

Article

Not peer-reviewed version

Information-Weight Interpretation of Probability: A Novel Information-Theoretic Perspective

[Hening Huang](#) *

Posted Date: 19 August 2025

doi: 10.20944/preprints202508.1430.v1

Keywords: information content; information interval; information spectrum; multiset; probability density



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Information-Weight Interpretation of Probability: A Novel Information-Theoretic Perspective

Hening Huang

Teledyne RD Instruments (retired), San Diego, California, USA; heninghuang1@gmail.com

Abstract

This paper proposes a novel information-theoretic perspective, where probability is interpreted as information weight and probability density is interpreted as information density. We define “information” as the state of a discrete system or the value of a continuous system. We model discrete systems using multisets. From this, we define information weight as the relative multiplicity of each state with respect to the size of the multiset. By relating a discrete system (multiset) to a discrete random variable, we establish a one-to-one correspondence between information weight and conventional probability. We then extend this framework to continuous systems. This information-theoretic perspective, distinct from personal belief, propensity, or frequency-based interpretations, emphasizes the contribution of probability to the informational structure of a system.

Keywords: information content; information interval; information spectrum; multiset; probability density

1. Introduction

Probability is virtually ubiquitous, playing a role in almost all scientific disciplines (Hájek, 2002). Mathematically, probability is defined as a function satisfying Kolmogorov’s axioms, providing a framework for quantifying uncertainty. However, as D’Agostini (2017) argues, probability is not just a number between 0 and 1 that satisfies the axioms, “because such a ‘definition’ says nothing about what we are talking about.” Probability requires a practical interpretation to connect mathematical formalism to real-world contexts. The question of what probability represents in real-world contexts has sparked decades of philosophical and scientific debate and remains an open issue (Galavotti, 2017). For comprehensive surveys of these interpretive debates, readers may refer to Bunge (1981), Seidenfeld (2015), D’Agostini (2017), Hennig (2024), and Spiegelhalter (2024).

Existing interpretations of probability generally fall into two categories: subjective and objective. In the subjective interpretation (or Bayesian interpretation), probability reflects personal beliefs about a parameter (treated as a constant) or hypothesis, often referred to as the “degree of belief” (e.g., D’Agostini, 2017; Willink, 2025). In contrast, objective interpretations view probability as an inherent physical property of a system. Objective interpretations include frequency interpretation and propensity interpretation. The frequency interpretation defines probability as the limiting relative frequency in repeated trials. The propensity interpretation describes probability as a system’s tendency to produce an outcome, and therefore probability can be called the “degree of propensity” (Huang, 2023). While both subjective and objective interpretations provide valuable insights, they often fail to fully capture the intuitive ways people perceive probability in terms of information and informativeness.

In this paper, we propose an information-theoretic perspective, which defines probability as “information weight” to align with the intuitive and familiar concept “weight”. The following sections are structured as follows. Section 2 describes the proposed information-theoretic perspective on probability and probability density. Section 3 describes the relations of the proposed perspective with Shannon information theory (Shannon, 1948) and informity theory (Huang, 2025). Section 4

introduces information-theoretic terminology for probability concepts. Section 5 provides discussion and conclusion.

2. The Proposed Information-Theoretic Perspective on Probability and Probability Density

2.1. Discrete Systems

In the context of this paper, a discrete system is a system with a countable number of states. The state variable of a discrete system changes only at specific, distinct points (states). For example, tossing a coin is a discrete system with two specific, distinct states: "heads" and "tails."

Let X denote the state variable. Suppose X has n multiple states: $\{x_1, \dots, x_n\}$, which is its state space. Each state x_i occurs multiple times m_i . Mathematically, this discrete system can be modeled by a multiset (X, m) : $\{x_1^{m_1}, x_2^{m_2}, \dots, x_n^{m_n}\}$, where $X = \{x_1, \dots, x_n\}$ is the underlying set of distinct elements (states), n is the number of elements, m_i is the multiplicity, i.e., the number of times each element x_i appears in the multiset, and $\sum_{j=1}^n m_j$ is the size (i.e., the total multiplicities) of the multiset.

Definition 1. Each state x_i is a piece of *information* about the state variable X (or the discrete system). For instance, the state "heads" or "tails" is a piece of information about the coin-tossing system.

Definition 2. The multiplicity m_i is defined as the *information content* of x_i , and the size of the multiset $\sum_{j=1}^n m_j$ is defined as the total information content of the discrete system.

Definition 3. The information weight of x_i , denoted by $W(x_i)$, is defined as the ratio of m_i to $\sum_{j=1}^n m_j$

$$W(x_i) = \frac{m_i}{\sum_{j=1}^n m_j}. \quad (1)$$

That is, information weight is the *relative information content* of x_i . Note that the sum of the information weights of all the states of the discrete system is

$$\sum_{i=1}^n \left(\frac{m_i}{\sum_{j=1}^n m_j} \right) = 1. \quad (2)$$

That is, the *total relative information content* of the discrete system is 1.

Furthermore, the information weight function, denoted by $W(X)$, represents the distribution of information weight across all states of the discrete system.

Definition 4. Assume that the discrete system has a mean state and each state x_i is numerical. (e.g., measurement errors are real numbers). Let $\bar{x}$ denote the mean state, which is defined as

$$\bar{x} = \sum_{i=1}^n \frac{m_i x_i}{\sum_{j=1}^n m_j} = \sum_{i=1}^n W(x_i) x_i. \quad (3)$$

Therefore, the mean state $\bar{x}$ is a weighted average of all states x_i of the discrete system, explaining why we term $W(x_i)$ the information weight.

Now, we relate the discrete system modeled by the multiset to a discrete random variable X with a probability mass function (PMF) $P(X)$. Each state x_i in the discrete system corresponds to a possible outcome of X and $P(x_i)$ represents its probability. From this, it is easy to see that the probability $P(x_i)$ is equal to the information weight $W(x_i)$

$$P(x_i) = W(x_i) = \frac{m_i}{\sum_{j=1}^n m_j}, \quad (4)$$

and

$$\sum_{i=1}^n P(x_i) = \sum_{i=1}^n \left(\frac{m_i}{\sum_{j=1}^n m_j} \right) = 1. \quad (5)$$

Equation (4) establishes the equivalence between information weight and probability, while Eq. (5) confirms the normalization of probabilities. Therefore, the PMF $P(X)$ and the information weight function $W(X)$ are equivalent, representing the informational structure of a discrete system or discrete distribution.

The information weight $W(x_i)$ quantifies how strongly a state contributes to the system's informational structure. Higher information weight means the state (or outcome) is more prevalent or predictable. For example, consider a fair six-sided die rolled 60 times, with outcomes $\{1, 2, 3, 4, 5, 6\}$ and multiplicities $\{10, 10, 10, 10, 10, 10\}$, the information weight of each outcome is $W(x_i) = 10/60 \approx 0.167$. Therefore, each outcome has equal informational weight, reflecting a uniform distribution where no single outcome dominates the system's informational structure.

2.2. Continuous Systems

In the context of this paper, a continuous system is a system where its state variable varies continuously, taking any value within a range. For example, sediment particle sizes, ranging from pebbles to microscopic clay, form a continuous system (i.e., a continuous distribution of particle sizes).

Let Y denote the state variable, taking values $y \in [a, b]$, where $[a, b]$ is a continuous state space. To extend the discrete framework, we divide $[a, b]$ into $n(\Delta y)$ number of small bins of size $\Delta y = \frac{b-a}{n(\Delta y)}$, each bin i ($i = 1, 2, \dots, n(\Delta y)$) contains the element $\delta_i \in [y_i - \frac{\Delta y}{2}, y_i + \frac{\Delta y}{2}]$ centered at y_i . Then, this discretized continuous system can be modeled by a multiset (Y, m) : $\{\delta_1^{m_1(\Delta y)}, \delta_2^{m_2(\Delta y)}, \dots, \delta_{n(\Delta y)}^{m_{n(\Delta y)}(\Delta y)}\}$, where $Y = \{\delta_1, \dots, \delta_{n(\Delta y)}\}$ is the underlying set of distinct elements, $m_i(\Delta y)$ is the multiplicity, i.e., the number of data points that fall into the bin i , and $\sum_{i=1}^{n(\Delta y)} m_j(\Delta y)$ is the size of the multiset. This multiset corresponds to a histogram commonly used to visualize a set of continuous data, with δ_i on the x axis and $m_i(\Delta y)$ on the y axis, such as data for sediment particle sizes. Smaller bin size Δy increases the bin number $n(\Delta y)$, but reduces the multiplicity $m_i(\Delta y)$ of each δ_i , thereby refining the histogram's approximation to the continuous distribution.

Definition 5. Each value $y_i \in [a, b]$ is a piece of *information* about the continuous state variable Y . For example, the sediment particle size $y_i = 0.003\text{mm}$ is a piece of information about Y .

Definition 6. The multiplicity $m_i(\Delta y)$ divided by the bin size Δy is the *information content* per unit state space, and $\sum_{i=1}^{n(\Delta y)} m_j(\Delta y)$ is the information content of the discretized continuous system. Then, the information weight per unit state space is given by

$$\frac{W(\delta_i)}{\Delta y} = \frac{m_i(\Delta y)/\Delta y}{\sum_{j=1}^{n(\Delta y)} m_j(\Delta y)}. \quad (6)$$

Definition 7. The information density at y_i , denoted by $D(y_i)$, is defined as

$$D(y_i) = \lim_{\Delta y \rightarrow 0} \left(\frac{W(\delta_i)}{\Delta y} \right) = \lim_{\Delta y \rightarrow 0} \left(\frac{m_i(\Delta y)/\Delta y}{\sum_{j=1}^{n(\Delta y)} m_j(\Delta y)} \right), \quad (7)$$

which reflects the informational contribution per unit state space to the informational structure of the continuous system. Furthermore, the information density function, denoted by $D(Y)$, represents the distribution of information density across the values $y \in [a, b]$.

Now, we relate the continuous system to a continuous random variable Y with probability density function (PDF) $p(Y)$ with the support $[a, b]$. Then, it is easy to see that the probability density $p(y_i)$ equals the information density $D(y_i)$

$$p(y_i) = D(y_i) = \lim_{\Delta y \rightarrow 0} \left(\frac{m_i(\Delta y)/\Delta y}{\sum_{j=1}^{n(\Delta y)} m_j(\Delta y)} \right), \quad (8)$$

with the normalization

$$\int_a^b p(y) dy = \lim_{\Delta y \rightarrow 0} \left(\sum_{i=1}^{n(\Delta y)} p(y_i) \Delta y \right) = 1. \quad (9)$$

Thus, the PDF $p(Y)$ and the information density function $D(Y)$ are equivalent, representing the informational structure of a continuous system.

Moreover, the information weight of an interval $[y_1, y_2]$ is given by

$$W(y_1, y_2) = \int_{y_1}^{y_2} y D(y) dy, \quad (10)$$

which corresponding to the coverage probability of the interval.

3. Relations with Shannon Information Theory and Informity Theory

This section explores how the information-weight interpretation relates to Shannon information theory (Shannon, 1948) and informity theory (Huang, 2025), highlighting correspondences and distinctions.

3.1. Relations with Shannon Information Theory

In discrete systems, each state x_i has information weight $D(x_i) = P(x_i)$, representing its *relative information content*. This should not be confused with *Shannon information content*, defined as (Shannon, 1948)

$$I_S(x_i) = -\log_2 P(x_i), \quad (11)$$

which is the logarithm of the probability's reciprocal. We term that reciprocal

$$R(x_i) = \frac{1}{P(x_i)} = \frac{1}{W(x_i)} \quad (12)$$

the *information rarity*. $R(x_i)$ quantifies how uncommon the state x_i is relative to the total occurrence of all states in the discrete system. Notably, information rarity and information weight are inversely related: as information rarity increases, information weight decreases, and vice versa.

In the multiset framework, information rarity can be expressed as

$$R(x_i) = \frac{\sum_{j=1}^n m_j}{m_i}. \quad (13)$$

Then, Shannon **information content** can be written in terms of information rarity

$$I_S(x_i) = \log_2 R(x_i). \quad (14)$$

Accordingly, the discrete **entropy** can be written as

$$H(X) = \sum_{i=1}^n P(x_i) \log_2 R(x_i), \quad (15)$$

which means that the discrete **entropy** is the expectation of the logarithm of information rarity.

Intuitively, information rarity is related to uncertainty. Higher rarity indicates greater uncertainty, explaining Shannon entropy as a measure of the uncertainty of a system.

Moreover, in continuous systems, the differential entropy (which is also called continuous entropy) can be written as

$$H(Y) = - \int_a^b p(y) \log_2 p(y) dy = \int_a^b p(y) \log_2 R(y) dy, \quad (16)$$

where $R(y) = 1/p(y)$, which can be called information rarity density. Thus, the differential entropy is the expectation of the logarithm of information rarity density.

3.2. Relations with Informity Theory

Informity theory, proposed by Huang (2025), is an information-theoretic framework where probability is viewed as relative information content, and informity quantifies the mean relative information content of a system. The proposed information-weight interpretation operationalizes relative information content using the multiset framework, equating probability, information weight, and relative information content as $P(x_i) = W(x_i) = \frac{m_i}{\sum_{j=1}^n m_j}$. Accordingly, the discrete informity $\beta(X)$ can be expressed in terms of information weight

$$\beta(X) = \sum_{i=1}^n P(x_i)^2 = \sum_{i=1}^n P(x_i) W(x_i) = E[W(X)]. \quad (17)$$

That is, the discrete informity is the expectation of the information weight.

Moreover, the continuous informity $\beta(Y)$ can be expressed in terms of information density

$$\beta(Y) = \int_a^b p(y)^2 dy = \int_a^b p(y) D(y) dy = E[D(Y)]. \quad (18)$$

That is, the continuous informity is the expectation of the information density.

4. Information-Theoretic Terminology for Probability Concepts

This section introduces terminology for probability concepts within the proposed information-theoretic framework, building on the information-weight and information-density interpretations established in Sections 2.1 and 2.2. By comparing the information-theoretic perspective with Bayesian and propensity perspectives, we define “degree of information,” “information spectrum,” and “information interval” to reflect the philosophical emphasis on informational structure of a system, contrasting these with subjective belief and causal propensity.

In various perspectives on probability, the mathematical concepts of probability, probability distribution, and probability interval are assigned distinct terminologies that reflect their philosophical underpinnings. In the Bayesian perspective, probability is termed the “degree of belief” (e.g., D’Agostini, 2017; Willink, 2025), the probability distribution can be named the “belief spectrum,” and the probability interval is often known as the “credible interval.” In the propensity perspective, these are referred to as the “degree of propensity,” “propensity spectrum,” and “propensity interval,” respectively (Huang, 2023). Similarly, in the proposed information-theoretic perspective, probability can be called the “degree of information,” the probability distribution can be termed the “information spectrum,” and the probability interval can be called the “information interval.” These terminologies are not mere labels; they align with distinct philosophical views. The “degree of belief” emphasizes subjective confidence in a parameter or hypothesis, the “degree of propensity” highlights the causal tendency of an outcome, and the “degree of information” underscores the informational weight of a state (or outcome). Similarly, the “belief spectrum” describes the distribution of personal beliefs, the “propensity spectrum” describes the distribution of propensity, and the “information spectrum” describes the distribution of information weight or density.

Furthermore, from the proposed information-theoretic perspective, a probability interval with a specified coverage probability can be termed the “information interval” with a specified “degree of information”, representing a range of values supported by the information weight equal to the coverage probability. For example, consider the sample mean (statistic) $\bar{Y}$ of n observations drawn from a normal distribution $Y \sim N(\mu, \sigma^2)$, where μ is the mean and σ is the standard deviation. Its sampling distribution, $N(\mu, \frac{\sigma^2}{n})$, serves as the information spectrum, describing the informational structure of the continuous system. The information interval covering sample means supported by an information weight of 0.95 is given by

$$\left(\mu - 1.96 \frac{\sigma}{\sqrt{n}}, \mu + 1.96 \frac{\sigma}{\sqrt{n}} \right), \quad (19)$$

which is equivalent to the probability interval with a coverage probability of 0.95.

When μ and σ are unknown, the sample mean $\bar{y}$ estimates μ and $s/c_{4,n}$ estimates σ , where s is the sample standard deviation and $c_{4,n}$ is the bias-correction factor for s . This factor is given by (e.g., Wadsworth, 1989)

$$c_{4,n} = \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}, \quad (20)$$

where $\Gamma(\cdot)$ represents the Gamma function.

Using these estimates, we can obtain an empirical information spectrum for the sample mean $\bar{Y}$, which follows a normal distribution $N(\bar{y}, \frac{s^2}{c_{4,n}^2 n})$ (Huang, 2018). Then, the empirical information interval covering sample means supported by a nominal information weight of 0.95 (i.e., nominal coverage probability of 0.95) can be constructed using this empirical information spectrum. It is given by

$$\left(\bar{y} - 1.96 \frac{s}{c_{4,n} \sqrt{n}}, \bar{y} + 1.96 \frac{s}{c_{4,n} \sqrt{n}} \right). \quad (21)$$

The half-width of this empirical information interval is $1.96 \frac{s}{c_{4,n}\sqrt{n}}$, which corresponds to the unbiased expanded uncertainty in measurement uncertainty analysis (e.g., Huang, 2018).

5. Discussion and Conclusion

The proposed information-weight interpretation of probability is intuitive because it maps probability onto the familiar concept of weight. This aligns with D'Agostini's (2017) view that we can understand probability statements by mapping them in "some 'categories' of our mind, as we do with space and time." For example, stating "the probability of rain tomorrow is 0.8" can be rephrased as "raining tomorrow has an information weight of 0.8". Thus, probability, as information weight, can be viewed as a "property" of a piece of information (e.g., a state, outcome, or hypothesis). This is analogous to "strength" is a "physical property" of a material, such as "AISI 302 stainless steel has a tensile strength of 585 MPa."

The proposed information-weight interpretation of probability provides a novel philosophical perspective. Rather than linking probability to personal beliefs (Bayesian), system tendencies (propensity), or long-run frequencies (frequentist), this framework emphasizes probability as the informational weight that supports a state, outcome, or hypothesis and reflects its contribution to the informational structure of a system. In addition, this **interpretation** holds **regardless** of how probabilities are derived, whether from professional judgment, physical models, or empirical frequencies, making it independent of probability assignment methods. Moreover, by defining probability distributions as information spectra and probability intervals as information intervals, this framework unifies the treatment of probability and probability density under a single information-theoretic perspective.

References

1. Bunge, M. (1981). Four concepts of probability. *Applied Mathematical Modelling*, 5(5), 306-312.
2. D'Agostini, G. (2017). Probability, propensity and probability of propensities (and of probabilities). *AIP Conference Proceedings*, 1853(1), 030001.
3. Galavotti, M. C. (2017). The interpretation of probability: still an open issue? *Philosophies*, 2(3), 20. <https://doi.org/10.3390/philosophies2030020>
4. Hájek, A. (2002). Interpretations of Probability. *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.) <https://plato.stanford.edu/entries/probability-interpret/>
5. Hennig, C. (2024). Probability Models in Statistical Data Analysis: Uses, Interpretations, Frequentism-as-Model. In Bharath Sriraman (ed.), *Handbook of the History and Philosophy of Mathematical Practice*. Cham: Springer. 1411-1458.
6. Huang, H. (2018). A unified theory of measurement errors and uncertainties. *Measurement Science and Technology*, 29, 125003. <https://doi.org/10.1088/1361-6501/aae50f>
7. Huang, H. (2023). A propensity-based framework for measurement uncertainty analysis. *Measurement*, 112693. <https://doi.org/10.1016/j.measurement.2023.112693>
8. Huang, H. (2025). The theory of informity: a novel probability framework. *Bulletin of Taras Shevchenko National University of Kyiv, Physics and Mathematics*, 80(1), 53-59. DOI: <https://doi.org/10.17721/1812-5409.2025/1.7>
9. Shannon, C. E. (1948). A mathematical theory of communications. *The Bell System Technical Journal*, 27, 379-423, 623-656.
10. Spiegelhalter D. (2024). Why probability probably doesn't exist (but it is useful to act like it does). *Nature*, 636(8043):560-563. doi: 10.1038/d41586-024-04096-5. PMID: 39681646.
11. Seidenfeld, T. (2015). Probability Theory: Interpretations. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, 37-42. <https://doi.org/10.1016/B978-0-08-097086-8.42068-4>
12. Wadsworth, Jr. H. M. (1989). Summarization and interpretation of data. In *Handbook of Statistical Methods for Engineers and Scientists*, 2.1-2.21, Ed. Harrison M Wadsworth Jr. McGRAW-HILL In.

13. Willink, R. (2025). On the role of probability in science, analytical measurement and QUAM. *Accred Qual Assur*, 30, 245–252. <https://doi.org/10.1007/s00769-025-01631-3>

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.