Uncertainty measures

Davide Ciucci

Department of Informatics, Systems and Communication
University of Milan-Bicocca

April 6, 2022

Outline

Classical Measures

Generalized measures

Shannon entropy

Intuition:

- X is a set of alternatives with probabilities $p(x)$. Only one alternative must occur (an outcome of an experiment, a received message, ...)

Shannon entropy

Intuition:

- X is a set of alternatives with probabilities $p(x)$. Only one alternative must occur (an outcome of an experiment, a received message, ...)
- How to quantify the information given by the fact that x occurs?

Shannon entropy

Intuition:

- X is a set of alternatives with probabilities $p(x)$. Only one alternative must occur (an outcome of an experiment, a received message, ...)
- How to quantify the information given by the fact that x occurs?
- if $x =$ "tomorrow the sun will rise", $p(x) = 0.9999....$ The fact that the sun rises is not a surprise hence we have no uncertainty in anticipating it

Shannon entropy

Intuition:

- X is a set of alternatives with probabilities $p(x)$. Only one alternative must occur (an outcome of an experiment, a received message, ...)
- How to quantify the information given by the fact that x occurs?
- if $x =$ "tomorrow the sun will rise", $p(x) = 0.9999...$. The fact that the sun rises is not a surprise hence we have no uncertainty in anticipating it
- if $x =$ "Fiorentina will won the Italian football championship", $p(x) = 0.0001$. If x occurs, it is unexpected, we have a huge amount of uncertainty in anticipating it

Shannon entropy (1948)

- Goal: define a function u such that it is
  - a decreasing function: the more likely the occurence of x, the less information is provided

Shannon entropy (1948)

- Goal: define a function u such that it is
  - a decreasing function: the more likely the occurence of x, the less information is provided
  - an additive function if $p(x, y) = p(x) \cdot p(y)$ then $u(p(x, y)) = u(p(x)) + u(p(y))$

Shannon entropy (1948)

- Goal: define a function u such that it is
  - a decreasing function: the more likely the occurence of x, the less information is provided
  - an additive function if $p(x, y) = p(x) \cdot p(y)$ then $u(p(x, y)) = u(p(x)) + u(p(y))$
- the solution is $u(p(x)) = K \log_b p(x)$.
  If we measure uncertainty in bits then $b = 2$, $u(\frac{1}{2}) = 1$ (normalization condition) and so $K = -1$
- hence the information is $\log_2 p(x)$ bits

Shannon entropy

- Let $p(x)$ be a probability distribution on a finite set U
- $H : \mathcal{P}(U) \mapsto [0, \infty)$

Shannon entropy

- Let $p(x)$ be a probability distribution on a finite set U
- $H : \mathcal{P}(U) \mapsto [0, \infty)$
- $H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$
- it measures the expected information content of X

Hartley entropy (1928)

Let set $p(x) = \frac{1}{|X|}$

- uncertainty associated with a choice among $n = |X|$ alternatives/possibilities

Hartley entropy (1928)

Let set $p(x) = \frac{1}{|X|}$

- uncertainty associated with a choice among $n = |X|$ alternatives/possibilities
- a measure of non-specificity: the fewer alternatives we have, the more specific is our choice

$$I(n) = \log_2 n$$

Logical (Ellerman) entropy

- the information of a partition π

Logical (Ellerman) entropy

- the information of a partition π
- the information is higher as higher is the ability to distinguish between two alternatives

Logical (Ellerman) entropy

- the information of a partition π
- the information is higher as higher is the ability to distinguish between two alternatives
  - 1 equivalence class $\rightarrow$ no ability to distinguish the elements, the lowest information

Logical (Ellerman) entropy

- the information of a partition π
- the information is higher as higher is the ability to distinguish between two alternatives
  - 1 equivalence class $\rightarrow$ no ability to distinguish the elements, the lowest information
  - n classes $\rightarrow$ all the elements are distinct, the highest information

Logical (Ellerman) entropy

- the information of a partition π
- the information is higher as higher is the ability to distinguish between two alternatives
  - 1 equivalence class $\rightarrow$ no ability to distinguish the elements, the lowest information
  - n classes $\rightarrow$ all the elements are distinct, the highest information
- based on the notion of dits (distinctions): ordered pairs (u_i, u_j) of elements in different blocks
  $dit(\pi) = \{(u_i, u_j) | u_i, u_j$ belong do different blocks of $\pi\}$

Logical (Ellerman) entropy

- the information of a partition π
- the information is higher as higher is the ability to distinguish between two alternatives
  - 1 equivalence class $\rightarrow$ no ability to distinguish the elements, the lowest information
  - n classes $\rightarrow$ all the elements are distinct, the highest information
- based on the notion of dits (distinctions): ordered pairs (u_i, u_j) of elements in different blocks
  $dit(\pi) = \{(u_i, u_j) | u_i, u_j \text{ belong do different blocks of } \pi\}$

$$h(\pi) = \frac{|dit(\pi)|}{|U \times U|}$$

Outline

Classical Measures

Generalized measures

Fuzzy entropy

- De Luca, Termini (1972)
- $f : X \mapsto [0, 1]$ a fuzzy susbset of X

Fuzzy entropy

- De Luca, Termini (1972)
- $f : X \mapsto [0,1]$ a fuzzy susbset of X

$$h(f) = -\sum_{x \in X} f(x)\log_2 f(x) + (1 - f(x))(\log_2(1 - f(x))$$

- It represents the degree of fuzziness: $h(f) = 0$ iff f is a Boolean set

Fuzzy entropy

- De Luca, Termini (1972)
- $f : X \mapsto [0,1]$ a fuzzy susbset of X

$$h(f) = -\sum_{x \in X} f(x) \log_2 f(x) + (1 - f(x))(\log_2(1 - f(x))$$

- It represents the degree of fuzziness: $h(f) = 0$ iff f is a Boolean set
- it has the max value in case of a maximally fuzzy set, i.e., $\forall x,\ f(x) = \frac{1}{2}$

Fuzzy entropy

- De Luca, Termini (1972)
- $f : X \mapsto [0,1]$ a fuzzy susbset of X

$$h(f) = -\sum_{x \in X} f(x)\log_2 f(x) + (1 - f(x))(\log_2(1 - f(x))$$

- It represents the degree of fuzziness: $h(f) = 0$ iff f is a Boolean set
- it has the max value in case of a maximally fuzzy set, i.e., $\forall x,\ f(x) = \frac{1}{2}$
- it captures the idea of less fuzzy then: if $f_1 \leq f_2$ then $h(f_1) \leq h(f_2)$

Conflict of Evidence

Between two body of evidence

- How are in conflict two mass distributions m_1, m_2?
- They conflict if $m_1(A) \neq 0, m_2(B) \neq 0$ and $A \cap B = \emptyset$

Conflict of Evidence

Between two body of evidence

- How are in conflict two mass distributions m_1, m_2?
- They conflict if $m_1(A) \neq 0, m_2(B) \neq 0$ and $A \cap B = \emptyset$

$$con(m_1, m_2) = -\log_2(1 - K)$$

where $K = \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B)$ and A, B are focal sets

Conflict of Evidence

Between two body of evidence

- How are in conflict two mass distributions m_1, m_2?
- They conflict if $m_1(A) \neq 0, m_2(B) \neq 0$ and $A \cap B = \emptyset$

$$con(m_1, m_2) = -\log_2(1 - K)$$

  where $K = \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B)$ and A, B are focal sets
- *con* is monotonic increasing with K
  - $con(m_1, m_2) = 0$ iff m_1 and m_2 do not conflict ($K = 0$)
  - $con(m_1, m_2) = \infty$ iff m_1 and m_2 totally conflict ($K = 1$)

Conflict of Evidence

Within a single body of evidence m

- Let $m_A(B) = \begin{cases} 1 & B = A \\ 0 & \text{otherwise} \end{cases}$

Conflict of Evidence

Within a single body of evidence m

- Let $m_A(B) = \begin{cases} 1 & B = A \\ 0 & \text{otherwise} \end{cases}$
- The amount of conflict of a focal set with all the other focal sets

Conflict of Evidence

Within a single body of evidence m

- Let $m_A(B) = \begin{cases} 1 & B = A \\ 0 & \text{otherwise} \end{cases}$
- The amount of conflict of a focal set with all the other focal sets

$$con(m, m_A) = -\log_2(1 - \sum_{B \cap A = \emptyset} m(B)) = -\log_2 Pl(A)$$

Conflict of Evidence

Within a single body of evidence m

- Let $m_A(B) = \begin{cases} 1 & B = A \\ 0 & \text{otherwise} \end{cases}$
- The amount of conflict of a focal set with all the other focal sets

$$con(m, m_A) = -\log_2(1 - \sum_{B \cap A = \emptyset} m(B)) = -\log_2 Pl(A)$$

- weigthed average over all focal sets:

$$E(m) = \sum_{A \in F} m(A) con(m, m_A) = -\sum_{A \in F} m(A) \log_2 Pl(A)$$

with F focal sets

Conflict of Evidence

- If m is a probability distribution (probability assigned to singletons), $E(m)$ is equivalent to Shannon entropy

Conflict of Evidence

- If m is a probability distribution (probability assigned to singletons), $E(m)$ is equivalent to Shannon entropy
- Conflict is also named dissonance
  If the focal sets are consonant (not dissonant),
  $A_1 \subseteq A_2 \subseteq \ldots A_n$, then $E(m) = 0$
  Possibility and necessity have no conflict

Confusion in Evidence

- From conflict $E(m) = -\sum_{A \in F} m(A) \log_2 Pl(A)$
  To confusion

$$C(m) = -\sum_{A \in F} m(A) \log_2 Bel(A)$$

Confusion in Evidence

- From conflict $E(m) = -\sum_{A \in F} m(A) \log_2 Pl(A)$
  To confusion

$$C(m) = -\sum_{A \in F} m(A) \log_2 Bel(A)$$

- $C(m) = 0$ iff $m(A) = 1$ for one particular A and $m(B) = 0$ for all other B

Confusion in Evidence

- From conflict $E(m) = -\sum_{A \in F} m(A) \log_2 Pl(A)$
  To confusion

$$C(m) = -\sum_{A \in F} m(A) \log_2 Bel(A)$$

- $C(m) = 0$ iff $m(A) = 1$ for one particular A and $m(B) = 0$ for all other B
- $C(m)$ characterizes
  - the uniformity of the distribution of the strength of evidence among the subsets
  - the multitude of subsets supported by evidence:
    the greater the number of subsets involved and the more uniform the distribution, the more we are confused

Specificity in possibility theory

- A generalization of Hartley entropy $I(n) = \log_2 n$

Specificity in possibility theory

- A generalization of Hartley entropy $I(n) = \log_2 n$
- Defined on ordered possibility distributions
  $\pi : \{x_1, x_2, \ldots, x_n\} \mapsto [0, 1]$
  $\forall i < j,\ \pi(x_i) \leq \pi(x_j)$

$$U(\pi) = \sum_{i=1}^{n} (\pi(x_i) - \pi(x_{i+1})) \log_2 i$$

Specificity in possibility theory

- A generalization of Hartley entropy $I(n) = \log_2 n$
- Defined on ordered possibility distributions
  $\pi : \{x_1, x_2, \ldots, x_n\} \mapsto [0, 1]$
  $\forall i < j,\ \pi(x_i) \leq \pi(x_j)$

$$U(\pi) = \sum_{i=1}^{n} (\pi(x_i) - \pi(x_{i+1})) \log_2 i$$

- Weigthed average of Hartley measure for all focal elements A_i
  $U(\pi) = \sum_{i=1}^{n} m(A_i) log_2 |A_i|$

Orthopartition: Ellerman Entropy

Partition π

$dit(\pi) = \{(u, u') \in U \times U |\ u, u' \text{ belongs to two different blocks of } \pi\}$

$$h(\pi) = \frac{|dit(\pi)|}{|U|^2}$$

Orthopartition: Ellerman Entropy

Partition π

$dit(\pi) = \{(u, u') \in U \times U|\ u, u' \text{ belongs to two different blocks of } \pi\}$

$$h(\pi) = \frac{|dit(\pi)|}{|U|^2}$$

Orthopartition $\mathcal{O}$

$\Pi_{\mathcal{O}}$ the set of all partitions consistent with $\mathcal{O}$

For $\pi \in \Pi_{\mathcal{O}}$ we can compute $h(\pi)$

Orthopartition: Ellerman Entropy

Partition π
$dit(\pi) = \{(u, u') \in U \times U |\ u, u' \text{ belongs to two different blocks of } \pi\}$

$$h(\pi) = \frac{|dit(\pi)|}{|U|^2}$$

Orthopartition $\mathcal{O}$
$\Pi_{\mathcal{O}}$ the set of all partitions consistent with $\mathcal{O}$
For $\pi \in \Pi_{\mathcal{O}}$ we can compute $h(\pi)$

lower entropy	$h_* = min\{h(\pi) \mid \pi \in \Pi_{\mathcal{O}}\}$
upper entropy	$h^* = max\{h(\pi) \mid \pi \in \Pi_{\mathcal{O}}\}$
mean entropy	$\hat{h} = \frac{h^*(\mathcal{O}) + h_*(\mathcal{O})}{2}$
average entropy	$h_A(\mathcal{O}) = \frac{1}{\lvert\Pi_{\mathcal{O}}\rvert} \sum_{\pi \in \Pi_{\mathcal{O}}} h(\pi)$

Entropy example

Consider the orthopartition $\mathcal{O} = \{O_1, O_2\}$, with $O_1 = \langle \{1\}, \{3\} \rangle$ and $O_2 = \langle \{3\}, \{1\} \rangle$, defined on universe $U = \{1, 2, 3, 4\}$.

Entropy example

Consider the orthopartition $\mathcal{O} = \{O_1, O_2\}$, with $O_1 = \langle\{1\}, \{3\}\rangle$ and $O_2 = \langle\{3\}, \{1\}\rangle$, defined on universe $U = \{1, 2, 3, 4\}$. The set $\Pi_{\mathcal{O}}$ of partitions consistent with $\mathcal{O}$ contains the following partitions:

- $\pi_1 = \{\{1, 2, 4\}, \{3\}\}$ with entropy $h = \frac{6}{16}$;
- $\pi_2 = \{\{1\}, \{2, 3, 4\}\}$ with entropy $h = \frac{6}{16}$;
- $\pi_3 = \{\{1, 2\}, \{3, 4\}\}$ with entropy $h = \frac{8}{16}$;
- $\pi_4 = \{\{1, 4\}, \{2, 3\}\}$ with entropy $h = \frac{8}{16}$;

Entropy example

Consider the orthopartition $\mathcal{O} = \{O_1, O_2\}$, with $O_1 = \langle\{1\},\{3\}\rangle$ and $O_2 = \langle\{3\},\{1\}\rangle$, defined on universe $U = \{1,2,3,4\}$. The set $\Pi_{\mathcal{O}}$ of partitions consistent with $\mathcal{O}$ contains the following partitions:

- $\pi_1 = \{\{1,2,4\},\{3\}\}$ with entropy $h = \frac{6}{16}$;
- $\pi_2 = \{\{1\},\{2,3,4\}\}$ with entropy $h = \frac{6}{16}$;
- $\pi_3 = \{\{1,2\},\{3,4\}\}$ with entropy $h = \frac{8}{16}$;
- $\pi_4 = \{\{1,4\},\{2,3\}\}$ with entropy $h = \frac{8}{16}$;

The lower entropy is thus equal to $h_* = \frac{6}{16}$
The upper entropy is equal to $h^* = \frac{1}{2}$

Orthopartition: Shannon entropy

- We need a probability distribution $p = \langle p_1, \ldots, p_n \rangle$

Orthopartition: Shannon entropy

- We need a probability distribution $p = \langle p_1, ..., p_n \rangle$
- The set of probability distributions compatible with an orthopartition $\mathcal{O}$:

$$\mathcal{P}_{\mathcal{O}} = \{\langle p_1, ..., p_n \rangle | p_i \in [\frac{|P_i|}{|U|}, \frac{|P_i \cup Bnd_i|}{|U|}] \text{ and } \sum_{i=1}^{n} p_i = 1\}$$

Orthopartition: Shannon entropy

- We need a probability distribution $p = \langle p_1, ..., p_n \rangle$
- The set of probability distributions compatible with an orthopartition $\mathcal{O}$:

$$\mathcal{P}_{\mathcal{O}} = \{\langle p_1, ..., p_n \rangle | p_i \in [\frac{|P_i|}{|U|}, \frac{|P_i \cup Bnd_i|}{|U|}] \text{ and } \sum_{i=1}^{n} p_i = 1\}$$

-

lower entropy	$H_{S*} = min\{H_S(p) \| p \in \mathcal{P}_{\mathcal{O}}\}$
upper entropy	$H_S^* = max\{H_S(p) \| p \in \mathcal{P}_{\mathcal{O}}\}$
mean entropy	$\hat{H}_S = \frac{H_S^*(\mathcal{O}) + H_{S*}(\mathcal{O})}{2}$
average entropy	$H_{SP}(\mathcal{O}) = \frac{1}{\|\Pi_{\mathcal{O}}\|} \sum_{\pi \in \Pi_{\mathcal{O}}} H_S(\pi)$

Computational complexity

We have polynomial algorithms for all Ellerman entropies

- lower, upper and mean: $O(|U| * n * \log_2 n)$
  n = number of orthopairs
- average (in the case that $P_i \neq \emptyset$): $\Theta(n * |U|^2)$

Computational complexity

We have polynomial algorithms for all Ellerman entropies

- lower, upper and mean: $O(|U| * n * \log_2 n)$
  n = number of orthopairs
- average (in the case that $P_i \neq \emptyset$): $\Theta(n * |U|^2)$

Shannon entropy: no simple algorithm to compute it

Orthopartition: Mutual Information

Orthopartition: Mutual Information

- Goal: quantify the common information of two orthopartitions

Orthopartition: Mutual Information

- Goal: quantify the common information of two orthopartitions
- The mutual information

$$m(\mathcal{O}_1, \mathcal{O}_2) = h(\mathcal{O}_1) + h(\mathcal{O}_2) - h(\mathcal{O}_1 \wedge \mathcal{O}_2)$$

where h is one of the entropies defined before

Orthopartition: Mutual Information

- Goal: quantify the common information of two orthopartitions
- The mutual information

$$m(\mathcal{O}_1, \mathcal{O}_2) = h(\mathcal{O}_1) + h(\mathcal{O}_2) - h(\mathcal{O}_1 \wedge \mathcal{O}_2)$$

  where h is one of the entropies defined before
- Meet orthopartition

$$\mathcal{O}_1 \wedge \mathcal{O}_2 = \{O_{i1} \sqcap_t O_{j2} | O_{i1} \in \mathcal{O}_1 \text{ and } O_{j2} \in \mathcal{O}_2\}$$

  with $\sqcap_t$ the intersection defined by truth ordering

$$O_1 \sqcap_t O_2 = (L_1 \cap L_2, U_1 \cup U_2)$$

- The choice of $\sqcap_t$ is due to the fact that it holds:

  $\Pi_{\mathcal{O}_1 \wedge \mathcal{O}_2} = \{\pi \wedge \sigma | \pi \text{ is consistent with } \mathcal{O}_1 \text{ and } \sigma \text{ is consistent with } \mathcal{O}_2\}$