Journal of Applied Mathematics and Physics, 2026, 14(11), 4640-4670
https://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

Beyond Uncertainty: Quantifying Doubt within the Certain

Ilija Barukčić

Independent Researcher, Horandstrasse, Jever, Germany
Email: Barukcic@t-online.de

How to cite this paper: Barukčić, I. (2026) Beyond Uncertainty: Quantifying Doubt within the Certain. *Journal of Applied Mathematics and Physics*, **14**, 4640-4670.
https://doi.org/10.4236/jamp.2026.1411228

Received: July 17, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

We present Hegel's distribution, a discrete uniform distribution tailored to the analysis of sure events or relationships where the theoretical probability is $p = 1$. The probability mass function $P(X = k) = 1/n$ and the right-tailed p-value $p_{\text{right}} = (n - k_{\text{events}} + 1)/n$ provide a straightforward and computationally efficient framework for testing deviations from certainty. Two examples—the causal relationship between BNT162b2 vaccination and COVID-19, and quality control in production lines—illustrate the practical utility of the method. The results demonstrate that Hegel's distribution enables objective, transparent inference regarding whether observed deviations from unity are attributable to random variation or disclose a systematic disturbance beneath the surface of observation. Due to its analytical simplicity, closed-form expressions, and minimal computational overhead, the framework is particularly well-suited for big data applications, where scalability and interpretability are paramount. At the same time, it bridges statistical rigor and theoretical reflection on the nature of certainty, imperfection, and doubt.

Keywords

Certain Events, P-Value, Causality

1. Introduction

Very often, observations—even those that appear to be random—are shaped and determined by the natural laws that structure reality. It is precisely under the constraining influence of such laws that certain expectations and predictions regarding observations we have not yet made become possible, thereby enabling us to move beyond what has actually been observed and into the realm of the unobserved. The question arises: how can we be absolutely certain that what we rea-

sonably suspect to be the case is indeed so, or will continue to be so? The enduring tension between theoretical certainty and probabilistic reality lies at the heart of any empirical inquiry. In theory, certain events are often assigned a probability of $p = 1$, elevating them to the status of virtual inevitability, in the hope that objective reality may be adequately described by the clean logic of classical certainty.

In this context, consider our sun. It has risen each morning of our lives with a constancy that seems to border on necessity, as though its rising were governed by an absolute, an unassailable certainty or a probability $p = 1$. Yet this certainty may not be the likeness of an eternal order, but rather a temporary product of material processes—shaped by events in space and time, like the traces we leave in the sand, which the shore erases with the next tide. For the sun, too, is subject to the flux of becoming; it may be altered, it may decay, and the laws that determine its existence are themselves not immune to transformation. They, too, may reach their end, and with them, the sun may one day cease to shine.

Herein lies the core of the tension between the apparent inevitability of the observed and the latent contingency of the real. To insist upon absolute epistemic certainty in theory, when the world in its concrete actuality affords us no such guarantee, appears logically untenable. And yet-precisely at this juncture—we must not succumb to *Hume's scepticism*. For while absolute certainty may forever elude our grasp, relative certainty remains within reach. Epistemic certainty, in the sense that matters for science and for practice, is not the absence of doubt, but the presence of logically sound reasoning in its presence. It is the quiet confidence that emerges not from denying the possibility of surprise, but from understanding and quantifying its degree.

The relativity or the fragility of human knowledge is rooted in the very nature of natural processes. In practice, therefore, we must attend to the measure of uncertainty that accompanies even the most seemingly certain events—for uncertainty accompanies every event. Any desire that objective reality should remain unchanged, that it must persist as it is, seems to rest upon a logical fallacy of immutability and one's limited temporal perspective onto the entire objective reality. For what is constant is change itself; and even our most trusted regularities, our most cherished certainties, therefore rest upon relative and provisional ground. Yet we have the means to quantify the degree of uncertainty that accompanies even our most secure knowledge. This quantification is the task of probability theory.

In this article, we turn our attention to the discrete uniform distribution—which we designate as Hegel's distribution—and to the p-values that arise from it values [1] [2]. What presents itself at first glance as a construct of utmost simplicity proves, upon deeper analysis, to harbour a surprising wealth of theoretical and statistical complexity. Beyond this, it opens the way toward a novel response to *David Hume's problem of induction* [3]—a problem that continues to delineate the horizons of scientific reasoning, probabilistic thought, and the very limits of human knowledge.

2. Material and Methods

Definition 1 (Bernoulli Trial). A Bernoulli trial [4] is a random experiment with exactly two possible outcomes: success and failure. In this work, each trial t is represented by the binary random variables

$$(U_t, W_t) \in \{0,1\}^2, \tag{1}$$

where U_t denotes the condition variable (antecedent, exposure, or cause) and W_t denotes the conditioned variable (consequent, outcome, or effect).

The joint distribution is given by

$$P(U_t = u, W_t = w),\ u, w \in \{0,1\}. \tag{2}$$

Definition 2 (Logical Condition or Relationship). For each Bernoulli trial t, let C_t denote a logical condition or relationship defined on the binary variables U_t and W_t. Examples of logical relationships considered in this work are:

$$C_t^{(1)} = (U_t \vee \neg W_t), \tag{3}$$

$$C_t^{(2)} = (\neg U_t \vee W_t), \tag{4}$$

$$C_t^{(3)} = (\neg U_t \vee \neg W_t). \tag{5}$$

where $\vee$, $\wedge$, and $\neg$ denote logical disjunction, conjunction, and negation, respectively. Each logical condition or relationship C_t defines an event with probability

$$P(C_t). \tag{6}$$

Definition 3 (Indicator Random Variable). For a logical condition C_t, define the indicator random variable

$$X_t = \mathbf{1}_{\{C_t\}} = \begin{cases} 1, & \text{if } C_t \text{ occurs,} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Therefore,

$$P(X_t = 1) = P(C_t), \tag{8}$$

and

$$P(X_t = 0) = 1 - P(C_t). \tag{9}$$

Hence,

$$X_t \sim \text{Bernoulli}(P(C_t)). \tag{10}$$

Definition 4 (Aggregate Count Variable). For n independent Bernoulli trials, the total number of occurrences of the condition C_t is defined as

$$X = \sum_{t=1}^{n} X_t = \sum_{t=1}^{n} \mathbf{1}_{\{C_t\}}. \tag{11}$$

Thus, X represents the number of trials in which the logical condition C_t is satisfied. Thus far, consider the special case in which the logical condition or relationship C_t is certain for each individual trial: $P(C_t) = 1$, $t = 1, \cdots, n$. In this

case, every individual trial produces success with certainty:

$$X_t = \mathbf{1}_{\{C_t\}}. \tag{12}$$

Under these conditions, the resulting count variable X satisfies

$$X = \sum_{t=1}^{n} X_t = \sum_{t=1}^{n} \mathbf{1}_{\{C_t\}} = n. \tag{13}$$

Equation (13) shows that, under the assumption $P(C_t) = 1$ for every trial, the count variable X is not random: it is deterministically equal to n, with zero entropy.

Definition 5 (Hegel's Distribution). An event—a single trial—is something elementary admitting exactly two mutually exclusive states: either success or failure. Let $n \in \mathbb{N}$ be the number of such events. Unlike the classical binomial distribution, the events need not share the same success probability: each individual event t possesses its own, individual success probability $P(C_t) \in [0,1]$, which may be known or unknown, equal or different across events. Let X denote the total number of successes observed among the n events.

For every event, success and failure exhaust all possibilities by definition, so $\left(P(C_t) + \left(1 - P(C_t)\right)\right) = 1$ holds for every $t \in \{1, 2, \cdots, n\}$. We define the probability mass function (PMF) of Hegel's distribution on the possible outcomes $\{1, 2, \cdots, n\}$ as the parameter-free reference model:

$$P(X = k) = \frac{\left(P(C_t) + \left(1 - P(C_t)\right)\right)}{n}, \quad k = 1, 2, \cdots, n. \tag{14}$$

Justification. Hegel's distribution must be distinguished from the ordinary discrete uniform distribution $P(X = k) = 1/n$, which presupposes a fixed, finite support with no reference whatsoever to a Bernoulli process. Hegel's distribution, by contrast, is the reference model for a Bernoulli counting space, in which each event t carries its own, potentially unknown success probability $P(C_t)$. This is the decisive difference—a difference in the nature of the underlying space, not merely in the resulting numbers.

Equation (14) makes this origin explicit: for every event, the tautology $\left(P(C_t) + \left(1 - P(C_t)\right)\right) = 1$ holds. This identity is not a simplification but the expression of the fact that success and failure exhaust all possibilities by definition. Substituting this identity into the numerator of the uniform distribution makes visible that Hegel's distribution is not derived from any particular Bernoulli probability—on the contrary, the $P(C_t)$ cancel out completely, for every t, regardless of its value, known or unknown. Herein lies its strength: it is the parameter-free, maximum-entropy reference model for a counting space whose true sampling law is unknown. It serves as a neutral benchmark against which an observed count k_{events} can be evaluated. Its mean is $\mathbb{E}[X] = (n+1)/2$, and its variance $\mathrm{Var}(X) = \left(n^2 - 1\right)/12$. Normalization follows trivially from $\sum_{k=1}^{n} \left(P(C_t) + \left(1 - P(C_t)\right)\right)/n = 1/n = 1$. For any value k outside this support, we have

$$P(X=k)=0, \tag{15}$$

since such values are not in the support. No additional case distinction is necessary. By definition, Hegel's distribution assigns the value $\left(P(C_t)+(1-P(C_t))\right)/n$ to each of the n possible outcomes. Hence, $P(X=k)=\left(P(C_t)+(1-P(C_t))\right)/n$ for all $k\in\{1,\cdots,n\}$. At the end, this definition emphasizes that the result is independent of the specific value of $P(C_t)$ and holds for any probability $P(C_t)\in[0,1]$, including our special cases $P(C_t)=1$ and $P(C_t)=0$. Hegel's distribution applies in particular to the probability mass of *the necessary condition, i.e.*, $P(C_t)=p(U_t\leftarrow W_t)=1$, for *the sufficient condition*, or *conditio per quam*, $P(C_t)=p(U_t\rightarrow W_t)=1$, for *the exclusion condition* (*conditio exclusionis*), as well as for the causal relationship, $|k|=1$.

Parameters of Hegel's distribution.

$n=b-a+1$: Total number of equally likely outcomes.

a: Smallest possible value of X.

b: Largest possible value of X.

$\mu=\frac{a+b}{2}$: Mean (expected value).

$\sigma^2=\frac{n^2-1}{12}$: Variance.

$$\mathcal{X}=\{1,2,\cdots,n\}. \tag{16}$$

Probability Mass Function (PMF):

$$P(X=k)=\frac{\left(P(C_t)+(1-P(C_t))\right)}{n},\ k=1,2,\cdots,n. \tag{17}$$

Cumulative Distribution Function (CDF):

$$F(x)=P(X\le x)=\begin{cases}0, & x<1,\\ \frac{\lfloor x\rfloor}{n}, & 1\le x\le n,\\ 1, & x>n.\end{cases} \tag{18}$$

Expected Value (Mean):

$\mathbb{E}[X]=\frac{n+1}{2}$. (In the special case where $a=1$ and $b=n$.).

Variance:

$$\mathrm{Var}(X)=\frac{n^2-1}{12}. \tag{19}$$

Standard Deviation:

$$\sigma(X)=\sqrt{\frac{n^2-1}{12}}. \tag{20}$$

Moment Generating Function (MGF):

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \frac{1}{n}\sum_{k=1}^{n} e^{tk} = \frac{e^t\left(1-e^{tn}\right)}{n\left(1-e^t\right)},\ t \neq 0. \tag{21}$$

Probability Generating Function (PGF):

$$G_X(s) = \mathbb{E}\left[s^X\right] = \frac{1}{n}\sum_{k=1}^{n} s^k = \frac{s\left(1-s^n\right)}{n\left(1-s\right)},\ s \neq 1. \tag{22}$$

Characteristic Function:

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right] = \frac{1}{n}\sum_{k=1}^{n} e^{itk} = \frac{e^{it}\left(1-e^{itn}\right)}{n\left(1-e^{it}\right)},\ t \neq 0. \tag{23}$$

Skewness:

$$\text{Skewness} = 0. \tag{24}$$

(The distribution is perfectly symmetric.)

Excess Kurtosis:

$$\text{Kurtosis} = -\frac{6\left(n^2+1\right)}{5\left(n^2-1\right)}. \tag{25}$$

Shannon Entropy:

$$H(X) = -\sum_{k=1}^{n}\frac{1}{n}\log\frac{1}{n} = \log n. \tag{26}$$

The distribution maximizes entropy among all discrete distributions on n outcomes.

Mode:

$$\text{Mode} = \{1, 2, \cdots, n\}. \tag{27}$$

All values are equally likely, hence every point is a mode.

Median:

$$\text{Median} = \frac{n+1}{2}. \tag{28}$$

In this work, the term Hegel's distribution denotes a specific discrete uniform distribution defined by certainty events satisfying $P(C_t) = 1$. This terminology refers exclusively to this particular construction and does not replace the general concept of a discrete uniform distribution. Hegel's distribution is based on the principle of equiprobability. The principle of equiprobability stands as one of the oldest and most enduring concepts in probability theory, emerging from the study of gambling games in the 16th century. Early mathematicians such as Gerolamo Cardano (c. 1564) formalized this principle as one of the first. Cardano [5] described the principle of "equal conditions" for a fair die, stating that each face of a fair die has the same chance of appearing. This constitutes a conceptual cornerstone, but does not provide a mathematical formula for the probability mass function (PMF). Pascal and Fermat (1654) applied the principle of equiprobability to pioneer the concept of expected value [6] and fair wager distributions in their fa-

mous correspondence on the problem of points. Abraham de Moivre (1718) defined probability [7] in general terms as a fraction, *i.e.*, the number of favorable cases divided by the total number of cases. While this is a general definition applicable to any situation with equally likely outcomes, it does not specify the PMF of the uniform distribution. Thomas Bayes (1763) employed the *continuous* [8] uniform distribution on the interval $[0,1]$ as a prior for an unknown probability parameter in his seminal essay on the problem of chance. This is a continuous, not a discrete, uniform distribution. Moreover, Bayes did not formulate a PMF, as the term itself was coined much later. While Pierre-Simon Laplace (1812) provided the first rigorous definition of classical probability [9] based on equally likely outcomes, it was J. V. Uspensky (1937) who popularized the term uniform distribution [4]. However, Uspensky defined a continuous uniform distribution, stating that probabilities attached to two equal intervals are equal, and did not address the discrete PMF. Therefore, in modern mathematical literature, the discrete uniform distribution is not attributed to any single author, as this brief historical overview indicates.

Definition 6. Hegel's distribution and Null and Alternative Hypotheses. In the present analysis, we formalize the underlying statistical framework by distinguishing between two competing hypotheses regarding the certainty of the event under investigation.

In classical frequentist statistics, the null hypothesis typically represents a status quo or a point of no effect, and the p-value is defined as the probability of observing data as extreme as or more extreme than the observed data, assuming that the null hypothesis is true. The classical hypotheses are usually formulated similar to:

$$H_0 : p = 1 \text{ vs. } H_A : p < 1. \tag{29}$$

The present framework reverses this logical direction. Instead of testing whether the process deviates from an ideal state, we test whether the observed data are consistent with the ideal state itself.

This is not merely a reformulation of classical testing, but a deliberate conceptual inversion. The classical approach starts from the assumption that the ideal is false and looks for evidence against it; the proposed approach starts from the assumption that the ideal may be true and looks for evidence that the data are consistent with it. Metaphorically speaking, the classical view sees the glass as half empty—assuming the ideal is absent unless proven otherwise—whereas the present framework sees the glass as half full—assuming the ideal is present unless the data prove otherwise.

This inversion is not arbitrary but reflects the epistemological context of analyzing sure events. When events are expected to occur with probability $p = 1$, the burden of proof shifts: it is no longer reasonable to assume that the ideal is false and to test against it; rather, it is necessary to prove whether the observed data are compatible with the ideal. In other words, we turn the conventional hypothesis structure upside down and, by setting it upon its feet, reformulate it as:

$$H_0 : p < 1 \text{ vs. } H_A : p = 1. \tag{30}$$

The null hypothesis (H_0) posits that the event is not certain, meaning that the theoretical probability of success is strictly less than unity. Our H_0 is

$$H_0 : p < 1. \tag{31}$$

formalized as: **Null Hypothesis (H_0):** The observed data follow the reference distribution of maximal entropy (Hegel's distribution), implying the absence of a deterministic relationship ($p < 1$). Informally, this means that, in the complete absence of further data or information, every possible outcome between 1 and n is treated as equally likely—just as every face of a fair die is equally likely to appear—rather than any particular value being favored over another.

Conversely, the alternative hypothesis (H_A) asserts that the event is certain, with a theoretical probability of success equal to unity. Our H_A is

$$H_A : p = 1. \tag{32}$$

formalized as: **Alternative Hypothesis (H_A):** The observed data are consistent with certainty ($p = 1$).

The right-tailed p-value is defined as

$$p_{\text{right}} = \frac{n - k_{\text{events}} + 1}{n} = \frac{1}{n} \sum_{i=k_{\text{events}}}^{n} 1, \tag{33}$$

where n denotes the sample size and k_{events} the number of observed successes. Under this reversed formulation, a small right-tailed p-value indicates that the process is consistent with being error-free ($p = 1$). More precisely, if the process were not error-free ($p < 1$), the probability of observing a result as favorable as or more favorable than the observed data would be very small. In other words, the observed data would be unlikely under the assumption that the process is not error-free. Therefore, we are justified in concluding that the data are consistent with the hypothesis that the process is error-free ($p = 1$).

Such hypothesis tests controls the Type I error rate at a nominal significance level α by rejecting H_0 whenever $p_{\text{right}} \le \alpha$. The power of the test increases with sample size, as the p-value $1/n$ decreases monotonically with n when $k_{\text{events}} = n$. For finite samples, the test is conservative in the sense that it does not require distributional assumptions beyond the uniform count distribution over $\{1, \cdots, n\}$.

Under the alternative hypothesis $H_A : p = 1$, the process is error-free, and thus $k_{\text{events}} = n$ with probability 1. The right-tailed p-value is then

$$p_{\text{right}} = \frac{1}{n}. \tag{34}$$

The null hypothesis $H_0 : p < 1$ is rejected if $p_{\text{right}} \le \alpha$. Consequently, the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true (statistical power) is quantified as:

$$\text{Power} = P\left(\frac{1}{n} \le \alpha \mid H_A : p = 1\right). \tag{35}$$

Since $k_{\text{events}} = n$ almost surely under H_A, the rejection event is deterministic:

$$\text{Power} = \begin{cases} 1, & \text{if } n \geq \dfrac{1}{\alpha}, \\ 0, & \text{if } n < \dfrac{1}{\alpha}. \end{cases} \tag{36}$$

Thus, the test has unit power whenever the sample size is large enough to yield a *p*-value below the nominal significance level. This property reflects the logical structure of testing sure events: if the process is truly error-free, a sufficiently large sample will always provide decisive evidence in favor of H_A.

The consequences of the results are straightforward: a small *p*-value implies that the observed data are *inconsistent* with the assumption of a non-error-free process and are therefore *consistent* with the hypothesis that the process operates as intended ($p = 1$). This framework does not claim to prove causality or certainty, but provides a quantitative, statistically rigorous basis for assessing whether observed deviations from an ideal state can be attributed to random variation or point to critically, systematic disturbances or failures.

Definition 7 (Normalization). The probability mass function of Hegel's distribution satisfies the normalization condition implied by Kolmogorov's fourth [10] axiom. It is:

$$\sum_{k=1}^{n} P(X = k) = \sum_{k=1}^{n} \frac{1}{n} = 1. \tag{37}$$

This follows from Kolmogorov's fourth [10] axiom which demands that the probability assigned to the complete sample space is unity or $P(E) = P(\Omega) = 1$.

Remark 1 (Logical Special Cases). The three logical relationships introduced here correspond to: $C_t^{(1)} = (U_t \vee \neg W_t)$ necessary condition, $C_t^{(2)} = (\neg U_t \vee W_t)$ sufficient condition, and $C_t^{(3)} = (\neg U_t \vee \neg W_t)$ exclusion condition. Whenever the corresponding condition satisfies $P(C_t^{(i)}) = 1$, the certainty condition applies.

Definition 8 (Causal relationship *k*).

Statistical independence [10] describes a situation in which the occurrence of one event does not alter the probability of occurrence of another event. In contrast, causal relationship indicates a deviation from this relationship of independence. In the context of statistical analysis, an observable relationship between a putative cause U_t and an effect W_t can be causally determined. Let U_t and W_t be two events defined on a common probability space $(\Omega, \mathcal{F}, P)$. The events U_t and W_t are causally related [11]-[13] if and only if

$$|k| = \left| \frac{P(U_t \cap W_t) - (P(U_t)P(W_t))}{\sqrt[2]{(P(U_t)(1 - P(U_t)))(P(W_t)(1 - P(W_t)))}} \right| = |1|. \tag{38}$$

provided that $P(U_t) > 0$ and that $P(W_t) > 0$. The relationship between sample size *n* and *k* is illustrated by **Figure 1**.

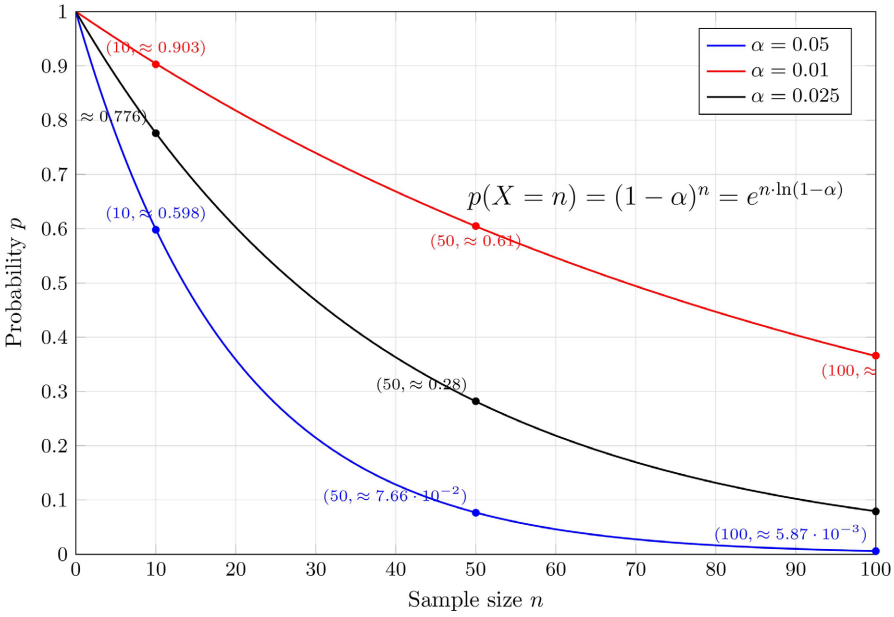

Figure 1. Probability $p(X=n)=(1-\alpha)^n$ versus sample size n for different a values.

Proposition 1. Let a,b,c,d denote the absolute cell frequencies in a 2×2 contingency table, and let the relative frequencies be determined as: $n(a)=\frac{a}{n}$, $n(b)=\frac{b}{n}$, $n(c)=\frac{c}{n}$, $n(d)=\frac{d}{n}$, with $n=a+b+c+d$. Furthermore, let $n(U_t)=\frac{a+b}{n}$, $n(W_t)=\frac{a+c}{n}$. Then the phi coefficient Φ, originally defined in terms of absolute frequencies as

$$\Phi_n=\frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \tag{39}$$

is determined entirely by relative frequencies as:

$$\Phi_n=\frac{n(a)\cdot n(d)-n(b)\cdot n(c)}{\sqrt{n(U_t)\cdot(1-n(U_t))\cdot n(W_t)\cdot(1-n(W_t))}}. \tag{40}$$

Proof. We begin with the definition of Φ in terms of absolute frequencies:

$$\Phi_n=\frac{n(a)\cdot n(d)-n(b)\cdot n(c)}{\sqrt{n(U_t)\cdot(1-n(U_t))\cdot n(W_t)\cdot(1-n(W_t))}}. \tag{41}$$

To express this in terms of relative frequencies, we use the substitutions:

$$a=n\cdot n(a),\ b=n\cdot n(b),\ c=n\cdot n(c),\ d=n\cdot n(d), \tag{42}$$

and

$$a=n\cdot n(a),\ b=n\cdot n(b),\ c=n\cdot n(c),\ d=n\cdot n(d), \tag{43}$$

and

$$a+c=n\cdot n(W_t),\ b+d=n\cdot(1-n(W_t)). \tag{44}$$

Substituting these relationships into the numerator yields:

$$\begin{aligned} ad-bc &= \left(n\cdot n(a)\right)\cdot\left(n\cdot n(d)\right)-\left(n\cdot n(b)\right)\cdot\left(n\cdot n(c)\right) \\ &= n^2\left(n(a)\cdot n(d)-n(b)\cdot n(c)\right). \end{aligned} \tag{45}$$

For the denominator, we obtain:

$$\begin{aligned} &(a+b)(c+d)(a+c)(b+d) \\ &= \left(n\cdot n(U_t)\right)\cdot\left(n\cdot\left(1-n(U_t)\right)\right)\cdot\left(n\cdot n(W_t)\right)\cdot\left(n\cdot\left(1-n(W_t)\right)\right) \\ &= n^4\cdot n(U_t)\cdot\left(1-n(U_t)\right)\cdot n(W_t)\cdot\left(1-n(W_t)\right). \end{aligned} \tag{46}$$

Thus,

$$\begin{aligned} \Phi_n &= \frac{n^2\left(n(a)\cdot n(d)-n(b)\cdot n(c)\right)}{\sqrt{n^4\cdot n(U_t)\cdot\left(1-n(U_t)\right)\cdot n(W_t)\cdot\left(1-n(W_t)\right)}} \\ &= \frac{n^2\left(n(a)\cdot n(d)-n(b)\cdot n(c)\right)}{n^2\sqrt{n(U_t)\cdot\left(1-n(U_t)\right)\cdot n(W_t)\cdot\left(1-n(W_t)\right)}}. \end{aligned} \tag{47}$$

Canceling n^2 in the numerator and denominator yields the desired expression:

$$\Phi_n = \frac{n(a)\cdot n(d)-n(b)\cdot n(c)}{\sqrt{n(U_t)\cdot\left(1-n(U_t)\right)\cdot n(W_t)\cdot\left(1-n(W_t)\right)}}. \tag{48}$$

Thus, the phi coefficient is determined solely by the relative frequencies and is invariant with respect to the sample size n.

Quod erat demonstrandum.

A major drawback of the phi coefficient is its implicit assumption that the relative frequencies $n(a), n(b), n(c), n(d)$ are constant across trials. In practice, these quantities are derived from a single sample and may vary due to random fluctuation, measurement error, or systematic changes in the underlying process. As a result, Φ captures only one aspect of reality—a static or mechanistic aggregate relationship—and remains blind to temporal dynamics, event-level fluctuations, and other contextual variations that may be essential for causal inference. This limitation becomes particularly relevant when analyzing causal relationships that may evolve over time or across different conditions. More fundamentally, *the relative frequency* $n(U_t)$ *need not to be identical with the single-event probability* $p(U_t)$. While both quantities theoretically coincide under certain circumstances, they need not do so in finite samples due to random variation, measurement error, or systematic deviations. In practice, $n(U_t)$ serves as an empirical estimate of $p(U_t)$, but the two may diverge. This discrepancy is not a flaw, but a fundamental feature of statistical inference: it is precisely this gap that hypothesis tests and confidence intervals are designed to quantify and evaluate. Pearson's phi coefficient Φ, by treating relative frequencies as fixed constants through all trials, obscures this essential distinction and thereby limits its own utility for causal analysis, where the stability and variability of probabilities over time are of central importance. In contrast to Pearson's phi coefficient Φ, the causal relationship k is defined at every single event or trial, not only for the aggregate. For each indi-

vidual observation i, a causal relationship k_i can be computed based on the specific realization of U_i and W_i as:

$$k_i = \frac{\mathbb{I}(U_i = 1, W_i = 1) - P(U)P(W)}{\sqrt{P(U)(1-P(U)) \cdot P(W)(1-P(W))}} = 1, \tag{49}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Despite these fundamental differences, both measures of relationship share several formal identities. Both k and Φ lie in the interval $[-1,1]$, both are zero under perfect independence, both are symmetric in U and W, both measure relationship for binary variables, and both are invariant under linear transformations of the variables. These formal similarities, however, should not obscure the profound conceptual and practical distinctions between these two measures.

The phi coefficient Φ is expressed in terms of cell counts a,b,c,d, while k is expressed in terms of probabilities $P(U)$, $P(W)$, and $P(U \cap W)$. The phi coefficient explicitly includes the sample size n through the cell counts, whereas k is written as a population parameter. Moreover, Φ is defined only at the level of the entire contingency table, requiring aggregated cell counts from all n observations, and provides a single summary measure for the complete dataset. The formula for k more clearly reveals the covariance structure:

$$k = \frac{\mathrm{Cov}(U,W)}{\sqrt{\mathrm{Var}(U)\mathrm{Var}(W)}}. \tag{50}$$

The causal relationship k is not identical to Pearson's correlation coefficient r in the general case. While both share a similar mathematical form $\frac{\mathrm{Cov}}{\sqrt{\mathrm{Var} \cdot \mathrm{Var}}}$, Pearson's correlation coefficient r is defined for continuous, interval-scaled variables and assumes linearity. In contrast, k is specifically tailored for the relationship between cause and effect, going far beyond the special case of Pearson's correlation for dichotomous data.

First, Pearson's correlation coefficient requires the variables to be measured on at least an interval scale, whereas U and W are determined by their own probabilities. *Second*, the interpretation of k is explicitly causal, grounded in the probabilistic independence condition

$$P(U \cap W) = P(U)P(W), \tag{51}$$

while Pearson's r is merely descriptive of linear association without any causal claims.

Third, the phi coefficient ϕ is actually the direct application of Pearson's correlation to binary variables, and at the aggregate level, there are circumstances where $k = \phi$. However, Pearson's original formulation does not possess the event-level interpretability of k_i, which is crucial for causal (trajectory) analysis.

Fourth, Pearson's correlation is symmetric and does not distinguish between cause and effect, whereas k is embedded in a framework where U is explicitly treated as a potential cause and W as an effect, allowing for directional hypoth-

eses testing. *Fifth*, the causal relationship k is derived from the Hegelian dialectical framework of certainty and contingency, a philosophical foundation that has no analogue in Pearson's correlation.

Thus, while numerical identities with Pearson's phi coefficient Φ or Pearson's correlation coefficient r may arise under certain circumstances, k is theoretically, mathematically, and practically fundamentally distinct—by virtue of its nature given determination, event-level granularity, and logical foundations. This event-level derivation allows for trajectory analysis, time-varying causal relationships, and the detection of local dependencies that would be obscured by the aggregated phi coefficient. The phi coefficient Φ answers the question: "*What is the overall association in the entire dataset*?" The causal relationship k answers the question: "*How is the causal relationship determined between individual cause and effect*?"

Unlike Φ, k permits the analysis of causal dynamics over time by examining the sequence $k_1, k_2, \cdots, k_n$. Event-level k_i values can reveal outliers or sudden changes in the causal relationship that are completely invisible in the single aggregate Φ. Furthermore, the event-level definition allows for sequential Bayesian updating of k as new observations arrive, while Φ requires recomputation from the entire aggregated table. These capabilities make k a far more flexible and informative tool for causal analysis, particularly in contexts where relationships are expected to evolve over time or vary across individuals.

Proposition 2 (Empirical Attenuation Model). Assume that the theoretical event-level causal relationship is perfect, namely $|k|=1$, and that each trial is subject to an independent disturbance represented by an attenuation factor $1-\delta$, $0 \le \delta \le 1$. Then, after n trials, the effective causal strength is $|k_n| = (1-\delta)^n$. If the empirical phi coefficient Φ_n is computed from the same n observations and is taken as the aggregate measure of the remaining causal strength, then $\Phi_n = (1-\delta)^n$.

Hence,

$$\Phi_n = |k_n| = (1-\delta)^n. \tag{52}$$

Proof. By assumption, the theoretical event-level causal relationship is perfect:

$$|k| = 1. \tag{53}$$

Suppose that each observation is affected by an independent disturbance that preserves only a proportion $1-\delta$ of the causal relationship. After one trial,

$$|k_1| = 1-\delta. \tag{54}$$

Applying the same attenuation successively over n trials gives

$$|k_n| = (1-\delta)^n. \tag{55}$$

However, the empirical phi coefficient Φ_n is computed from the same set of n observations. As an example, the number 10 may be represented in many different ways: as 5 + 5, as 5*2, as $\log_{10}(10^{10})$ et cetera. Although these expressions

arise from distinct mathematical operations and embody different conceptual perspectives, they nevertheless refer to one and the same number 5. The multiplicity of representations does not alter the identity of the object itself; rather, it reveals that mathematical truth may manifest itself through different formal paths. An analogous situation arises in the study of causality. The event-level coefficient k and the aggregate measure Φ_n emerge from different constructions and are derived from distinct mathematical considerations. Yet, when expressed in terms of relative frequencies and sample size, both measures of relationship describe the same empirical reality, under certain circumstances. In particular, under the assumption of a perfect causal relationship, that is, $|k|=1$, both the empirical causal relationship k_n and the global measure Φ_n are governed by the same law of attenuation. Therefore, *under the modeling assumption*, there are circumstances where Φ_n is numerically equivalent with the empirical causal relationship $|k_n|$. In other words, it follows that

$$\Phi_n = |k_n|. \tag{56}$$

This identity is not a mere algebraic coincidence. Rather, it reflects a deeper unity concealed beneath apparently different formulations. The discrepancy δ between theoretical expectation and empirical observation acts upon both the event-level and the aggregate description in precisely the same multiplicative fashion. Thus, what appears as two distinct measures is revealed to be two different perspectives on a single underlying structure. More generally, the equivalence of these expressions illustrates an important epistemological principle: *the essence of a phenomenon appears to be independent of the particular route by which it is apprehended*. In so far, many roads may traverse the landscape of thought, and many different formal systems and statistical methods may attempt to capture the world. Yet, beneath this diversity, these methods are ultimately bound to express the same reality. Beyond the variety of forms and approaches, there persists an enduring unity of nature—a unity that is not negated by the plurality of scientific methods. The methods may vary, yet the truth they seek to express remains one and the same. It is precisely through this plurality that the underlying truth becomes manifest. The manifold forms of representation conceal an underlying unity, namely that empirical causal relationship $|k_n|$ is determined jointly by sample size and by the discrepancy between theory and observation. In this sense, mathematical reasoning does not merely manipulate symbols; it uncovers invariants that persist across different modes of description. The unity of the result survives the diversity of its representations, revealing that distinct formal languages may ultimately express the same empirical truth. In general, it is

$$\Phi_n = (1-\delta)^n, \tag{57}$$

and consequently,

$$\Phi_n = |k_n| = (1-\delta)^n. \tag{58}$$

Quod erat demonstrandum.

For sufficiently small δ, the term $(1-\delta)^n$ can be simplified to the exponential form as $|k_n| = (1-\delta)^n = \left(1-\frac{\delta \cdot n}{n}\right)^n$. Using the approximation $\ln(1-\delta) \approx -\delta$, we obtain $|k_n| = e^{-(n\cdot\delta)}$. Thus far, there are circumstances where the aggregate Φ_n coefficient and the empirical event-level causal relationship $|k_n|$ are related by the relationship

$$\Phi_n = |k_n| = (1-\delta)^n = (|k| - \delta)^n = e^{-(n\cdot\delta)}. \tag{59}$$

Taking the *n*-th root on both sides yields

$$|k| - \delta = (\Phi_n)^{1/n} = (|k_n|)^{1/n}. \tag{60}$$

Thus, the discrepancy-corrected causal relationship per trial is obtained as the *n*-th root of the aggregate phi coefficient, or equivalently, of the attenuated empirical event-level causal relationship $|k_n|$. This relationship is illustrated in Figure 1, with the discrepancy δ substituted by the significance level *α*.

Definition 9 (Shannon entropy). Let X be a discrete random variable with probability mass function $p(x)$ over a finite set $\mathcal{X}$. The Shannon entropy $H(X)$ is defined as $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x))$, and measures the average uncertainty associated with the outcomes of X.

Remark 2 (Shannon Entropy). Shannon entropy [14] was introduced by Claude E. Shannon in 1948 as a mathematical measure of uncertainty and information content in a message source. It was first published 1948 in the landmark two-part paper *A Mathematical Theory of Communication* in the *Bell System Technical Journal*. The entropy function is defined as $H(X) = -\sum p(x)\log(p(x))$, where X is a discrete random variable. The term "entropy" was proposed by John von Neumann, who noted its formal similarity to the thermodynamic entropy of physical systems. Shannon's work fundamentally transformed communication engineering and remains a cornerstone of modern information theory, data compression, statistical inference, and machine learning.

3. Results

3.1. The Probability Mass Function (PMF) of Hegel's Distribution

Theorem 3.1. The probability mass function (PMF) of Hegel's distribution is given as:

$$P(X=k) = \frac{\left(P(C_t) + (1-P(C_t))\right)}{n}, \quad k = 1,2,\cdots,n. \tag{61}$$

Proof. In general, based on the normalization axiom of probability theory (Kolmogorov's fourth [10] axiom), the total probability over all possible outcomes must equal 1. Therefore, we have

$$\sum_{k=1}^{n} P(X=k) = 1. \tag{62}$$

Substituting $P(X=k)=\frac{1}{n}\cdot\left(P(C_t)+\left(1-P(C_t)\right)\right)$ gives

$$\sum_{k=1}^{n}\frac{1}{n}\cdot\left(P(C_t)+\left(1-P(C_t)\right)\right)=\frac{1}{n}\cdot\sum_{k=1}^{n}1=\frac{1}{n}\cdot n=1. \tag{63}$$

Therefore,

$$P(X=k)=\frac{\left(P(C_t)+\left(1-P(C_t)\right)\right)}{n},\ k=1,2,\cdots,n. \tag{64}$$

Quod erat demonstrandum.

This is equivalent to the standard parametrization for an arbitrary interval $\{a,a+1,\cdots,b\}$, where $n=b-a+1$:

$$P(X=k)=\frac{1}{b-a+1}=\frac{1}{n},\ k=a,a+1,\cdots,b. \tag{65}$$

The figure below (Figure 2) elucidates the fundamental character of Hegel's distribution.

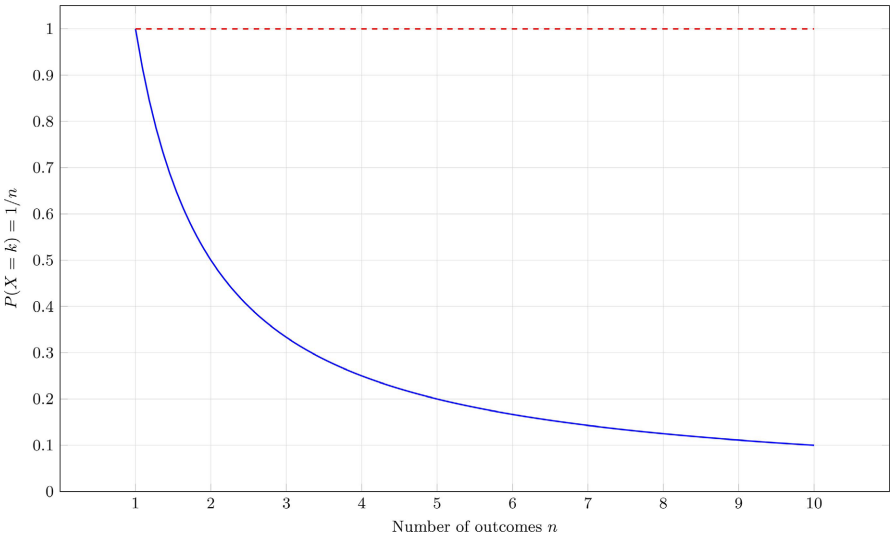

Figure 2. Hegel's distribution and sample size n.

3.2. The Expectation Value of Hegel's Distribution

Theorem 3.2. The expectation value of Hegel's distribution is given as:

$$\mathbb{E}[X]=\frac{a+b}{2}. \tag{66}$$

Proof. Using the definition of the expected value:

$$\mathbb{E}[X]=\sum_{k=a}^{b}k\cdot P(X=k)=\sum_{k=a}^{b}k\cdot\frac{1}{n}=\frac{1}{n}\sum_{k=a}^{b}k. \tag{67}$$

The sum of an arithmetic series is $\sum_{k=a}^{b}k=\frac{n(a+b)}{2}$, where $n=b-a+1$. Therefore,

$$\mathbb{E}[X]=\frac{1}{n}\cdot\frac{n(a+b)}{2}=\frac{a+b}{2}. \tag{68}$$

Quod erat demonstrandum.

3.3. The Expectation Value of X^2 of Hegel's Distribution

Theorem 3.3. The expectation value of X^2 of Hegel's distribution is given as:

$$\mathbb{E}\left[X^2\right]=\frac{b(b+1)(2b+1)-(a-1)a(2a-1)}{6n}. \tag{69}$$

Proof. By definition,

$$\mathbb{E}\left[X^2\right]=\sum_{k=a}^{b}k^2\cdot P(X=k)=\sum_{k=a}^{b}k^2\cdot\frac{1}{n}=\frac{1}{n}\sum_{k=a}^{b}k^2. \tag{70}$$

Using the identity $\sum_{k=a}^{b}k^2=\frac{b(b+1)(2b+1)}{6}-\frac{(a-1)a(2a-1)}{6}$, we obtain:

$$\mathbb{E}\left[X^2\right]=\frac{b(b+1)(2b+1)-(a-1)a(2a-1)}{6n}. \tag{71}$$

Quod erat demonstrandum.

3.4. The Variance of Hegel's Distribution

Theorem 3.4. The variance of Hegel's distribution is given as:

$$\mathrm{Var}(X)=\frac{n^2-1}{12}. \tag{72}$$

Proof. We exploit the translation invariance of the variance: $\mathrm{Var}(X)=\mathrm{Var}(X-a)$. Let $Y=X-a$. Then Y is uniformly distributed on $\{0,1,\cdots,n-1\}$. First, we compute the expected value of Y:

$$\mathbb{E}[Y]=\frac{0+(n-1)}{2}=\frac{n-1}{2}. \tag{73}$$

Next, we compute the second moment:

$$\mathbb{E}\left[Y^2\right]=\frac{1}{n}\sum_{k=0}^{n-1}k^2=\frac{1}{n}\cdot\frac{(n-1)n(2n-1)}{6}=\frac{(n-1)(2n-1)}{6}. \tag{74}$$

Using the identity $\mathrm{Var}(Y)=\mathbb{E}\left[Y^2\right]-\left(\mathbb{E}[Y]\right)^2$, we obtain:

$$\mathrm{Var}(Y)=\frac{(n-1)(2n-1)}{6}-\left(\frac{n-1}{2}\right)^2. \tag{75}$$

To simplify, we bring both terms to the common denominator 12:

$$\mathrm{Var}(Y)=\frac{2(n-1)(2n-1)}{12}-\frac{3(n-1)^2}{12}. \tag{76}$$

Factoring out $(n-1)$ from the numerator:

$$\mathrm{Var}(Y)=\frac{(n-1)\left(2(2n-1)-3(n-1)\right)}{12}. \tag{77}$$

Simplifying the term in parentheses:

$$2(2n-1)-3(n-1)=4n-2-3n+3=n+1. \tag{78}$$

Thus, we have:

$$\mathrm{Var}(Y)=\frac{(n-1)(n+1)}{12}=\frac{n^2-1}{12}. \tag{79}$$

Since $\mathrm{Var}(X)=\mathrm{Var}(Y)$, it follows that:

$$\mathrm{Var}(X)=\frac{n^2-1}{12}. \tag{80}$$

Quod erat demonstrandum.

3.5. Right-Tailed P-Value of Hegel's Distribution

Theorem 3.5. (Right-Tailed P-Value of Hegel's Distribution). Let K be a discrete random variable following the discrete uniform distribution (Hegel's distribution) on the set $\{1,2,\cdots,n\}$, with probability mass function $P(K=j)=\frac{1}{n}$, $j=1,2,\cdots,n$. For an observed value $k_{\text{events}}\in\{1,2,\cdots,n\}$, the right-tailed p-value is given by

$$p_{\text{right}}=\frac{n-k_{\text{events}}+1}{n}. \tag{81}$$

Proof. By definition, the right-tailed p-value is the probability of observing a result as extreme as, or more extreme than, the observed value k_{events}:

$$p_{\text{right}}=P(K\geq k_{\text{events}}). \tag{82}$$

Since K follows the discrete uniform distribution, we have

$$P(K=j)=\frac{1}{n},\ j=1,2,\cdots,n. \tag{83}$$

The right-tailed p-value is defined as the probability of observing a result as extreme as, or more extreme than, the observed value k_{events}:

$$p_{\text{right}}=P(K\geq k_{\text{events}}). \tag{84}$$

By the law of total probability, this probability can be expressed as the sum over all possible outcomes j from k_{events} to n:

$$p_{\text{right}}=\sum_{j=k_{\text{events}}}^{n}P(K=j). \tag{85}$$

Under the discrete uniform distribution (Hegel's distribution), each individual outcome has equal probability $1/n$. Hence, substituting $P(K=j)=1/n$ yields:

$$p_{\text{right}}=\sum_{j=k_{\text{events}}}^{n}\frac{1}{n}. \tag{86}$$

Since the term $1/n$ is constant with respect to j, it can be factored out of the summation:

$$p_{\text{right}}=\frac{1}{n}\cdot\sum_{j=k_{\text{events}}}^{n}1. \tag{87}$$

The summation $\sum_{j=k_{\text{events}}}^{n}1$ simply counts the number of integer values from

k_{events} to n, which is $n-k_{\text{events}}+1$:

$$p_{\text{right}}=\frac{1}{n}\cdot\left(n-k_{\text{events}}+1\right). \tag{88}$$

Thus, the right-tailed *p*-value simplifies to:

$$p_{\text{right}}=\frac{n-k_{\text{events}}+1}{n}. \tag{89}$$

This completes the proof.

Quod erat demonstrandum.

3.6. Left-Tailed P-Value of Hegel's Distribution

Theorem 3.6. (Left-Tailed P-Value of Hegel's Distribution). Let K be a discrete random variable following the discrete uniform distribution (Hegel's distribution) on the set $\{1,2,\cdots,n\}$, with probability mass function $P\left(K=j\right)=\frac{1}{n}$, $j=1,2,\cdots,n$. For an observed value $k_{\text{events}}\in\{1,2,\cdots,n\}$, the left-tailed *p*-value is given by

$$p_{\text{left}}=\frac{k_{\text{events}}}{n}. \tag{90}$$

Proof. The left-tailed *p*-value is defined as the probability of observing a result as extreme as, or less extreme than, the observed value k_{events}:

$$p_{\text{left}}=P\left(K\le k_{\text{events}}\right). \tag{91}$$

By the law of total probability, this probability can be expressed as the sum over all possible outcomes j from 1 to k_{events}:

$$p_{\text{left}}=\sum_{j=1}^{k_{\text{events}}}P\left(K=j\right). \tag{92}$$

where $P\left(K=j\right)$ is the probability mass function. Under the discrete uniform distribution (Hegel's distribution), each individual outcome has equal probability $1/n$. Hence, substituting $P\left(K=j\right)=1/n$ yields:

$$p_{\text{left}}=\sum_{j=1}^{k_{\text{events}}}\frac{1}{n}. \tag{93}$$

Since the term $1/n$ is constant with respect to j, it can be factored out of the summation:

$$p_{\text{left}}=\frac{1}{n}\cdot\sum_{j=1}^{k_{\text{events}}}1. \tag{94}$$

The summation $\sum_{j=1}^{k_{\text{events}}}1$ simply counts the number of integer values from 1 to k_{events}, which is k_{events}:

$$p_{\text{left}}=\frac{1}{n}\cdot k_{\text{events}}. \tag{95}$$

Thus, the left-tailed *p*-value for Hegel's distribution simplifies to:

$$p_{\text{left}} = \frac{k_{\text{events}}}{n}. \tag{96}$$

This completes the proof.

Quod erat demonstrandum.

3.7. Maximum Entropy of the Discrete Uniform Distribution

Theorem 3.7. (Maximum Entropy of the Discrete Uniform Distribution). Let X be a discrete random variable with a finite set of possible outcomes $\mathcal{X} = \{1, 2, \cdots, n\}$. Among all probability distributions on $\mathcal{X}$, the Shannon entropy $H(X)$ is maximized if and only if X follows the discrete uniform distribution, *i.e.*, if $P(X = k) = \frac{1}{n}$, $k = 1, 2, \cdots, n$.

Proof. The entropy of a discrete uniform distribution U over n outcomes is given by

$$H(U) = -\sum_{k=1}^{n} \frac{1}{n} \log \frac{1}{n} = -\log \frac{1}{n} = \log(n). \tag{97}$$

For any other probability distribution P on the same set, we have

$$H(P) \leq \log(n). \tag{98}$$

This follows from the concavity of the logarithm and the normalization constraint $\sum_{k=1}^{n} p_k = 1$:

$$H(P) = \sum_{k=1}^{n} p_k \log \frac{1}{p_k} \leq \log\left(\sum_{k=1}^{n} p_k \cdot \frac{1}{p_k} \right) = \log(n). \tag{99}$$

Equality holds if and only if $p_k = 1/n$ for all k. Therefore, the uniform distribution maximizes entropy.

Quod erat demonstrandum.

Corollary 2. For the discrete uniform distribution on n outcomes, the maximum possible entropy is

$$H_{\max} = \log(n). \tag{100}$$

3.8. Theoretical Probability per Trial

Theorem 3.8. (Theoretical probability per trial). Let $p = 1$ be the true probability of an event in a single trial. Suppose we accept an error rate $\alpha \in (0,1)$ due to various causes. Then the theoretical probability per trial, denoted by p^*, is given by

$$p^* = (p - \alpha) = 1 - \alpha. \tag{101}$$

After n independent trials, the probability of observing all successes ($k = n$) under the theoretical probability is

$$(p^*)^n = (p - \alpha)^n = (1 - \alpha)^n. \tag{102}$$

Proof. Our assumption is that the true probability of the event in each trial is

$$p = 1. \tag{103}$$

Thus, if no error is allowed, the probability of observing $k = n$ successes in n trials is

$$P(K = n) = p^n = 1^n = 1. \tag{104}$$

Now suppose we accept an error rate α at each single trial *t*. Then the theoretical probability per trial, denoted by p^*, is the true probability of an event in a single trial reduced by α or:

$$p^* = p - \alpha = 1 - \alpha. \tag{105}$$

This theoretical probability applies to each single trial *t*. After n independent trials, the probability of observing all successes $k = n$ in n trials is

$$\left(p^*\right)^n = \left(p - \alpha\right)^n = \left(1 - \alpha\right)^n. \tag{106}$$

Equivalently, using the exponential identity $x = \mathrm{e}^{\ln x}$ for $x > 0$, we may write

$$\left(p^*\right)^n = \mathrm{e}^{n \cdot \ln(p-\alpha)} = \mathrm{e}^{n \cdot \ln(1-\alpha)}. \tag{107}$$

Thus, the theoretical probability per trial accounts for the allowable error rate α over n trials. This completes the proof.

Quod erat demonstrandum.

Theorem 3.8. and the relationship between *k*, the theoretical probability per trial and the sample size n is illustrated by Figure 3.

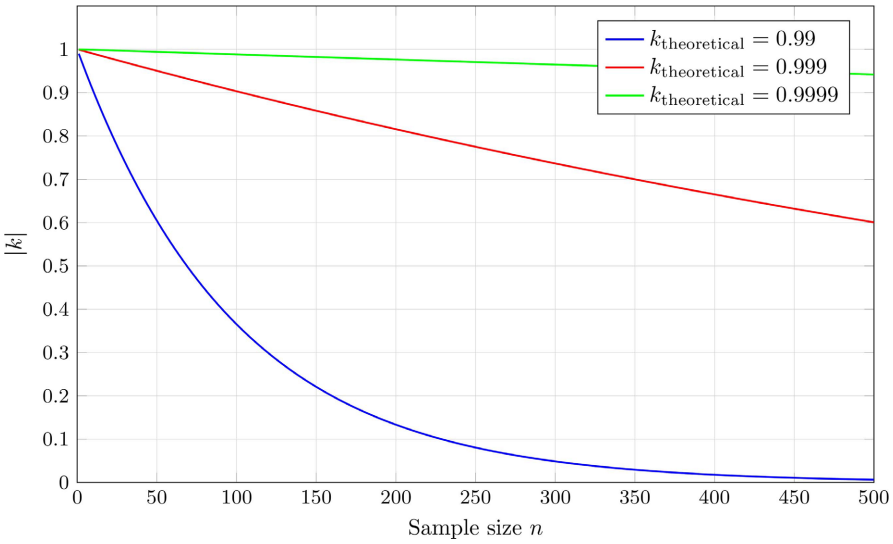

Figure 3. Theoretical causal relationship $k_{\text{theoretical}}$ as a function of sample size *n*.

3.9. Recovery of the Theoretical Probability per Trial

Theorem 3.9. (Recovery of the Theoretical Probability per Trial). Let $|x| = \left(p^*\right)^n$ be the observed probability of obtaining all successes in n independent trials, where $p^* = p - \alpha$ is the theoretical probability per trial reduced by an error rate α. Then the theoretical probability per trial p^* can be recov-

ered from the observed probability $|x|$ by

$$p^* = |x|^{1/n}. \tag{108}$$

Equivalently, using the exponential identity,

$$p^* = \mathrm{e}^{\frac{1}{n}\ln|x|}. \tag{109}$$

Proof. We begin with the observed probability of all successes after n trials:

$$|x| = \left(p^*\right)^n. \tag{110}$$

Our objective is to solve for the theoretical probability per trial p^*. Taking the n-th root on both sides yields

$$p^* = |x|^{1/n}. \tag{111}$$

Using the exponential identity $|x| = \mathrm{e}^{\ln|x|}$ for $x > 0$, we obtain

$$p^* = \mathrm{e}^{\ln\left(|x|^{1/n}\right)} = \mathrm{e}^{\frac{1}{n}\ln|x|}. \tag{112}$$

Since $p^* = p - \alpha$, we can also express the result as

$$p - \alpha = |x|^{1/n}. \tag{113}$$

Thus, from the observed value $|x|$, the theoretical probability per trial p^* is recovered. This completes the proof.

Quod erat demonstrandum.

3.10. Causal Relationship k and Hypothesis Testing

Theorem 3.10. (Causal Relationship and Hypothesis Testing: BNT162b2 vaccine and Covid-19 infection). Polak *et al.* [15] investigated the efficacy of the BNT162b2 vaccine against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a total of 43,448 participants. In the treatment group, 21,720 individuals received BNT162b2, while 21,728 were assigned to the placebo group. There were 8 cases of COVID-19 in the BNT162b2 group and 162 cases among those assigned to placebo.

Table 1. BNT162b2 vaccine efficacy against COVID-19.

		COVID 19 *positive*		
		YES	NO	Total
BNT162b2	YES	**8**	21712	21720
vaccine	No	162	21566	21728
Total		170	43278	**43448**

We reanalyze the data before (**Table 1**) and test the following hypotheses regarding the causal relationship k:

Null hypothesis: $H_0 : |k| < 1$, indicating that the causal relationship is not perfect and that there is some degree of uncertainty or variability.

Alternative hypothesis: $H_A:|k|=1$, indicating that the causal relationship is absolute, deterministic, and perfectly certain.

Proof. It is: $\big((a+b+c+d)\cdot a\big)-\big((a+b)\cdot(a+c)\big)=(a\cdot d)-(b\cdot c)$. The empirical causal relationship k_n is calculated from data before as:

$$k_n=\frac{(43448\cdot 8)-(21720\cdot 170)}{\sqrt{170\cdot 43278\cdot 21728\cdot 21720}}=-0.05676418323. \tag{114}$$

The theoretical probability per trial, reduced by an error rate, is obtained from k_n as:

$$\begin{aligned} p^* &= |k_n|^{1/n} \\ &= 0.05676418323^{\frac{1}{43448}} \\ &= e^{\frac{1}{43448}\cdot\ln(0.05676418323)} \\ &= 0.9999339727. \end{aligned} \tag{115}$$

Let k_{events} denote the number of events observed in the sample. The number of events observed in the sample is given as:

$$\begin{aligned} k_{\text{events}*} &= n\cdot p^* \\ &= 43448\cdot 0.9999339727 \\ &= 43445. \end{aligned} \tag{116}$$

The right-tailed *p*-value of the empirical causal relationship is $|k_n|$ is:

$$\begin{aligned} p_{\text{right}} &= \frac{n-k_{\text{events}*}+1}{n} \\ &= \frac{43448-43445+1}{43448} \\ &= 9.20640766\times 10^{-5} \\ &= 0.0000920640766. \end{aligned} \tag{117}$$

Decision. Since $p_{\text{right}}=0.0000920640766<0.05$, we reject the null hypothesis $H_0:|k|<1$ at the $\alpha=0.05$ significance level. There is a statistically significant causal relationship between BNT162b2 vaccination and the reduction of COVID-19 cases. However such a rejection of $H_0:|k|<1$ should not be interpreted as an absolute proof that the causal relationship between BNT162b2 vaccine and the prevention of a Covid-19 infection is $|k|=1$. Instead, it represents a conclusion drawn within the reference framework of Hegel's distribution, based on the data available and holding at the specified significance level *α*. That is to say, were the null state of non-causality (H_0) actually the case, an outcome this extreme would occur in at most a proportion *α* of instances. Under this controlled error rate, the data are judged to be consistent with $|k|=1$. The observed empirical causal relationship is $|k_n|=0.05676418323$, is less than 1 (*i.e.* $p^*=0.9999339727$), while highly statistically significant ($p_{\text{right}}=9.20640766\times 10^{-5}$). This suggests that the vaccine provides strong protection, but does not guarantee an absolute and complete immunity for all individuals. The deviation from unity may be due to factors such as random variation, individual immune response, or errors in vaccine pro-

duction or errors in vaccine storage or errors in vaccine administration et cetera. Nonetheless, these findings support the continued use of the vaccine, while emphasizing the need for quality assurance and further research into the determinants of vaccine effectiveness.

Quod erat demonstrandum.

Figure 4 illustrates the relationship between sample size and k.

Exclusion relationship. Using real observed data—such as the 8 COVID-19 cases in the investigated group out of 43,448—the exclusion condition relationship with $k_{\text{events}} = b + c + d = 21712 + 162 + 21566 = 43440$ can be tested in detail within the same framework. The null and alternative hypotheses are formulated as:

$$H_0 : p\left(\text{BNT162b2 vaccination excludes COVID-19 infection}\right) < 1,$$
$$H_A : p\left(\text{BNT162b2 vaccination excludes COVID-19 infection}\right) = 1.$$

The right-tailed p-value for the exclusion relationship is then given by:

$$p_{\text{right}} = \frac{n - k_{\text{events}} + 1}{n} = \frac{43448 - 43440 + 1}{n} = \frac{8+1}{43448} = \frac{9}{43448} \approx 0.000207. \quad (118)$$

This implies that, assuming the exclusion condition holds with probability $p = 1$, the probability of observing 8 or fewer events in a sample of 43,448 is approximately 0.0207%. In other words, we would be wrong in only about 0.0207% of cases if we were to assume that the exclusion condition holds. This strongly supports the conclusion that the exclusion relationship is almost certainly valid, and we may reject the null hypothesis $H_0 : p < 1$ in favor of $H_A : p = 1$. Importantly, this calculation pertains to the exclusion condition, not to the causal relationship itself. It is almost certain that **a BNT162b2 vaccination excludes a COVID-19 infection**, corresponding to a right-tailed p-value of p_{right} = 0.000207144172 (approximately 0.0207%).

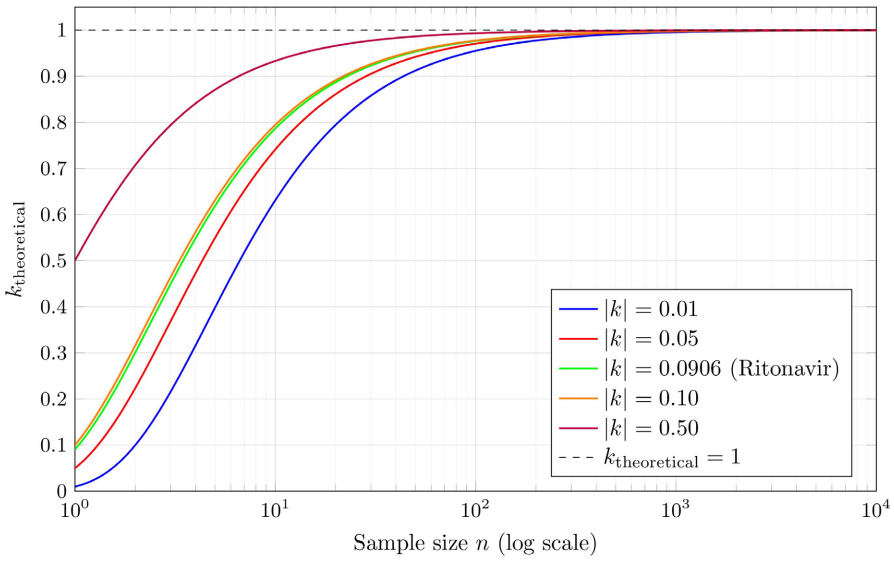

Figure 4. Comparison with a binomial sampling model.

The right-tailed p-value of Hegel's distribution, $p_{\text{right}} = 9.20640766 \times 10^{-5}$, evaluates the observed count $k_{\text{events}} = 43{,}445$ (out of $n = 43{,}448$) against the reference distribution of complete ignorance about p (Definition 5). A complementary assessment can be obtained directly from the Binomial(n, p) sampling model implied by Definitions 1 - 4. The exact one-sided (Clopper-Pearson) lower confidence bound for p at the 95% confidence level is $p_L = 0.99982$; that is, based directly on the binomial sampling distribution, we may state with 95% confidence that the underlying probability of the exclusion relationship holding exceeds 0.9998. For illustration, the exact probability of observing $k_{\text{events}} = 43{,}445$ or more is highly sensitive to the assumed value of p under the binomial model:
$P\left(K \geq 43445 \mid p = 0.999\right) \approx 0$, $P\left(K \geq 43445 \mid p = 0.9999\right) \approx 36.9\%$,
$P\left(K \geq 43445 \mid p = 0.99993\right) \approx 63.8\%$. This illustrates that, unlike the ignorance-based p-value of Hegel's distribution, a binomial-model assessment requires either a specific hypothesized p or an estimated confidence bound; the Clopper-Pearson bound above provides the latter without requiring such a specific assumption, and corroborates the conclusion of near-certainty reached above.

It should be noted that computing the exact Clopper-Pearson bound for such a large sample size ($n = 43{,}448$) requires either substantial computational effort or the use of approximation methods, each of which carries its own associated error. In point of fact, exact Clopper-Pearson limits are, as a rule, numerically precarious. At $p = 1$, they rest upon the inversion of the Beta distribution at the boundary of a degenerate binomial model, which in turn requires high-precision computation or approximations carrying their own error. This circumstance illustrates a substantive weakness inherent to the binomial distribution—one that Hegel's distribution is specifically constructed to circumvent. Furthermore, when $p = 1$, the binomial distribution degenerates into a deterministic point distribution, since the probability of success is absolutely certain for all n trials. The variance $np\left(1-p\right)$ becomes exactly zero, so that all dispersion around the expected value n vanishes and the distribution loses its characteristic bell shape. The probability mass concentrates entirely on the single possible outcome $k = n$, while every other outcome receives probability zero. The classical binomial model thus collapses completely as a statistical tool for modeling uncertainty, since it can no longer convey any information about alternative sources of error or unforeseen deviations. It is precisely at this boundary of perfect certainty that Hegel's distribution takes over, restoring the missing uncertainty through a maximum-entropy reference distribution. It is worth noting that this sensitivity is confined to the binomial-model comparison: because p_{right} is linear in k_{events}, with $p_{\text{right}} = (n - k_{\text{events}} + 1)/n$, the conclusion drawn from Hegel's distribution remains robust against exactly this kind of modeling uncertainty. At $n = 43{,}448$, k_{events} would have to fall by more than 2000 units—from 43,445 to roughly 41,276—before p_{right} would cross the $\alpha = 0.05$ threshold. Hegel's distribution is therefore insensitive to the precise functional form assumed for the attenuation mechanism in Proposition 2, a **robustness** the binomial-model comparison above does not share.

3.11. Exact Type I Error Control and Unit Power

Finally, it is necessary to establish by formal proof that the procedures introduced in this publication permits exact, non-asymptotic Type I error control and has unit power.

Theorem 3.11 (*Exact Type I Error Control and Unit Power*). Let X denote the count variable of Definition 4, and consider the test that rejects H_0 whenever

$$p_{\text{right}} = \frac{n - X + 1}{n} \leq \alpha, \tag{119}$$

for a fixed significance level $\alpha \in (0,1]$. Then:

Under H_0 (Definition 6: $X \sim \text{Uniform}\{1,\cdots,n\}$),

$$P(\text{reject } H_0 \mid H_0) = \frac{\lfloor n\alpha \rfloor}{n} \leq \alpha, \tag{120}$$

so the test controls the Type I error rate exactly at or below the nominal level α for every finite n.

Under H_A ($X = n$ with probability 1), $p_{\text{right}} = 1/n$ deterministically, so the test has power

$$P(\text{reject } H_0 \mid H_A) = 1 \text{ whenever } n \geq 1/\alpha, \tag{121}$$

and power 0 otherwise.

Proof. 1) $p_{\text{right}} = (n - X + 1)/n$ is a strictly decreasing bijection of X on $\{1,\cdots,n\}$, mapping $X = n$ to $p_{\text{right}} = 1/n$ and $X = 1$ to $p_{\text{right}} = 1$. Since $X \sim \text{Uniform}\{1,\cdots,n\}$ under H_0, p_{right} is itself uniformly distributed on $\{1/n, 2/n, \cdots, n/n\}$. Hence

$$P(p_{\text{right}} \leq \alpha) = P(p_{\text{right}} \in \{1/n,\cdots,\lfloor n\alpha \rfloor/n\}) = \frac{\lfloor n\alpha \rfloor}{n} \leq \alpha, \tag{122}$$

with equality whenever $n\alpha$ is an integer. This holds exactly, for every finite n not merely asymptotically precisely because H_0 is a single, fully specified distribution (Definition 6), not a composite family indexed by an unknown parameter p.

2) Under H_A, $X = n$ with probability 1 (Definition 5, Equation 13), so $p_{\text{right}} = (n - n + 1)/n = 1/n$ with probability 1. The rejection condition $p_{\text{right}} \leq \alpha$ therefore holds with probability 1 whenever $1/n \leq \alpha$, *i.e.*, $n \geq 1/\alpha$, giving power 1; for $n < 1/\alpha$, $p_{\text{right}} = 1/n > \alpha$ with probability 1, giving power 0.

Quod erat demonstrandum.

This exact control is a property of the reference model itself: the theorem establishes that the test is internally consistent under the assumption $X \sim \text{Uniform}\{1,\cdots,n\}$, while the broader epistemic question of whether this assumption adequately represents complete ignorance about $p < 1$ is addressed separately above (see the comparison with the binomial sampling model) and in the Discussion below.

4. Discussion

A particularly important consideration concerns the relativity of confounders

(third variables) for the relationship measures introduced in this work. Confounders—such as a participant's age, wealth, health behavior, sensitivity of the laboratory kit, access to testing et cetera—can distort the observed statistical relationship between two variables by influencing both the putative risk factor and the outcome; the extent to which this occurs depends, not least, on the statistical model employed.

The relevance of confounders, however, is not absolute but relative, and depends on a variety of factors. Consider, for example, a claim concerning a necessary condition, such as "no human life without gaseous oxygen." This relationship holds for every individual trial—regardless of whether the individuals observed are wealthy, healthy, vaccinated, or differ in any other respect (*i.e.*, with respect to confounders, or third variables).

In conducted studies, however, apparent violations of even a genuine necessary condition can be observed. Such violations do not arise from confounding, but typically from deficiencies in measurement or study design. For instance, a study examining the absence of oxygen for only one minute might erroneously conclude that oxygen is not a necessary condition for life, while another study observing the absence of gaseous oxygen over several hours correctly confirms the necessity — such a discrepancy reflects an inadequate observation window, not confounding by a third factor.

A fundamental distinction therefore becomes necessary: that between objective reality (the ontological level—the world as it actually is, independent of any observer, model, or dataset) and our representation of that reality (the epistemological level—what we measure, infer, and record as data). A law of nature cannot itself be "in error"; it simply is what it is. Errors arise, rather, predominantly in the process by which we map objective reality onto data through observation, measurement, and inference—that is, at the epistemological level.

When studies of the same matter arrive at contradictory conclusions, it must be examined where the error is located: in the mapping—an inadequate observation window, a flawed measurement, or a mistaken inference—in the mathematical methods employed, or elsewhere. The framework developed in this work is particularly suited to testing claims about necessary, sufficient, and exclusion conditions (Definition 2), as well as other relationship measures with p = 1, and to determining whether the collected data are consistent with such a relationship.

Regrettably, methods known and used today—such as the Clopper-Pearson bound illustrated above—do not guarantee freedom from systematic error. Hegel's distribution provides a sharp measure of statistical incompatibility with ignorance, but, owing to its methodological nature, it cannot by itself prove epidemiological causal certainty. A consequence should therefore be noted in this regard: a statistically established significance—what p_{right} measures, namely incompatibility with the reference distribution—is not automatically to be identified with substantive, epidemiological certainty, which requires external validation. Statistical significance should be regarded as provisional until a significant rela-

tionship can be confirmed through further external validation.

In this regard, reference should also be made to other attempts to develop methods of causal inference, such as directed acyclic graphs (DAGs), propensity-score matching, or instrumental variables. The relationship between such methods and Hegel's distribution, however, is not the subject of the present investigation and may be reserved for further work.

Thus far, even events that are deemed certain in theory, *i.e.*, those with $p = 1$, may be subject to uncertainty or doubt in practice. While the Binomial, Poisson, and other distributions offer one approach to quantifying this uncertainty, the present work addresses this problem from the perspective of Hegel's distribution. As we have demonstrated, Hegel's distribution is particularly well-suited for analyzing almost sure events, as it directly captures the deviation from certainty in terms of sample size n and the observed number of successes k_{events}.

Example 1. Suppose a production process is assumed not to be error-free, *i.e.*, $p < 1$. In a random sample of $n = 30$ items, no defective items are discovered, meaning that $k_{\text{events}} = 30$ non-defective items were observed. We formulate the hypotheses: H_0: $p < 1$ (the process is not error-free, *i.e.*, defects exist), and H_A: $p = 1$ (the process is error-free). Under Hegel's distribution, the right-tailed p-value is: $p_{\text{right}} = (n - k_{\text{events}} + 1)/n = (30 - 30 + 1)/30 = 1/30 \approx 0.0333$ (3.33%). This result implies that, if the process were truly not error-free ($p < 1$), the probability of observing 30 or more non-defective items in a sample of 30 is only 3.33%. More generally, if $k_{\text{events}} = n$, the right-tailed p-value simplifies to: $p_{\text{right}} = (n - n + 1)/n = 1/n$. Thus, as the sample size increases, the p-value decreases correspondingly: for $n = 30$, $p_{\text{right}} \approx 3.33\%$; for $n = 50$, $p_{\text{right}} = 2\%$; and for $n = 100$, $p_{\text{right}} = 1\%$. This demonstrates that larger samples provide stronger evidence against the null hypothesis when no defects are observed. Since the p-value is small ($p < 0.05$), we reject the null hypothesis H_0: $p < 1$ in favor of the alternative H_A: $p = 1$. In other words, the observed data are statistically consistent with the hypothesis that the production process is error-free, and the absence of defects is best explained by the process operating as intended, rather than by systematic error.

Comparison with a binomial sampling model. Hegel's distribution yields $p_{\text{right}} = 1/30 \approx 3.33\%$ for the observation of $k_{\text{events}} = 30$ non-defective items out of $n = 30$. Under the Binomial(n,p) sampling model implied by the stated trial process, the exact one-sided (Clopper-Pearson) 95% lower confidence bound for the true non-defect probability p, given zero observed defects in 30 trials, is $p_L \approx 0.9050$. That is, the data are statistically compatible, at the 95% confidence level, with a true defect rate of up to roughly 9.5%—a materially weaker statement than the small p-value of Hegel's distribution alone might suggest, reflecting the limited information available from a sample of only $n = 30$ trials. This comparison illustrates the complementary roles of the two approaches: Hegel's distribution quantifies how extreme the observed count is relative to complete ignorance of p, while the binomial-derived bound quantifies the range of p values statistically compatible with the data at a stated confidence level.

Example 2. The same line of reasoning is equally applicable to human medicine. In an illustrative application, we reanalyzed the relationship between the BNT162b2 vaccination and COVID-19 infection. The simple computation of the right-tailed p-value via Hegel's distribution yielded a highly significant result, confirming the well-established efficacy of the vaccine. Specifically, the observed causal relationship $|k| = 0.05676418323$ led to a theoretical probability per trial of $p^* = 0.9999339727$, resulting in an expected number of cause-effect events $k_{\text{events}} = 43445$ and a right-tailed p-value of $p_{\text{right}} = 0.0000920640766$. These findings are consistent with the original trial results and demonstrate the practical utility of the proposed framework.

Example 3 (Illustrative, hypothetical). Another example from astrophysics concerns the search for black holes in galaxies. The underlying assumption is that not every galaxy contains a black hole, *i.e.*, $p < 1$. To illustrate the framework, the following purely theoretical data are considered: for $n = k_{\text{events}} = 10$, the right-tailed p-value is 0.1; for $n = k_{\text{events}} = 20$, it is 0.05 and for $n = k_{\text{events}} = 40$, it is 0.025.

More generally, the analysis illustrates how formal hypothesis testing resolves the ambiguity between random variation and structural deviation from a process specification. Without such a test, one might be tempted to accept failures as inherent evidence against the alternative hypothesis. However, the extremely small p-value obtained here shows that such an explanation is statistically untenable: the data are entirely inconsistent with the null hypothesis and sharply consistent with the alternative. This statistical framework thereby has the potential to transform intuitive doubt into rigorous inference, compelling a reassessment of the operational assumptions governing a given relationship.

With Hegel's distribution, we now have a simple and reliable method to test the significance of causal relationships, sufficient conditions, necessary conditions, and other almost sure events. The distribution provides a unified approach that is both computationally straightforward and conceptually transparent, making it a valuable tool for quality control, diagnostic testing, and reliability engineering. Future work may extend this framework to more complex causal structures and further explore its relationship to other probabilistic models.

5. Conclusions

This inquiry has compellingly revealed that a deviation from unity is not a failure of the principle, but a guide to that boundary at which certainty might pass over into doubt—a boundary that encloses our knowledge, sets its limits, and yet gives it its depth.

In this context, the application of a right-tailed p-value offers a robust and intuitive framework for assessing the credibility of sure events when confronted with contradictory empirical evidence. By testing H_0: $p < 1$, researchers may transcend mere anecdotal observation and determine, with objective clarity, whether an observed frequency aligns with the alternative hypothesis H_A: $p = 1$, despite the natural imperfection inherent in all worldly phenomena.

This echoes Confucius's famous dictum: to know that one does not know is itself a form of knowing. In this sense, the right-tailed *p*-value does not pretend to abolish doubt; it renders doubt measurable, turning the boundary between certainty and uncertainty into the very condition under which knowledge becomes reflexive.

Thus, this apparent divergence from unity may be understood as the natural echo of the tension between identity and difference—a gap that wisdom does not mourn, but silently takes into itself as a necessary contradiction inherent in the very structure of objective reality.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Arbuthnot, J. (1710) An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. *Philosophical Transactions of the Royal Society of London*, **27**, 186-190. https://doi.org/10.1098/rstl.1710.0011

[2] Pearson, K. (1900) On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **50**, 157-175. https://doi.org/10.1080/14786440009463897

[3] Hume, D. (1739) A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects. Volume, 1, John Noon.

[4] Uspensky, J.V. (1937) Introduction to Mathematical Probability. McGraw-Hill Book Company.

[5] Cardanus, H. (1663) Liber de ludo aleae. Ioannes Antonius Huguetan & Marcus Antonius Ravaud.

[6] Pascal, B. (1665) Traité du triangle arithmétique, avec quelques autres petits traitez sur la mesme matière. https://zenodo.org/doi/10.5281/zenodo.21299271

[7] Moivre, A. (1756) De: The Doctrine of Chances: Or, a Method of Calculating the Probabilities of Events in Play. Printed for A. Millar.

[8] Bayes, T. (1763) LII. An Essay Towards Solving a Problem in the Doctrine of Chances. by the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, No. 53, 370-418. https://doi.org/10.1098/rstl.1763.0053

[9] Laplace, P.S. (1812) Théorie analytique des probabilités. V. Courcier.

[10] Kolmogorov, A.N. (1933) Foundations of the Theory of Probability. Translation Edited 1950 by Nathan Morrison. https://zenodo.org/doi/10.5281/zenodo.7883088

[11] Barukčić, I. (2016) The Mathematical Formula of the Causal Relationship K. *International Journal of Applied Physics and Mathematics*, **6**, 45-65. https://doi.org/10.17706/ijapm.2016.6.2.45-65

[12] Barukčić, I. (2021) The Causal Relationship K. *MATEC Web of Conferences*, **336**, Article 09032. https://doi.org/10.1051/matecconf/202133609032

[13] Barukčić, I. (2025) Theoriae Causalitatis Principia Mathematica. 3rd Edition, Books

on Demand.

[14] Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[15] Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., *et al.* (2020) Safety and Efficacy of the BNT162b2 mRNA COVID-19 Vaccine. *New England Journal of Medicine*, **383**, 2603-2615. https://doi.org/10.1056/nejmoa2034577