



Author date: 2022-03-28

Copyright:

creativecommons.org/licenses/by/4.0/
2022 © The Authors. This document is distributed under a Creative Commons Attribution 4.0 International license.

Scoped Entropy

E. Castedo Ellerman  (castedo@castedo.com)

Abstract

This WORKING DRAFT defines *scoped entropy*.

| WORKING DRAFT

Both variance and entropy are measures of uncertainty and thus information. *Scoped entropy* is a generalization of both statistical variance and Shannon entropy [1]. It jointly generalizes by using a construct of *scope of relevance*. *Scoped entropy* can be applied to both discrete and continuous variables.

A *random object* is a function with a domain of a probability space [2]. When the function values are real numbers, it is a *random variable* (also called *real random object* in this document). In this document, a *finite random object* means a random object that takes on finitely many values. Or in other words, the range of a *finite random object* is a set of finite size.

Both entropy and variance are functions of random objects. In the case of Shannon entropy, the random object is finite (with values often referred to as symbols). In the case of variance, the random object is a (real) random variable. The distances between values of a finite random variable affect variance, but not Shannon entropy.

Given a scope of relevance θ and a finite or real random object X , *scoped entropy* is denoted as the function

$$V_{\theta}(X)$$

There is a *unirelevant decomposition* function \mathfrak{h} which given any finite random object X generates a scope $\mathfrak{h}(X)$ such that

$$V_{\mathfrak{h}(X)}(X) = H(X)$$

where $H(X)$ is Shannon entropy.

Similarly, there is a *distribution decomposition* function \mathfrak{d} which given any random variable X generates a scope $\mathfrak{d}(X)$ such that

$$V_{\mathfrak{d}(X)}(X) = \text{Var}(X)$$

Entropy Scope of Relevance

Let $\text{dom } f$ denote the domain of a function f . Given π a set of sets, let $\cup\pi$ denote the union of all member sets in π . When π is a partition, π covers $\cup\pi$.

In a *scope of relevance*, partitions represent questions with each member set (part) an answer. Each answer represents an event of a probability space.

Given a function over partitions of events of probability space Ω , define levels of the domain:

$$\begin{aligned}\text{dom}_0 \theta &:= \{\pi \in \text{dom } \theta : \cup\pi = \Omega\} \\ \text{dom}_{i+1} \theta &:= \{\pi \in \text{dom } \theta : \cup\pi \in \rho, \rho \in \text{dom}_i \theta, \pi \notin \text{dom}_i \theta\}\end{aligned}$$

A *scope of relevance* θ is a non-negative real-valued function over partitions of events (subsets) of a probability space Ω with the following conditions:

1. At most one partition can cover any event. Formally, $\cup\pi = \cup\rho$ implies $\pi = \rho$ for any partitions π and ρ in $\text{dom } \theta$.
2. $\text{dom } \theta = \bigcup_{i=0}^{\infty} \text{dom}_i \theta$

The domain of a scope of relevance is all relevant questions. Each real value assigned to a question is a degree of relevance. The second condition in the definition means the partitions (questions) are dividing the probability space Ω in a nested hierarchical manner. Or in other words, either the event of all outcomes (Ω) or the event of an answer to a question is the event under which another question can be asked.

Scope product

$$\begin{aligned}\pi \times \rho &:= \{A \cap B : A \in \pi, B \in \rho\} \\ \pi \times B &:= \{A \cap B : A \in \pi\}\end{aligned}$$

Consider any two scopes θ and ϕ sharing probability space Ω . Define

$$\theta \times \phi := \bigcup_{i=0}^{\infty} \bigcup_{\substack{\pi \in \text{dom}_i \theta \\ \rho \in \text{dom}_i \phi \\ (\cup\pi) \cap (\cup\rho) \neq \emptyset}} \mathcal{C}(\pi, \rho)$$

where

$$\mathcal{C}(\pi, \rho) = \begin{cases} \{(\pi \times \rho) \mapsto \theta(\pi)\} & \text{when } \theta(\pi) = \phi(\rho) \\ \{\pi \mapsto \theta(\pi)\} \cup \bigcup_{A \in \pi} \{\rho \times A \mapsto \phi(\rho)\} & \text{when } \theta(\pi) > \phi(\rho) \\ \{\rho \mapsto \phi(\rho)\} \cup \bigcup_{A \in \rho} \{\pi \times A \mapsto \theta(\pi)\} & \text{when } \theta(\pi) < \phi(\rho) \end{cases}$$

Theorem (Scope product is a scope of relevance)

For any two scopes of relevance θ and ϕ sharing probability space Ω , $\theta \times \phi$ is also a scope of relevance.

Proof

TBD

Distribution Decomposition

Given a set of random variables $\{Z_i\}_{i \in I}$ with countable index set I , and probability space Ω , for each $i \in I$, define functions $B_i(s)$ and $c_i(s)$ over strings of alphabet $\{0, 1\}$ [3].

$$B_i(\epsilon) := \Omega$$

For any string s with $P(B_i(s)) > 0$,

$$\begin{aligned} c_i(s) &:= \mathbf{E}(Z_i | B_i(s)) \\ B_i(s0) &:= B_i(s) \cap \{Z_i < c_i(s)\} \\ B_i(s1) &:= B_i(s) \cap \{Z_i > c_i(s)\} \end{aligned}$$

otherwise

$$B_i(s\ell) := \emptyset$$

A new probability space Ω' is defined by extending Ω such that for each $i \in I$ and string $s \in \{0, 1\}^*$ with $P(B_i(s)) - P(\{Z_i = c_i(s)\}) > 0$, there are new disjoint events $C_{i,0}(s)$ and $C_{i,1}(s)$ such that

$$\{Z_i = c_i(s)\} = C_{i,0}(s) \cup C_{i,1}(s)$$

and for each $\ell \in \{0, 1\}$ and all events S in Ω

$$P(S \cap C_{i,\ell}(s)) = P(S \cap \{Z_i = c_i(s)\}) \frac{P(B_i(s\ell))}{P(B_i(s)) - P(\{Z_i = c_i(s)\})}$$

Define

$$\begin{aligned} B'_i(s) &:= B_i(s) \cup C_{i,\ell}(s) \\ v_i(s) &= \sum_{\ell \in \{0,1\}} P(B_i(s\ell) | B_i(s)) (c_i(s\ell) - c_i(s))^2 \\ h_i(s) &= \sum_{\ell \in \{0,1\}} h(B_i(s\ell) | B_i(s)) \\ \theta_i &:= \{\{B'_i(s0), B'_i(s1)\} \mapsto v_i(s)/h_i(s) : P(B_i(s)) > 0\} \end{aligned}$$

The scope product across all θ_i is the distribution decomposition of $\{Z_i\}_{i \in I}$.

TODO: scope product across countable index set

Scoped Entropy

Given a scope of relevance θ , let:

$$\begin{aligned}
 \mathcal{A}_\theta(0) &:= \{\Omega\} \\
 \mathcal{A}_\theta(i+1) &:= \{A \cap B : A \in \mathcal{A}_\theta(i), B \in \pi, \pi \in \text{dom}_i \theta\} \\
 h(A|B) &:= P(A|B) \log_2 \frac{1}{P(A|B)} \\
 U_\pi(Q) &:= P(\cup \pi | Q) \sum_{S \in \pi} h(S | \cup \pi \cap Q) \\
 U_\theta(Q) &:= \sum_{i=0}^{\infty} \sum_{A \in \mathcal{A}_\theta(i)} P(A|Q) \sum_{\pi \in \text{dom}_i \theta} \theta(\pi) U_\pi(A \cap Q) \\
 U_\theta(\pi) &:= \sup \left\{ \sum_{Q \in \rho} P(Q) U_\theta(Q) : \text{finite partition } \rho \text{ coarser than } \pi \right. \\
 &\quad \left. \text{(or equal to)} \right\} \\
 V_\theta(Q) &:= U_\theta(\Omega) - U_\theta(Q) \\
 V_\theta(\pi) &:= U_\theta(\{\Omega\}) - U_\theta(\pi) \\
 \ker X &:= \{\{\omega \in \Omega : X(\omega) = v\} : v \text{ in the range of } X\} \\
 U_\theta(X) &:= U_\theta(\ker X) \\
 V_\theta(X) &:= V_\theta(\ker X)
 \end{aligned}$$

Shannon Entropy Equality

Given any finite random object X , a *unirelevant decomposition* is the trivial mapping of X to a scope θ assigning 1 to a single partition consisting of the events for each value taken by X :

$$\begin{aligned}
 \text{dom } \theta &= \{\ker X\} \\
 \theta(\ker X) &= 1
 \end{aligned}$$

Theorem: Unirelevant Scoped Entropy equals Shannon Entropy

Given any finite random objects X and Y and scope θ equal to the *unirelevant decomposition* of X ,

$$U_\theta(Y) = H(X|Y)$$

It follows as a corollary that

$$V_\theta(X) = H(X)$$

Proof

Consider any single partition scope $\theta = \{\pi \mapsto 1\}$. Both $\mathcal{A}_\theta(i)$ and $\text{dom}_i \theta$ are only non-empty at $i = 0$ and equal $\{\Omega\}$ and $\{\pi\}$ respectively. Thus by definition

$$\begin{aligned}
 U_\theta(Y) &= \sum_{Q \in \ker Y} P(Q) P(\Omega|Q) \theta(\pi) P(\cup \pi | \Omega, Q) \sum_{B \in \pi} h(B | \cup \pi, \Omega, Q) \\
 &= \sum_{Q \in \ker Y} P(Q) \sum_{B \in \pi} h(B|Q) \\
 &= \sum_{Q \in \ker Y} \sum_{B \in \pi} P(Q \cap B) \log_2 \frac{1}{P(B|Q)} \\
 &= H(X|Q)
 \end{aligned}$$

Proof of the corollary is:

$$V_\theta(X) = U_\theta(\Omega) - U_\theta(X) = H(X|\{\Omega\}) - H(X|X) = H(X)$$

Benefit of scoped entropy over variance

A feature analogous to mutual information is the following:

For any random variable X and random object Y of finitely many values,

$$V_{\partial(X)}(Y) = 0 \quad \text{if and only if} \quad X \text{ and } Y \text{ are independent}$$

References

1. Shannon CE, Weaver W. The mathematical theory of communication. Urbana: Univ. of Illinois Press; 1998.
2. Ash RB, Doléans-Dade C, Ash RB. Probability and measure theory. 2nd ed. San Diego: Harcourt/Academic Press; 2000.
3. Hopcroft JE, Ullman JD. Introduction to automata theory, languages, and computation. Reading, Mass: Addison-Wesley; 1979.