The Measure of Information

Uniqueness of the Logarithmic Uncertainty Measure

Almudena Colacito

Information Theory
Fall 2014

December 17th, 2014

The Measurement of Information

[R.V.L. Hartley, 1928]

"A quantitative measure of "information" is developed which is based on physical as contrasted with psychological considerations."

How much "choice" is involved?

- Shannon's set of axioms;
- Proof: we are talking about the *entropy* indeed;
- Other sets of axioms: comparisons and consequences;
- Logarithm: why?

How much "choice" is involved?

- Shannon's set of axioms;
- Proof: we are talking about the *entropy* indeed;
- Other sets of axioms: comparisons and consequences;
- Logarithm: why?

How much "choice" is involved?

- Shannon's set of axioms;
- Proof: we are talking about the *entropy* indeed;
- Other sets of axioms: comparisons and consequences;
- Logarithm: why?

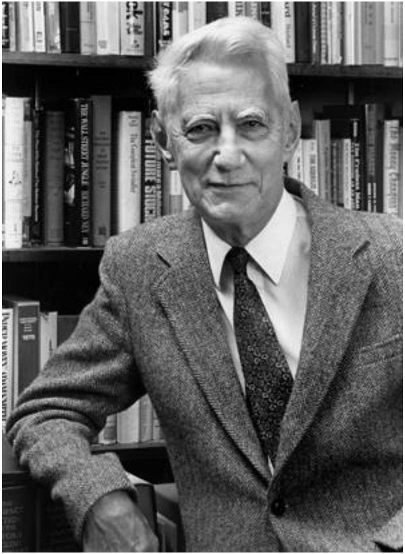
How much "choice" is involved?

- Shannon's set of axioms;
- Proof: we are talking about the *entropy* indeed;
- Other sets of axioms: comparisons and consequences;
- Logarithm: why?

How much “choice” is involved?

- Shannon’s set of axioms;
- Proof: we are talking about the *entropy* indeed;
- Other sets of axioms: comparisons and consequences;
- Logarithm: why?

Shannon's axioms



Shannon's axioms

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$:

1. H is continuous in pi, for any i;
2. If $pi = \frac{1}{n}$, for any i, then H is a *monotonic increasing function* of n;
3. If a choice be broken down into two successive choices, the original H is the weighted sum of the individual values of H.

Shannon's axioms

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$:

1. H is continuous in pi, for any i;
2. If $pi = \frac{1}{n}$, for any i, then H is a *monotonic increasing function* of n;
3. If a choice be broken down into two successive choices, the original H is the weighted sum of the individual values of H.

Shannon's axioms

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$:

1. H is continuous in pi, for any i;
2. If $pi = \frac{1}{n}$, for any i, then H is a *monotonic increasing function* of n;
3. If a choice be broken down into two successive choices, the original H is the weighted sum of the individual values of H.

Shannon's axioms

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$:

1. H is continuous in pi, for any i;
2. If $pi = \frac{1}{n}$, for any i, then H is a *monotonic increasing function* of n;
3. If a choice be broken down into two successive choices, the original H is the weighted sum of the individual values of H.

Uniqueness of Uncertainty Measure

Theorem

There exists a unique H satisfying the three above assumptions. In particular, H is of the form:

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

Proof: Consider $A(n) := H(\frac{1}{n}, \dots, \frac{1}{n})$.

Uniqueness of Uncertainty Measure

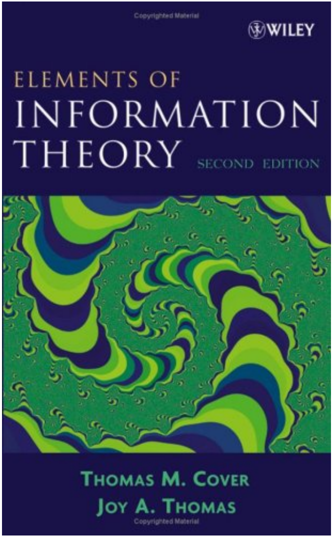
Theorem

There exists a unique H satisfying the three above assumptions.
In particular, H is of the form:

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

Proof: Consider $A(n) := H(\frac{1}{n}, \dots, \frac{1}{n})$.

Alternative set of axioms [CT]



Alternative set of axioms [CT]

Let $H_m(p_1, p_2, \ldots, p_m)$ be a sequence of symmetric functions, then it satisfies the following properties:

1. *Normalization:* $H_2(\frac{1}{2}, \frac{1}{2}) = 1$;
2. *Continuity:* $H(p, 1-p)$ is a continuous function in p;
3. *Grouping:* $H_m(p_1, p_2, \ldots, p_m) =$
   $= H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Alternative set of axioms [CT]

Let $H_m(p_1, p_2, \ldots, p_m)$ be a sequence of symmetric functions, then it satisfies the following properties:

1. *Normalization:* $H_2(\frac{1}{2}, \frac{1}{2}) = 1$;
2. *Continuity:* $H(p, 1-p)$ is a continuous function in p;
3. *Grouping:* $H_m(p_1, p_2, \ldots, p_m) =$
   $= H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Alternative set of axioms [CT]

Let $H_m(p_1, p_2, \ldots, p_m)$ be a sequence of symmetric functions, then it satisfies the following properties:

1. *Normalization:* $H_2(\frac{1}{2}, \frac{1}{2}) = 1$;
2. *Continuity:* $H(p, 1-p)$ is a continuous function in p;
3. *Grouping:* $H_m(p_1, p_2, \ldots, p_m) =$
   $= H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Alternative set of axioms [CT]

Let $H_m(p_1, p_2, \ldots, p_m)$ be a sequence of symmetric functions, then it satisfies the following properties:

1. *Normalization:* $H_2(\frac{1}{2}, \frac{1}{2}) = 1$;
2. *Continuity:* $H(p, 1-p)$ is a continuous function in p;
3. *Grouping:* $H_m(p_1, p_2, \ldots, p_m) =$
   $= H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2) H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Alternative set of axioms [Carter]

Let $I(p)$ be an information measure and let p indicate a probability measure.

1. $I(p) \geq 0$ (*non-negative*);
2. $I(1) = 0$,
   (we don't get any information from an event with probability 0);
3. let p_1 and p_2 be the probabilities of two independent events. Then, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (!);
4. I is a continuous and monotonic function of the probability (slight changes in probability-slight changes in information).

Alternative set of axioms [Carter]

Let $I(p)$ be an information measure and let p indicate a probability measure.

1. $I(p) \geq 0$ (*non-negative*);
2. $I(1) = 0$,
   (we don't get any information from an event with probability 0);
3. let p_1 and p_2 be the probabilities of two independent events. Then, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (!);
4. I is a continuous and monotonic function of the probability (slight changes in probability-slight changes in information).

Alternative set of axioms [Carter]

Let $I(p)$ be an information measure and let p indicate a probability measure.

1. $I(p) \geq 0$ (*non-negative*);
2. $I(1) = 0$,
   (we don't get any information from an event with probability 0);
3. let p_1 and p_2 be the probabilities of two independent events. Then, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (!);
4. I is a continuous and monotonic function of the probability (slight changes in probability-slight changes in information).

Alternative set of axioms [Carter]

Let $I(p)$ be an information measure and let p indicate a probability measure.

1. $I(p) \geq 0$ (*non-negative*);
2. $I(1) = 0$,
   (we don't get any information from an event with probability 0);
3. let p_1 and p_2 be the probabilities of two independent events. Then, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (!);
4. I is a continuous and monotonic function of the probability (slight changes in probability-slight changes in information).

Alternative set of axioms [Carter]

Let $I(p)$ be an information measure and let p indicate a probability measure.

1. $I(p) \geq 0$ (*non-negative*);
2. $I(1) = 0$,
   (we don't get any information from an event with probability 0);
3. let p_1 and p_2 be the probabilities of two independent events. Then, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (!);
4. I is a continuous and monotonic function of the probability (slight changes in probability-slight changes in information).

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Comparisons between axiomatisations

- $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$ by axiom (3);
- by induction on n, we get: $I(p^n) = I(p \cdot \dots \cdot p) = n \cdot I(p)$;
- $I(p) = I((p^{\frac{1}{m}})^m) = m \cdot I(p^{\frac{1}{m}})$, then: $I(p^{\frac{1}{m}}) = \frac{1}{m} I(p)$;
- by continuity, for any $0 < p \leq 1$ and $0 < a$: $I(p^a) = a \cdot I(p)$.

We get, again, $I(p) = log(\frac{1}{p})$ as measure of information.

Logarithm: why?

"The most natural choice is the logarithmic function." (Shannon, 1948)

- It is practically more useful;
- It is nearer to our intuitive feelings;
- It is mathematically more suitable.

Logarithm: why?

"The most natural choice is the logarithmic function." (Shannon, 1948)

- It is practically more useful;
- It is nearer to our intuitive feelings;
- It is mathematically more suitable.

Logarithm: why?

"The most natural choice is the logarithmic function." (Shannon, 1948)

- It is practically more useful;
- It is nearer to our intuitive feelings;
- It is mathematically more suitable.

Logarithm: why?

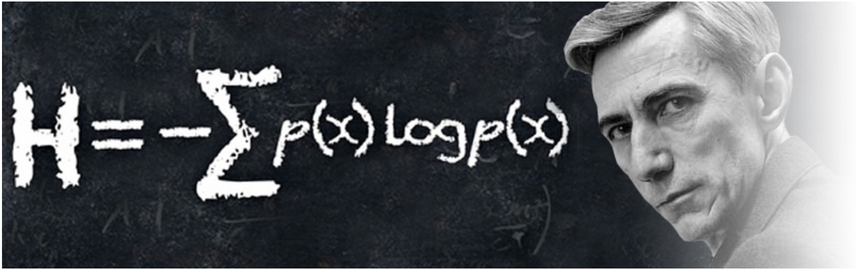
"The most natural choice is the logarithmic function." (Shannon, 1948)

- It is practically more useful;
- It is nearer to our intuitive feelings;
- It is mathematically more suitable.

Here it is: the Entropy!

[John von Neumann]

"You should call it entropy, for two reasons. In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage."



"Claude Shannon invented a way to measure the 'amount of information' in a message without defining the word *information* itself, nor even addressing the question of the meaning of the message."
(Hans Christian von Baeyer)

THANK YOU!

References

1 C. E. Shannon, *A Mathematical Theory of Communication*, *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

2 R. V. Hartley, *Transmission of Information*, 1928.

3 T. M. Cover, J. A. Thomas, *Elements of Information Theory*, second edition, Wiley-Interscience, New York, 2006.

4 T. Carter, *An Introduction to Information Theory and Entropy*, http://astarte.csustan.edu/ tom/SFI-CSSS, California State University Stanislaus.

Baudot System