes in the entropy value. For generalized entropies, this property is important for the entropy to be used in a way that is experimentally and statistically meaningful. Amigó and colleagues [13] note that the Tsallis entropy is Lesche-stable for $q > 0$. This places the Tsallis entropy on a more solid mathematical foundation compared to some other generalized entropy measures.

The Tsallis entropy is also closely related to the Rényi entropy. Both entropies use the common core expression

$$\sum_{i=1}^{n} p_i^q \tag{28}$$

However, while Rényi entropy transforms this expression into the entropy scale using the natural logarithm, Tsallis entropy is defined via a deformed logarithmic structure. Therefore, while Rényi entropy preserves classical additivity in independent systems, Tsallis entropy exhibits a pseudo-additive composition rule [13,15]. This difference demonstrates that generalized entropy measures differ not only at the formula level but also in terms of system composition and the interpretation of independence.

Finally, Tsallis entropy has also served as the foundation for the development of newer information measures. For example, Tsallis extropy has been defined as an extension of the Shannon extropy concept related to Tsallis entropy, and its mathematical relationships

with Tsallis entropy have been investigated [16]. Such studies demonstrate that Tsallis entropy serves as a productive foundation not only in the context of non-extensive statistical mechanics but also in the literature on modern information measures.

In conclusion, Tsallis entropy is a generalized, non-additive, and quasi-additive form of Shannon entropy with parameter q. It holds a fundamental place in the literature on generalized entropies because it recovers Shannon entropy in the limit $q \to 1$, exhibits concavity and stability for $q > 0$, and incorporates the classical additivity for independent systems within a more general composition rule. With these properties, Tsallis entropy serves as an important intermediate step in the transition to Sharma–Mittal entropy.

5. Sharma–Mittal Entropy

5.1. Definition of Sharma–Mittal Entropy

Sharma–Mittal entropy is an important family of entropies that stands out among generalized entropy measures due to its two-parameter structure. Proposed by Sharma and Mittal [9], this framework is regarded as a unifying framework in the literature on generalized entropy because of the special cases and limit relationships it establishes with classical Shannon entropy as well as Rényi and Tsallis entropies. In their study on entropy statistics, Esteban and Morales [10] also demonstrated that different entropy measures—such as Shannon, Rényi, Havrda–Charvát, and Sharma–Mittal—can be addressed under more general functional frameworks.

For a probability distribution P, as in Equations (7) and (8), and for $\alpha > 0$, $\beta > 0$, $\alpha \neq 1$, and $\beta \neq 1$, the Sharma–Mittal entropy can be defined as follows:

$$S_{\alpha,\beta}^{SM}(P) = \frac{1}{1-\beta}\left[\left[\sum_{i=1}^{n} p_i^{\alpha}\right]^{\frac{1-\beta}{1-\alpha}} - 1\right] \tag{29}$$

Here, the parameter α determines the order of weighting of the probabilities, while the parameter β determines which non-additive transformation is used to map this weighted sum to the entropy scale. Therefore, unlike the Rényi and Tsallis entropies, which have a single deformation parameter, the Sharma–Mittal entropy offers a more general framework that evaluates uncertainty through two separate parameters [17].

In this formulation, the fundamental kernel is the sum $\sum_{i=1}^{n} p_i^{\alpha}$ of Equation (10), which represents the common mathematical structure found in both the Rényi and Tsallis entropies. However, the Sharma–Mittal entropy transforms this core through a two-parameter transformation, thereby presenting a form capable of unifying both Rényi's logarithmic structure and Tsallis's quasi-additive structure under a single framework. Masi [17] notes that the Rényi and Tsallis entropies are not direct generalizations of one another; rather, the Sharma–Mittal entropy offers a potential overarching framework that unifies these two structures within a consistent two-parameter framework.

Therefore, the Sharma–Mittal entropy should be regarded not merely as another generalization of Shannon entropy, but as an entropy family that expresses the common mathematical core of the Rényi and Tsallis entropies within a broader parameter space. In the classification of generalized entropic forms by Ilić and colleagues [18], the Sharma–Mittal entropy is also considered among the fundamental entropy families defined by two deformation parameters.

5.2. Relationship with Rényi and Tsallis Entropy

The most important property of the Sharma–Mittal entropy is that, under appropriate parameter choices or as limits are taken, it encompasses the Rényi and Tsallis en-

tropies. This property places the Sharma–Mittal entropy in a central position among generalized entropies.

First, the Rényi entropy is obtained as the limit of the Sharma–Mittal entropy as $\beta \to 1$. Writing $A = \sum_{i=1}^{n} p_i^{\alpha}$ for the kernel of Equation (10), Equation (29) becomes

$$S_{\alpha,\beta}^{SM}(P) = \frac{A^{\frac{1-\beta}{1-\alpha}} - 1}{1-\beta} \tag{30}$$

As β to 1, the expression converges to the indeterminate form $0/0$. Applying L'Hôpital's rule,

$$\lim_{\beta \to 1} S_{\alpha,\beta}^{SM}(P) = \frac{1}{1-\alpha} log(\sum_{i=1}^{n} p_i^{\alpha}) \tag{31}$$

is obtained. This expression is exactly the Rényi entropy in Equation (9).

$$\lim_{\beta \to 1} S_{\alpha,\beta}^{SM}(P) = H_{\alpha}^{R}(P) \tag{32}$$

This result shows that the Rényi entropy is a limit case of the Sharma–Mittal entropy [17,18].

Second, the Tsallis entropy is obtained as a special case of the Sharma–Mittal entropy where $\beta = \alpha$. If $\beta = \alpha$ is substituted into the Sharma–Mittal formula, the exponent $(1-\beta)/(1-\alpha)$ equals 1, and the expression reduces to $\left(\sum_{i=1}^{n} p_i^{\alpha} - 1\right)/(1-\alpha)$, which is the Tsallis entropy of Equation (18) with $q = \alpha$; therefore,

$$S_{\alpha,\alpha}^{SM}(P) = S_{\alpha}^{T}(P) \tag{33}$$

holds. This reduction demonstrates that the Sharma–Mittal entropy encompasses the Tsallis entropy not only via a limit but also directly through a specific choice of parameters [17].

When these two transitions are considered together, the meaning of the two-parameter structure of the Sharma–Mittal entropy becomes clearer. The α parameter represents the degree of probability weighting common to both the Rényi and Tsallis entropies. The β parameter, on the other hand, is the second deformation parameter that determines how this weighted sum is transformed into the entropy scale. Therefore, the Sharma–Mittal entropy establishes a mathematical bridge between Rényi's summative logarithmic structure and Tsallis's quasi-summative structure.

According to Masi's assessment [17], although the Rényi and Tsallis entropies use the same probability kernel, they have different transformation mechanisms. While the Rényi entropy processes this kernel using the classical logarithm, the Tsallis entropy is expressed through a deformed logarithmic structure. The Sharma–Mittal entropy, on the other hand, combines these two approaches into a two-parameter form, creating a broader family.

5.3. Shannon Entropy as a Limiting Case

The Sharma–Mittal entropy includes not only the Rényi and Tsallis entropies but also the Shannon entropy in the limit. As shown in previous sections, the Rényi entropy reduces to the Shannon entropy in the limit as $\alpha \to 1$. Similarly, the Tsallis entropy yields the Shannon entropy in the limit as $q \to 1$. The Sharma–Mittal entropy, however, due to its two-parameter structure, regains the Shannon entropy in the joint limit as $\alpha \to 1$ and $\beta \to 1$.

This can be symbolically expressed as follows:

$$\lim_{\alpha \to 1, \beta \to 1} S_{\alpha,\beta}^{SM}(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{34}$$

This limit relation demonstrates that the Sharma–Mittal entropy does not exclude the Shannon entropy; rather, it encompasses the classical Shannon entropy as a special limit point within a broader two-parameter structure. Therefore, the Shannon entropy can be interpreted as the classical case corresponding to the point $(\alpha, \beta) = (1, 1)$ in the Sharma–Mittal parameter space.

5.4. Main Mathematical Properties

The main mathematical feature of Sharma–Mittal entropy is that it is a family of two-parameter generalized entropies. This structure allows for separate control of both the degree of probability weighting and the non-additive transformation form of the entropy. Therefore, Sharma–Mittal entropy offers a more flexible mathematical structure compared to the one-parameter Rényi and Tsallis entropies [17,18].

The first important feature is the generalizing nature of Sharma–Mittal entropy. While the Rényi and Tsallis entropies are obtained through appropriate parameter choices, the Shannon entropy is recovered in the limit. This transforms Sharma–Mittal entropy from merely a standalone entropy measure into a framework that establishes connections among generalized entropy families. Ilić and Djordjević [19] also note that the Sharma–Mittal entropy class can encompass the Rényi, Tsallis, Landsberg–Vedral, and Gaussian entropies as special cases.

Second, the Sharma–Mittal entropy does not generally satisfy the classical Shannon additivity for independent systems. Instead, it exhibits a parameter-dependent non-additive or pseudo-additive composition behavior. In this respect, it shares a structure similar to the Tsallis entropy; however, unlike the Tsallis entropy, this composition behavior arises under the influence of a two-parameter structure. Beck [20] emphasizes that properties such as composability and Lesche stability are important for generalized measures of information to carry physical meaning; he also discusses Sharma–Mittal entropies among generalized entropy families.

Third, Sharma–Mittal entropy has a structure that is sensitive to the parameter space. This is both an advantage and a point requiring attention. It is an advantage because different (α, β) values can highlight different uncertainty structures of the distribution. However, it also requires caution because properties such as concavity, stability, and physical interpretation may not hold in the same way across the entire parameter space. Aktürk, Bağcı, and Sever [21] have specifically discussed, particularly in a thermostatistical context, whether Sharma–Mittal entropy truly goes beyond the Rényi and Tsallis entropies, and have noted that the physical interpretation of the Sharma–Mittal relative entropy is more reliable within certain limits. Therefore, when using Sharma–Mittal entropy, the mathematical and contextual meaning of the parameters must be clearly specified.

Fourth, Sharma–Mittal entropy can be expressed using different parameterizations. Various notations, such as (α, β), (r, q), or (α, q), are used in the literature. These differences are often merely different notations for the same underlying mathematical structure. However, since the specific case that corresponds to either Rényi or Tsallis may vary from study to study, it is important to clearly define the parameterization used [18,19].

6. Comparative Mathematical Evaluation of Generalized Entropy Measures

6.1. Additivity and Pseudo-Additivity

A fundamental axis in comparing entropy measures is their compositional behavior under independent systems. For Shannon entropy, the additivity of Equation (3) holds for independent systems A and B.

Rényi entropy also preserves additivity for independent systems. This is because the logarithmic structure of Rényi entropy transforms the multiplicative probability sum that arises in independent distributions into an additive form [7]. Therefore, although Shannon and Rényi entropies have different parameter structures, they align in terms of classical additivity under independence.

Tsallis entropy, on the other hand, exhibits pseudo-additivity rather than classical additivity for independent systems, with the composition rule given in Equation (25).

When $q = 1$, the third term vanishes, and classical Shannon additivity is restored. Therefore, the Tsallis entropy does not completely rule out classical additivity; rather, it includes it as a special case in the limit $q = 1$ [8,13].

The Sharma–Mittal entropy also generally does not satisfy classical additivity; instead, it exhibits a two-parameter non-additive composition structure. This structure generalizes the pseudo-additive behavior of the Tsallis entropy to a broader parameter space. However, the composition properties of the Sharma–Mittal entropy must be carefully considered depending on the parameterization and parameter region used. Beck [20] emphasizes that properties such as composability and Lesche stability are important for the physical and mathematical meaningfulness of generalized entropies. Therefore, the Sharma–Mittal entropy should be evaluated not merely because it is a more general formula, but by specifying under which composition conditions it is meaningful [20,21].

6.2. Limiting Relationships

The most important mathematical connection among these four measures of entropy lies in their limit and special-case relationships. Shannon entropy serves as the classical reference point for the other three families of entropy. For Rényi entropy,

$$\lim_{\alpha \to 1} H_\alpha^R(p) = -\sum_{i=1}^{n} p_i \log p_i \tag{35}$$

since Shannon entropy is the limit of Rényi entropy as α approaches 1 [7]. Similarly, for Tsallis entropy,

$$\lim_{q \to 1} S_q^T(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{36}$$

holds [8]. These two results show that, although the Rényi and Tsallis entropies have different mathematical structures, they share the same classical limit.

As shown in Section 5, the Sharma–Mittal entropy reduces to the Rényi entropy as $\beta \to 1$ (Equation (32)), to the Tsallis entropy for $\beta = \alpha$ (Equation (33)), and to the Shannon entropy in the joint limit $(\alpha,\beta) \to (1, 1)$ (Equation (34)), which clearly demonstrates its unifying role within the family of generalized entropies [17–19].

The thermodynamic origin of entropy provides a useful consistency check for these relationships. Using natural logarithms, for W equiprobable microstates, $p_i = 1/W$, Shannon entropy reduces to $\ln W$, which corresponds to Boltzmann's principle up to the factor k_B. Rényi entropy also gives $\ln W$ for every admissible α, the Tsallis entropy reduces to the q-logarithm $\ln_q W$ defined in Equation (19), and the Sharma–Mittal entropy reduces to $\ln_\beta W$, independently of α. This is consistent with the Rényi limit ($\beta \to 1$) and the Tsallis case ($\beta = \alpha$). The statement is purely formal and restricted to the equiprobable case; it does not by itself assign a thermodynamic interpretation to every parameter pair (α, β).

6.3. Mathematical Positioning of Sharma–Mittal Entropy

The Sharma–Mittal entropy incorporates both the additive, logarithmic structure of the Rényi entropy and the pseudo-additive, deformed-logarithmic structure of the Tsallis entropy simultaneously. Masi [17] notes that the Rényi and Tsallis entropies are not direct generalizations of one another, but that the Sharma–Mittal entropy unifies these two structures within a broader and more consistent two-parameter form. Therefore, the Sharma–Mittal entropy should be regarded not as a simple sum in the form of "Rényi + Tsallis," but as a more general mathematical structure that encompasses the common core of both entropies through different transformation mechanisms. Nevertheless, this

mathematical unification does not, by itself, establish a thermodynamic interpretation for every parameter pair (α, β).

The geometric interpretation also extends to the generalized entropy framework through the probability power sum $\sum_{i=1}^{n} p_i^{\alpha}$, which underlies both Rényi and Sharma–Mittal entropies. Using natural logarithms, Equations (9) and (29) imply that, for fixed $\alpha > 0$ with $\alpha \neq 1$ and $\beta \neq 1$, the Sharma–Mittal entropy can be written as $S_{\alpha,\beta}^{SM}(P) = \frac{\exp\left((1-\beta)H_{\alpha}^{R}(P)\right)-1}{1-\beta}$. Since this transformation is strictly increasing in $H_{\alpha}^{R}(P)$, varying β for a fixed α changes the entropy scale without altering the corresponding families of level sets on the probability simplex. In contrast, varying α can modify the geometry of these level sets by changing the relative weighting of probability components. This interpretation does not require a general concavity assumption for the Sharma–Mittal family.

However, the two-parameter structure of the Sharma–Mittal entropy does not always imply automatic superiority. While an increase in the number of parameters enhances modeling flexibility, it also introduces interpretive challenges. Aktürk, Bağcı, and Sever [21], in discussing whether the Sharma–Mittal entropy truly goes beyond Rényi and Tsallis in a thermostatistical context, have noted that certain physical interpretations of this entropy are particularly robust in limit cases. Therefore, when using the Sharma–Mittal entropy, the selected parameters must be determined not only based on computational convenience but also on mathematical and contextual justifications.

Concavity and Schur-concavity conditions have been studied systematically for generalized entropies. Kvålseth [22] introduced a three-parameter entropy that contains several proposed generalized entropies as special cases, derived the conditions under which it is concave and Schur-concave, and applied these results to one- and two-parameter family members. Such conditions provide a rigorous basis for the parameter-region qualifications noted above for the Rényi and Sharma–Mittal entropies, and we refer the reader to that work for the precise conditions.

In this context, the Sharma–Mittal entropy can be regarded as an "umbrella" among generalized entropy measures; however, this assessment must be made with caution. The unifying role of the Sharma–Mittal entropy does not imply that it is the most appropriate entropy measure in every situation.

The hierarchical mathematical relationships and limiting transitions among Shannon, Rényi, Tsallis, and Sharma–Mittal entropies are illustrated in Figure 1.

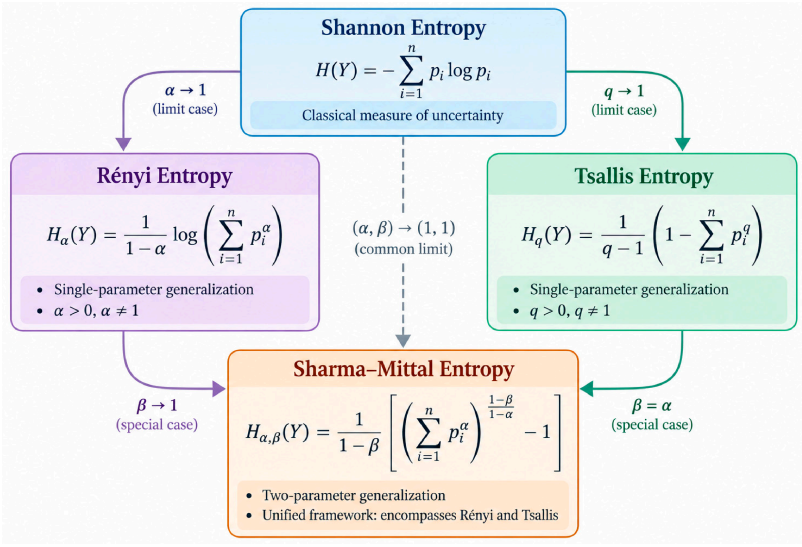


Figure 1. Mathematical hierarchy and limiting relationships among Shannon, Rényi, Tsallis, and Sharma–Mittal entropies.

The fundamental mathematical differences between the Shannon, Rényi, Tsallis, and Sharma–Mittal entropies—in terms of parameter structure, additivity behavior, limit relationships with Shannon entropy, and their roles within the generalized entropy framework—are summarized in Table 1.

Table 1. Comparative mathematical structure of Shannon, Rényi, Tsallis, and Sharma–Mittal entropies.

Entropy Measure	Parameter Structure	Additivity Behavior	Shannon Limit	Mathematical Role
Shannon entropy	Without parameters	Additive	Basic form	The classical measure of uncertainty
Rényi entropy	Single parameter (α)	Additive	$\alpha \to 1$	Parameterizes the probability weighting
Tsallis entropy	Single parameter (q)	Pseudo-additive	$q \to 1$	Non-additive/non-extensive offers a structure
Sharma–Mittal entropy	Two-parameter (α, β)	Two-parameter non-additive structure	$(\alpha, \beta) \to (1, 1)$	A general framework combining the Rényi and Tsallis entropies

7. From Theory to Practice: A Parameter-Integrated Volumetric Extension of Sharma–Mittal Entropy

7.1. Motivation: Beyond a Single Parameter Pair

The relationships established in Sections 2–6 show that Sharma–Mittal entropy is not merely a two-parameter formula but a coherent mathematical bridge that connects the Rényi and Tsallis families and reduces to the Shannon entropy in the appropriate common limit. A structure of this kind raises a natural applied question: if α and β index an entire family of admissible uncertainty measures rather than a single one, which member of that family should an empirical analysis actually report? In many applications of Sharma–Mittal entropy, this question is resolved by fixing α and β at a single, often arbitrarily chosen, pair; the resulting value is then treated as though it were a unique descriptor of the underlying uncertainty, even though a different, equally admissible parameter pair may yield a materially different value, and in a feature-selection context, even a different ranking of candidate variables [9,17,18]. Classical filter-based feature-selection criteria face a related difficulty: correlation coefficients are sensitive to a single form of dependence, and many conventional entropy- or mutual-information-based feature-selection implementations for continuous variables rely on discretization, making the resulting estimates potentially sensitive to the chosen binning scheme [23,24]. Parametric generalized entropies have also been incorporated directly into applied information-theoretic pipelines—such as random-forest-based classification and regression [25] and topic-model performance evaluation [26]—but typically by fixing the entropy parameter at a single value throughout the analysis; the volumetric approach developed below instead removes this dependence on a single parameter choice.

Figure 2 contrasts the two approaches on the same α-β plane: rather than committing to a single, arbitrarily chosen point such as the one marked in red, the conditional entropy surface is evaluated everywhere inside the shaded region and integrated to obtain a single, parameter-integrated score that does not depend on the selection of a particular (α, β) pair.

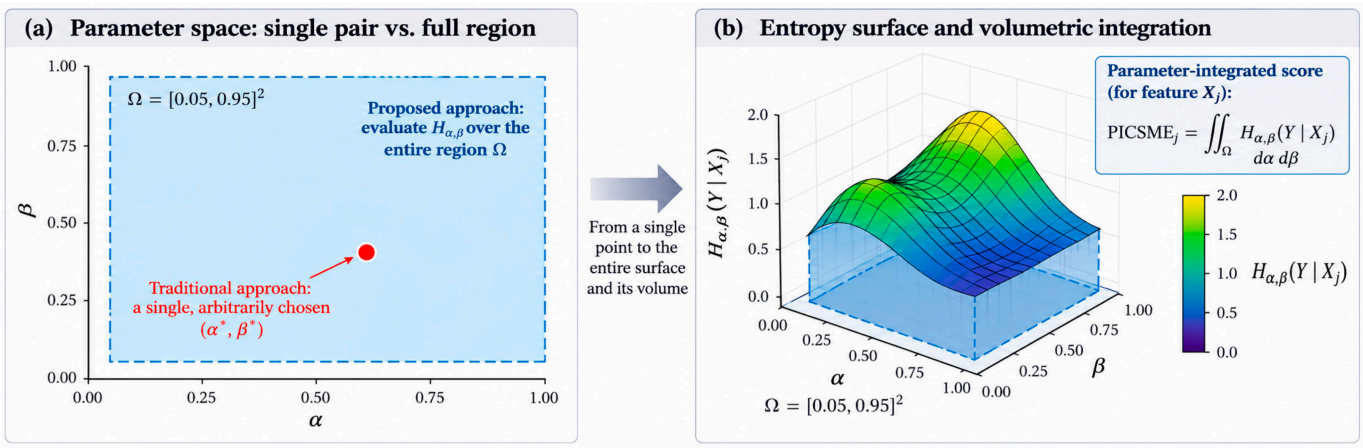


Figure 2. Single-parameter selection versus volumetric integration over the α-β parameter space. (**a**) The conventional approach evaluates the conditional entropy at a single, arbitrarily selected parameter pair (α^*, β^*), whereas the parameter-integrated approach considers the entire parameter region $\Omega = [0.05,0.95]^2$. (**b**) For each feature X_j, the Sharma–Mittal conditional entropy $H_{\alpha,\beta}((Y|X_J)$ forms a surface over Ω. Integrating this surface over the continuous parameter region yields the parameter-integrated conditional entropy score, $PICSME_j$, corresponding to the volume under the entropy surface. Thus, the resulting score summarizes the information contained across the entire parameter region rather than depending on a single predefined (α, β) pair.

Building on the framework previously developed and empirically validated by the present authors [27], this section shows how the theoretical unification established above can be translated into a practical feature-ranking criterion: rather than evaluating the conditional Sharma–Mittal entropy at one point of the α-β plane, the entire admissible region is integrated, and the importance of a candidate explanatory variable is defined through the resulting volumes rather than through its value at an arbitrarily selected parameter pair.

7.2. Conditional Sharma–Mittal Entropy for Continuous Variables

For a continuous target Y and a candidate explanatory variable X_j, the conditional Sharma–Mittal entropy is obtained by applying the transformation φ_β pointwise to the conditional Rényi entropy and averaging the result over the marginal distribution of X_j:

$$H_{\alpha,\beta}\left(Y \mid X_j\right) = \int f_{X_j}(x) \left\{ \frac{\left[\int \left(\frac{f_{Y,X_j}(y,x)}{f_{X_j}(x)} \right)^{\alpha} dy \right]^{\frac{1-\beta}{1-\alpha}} - 1}{1-\beta} \right\} dx \tag{37}$$

Because the marginal, joint, and conditional densities entering Equation (37) are unknown for empirical data, they are estimated nonparametrically through a Gaussian kernel density estimator, with the bandwidth fixed by Scott's rule rather than by manual tuning, so that the same estimation protocol applies uniformly to every dataset and every candidate variable [28,29]. Keeping the bandwidth in the direction of Y identical in the marginal and joint estimators guarantees that marginalizing the joint density over X_j exactly reproduces the marginal density of Y; this consistency condition, together with a renormalization of the conditional density at each conditioning point, removes a spurious source of negative information gain that would otherwise arise when kernel density estimation is applied to continuous variables without these safeguards [27].

Figure 3 summarizes the full computational pipeline, from the raw observations Y and Xj to the three reported indices, showing where kernel density estimation, the consistency

and renormalization safeguards, and volumetric integration each enter the calculation. Importantly, the observed variables are not discretized at any stage; the 60 × 60 grid is used solely for numerical integration over the continuous α-β parameter domain.

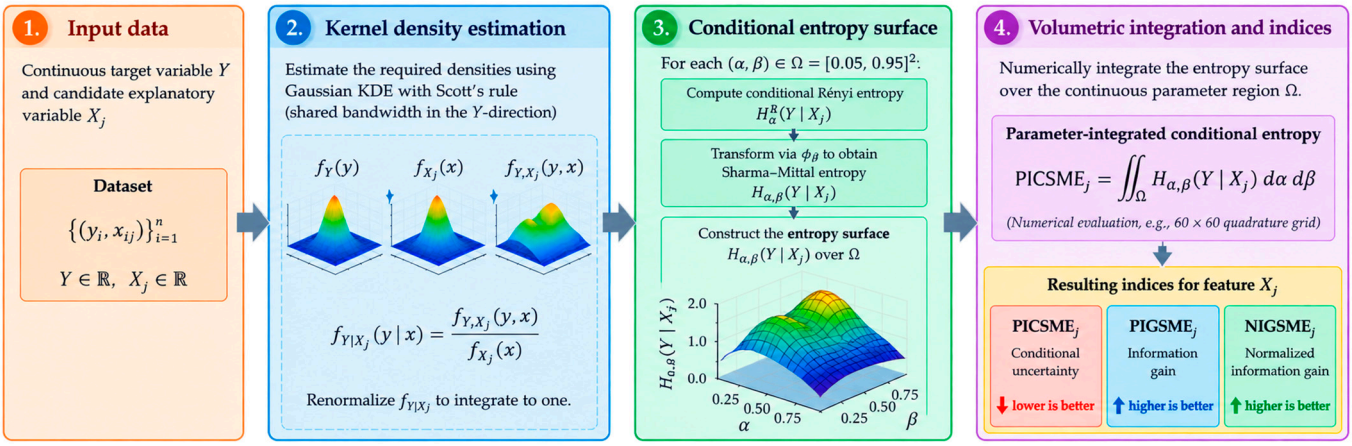


Figure 3. Computational pipeline of the parameter-integrated Sharma–Mittal framework, from kernel density estimation to the PICSME, PIGSME, and NIGSME indices.

7.3. Three Volume-Based Indices over the Parameter Plane

Instead of reporting $H_{\alpha,\beta}(Y|X_j)$ at a single (α, β), the conditional entropy surface is integrated over the rectangular region $\Omega = [0.05, 0.95]^2$ of the α-β plane, a region bounded away from the singularities at $\alpha = 1$ and $\beta = 1$ while still spanning a broad deformation range. Specifically, $\alpha = 1$ and $\beta = 1$ are excluded by the upper bound 0.95, since Equation (29) is defined directly only for $\alpha \neq 1$ and $\beta \neq 1$, and these cases are obtained as limits; the same margin, 0.05, is applied at the lower end of both parameters, which keeps the region away from the boundaries $\alpha \to 0$ and $\beta \to 0$. Let $D \subset (0,1)^2$ denote the general integration domain in the α-β parameter plane; following [27], the present analysis adopts the specific domain $D = \Omega = [0.05, 0.95]^2$, which is used throughout the derivations below [27]. In that study, sensitivity analyses conducted using alternative integration boundaries and grid resolutions yielded completely stable feature rankings across all evaluated datasets ($\rho = 1.00$), supporting the robustness of the ranking criterion to these numerical design choices within the investigated settings; the chosen domain therefore provides a representative integration region while maintaining a safe distance from the singularities noted above. Three indices are obtained from this integration. The Parameter-Integrated Conditional Sharma–Mittal Entropy (PICSME) represents the residual-uncertainty volume that remains about Y once X_j has been observed:

$$\mathrm{PICSME}_j = \frac{1}{|D|}\int_{\alpha_L}^{\alpha_U}\int_{\beta_L}^{\beta_U} H_{\alpha,\beta}\big(Y \mid X_j\big)\, d\beta\, d\alpha \tag{38}$$

with D representing the integration region and $|D|$ its area; a smaller PICSME indicates a larger reduction in uncertainty about Y. The Parameter-Integrated Gain of Sharma–Mittal Entropy (PIGSME) is the volume of the difference between the marginal and conditional entropy surfaces,

$$\mathrm{PIGSME}_j = \frac{1}{|D|}\int_{\alpha_L}^{\alpha_U}\int_{\beta_L}^{\beta_U}\Big[H_{\alpha,\beta}(Y) - H_{\alpha,\beta}\big(Y \mid X_j\big)\Big]d\beta\, d\alpha \tag{39}$$

And the Normalized Integrated Gain of Sharma–Mittal Entropy (NIGSME) rescales this gain volume by the target's total entropy volume so that the index is dimensionless and comparable across datasets,

$$\mathrm{NIGSME}_j = \frac{\int_{\alpha_L}^{\alpha_U} \int_{\beta_L}^{\beta_U} \left[H_{\alpha,\beta}(Y) - H_{\alpha,\beta}\left(Y \mid X_j\right) \right] d\beta \, d\alpha}{\int_{\alpha_L}^{\alpha_U} \int_{\beta_L}^{\beta_U} H_{\alpha,\beta}(Y) \, d\beta \, d\alpha} \tag{40}$$

Table 2 summarizes the interpretation and ranking direction of the three indices.

Table 2. Interpretation of the three parameter-integrated Sharma–Mittal information indices.

Index	Interpretation	Ranking Direction
PICSME	Absolute volume of residual uncertainty about *Y* once *Xj* is known	Lower value indicates a more informative variable
PIGSME	Absolute volume of uncertainty reduction attributable to *Xj*	Higher value indicates a more informative variable
NIGSME	Proportional share of the target's uncertainty eliminated by *Xj*; dimensionless and, under the non-negativity conditions established in [27], bounded above by one	Higher value indicates a more informative variable

A numerical illustration of PICSME, PIGSME, and NIGSME applied to real datasets, together with their empirical rank correlations, is provided in [27].

7.4. Theoretical Properties Inherited from the Sharma–Mittal Structure

The volumetric construction preserves, rather than replaces, the mathematical properties of the Sharma–Mittal entropy discussed in Sections 5 and 6. If X_j and Y are independent, the conditional Sharma–Mittal entropy coincides with the marginal entropy at every $(\alpha, \beta) \in D$, so that PIGSME and NIGSME vanish exactly, and PICSME reduces to the target's own entropy volume; negative values can therefore never arise from independence itself, but only from the interaction between a weak dependency and the convexity of the transformation φ_β [27]. As $(\alpha, \beta) \to (1, 1)$, the gain volume converges to the classical Shannon mutual information, which is non-negative by the Kullback–Leibler property, so the volumetric criterion recovers the familiar information-gain filter as a limiting case rather than replacing it [5]. Because the target's marginal entropy volume $V_Y = \int_{\alpha_L}^{\alpha_U} \int_{\beta_L}^{\beta_U} H_{\alpha,\beta}(Y) \, d\beta \, d\alpha$, the denominator of NIGSME, does not depend on the candidate variable within a fixed dataset, PIGSME and NIGSME are exact, strictly decreasing affine transformations of PICSME,

$$\mathrm{NIGSME}_j = 1 - \frac{|D| \cdot \mathrm{PICSME}_j}{V_Y} \tag{41}$$

For a fixed dataset and integration region D, PIGSME and NIGSME are therefore exact decreasing affine transformations of PICSME and induce the same feature ranking once the corresponding ranking directions are taken into account—an ordinal equivalence that follows as a mathematical necessity rather than an empirical coincidence [27]. Read together with Sections 3–6, this means that PICSME, PIGSME, and NIGSME should not be understood as three competing selection criteria, but as three reporting scales—absolute, differential, and normalized—of a single variable-importance criterion that is grounded in the same two-parameter structure that makes the Sharma–Mittal entropy a unifying framework in the first place.

8. Concluding Comments

This study has provided a unified mathematical synthesis of the Shannon, Rényi, Tsallis, and Sharma–Mittal entropies, focusing on their axiomatic foundations, parameterizations, compositional properties, and limiting transitions, and has shown how they are connected within a common structural hierarchy. Shannon entropy [5] serves as the parameter-free classical reference point, while the Rényi and Tsallis entropies are two distinct single-parameter generalizations that, respectively, preserve and relax classical additivity [7,8]. Both act on the same probability kernel but map it to the entropy scale through different transformations: a logarithm for Rényi and a deformed logarithm for Tsallis [13]. Sharma–Mittal entropy encompasses both under appropriate parameter choices and reduces to Shannon entropy in the joint limit, which supports its unifying mathematical position rather than that of a merely more general formula [9,17,18]. At the same time, greater generality does not make it the most suitable choice for every application: concavity, stability, physical interpretability, and axiomatic consistency must be assessed for the problem at hand.

The two-parameter flexibility of the Sharma–Mittal entropy also makes any derived quantity depend on the chosen parameter pair. The parameter-integrated volumetric framework previously developed and empirically validated by the present authors [27], summarized in Section 7, addresses this dependence in feature selection by integrating the conditional Sharma–Mittal entropy and information gain over an admissible α-β region. This yields the indices PICSME, PIGSME, and NIGSME, which are ordinally equivalent for a fixed dataset and integration region. Thus, the same two-parameter structure that creates the dependence also provides the basis for integrating over it.

Several limitations should be acknowledged. The synthesis focuses on four major entropy families and does not aim to exhaust the broader landscape of generalized entropies. The volumetric extension is formulated for continuous target and explanatory variables and relies on kernel density estimation, numerical integration, and a predefined parameter domain; its robustness to alternative parameter ranges and grid resolutions was examined in [27], but the absolute values of the indices may still depend on the density estimator and the integration scheme. Future work may extend the framework to discrete or classification targets, alternative density estimators, other generalized entropy families, and parameter-integration strategies beyond feature selection.

Author Contributions: Conceptualization, N.O.Ü., M.G. and D.Y.; methodology, N.O.Ü., M.G. and D.Y.; software, N.O.Ü.; validation, N.O.Ü., M.G. and D.Y.; formal analysis, N.O.Ü. and M.G.; investigation, N.O.Ü. and M.G.; resources, N.O.Ü. and M.G.; writing—original draft preparation, N.O.Ü. and M.G.; writing—review and editing, N.O.Ü. and M.G.; visualization, N.O.Ü. and M.G.; supervision, D.Y.; project administration, N.O.Ü. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable, as no new data were created in this study.

Acknowledgments: During the preparation of this manuscript, the authors utilized ChatGPT (OpenAI; GPT-6, high) to assist in the creation of visuals and Claude (Anthropic; Claude Sonnet 5.5) to help refine English phrasing. The authors have reviewed and edited all outputs and take full responsibility for the content of this publication.

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

References

1. Clausius, R. Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Ann. Der Phys.* **1865**, *125*, 353–400. [CrossRef]
2. Boltzmann, L. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respective den Sätzen über das Wärmegleichgewicht. *Sitzungsber. Kaiserl. Akad. Wiss. Math.-Naturwiss. Cl.* **1877**, *76*, 373–435.
3. Gibbs, J.W. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*; Charles Scribner's Sons: New York, NY, USA, 1902.
4. Ben-Naim, A. Entropy, Shannon's Measure of Information and Boltzmann's H-Theorem. *Entropy* **2017**, *19*, 48. [CrossRef]
5. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656. [CrossRef]
6. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover Publications: New York, NY, USA, 1957.
7. Rényi, A. On Measures of Entropy and Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
8. Tsallis, C. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
9. Sharma, B.D.; Mittal, D.P. New Nonadditive Measures of Entropy for Discrete Probability Distributions. *J. Math. Sci.* **1975**, *10*, 28–40.
10. Esteban, M.D.; Morales, D. A Summary on Entropy Statistics. *Kybernetika* **1995**, *31*, 337–346.
11. Faddeev, D.K. Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas. In *Arbeiten zur Informationstheorie I*; Deutscher Verlag der Wissenschaften: Berlin, Germany, 1957; pp. 85–90.
12. Bormashenko, E.; Legchenkova, I.; Frenkel, M.; Shvalb, N.; Shoval, S. Shannon (Information) Measures of Symmetry for 1D and 2D Shapes and Patterns. *Appl. Sci.* **2022**, *12*, 1127. [CrossRef]
13. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*, 813. [CrossRef] [PubMed]
14. Fuentes, J.; Gonçalves, J. Rényi Entropy in Statistical Mechanics. *Entropy* **2022**, *24*, 1080. [CrossRef] [PubMed]
15. Anastasiadis, A. Special Issue: Tsallis Entropy. *Entropy* **2012**, *14*, 174–176. [CrossRef]
16. Balakrishnan, N.; Buono, F.; Longobardi, M. On Tsallis Extropy with an Application to Pattern Recognition. *arXiv* **2021**, arXiv:2103.07168.
17. Masi, M. A Step Beyond Tsallis and Rényi Entropies. *Phys. Lett. A* **2005**, *338*, 217–224. [CrossRef]
18. Ilić, V.M.; Korbel, J.; Gupta, S.; Scarfone, A.M. An Overview of Generalized Entropic Forms. *Europhys. Lett.* **2021**, *133*, 50005. [CrossRef]
19. Ilić, V.M.; Djordjević, I.B. On the α-q-Mutual Information and the α-q-Capacities. *Entropy* **2021**, *23*, 702. [CrossRef] [PubMed]
20. Beck, C. Generalised Information and Entropy Measures in Physics. *Contemp. Phys.* **2009**, *50*, 495–510. [CrossRef]
21. Aktürk, E.; Bağcı, G.B.; Sever, R. Is Sharma-Mittal Entropy Really a Step Beyond Tsallis and Rényi Entropies? *arXiv* **2007**, arXiv:cond-mat/0703277.
22. Kvålseth, T.O. Entropies and Their Concavity and Schur-Concavity Conditions. *IEEE Access* **2022**, *10*, 96006–96015. [CrossRef]
23. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]
24. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]
25. Ignatenko, V.; Surkov, A.; Koltcov, S. Random Forests with Parametric Entropy-Based Information Gains for Classification and Regression Problems. *PeerJ Comput. Sci.* **2024**, *10*, e1775. [CrossRef] [PubMed]
26. Koltcov, S.; Ignatenko, V.; Koltsova, O. Estimating Topic Modeling Performance with Sharma-Mittal Entropy. *Entropy* **2019**, *21*, 660. [CrossRef] [PubMed]
27. Oruç Ünal, N.; Göztaş, M.; Yıldız, D. Parameter-Independent Feature Ranking with Volume-Integrated Sharma-Mittal Entropy: Kernel-Based Estimation, Theoretical Properties and Empirical Validation. *Entropy* **2026**, *28*, 933. [CrossRef] [PubMed]
28. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2015.
29. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.