How do I entropy thee? Let me count the ways!

Measuring Statistical Bias in Data Using Entropy

June 22, 2026

About the Academy

Mission:
To serve the public and the U.S. actuarial profession

Community:
Serving over 20K MAAAs & public stakeholders for 60 years

Standards:
Setting qualification, practice, and professionalism standards

Impact:
Delivering over 300 insight-driven publications & resources annually

Visit www.actuary.org to learn more.

Information About This Presentation

- The presenters' statements and opinions are their own and do not necessarily represent the official statements or opinions of the Actuarial Board for Counseling and Discipline (ABCD), Actuarial Standards Board (ASB), any boards or committees of the American Academy of Actuaries, or any other actuarial organization, nor do they necessarily express the opinions of their employers.
- The Academy operates in compliance with the requirements of applicable law, including federal antitrust laws. The Academy's antitrust policy is available online at https://www.actuary.org/content/academy-antitrust-policy.
- Academy members and other individuals who serve as members or interested parties of any of its boards, councils, committees, etc., are required to annually acknowledge the Academy's Conflict of Interest Policy, available online at https://www.actuary.org/content/conflict-interest-policy-1.
- The Data Science and Analytics Committee (DSAC) has undergone a change and is now the Artificial Intelligence, Data Science, and Analytics Committee (ADAC)

Speakers

Moderator

David Sandberg, MAAA FSA

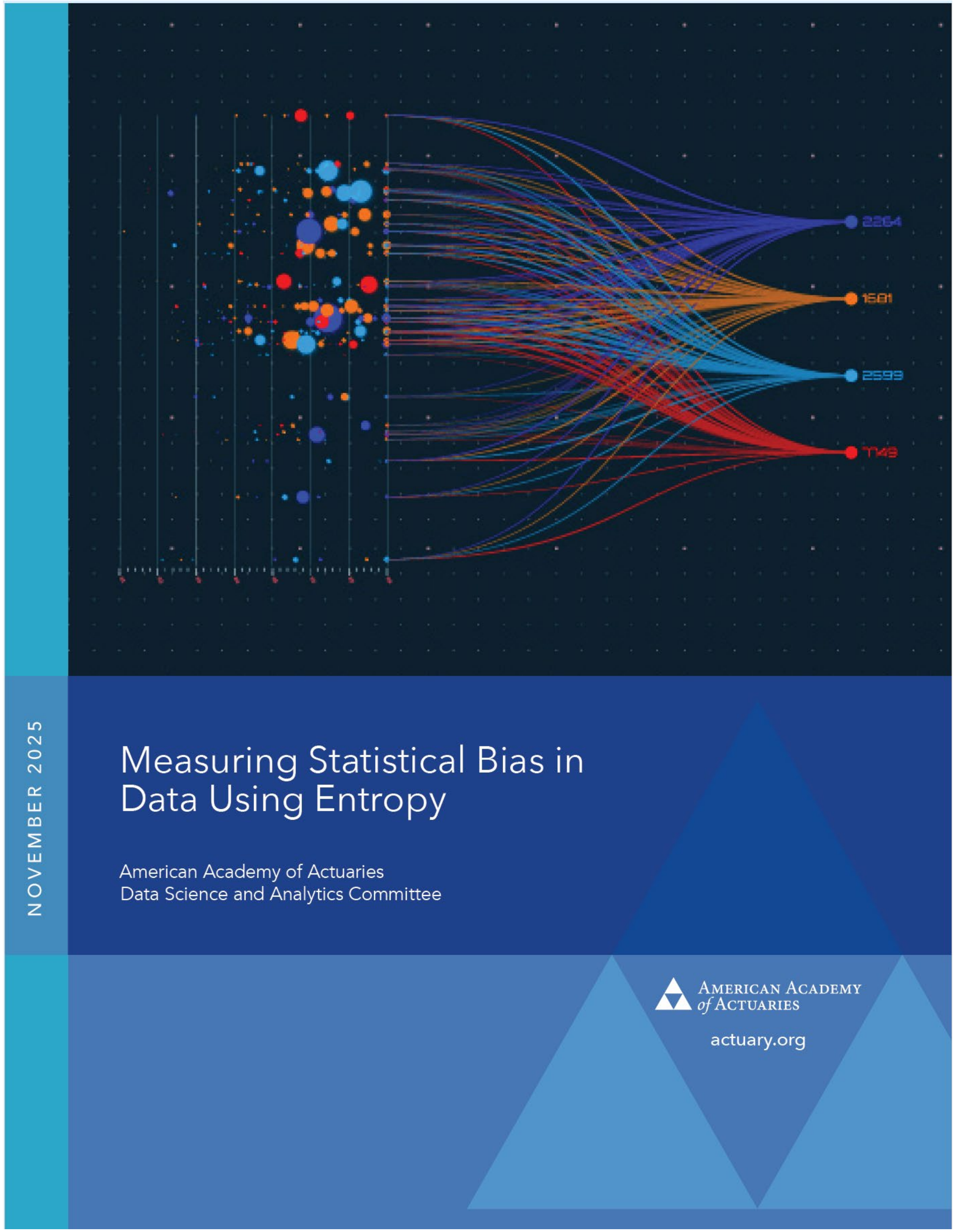
Presenters

Dorothy L. Andrews, MAAA, PhD, ASA, CSPA

Paul Miexler, MAAA, EA FCA

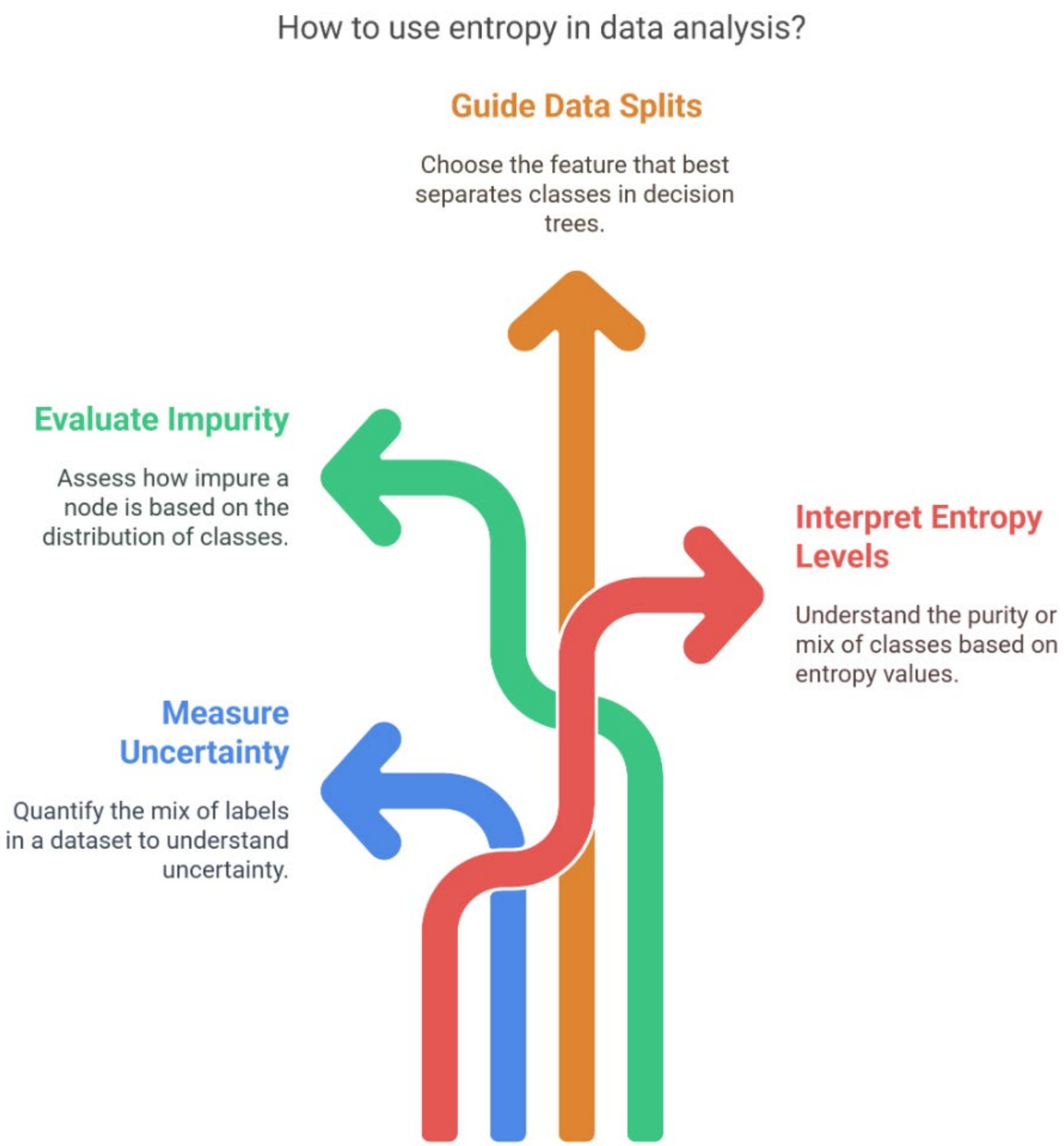
Alexander Esche, MAAA, ASA

Measuring Statistical Bias in Data Using Entropy



Agenda

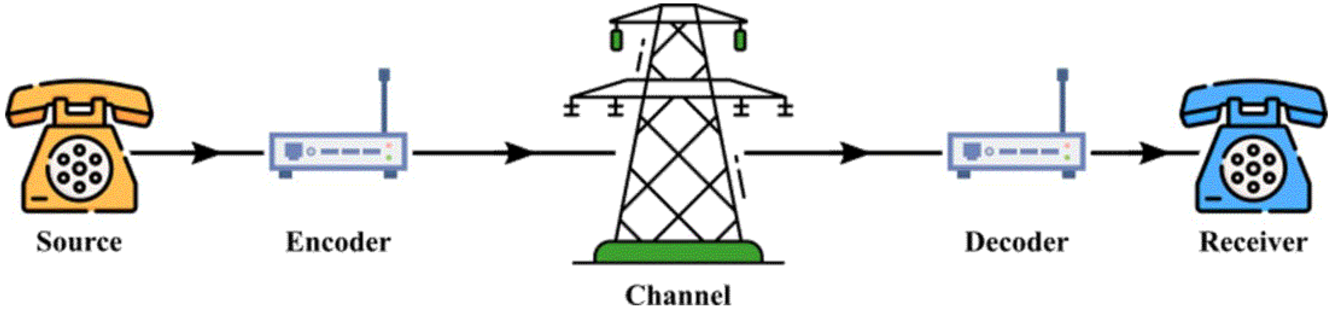
- Foundations
- Entropy Defined
- Entropy in Cardiac Research
- Semantic Entropy
- Approximate Entropy
- Runner’s Entropy
- Entropy in Machine Learning
- Entropy Balancing Study
- Entropy in Thermodynamics
- Entropy & Statistics
- Maximum Entropy Classifier Model
- Life Table Entropy
- Bias vs. Diversity
- Conclusion
- Question & Answer



Information Theory
The Foundational Basis of Entropy

Entropy Measure to Detect Bias in Data

The Transmission of Information.



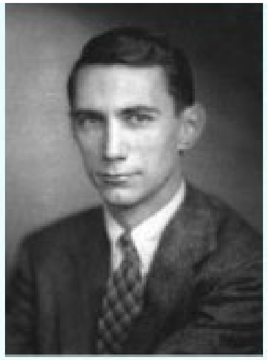
Binary Digit or Bit: Smallest unit to encode a “0” or “1”
Claude Shannon is credited with coining the term “bit.”

Fathers of Information Theory

Ralph V.L. Hartley
1888 – 1970
Bell Labs

Harry Nyquist
1889 – 1976
Bell Labs

Norbert Wiener
1894 – 1964
MIT



Claude Shannon
1916 – 2001
Bell Labs, MIT

Information Theory

Four Papers Establishing the Field of Information Theory

where K_0 is the same for all systems. Since K_0 is arbitrary, we may omit it if we make the logarithmic base arbitrary. The particular base selected fixes the size of the unit of information. Putting this value of K in (4),

$$H = n \log s \tag{9}$$

$$= \log s^n. \tag{10}$$

What we have done then is to take as our practical measure of information the logarithm of the number of possible symbol sequences.

Source: Hartly, *Transmission of information (1928)*

Ralph V.L. Hartley
1888 – 1970
Bell Labs

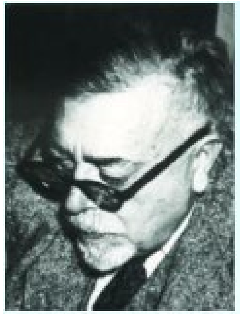
Harry Nyquist
1889 – 1976
Bell Labs

Hartly, *Transmission of information (1928)*

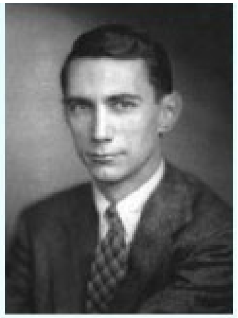
Nyquist, *Certain Factors Affecting Telegraph Speed* (1924)

Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine (1948)*

Shannon, *A Mathematical Theory of Communication (1948)*



Norbert Wiener
1894 – 1964
MIT



Claude Shannon
1916 – 2001
Bell Labs, MIT

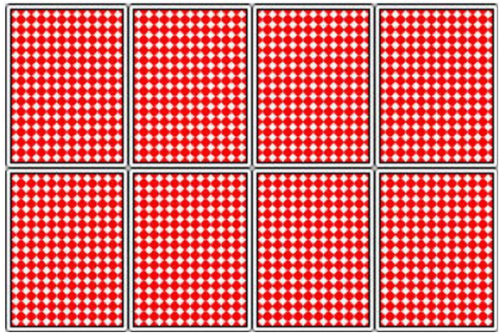
Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

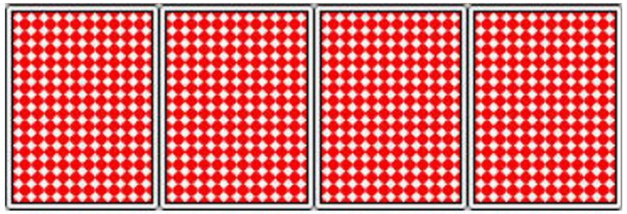
where K is a positive constant.

Yes or No Encoding Questions

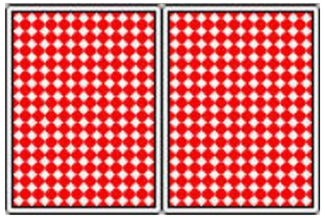
Minimum Number of Questions to Find the King



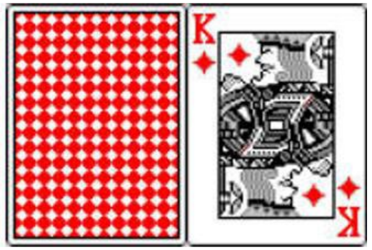
Is the King in the top row? If the answer is "No," then it must be in the bottom row.



Is the King one of the two right most cards? If the answer is "No," then it must be one of the two left most cards.



Is the King one of the two right most cards? If the answer is "No," then it must be one of the two left most cards.



We started out with eight cards and asked three "yes" or "no" questions before determining where the King was with certainty.

Eight Possibilities – Three Yes/No Questions =>

$$8 = 2^3 \text{ or } \log_2(8) = \log_2(2^3) \text{ or } \log_2(8) = 3 = \text{Information } H^{artley}$$

AMERICAN ACADEMY of ACTUARIES

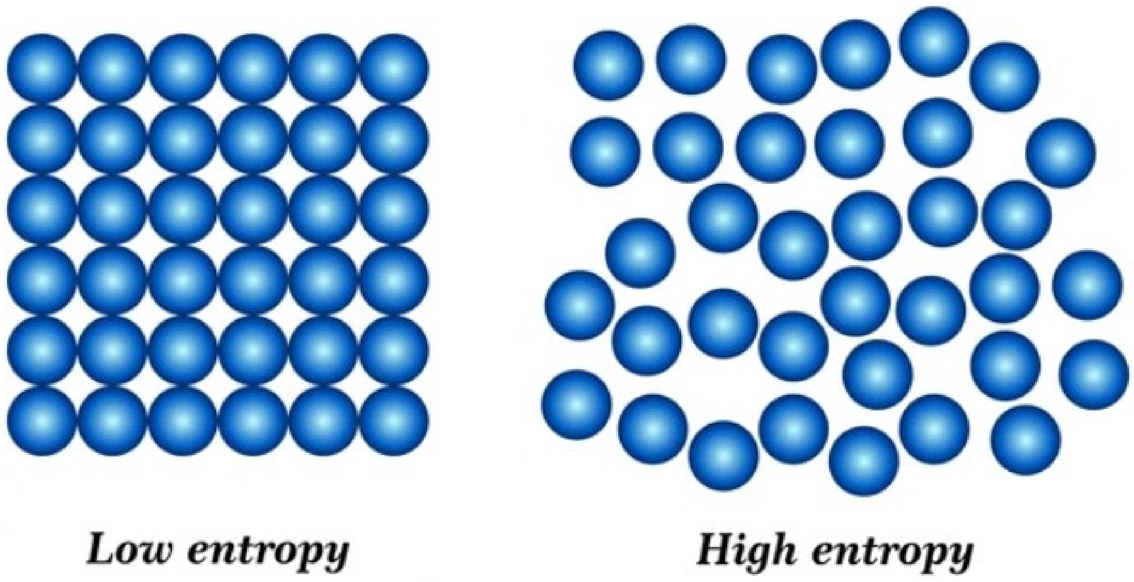
Entropy Defined
The Foundational Basis of Entropy

What is Entropy?
A measure of disorder or uncertainty.

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

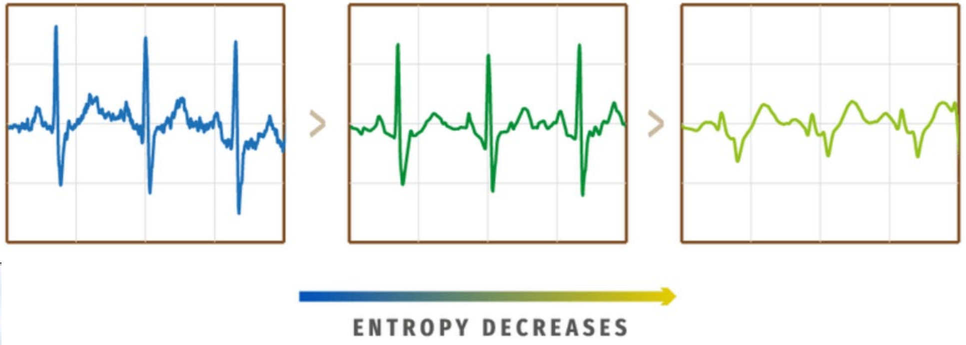
p_i → probability of class i
C → Number of Classes
– $\log_2(p_i)$ → Number of Bits
Entropy = 0 → Order/Certainty
Entropy > 0 → Disorder/Uncertainty

Order vs. Disorder



Entropy in Cardiac Research

Let's take math to heart.



Source: https://scienceispeople.mug.edu.pl/entropy-in-cardiovascular-research/

Cardiovascular diseases are a significant social problem. The main tool used to diagnose them is electrocardiography – a non-invasive test, the recording of which provides doctors and scientists with information about the bioelectric activity of the heart.

"A healthy heart dances, and a seriously ill one marches steadily."

– Dr. Ary Louis Goldberger

American Academy of Actuaries

What is Semantic Entropy?

Semantic entropy measures the uncertainty in the meaning of words and sentences within a given text, rather than focusing on the probability of word occurrences. It evaluates how well the words and phrases fit together to convey coherent meaning.

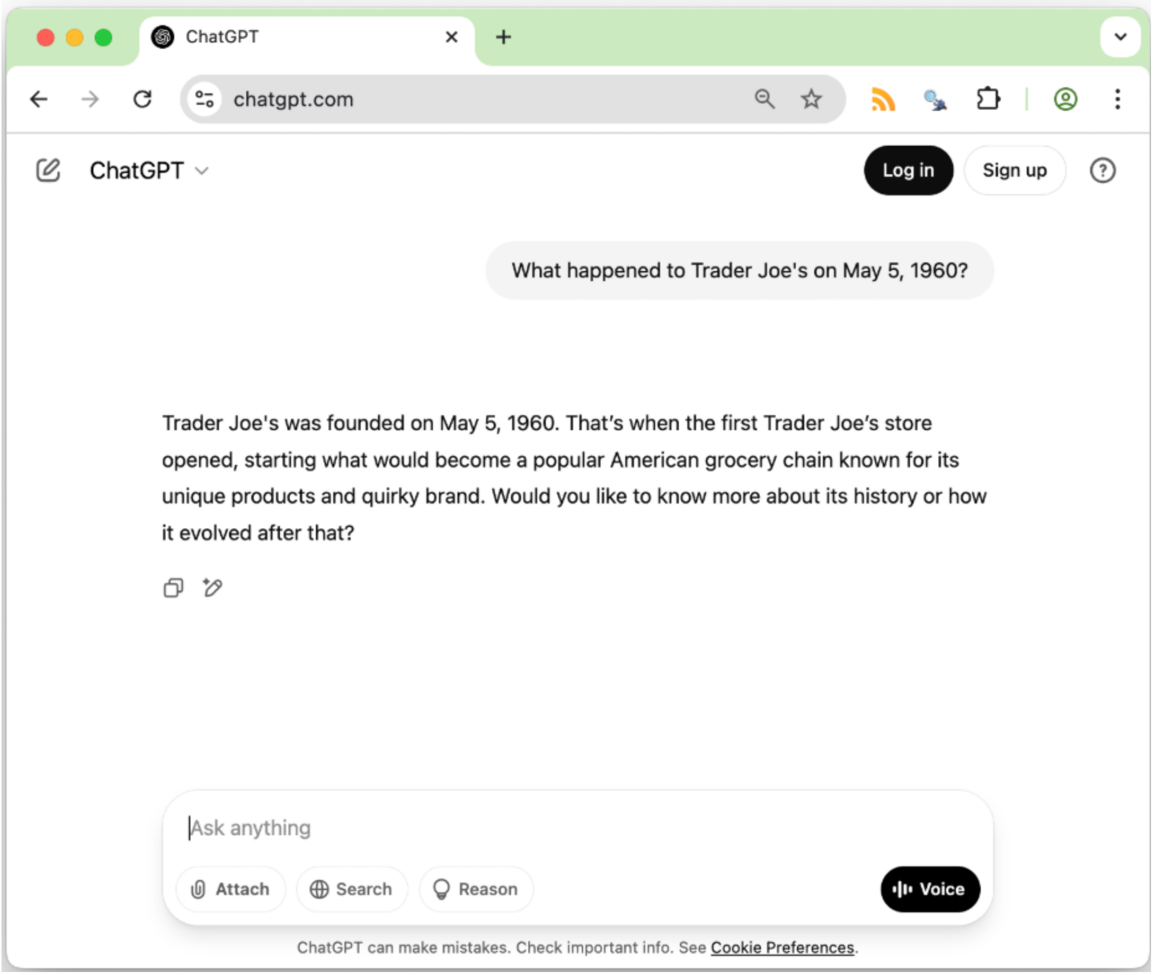
High semantic entropy indicates a lack of coherence and greater uncertainty, suggesting the text might be unreliable or incorrect. This approach helps in identifying AI-generated hallucinations by flagging text that deviates from expected semantic coherence, ensuring more accurate and meaningful outputs.

Source: https://www.nature.com/articles/s41586-024-07421-0

Semantic Entropy – Hallucinations vs. Confabulations

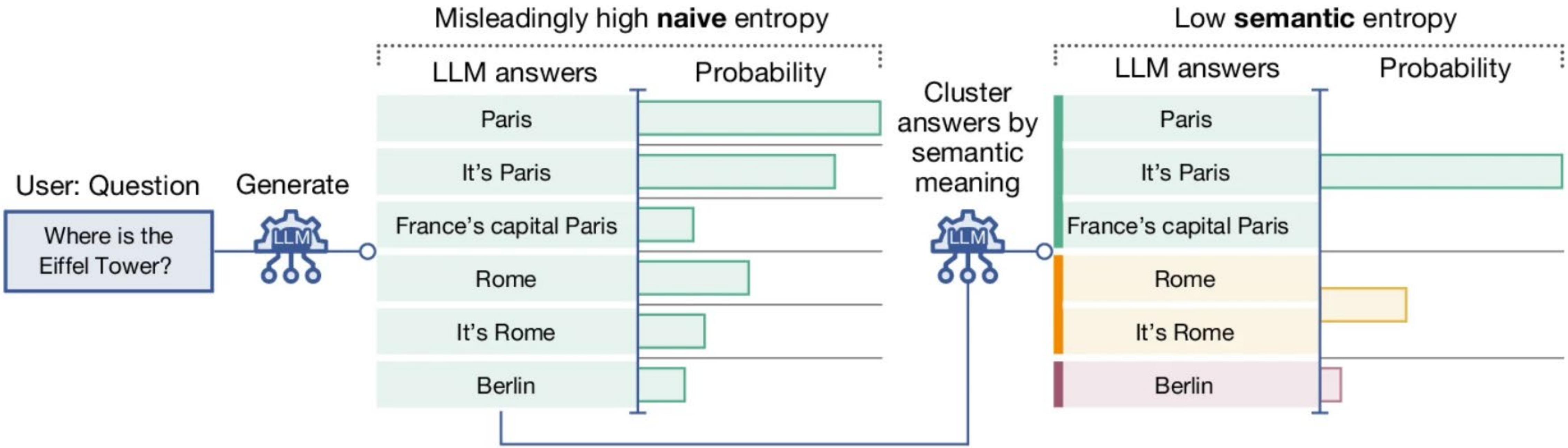
According to the ICD-11, the eleventh edition of the International Classification of Diseases and the global standard for documenting health information, **hallucinations** are described as “sensory perceptions of any modality occurring in the absence of the appropriate (external) stimulus. The person may or may not have insight into the unreal nature of the perception.”

Confabulation, by contrast, is defined as “the filling of memory gaps with fabricated, distorted, or misinterpreted memories about oneself or the world, without the conscious intention to deceive.”



Source: https://cosimameyer.com/post/are-ai-models-hallucinating-or-just-confabulating/

Semantic Entropy – Helps Detect Confabulations



High semantic entropy = likely confabulation

Low semantic entropy = likely not confabulation

Source: https://www.nature.com/articles/s41586-024-07421-0

Approximate Entropy
Identifying Patterns in Time Series

$$\mathrm{ApEn}(m, r, N)(u) = \phi^m(r) - \phi^{m+1}(r)$$

Example

$$S_N = \{85, 80, 89, 85, 80, 89, \ldots\}$$

Form a sequence of vectors:

$$\begin{aligned}\mathbf{x}(1) &= [u(1)\ u(2)] = [85\ 80]\\ \mathbf{x}(2) &= [u(2)\ u(3)] = [80\ 89]\\ \mathbf{x}(3) &= [u(3)\ u(4)] = [89\ 85]\\ \mathbf{x}(4) &= [u(4)\ u(5)] = [85\ 80]\\ &\vdots\end{aligned}$$

Calculate Distances

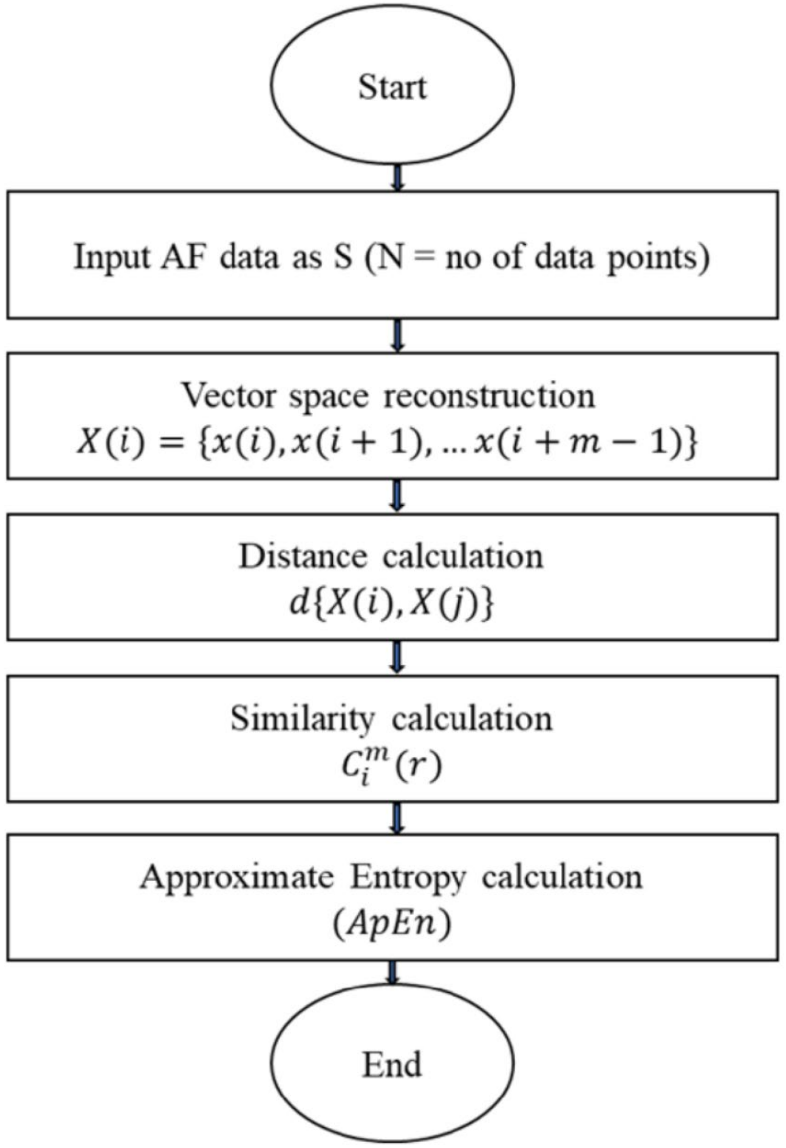
$$d[\mathbf{x}(i), \mathbf{x}(j)] = \max_k \left(|\mathbf{x}(i)_k - \mathbf{x}(j)_k|\right)$$

Define a count C_i^m as

$$C_i^m(r) = \frac{(\text{number of } j \text{ such that } d[\mathbf{x}(i), \mathbf{x}(j)] \leq r)}{n}$$

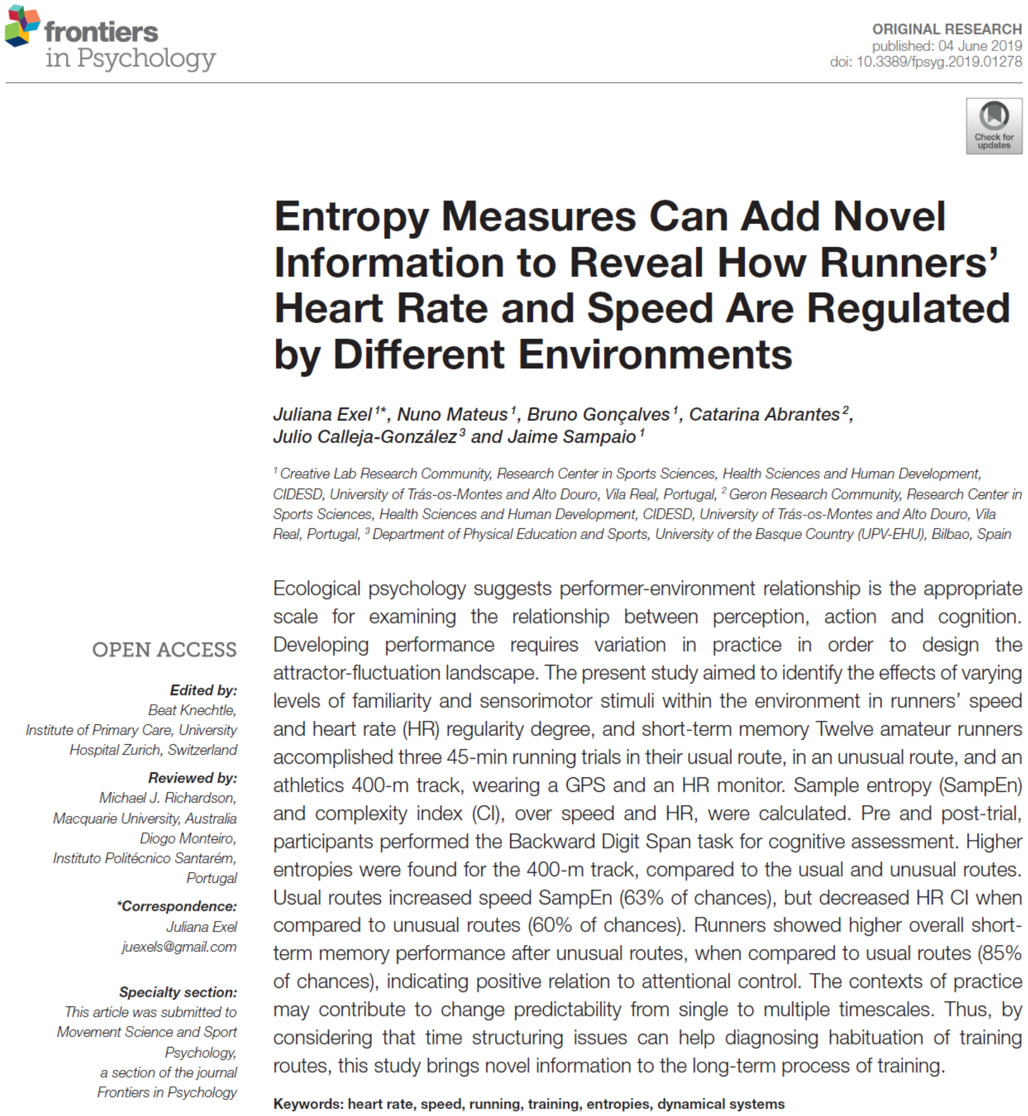
Define

$$\phi^m(r) = \frac{1}{n}\sum_{i=1}^{n} \log(C_i^m(r))$$



Source: https://en.wikipedia.org/wiki/Approximate_entropy

Runner's Entropy
Terrain Impacts on Heart Health



frontiers in Psychology
ORIGINAL RESEARCH
published: 04 June 2019
doi: 10.3389/fpsyg.2019.01278

Check for updates

Entropy Measures Can Add Novel Information to Reveal How Runners' Heart Rate and Speed Are Regulated by Different Environments

Juliana Exel[1], Nuno Mateus[1], Bruno Gonçalves[1], Catarina Abrantes[2], Julio Calleja-González[3] and Jaime Sampaio[1]*

[1] Creative Lab Research Community, Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, [2] Geron Research Community, Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, [3] Department of Physical Education and Sports, University of the Basque Country (UPV-EHU), Bilbao, Spain

OPEN ACCESS

Edited by:
Beat Knechtle,
Institute of Primary Care, University Hospital Zurich, Switzerland

Reviewed by:
Michael J. Richardson,
Macquarie University, Australia
Diogo Monteiro,
Instituto Politécnico Santarém, Portugal

****Correspondence:***
Juliana Exel
juexels@gmail.com

Specialty section:
This article was submitted to Movement Science and Sport Psychology, a section of the journal Frontiers in Psychology

Ecological psychology suggests performer-environment relationship is the appropriate scale for examining the relationship between perception, action and cognition. Developing performance requires variation in practice in order to design the attractor-fluctuation landscape. The present study aimed to identify the effects of varying levels of familiarity and sensorimotor stimuli within the environment in runners' speed and heart rate (HR) regularity degree, and short-term memory Twelve amateur runners accomplished three 45-min running trials in their usual route, in an unusual route, and an athletics 400-m track, wearing a GPS and an HR monitor. Sample entropy (SampEn) and complexity index (CI), over speed and HR, were calculated. Pre and post-trial, participants performed the Backward Digit Span task for cognitive assessment. Higher entropies were found for the 400-m track, compared to the usual and unusual routes. Usual routes increased speed SampEn (63% of chances), but decreased HR CI when compared to unusual routes (60% of chances). Runners showed higher overall short-term memory performance after unusual routes, when compared to usual routes (85% of chances), indicating positive relation to attentional control. The contexts of practice may contribute to change predictability from single to multiple timescales. Thus, by considering that time structuring issues can help diagnosing habituation of training routes, this study brings novel information to the long-term process of training.

Keywords: heart rate, speed, running, training, entropies, dynamical systems

[Source: Frontiers | Entropy Measures Can Add Novel Information to Reveal How Runners' Heart Rate and Speed Are Regulated by Different Environments]()

Entropy in Machine Learning
Using Entropy to Define Information Gain

The following data that reflects three features (Age, Mileage, Road Tested) and one outcome (Buy Decision) to build a decision tree for the purchase decision of a used car.

Age	Mileage	Road Tested	Buy Decision
Recent	Low	Yes	Buy
Recent	High	Yes	Buy
Old	Low	No	Don't Buy
Recent	High	No	Don't Buy

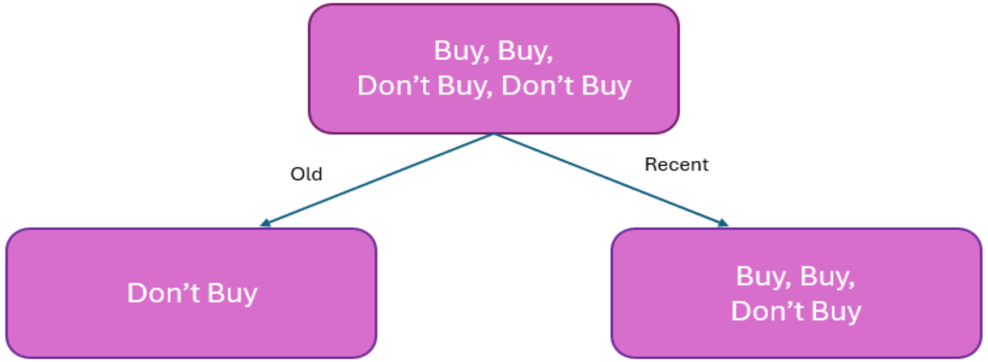
$Parent\ Node\ Entropy = -Prob(Buy) * log_2\,(Buy) - Prob(Don'tBuy) * log_2(Don'tBuy) = 1.0$

Information Gain =
Parent Node Entropy
Minus
Child Node Entropy

Split Decision

Maximum Information Gain

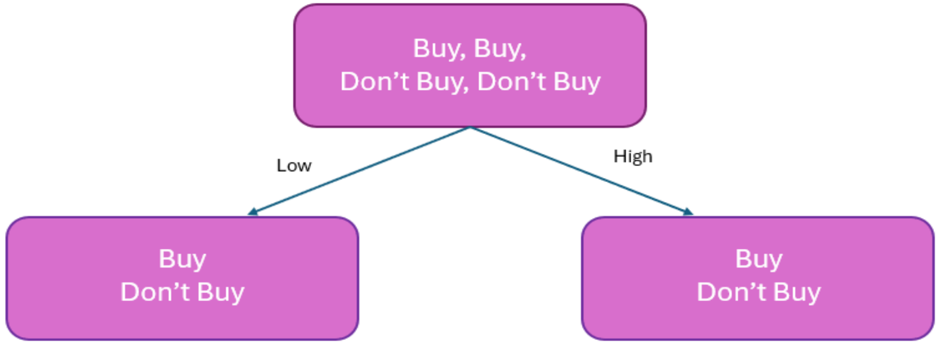
Split on Age



$$ChildEntropy = \frac{1}{4} * (0) + \frac{3}{4} * (0.918) = 0.688$$

$$InformationGain = 1 - 0.688 = 0.3112$$

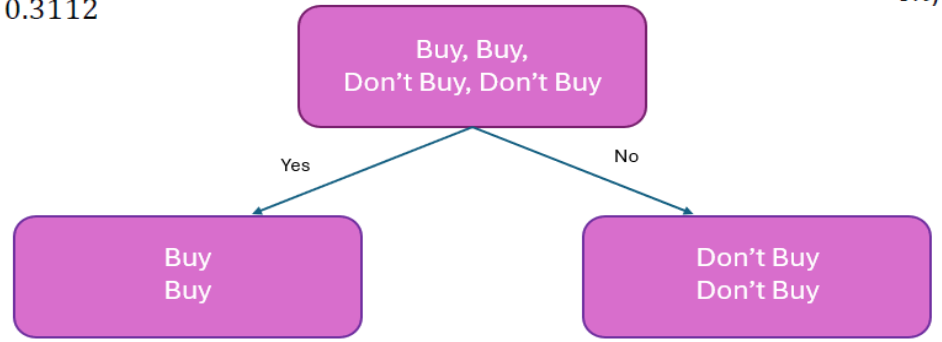
Split on Mileage



$$ChildEntropy = \frac{1}{2} * (1) + \frac{1}{2} * (1) = 1.00$$

$$InformationGain = 1 - 1 = 0.0$$

Split on Road Tested



$$Child\ Entropy = \frac{1}{2} * (0) + \frac{1}{2} * (0) = 0.00$$

$$Information\ Gain = 1 - 0 = 1.0$$

← **The Winner: Road Tested Variable**

Entropy Balancing

- Entropy Balancing is a data adjustment technique used when there is a binary (Yes/No) division in a set of data.
- The data are adjusted so that certain features are the same for both subsets of the data.
- This adjustment is done to identify the relationship of the binary characteristic to other characteristics in the data independent of already known relationships.
- Based on a paper by J. Hainmueller.

Entropy Balancing Example

- W. Zhong applied entropy balancing to pension buyout data from Form 5500 filings.
- Zhong found that PE buyouts increase the likelihood of DB plan termination or freeze by 14.2 percentage points.
- Following a buyout, Zhong found an increase in the pension liability discount rate and decreases in the projected benefit obligations, pension assets, and contributions. There were no significant effects on funding ratio.
- Following a buyout, investment strategies for these plans become riskier, with a higher allocation to equities and lower allocations to cash, government securities, insurance accounts, and mutual funds.

Entropy Balancing Example – Pension Buyout

W. Zhong used entropy balancing to analyze the impact of private equity buyouts on defined benefit (DB) plans

Illustration

Assumptions

- A known difference in the likelihood of a DB plan termination or freeze for mature companies with a small or large pension plan
- Small pension plans are significantly more likely than large plans to have a plan termination
- There are many more small plans then there are large plans

When looking at the data for pension plans that terminated, the data would mostly be for small pension plans.

Without entropy balancing, the difference in the rates of DB termination that had a buyout and those that did not could be due to more small pension plans that terminated.

Entropy balancing adjusts the data, producing a set of data that has the same proportion of small pension plans in both the subset that had a buyout and the subset that did not. This may be done by applying different weights to the smaller plans than the larger plans.

Hypothetical Data

Plan Size	Number of Plans	Number of Plan Failures	Number of PE Buyouts
Small	100	40	50
Large	10	1	5

Entropy Balanced Data	Number of Plans	Number of Plan Failures	Number of PE Buyouts
Small	10	4	5
Large	10	1	5

Thermodynamics
The Foundational Basis of Entropy

The Second Law of Thermodynamics states that the total entropy of a closed system always increases over time, naturally moving toward equilibrium.

Boltzmann Entropy Equation.

$$S = k_B \ln W$$

Gibbs Entropy Equation.

$$S = -k_B \sum p_i \ln p_i$$

Gibbs Free Energy Equation.

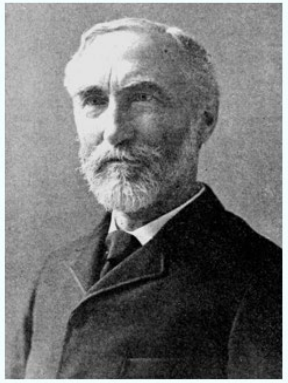
$$\Delta G = \Delta H - T\Delta S,$$

where Enthalpy (H) is the total heat content and Tempature (T).

Nicolas Léonard Sadi Carnot
1796-1832

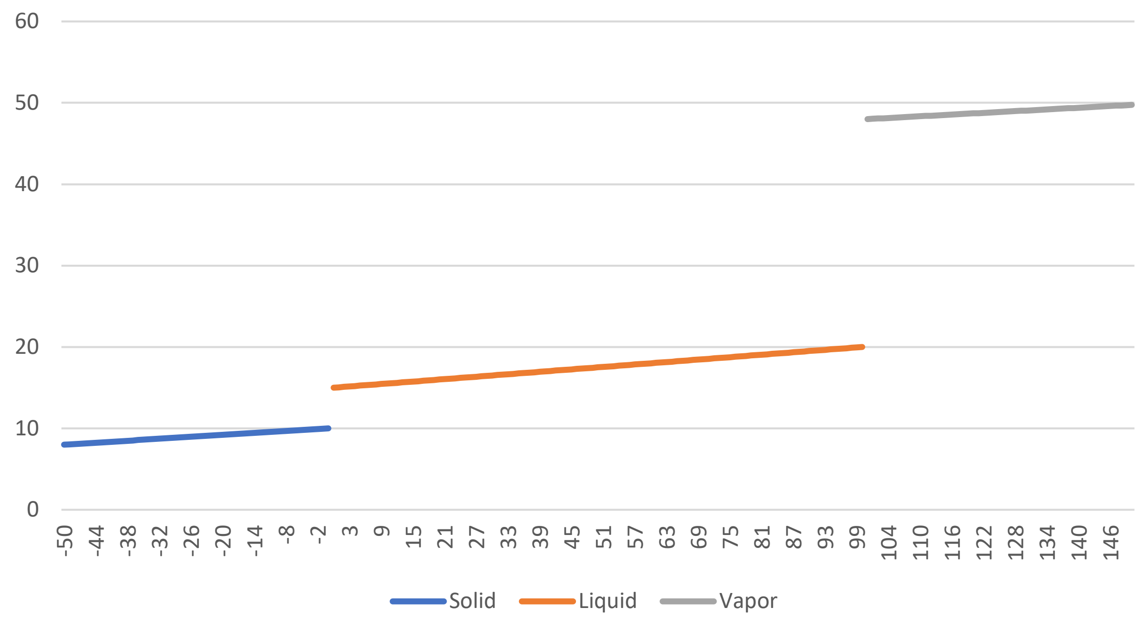
Rudolph Clausius
1822-1888

Ludwig Eduard Boltzmann
1844- 1906



Josiah Willard Gibbs
1839- 1903

Thermodynamics
Entropy States of Water



Ice Melting: Endothermic ($\Delta H > 0$, unfavorable) but increases disorder ($\Delta S > 0$, favorable). Occurs spontaneously (TDP) at high temperatures (above 0°C).

Water Freezing: Exothermic ($\Delta H < 0$, favorable) but decreases disorder ($\Delta S < 0$, unfavorable). Occurs spontaneously (TDP) at low temperatures.

Entropy and Statistics
The Foundational Basis of Entropy in Statistics

Entropy as an Objective Measure.

Entropy is an objective, measurable property of a system, like temperature or pressure, rather than just an epistemic (subjective) "information" index.

Measurement of Microscopic Disorder.

Entropy as a measure of the microscopic disorder of a system (N elements).

Reconciling Micro and Macro States.

Entropy reconciles the fine descriptions of a system (micro) with thermodynamically-favored processes (macro).

Rudolf Carnap
1891-1970

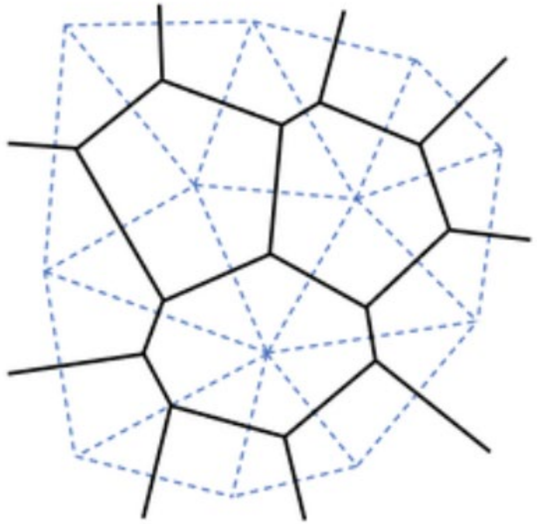
Teleonomic Entropy

Extends the concept of entropy to complex biological and social systems with "purpose" (teleonomy)

Method.
Monitors the phase-space of internal variables within a system to map how these variables evolve during learning or adaptation.
Complexity.
Analyze the number of paths that systems use to achieve goals.
Transfrence and Transformation.
Entropy can be transferred between subsystems or transformed, where the "entropic" (disorder-increasing) input from one source is managed by "negentropic" (order-increasing) inputs from another

Voronoi

Source: https://en.wikipedia.org/wiki/Delaunay_triangulation



$$d_\alpha^2(\boldsymbol{r}) = |\boldsymbol{r} - \boldsymbol{q}_\alpha|^2$$

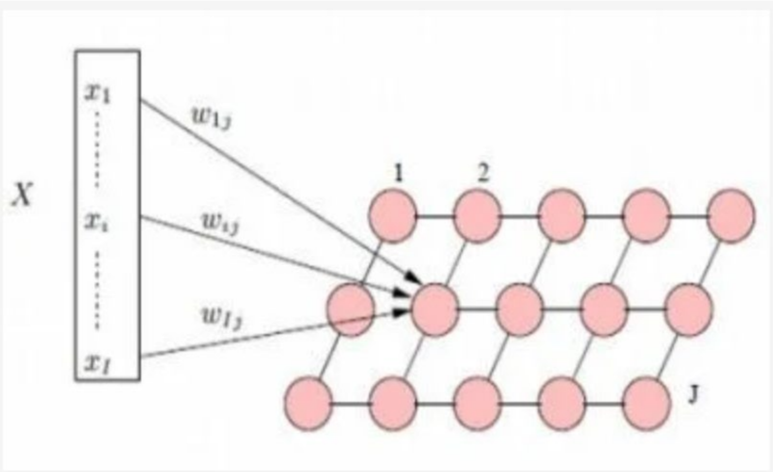
Maximum Entropy Classifier Model

Probabilistic, discriminative model that estimates the probability distribution of class labels

Principle of Maximum Entropy:
Selects the model with the largest entropy, often referred to as being "least informative" or most uniform.

Discriminative Model:
Computes the conditional probability P(c|d) of the class c given data d.

Multinomial Logistic regression:
The Maximum Entropy classifier is equivalent to multinomial logistic regression when applied to categorical data.



Life Table Entropy

Premature entropy peaking

- Leonard Hayflick in *Longevity Determination and Aging* states a thermodynamic type viewpoint that “the aging of living things is not unlike the aging of everything in the universe including the universe itself. The molecular disorder that defines biological aging might occur passively by increasing decrements in the energy necessary to maintain molecular fidelity or actively through, for example, the action of reactive oxygen species

	Life Table for the Projected Total Population in the United States: 2020			IRS Notice 2019-26 417(e)(3) 2020
	Normalized Entropy			Normalized Entropy
Age	Male	Female		Unisex
0	0.866022	0.8307002		
10	0.881290	0.8454024		
20	0.900687	0.8655545		0.81325011
30	0.918786	0.8872633		0.82940574
40	0.940104	0.9115660		0.84743485
50	0.961831	0.9369014		0.86670623
60	0.977194	0.9607660		0.88099269
70	0.985273	0.9817410		0.87610506
80	0.980782	0.9912299		0.83852813
90	0.971539	0.9827134		0.74455053

Bias v. Diversity

Using entropy as a measure

An Example Measuring Entropy

Suppose we are given the attributes such as Salary, Age, and Purchase and we want to calculate the entropy for the purchase variable, whether a customer will make a purchase or not.

Salary	Age	Purchase
20,000	21	No
10,000	45	No
60,000	27	No
15,000	31	No
12,000	18	No

p_i → probability of class i

C = 2: Yes & No

Probability (Yes) = 0

Probability (No) = 1

$$Entropy = -0 * log_2(0) - 1 * log_2(1) = 0$$

The answer makes sense. All outcomes are the same. There is not uncertainty about the purchase. Zero entropy. Is the result biased?

Note: $\log_2(0)$ is undefined and ignored.

Bias v. Diversity

Using entropy as a measure

An Example Measuring Entropy

		Purchase Scenarios					
Salary	**Age**	**1**	**2**	**3**	**4**	**5**	**6**
20,000	21	No	Yes	Yes	Yes	Yes	Yes
10,000	45	No	No	Yes	Yes	Yes	Yes
60,000	27	No	No	No	Yes	Yes	Yes
15,000	31	No	No	No	No	Yes	Yes
12,000	18	No	No	No	No	No	Yes
Entropy Metrics		0	0.72	**0.97**	**0.97**	0.72	0

Bias v. Diversity

Using entropy as a measure

Interpreting the Results

Yes	No	Entropy
0	5	0.00
1	4	0.72
2	3	0.97
3	2	0.97
4	1	0.72
5	0	0.00

Most Class Balance

Conclusion:

Low Entropy → High Bias → Low Diversity
High Entropy → Low Bias → High Diversity

Conclusion

Measuring Bias & Diversity in Data

Statistical Definition of Bias

$$Bias = E\left[\hat{\theta}\right] - \theta$$

Entropy Definition of Bias

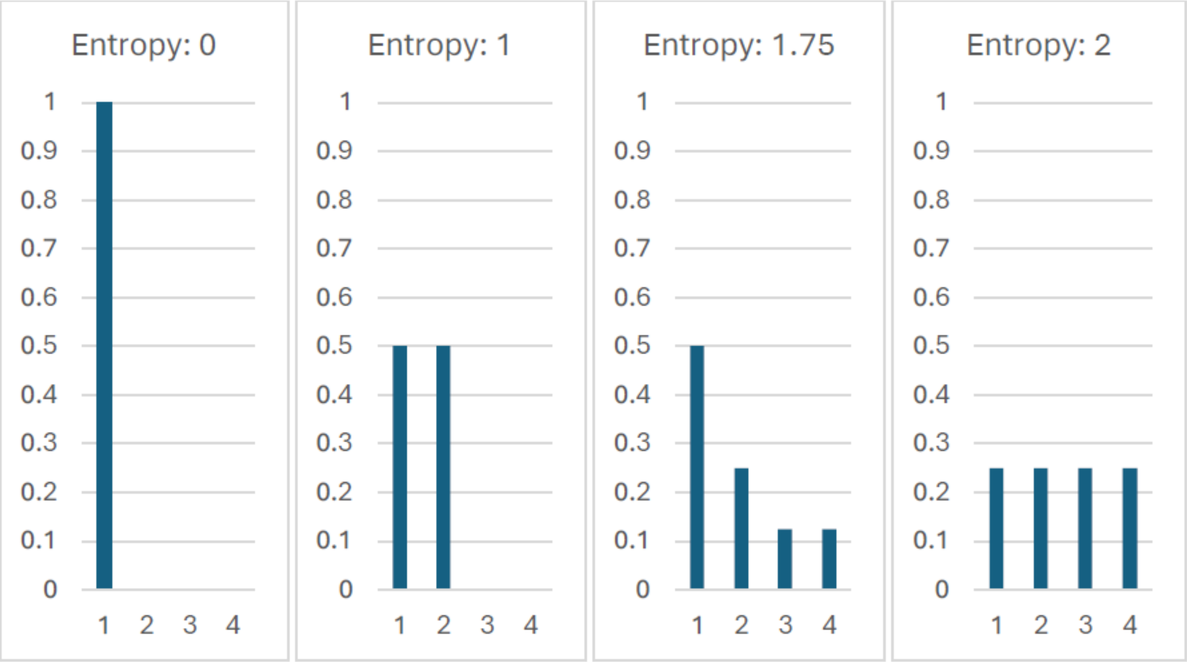
$$H(\boldsymbol{p}) = \sum_{i=1}^{n} \mathrm{p}_i \log\left(\frac{1}{p_i}\right)$$

The Bias-Diversity Continuum

- As bias increases, diversity decreases
- Entropy can mathematically define both
- Useful measures for large data sets
- Compare to maximum entropy/diversity

$$Log_2\ (N)$$

where, N = Number of Distinct Elements



Decreasing Entropy | Increasing Entropy

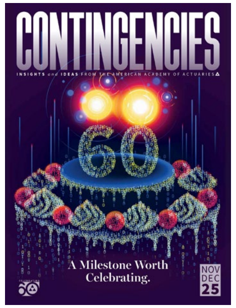
Decreasing Entropy	Increasing Entropy
Certainty	Uncertainty
Order	Disorder
High Bias	Low Bias
Non-Randomness	Randomness
Higher Information	Lower Information

American Academy of Actuaries

Other Academy Resources

Follow the Academy on LinkedIn

Access the Following Resources:



Contingencies Magazine

Policy Forum

Actuarially Sound Blog

Actuary Voices Podcast

Sources

- Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies
- How Do Private Equity Buyouts Affect Employee Pension Plans?
- https://www.soa.org/globalassets/assets/files/resources/essays-monographs/2002-living-to-100/mono-2002-m-li-02-1-hayflick.pdf
- https://build5nines.com/detecting-ai-hallucinations-using-semantic-entropy/

AMERICAN ACADEMY of ACTUARIES

Questions?

For more information, please contact

Will Behnke

Policy Project Manager, Risk Management and Financial Reporting

behnke@actuary.org