The Cost of Information: The Case of Constant Marginal Cost
Pomatto, Strack, & Tamuz 2023

Willy Chen

May 2, 2024

Motivation

While the literature on the value of information is well established, modelling the cost of producing information has not gotten the same attention.

The most common measure of cost is *Mutual Information Costs*, defined as the expected change of Shannon's entropy:

$$E\left[H(q) - H(p)\right]$$

where q is the distribution of prior beliefs, p is the distribution of posterior beliefs, and

$$H(p) = -\sum_{i \in \Theta} p_i ln(p_i)$$

In this Paper

This paper considers the case of constant marginal cost in information production and axiomatizes a family of cost functions satisfying:

In this Paper

This paper considers the case of constant marginal cost in information production and axiomatizes a family of cost functions satisfying:

- Experiments that are more informative in the sense of Blackwell order are more costly

In this Paper

This paper considers the case of constant marginal cost in information production and axiomatizes a family of cost functions satisfying:

- Experiments that are more informative in the sense of Blackwell order are more costly
- The cost of generating two independent experiments is the sum of the cost of generating each of the two experiments

In this Paper

This paper considers the case of constant marginal cost in information production and axiomatizes a family of cost functions satisfying:

- Experiments that are more informative in the sense of Blackwell order are more costly
- The cost of generating two independent experiments is the sum of the cost of generating each of the two experiments
- Cost of generating an experiment with probability half equals half the cost of generating it with probability one.

Main Result

Take a finite set Θ of states, and an experiment μ that produces a signal $s \in S$ with probability $\mu_i(s)$ in state $i \in \Theta$. Any continuous cost function that satisfy the three axioms can be uniquely represented by the *Log-Likelihood Ratio Cost* function (LLR):

$$C(\mu) = \sum_{i,j \in \Theta} \beta_{ij} \underbrace{\left(\sum_{s \in S} \mu_i(s) ln \left(\frac{\mu_i(s)}{\mu_j(s)} \right) \right)}_{\text{Kullback-Leibler Divergence}}$$

If Θ is one-dimensional, then $\exists\, \kappa \in \mathbb{R}_+$ such that

$$\beta_{ij} = \frac{\kappa}{(i-j)^2}$$

Equivalence Axiom

Definition: Let μ and ν be two experiments inducing the distributions π_μ and π_ν over posteriors p given the uniform prior q. Then μ **dominates** ν in Blackwell order if

$$\int_{P(\Theta)} f(p) d\pi_\mu(p) \geq \int_{P(\Theta)} f(p) d\pi_\nu(p)$$

for every convex function $f : P(\Theta) \to \mathbb{R}$.

Equivalence Axiom

Definition: Let μ and ν be two experiments inducing the distributions π_μ and π_ν over posteriors p given the uniform prior q. Then μ **dominates** ν in Blackwell order if

$$\int_{P(\Theta)} f(p) d\pi_\mu(p) \geq \int_{P(\Theta)} f(p) d\pi_\nu(p)$$

for every convex function $f : P(\Theta) \to \mathbb{R}$.

Axiom 1: If μ and ν dominates each other in Blackwell order, then the cost function is such that $C(\mu) = C(\nu)$.

This implies that if the experiments are chosen optimally by the decision maker, then the cost function can be identified with its lower envelope.

Independence Axiom

Definition: Given two experiments $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$, the product of these two experiments is:

$$\mu \otimes \nu = (S \times T, (\mu_i \times \nu_i))$$

Independence Axiom

Definition: Given two experiments $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$, the product of these two experiments is:

$$\mu \otimes \nu = (S \times T, (\mu_i \times \nu_i))$$

Axiom 2: The cost of performing two independent experiments is the sum of their costs:

$$C(\mu \otimes \nu) = C(\mu) + C(\nu)$$

Axiom 1 and 2 together implies that any uninformative experiment has cost 0 and any completely informative experiment has infinite cost.

Homogeneity Axiom

Definition: Given an experiment $\mu = (S, (\mu_i))$, an uninformative signal $o \notin S$, and a probability $\alpha \in [0, 1]$, the dilution ν of the experiment μ is

$$\nu = \alpha \cdot \mu = (S \cup \{o\}, (\nu_i))$$

where $\nu_i(E) = \alpha \mu_i(E)$ for every measurable $E \subseteq S$ and $\nu_i(\{o\}) = 1 - \alpha$.

Homogeneity Axiom

Definition: Given an experiment $\mu = (S, (\mu_i))$, an uninformative signal $o \notin S$, and a probability $\alpha \in [0, 1]$, the dilution ν of the experiment μ is

$$\nu = \alpha \cdot \mu = (S \cup \{o\}, (\nu_i))$$

where $\nu_i(E) = \alpha\mu_i(E)$ for every measurable $E \subseteq S$ and $\nu_i(\{o\}) = 1 - \alpha$.

Axiom 3: A diluted experiment $\alpha \cdot \mu$ costs $\alpha \cdot C(\mu)$

Continuity

For every experiment μ, let $l_{ij} = \frac{d\mu_i(s)}{d\mu_j(s)}$ be the log-likelihood ratio and $\bar{\mu}$ be the distribution of log-likelihood ratios conditional on state i.

Let $d_{tv}(\bar{\mu}_i, \bar{\nu}_i) = \sup \mid \bar{\mu}_i(A) - \bar{\nu}_i(A) \mid$ be the total-variation distance.

Given a vector $\alpha \in \mathbb{N}^{|\Theta|}$, let $M_i^{\mu}(\alpha) = \int\limits_S \mid \prod\limits_{k \neq i} l_{ik}^{\alpha_k} \mid d\mu_i(s)$ be the $\alpha-$moment of the vector of log-likelihood ratios $(l_{ik})_{i \neq k}$. Given an upper-bound $N \geq 1$, the total-variation distance d_N is:

$$d_N(\mu, \nu) = \max_{i \in \Theta} d_{tv}(\bar{\mu}_i, \bar{\nu}_i) + \max_{i \in \Theta} \max_{\alpha \in \{0,\ldots,N\}^{|\Theta|}} \mid M_i^{\mu}(\alpha) - M_i^{\nu}(\alpha) \mid$$

Axiom 4: For some $N \geq 1$, the function C is uniformly continuous with respect to d_N.

Limiting the Domain

The setup here rules out experiments that, with positive probability, allow the decision-maker to be certain that a state did not happen because such experiments would have infinite cost under these axioms.

Consider the scenario with 100 balls in an urn and they are either all red or all blue. In this case, drawing a single ball from the urn perfectly reveals the state, and would have infinite LLR cost.

If instead, 99 balls are blue (red) and only 1 is red (blue), than drawing a single ball becomes feasible as it does not perfectly reveal the state.

Unique Representation

Theorem 1: An information cost function C satisfies Axioms 1-4 if and only if there exists a collection $\{\beta_{ij} \in \mathbb{R}_+ \mid i, j \in \Theta,\ i \neq j\}$ such that for every experiment $\mu = (S, (\mu_i))$,

$$C(\mu) = \sum_{i,j} \beta_{ij} \int_S ln \left(\frac{d\mu_i(s)}{d\mu_j(s)} \right) d\mu_i(s)$$

and the collection of β_{ij} is unique given C.

Unique Representation

Theorem 1: An information cost function C satisfies Axioms 1-4 if and only if there exists a collection $\{\beta_{ij} \in \mathbb{R}_+ \mid i, j \in \Theta,\ i \neq j\}$ such that for every experiment $\mu = (S, (\mu_i))$,

$$C(\mu) = \sum_{i,j} \beta_{ij} \int_S ln \left(\frac{d\mu_i(s)}{d\mu_j(s)} \right) d\mu_i(s)$$

and the collection of β_{ij} is unique given C.

The coefficient β_{ij} measures the marginal cost of increasing the expected log-likelihood ratio between states i and j, conditional on the state being i.

One-Dimensional States

Let T be a finite subset of W, a nonempty open interval of $\mathbb{R}$.

Axiom A: For all $\Theta, \Xi \in T$ such that $|\Theta| = |\Xi|$ and for all $i, j \in \Theta$ and $k, l \in \Xi$,

$$|i - j| = |k - l| \;\Rightarrow\; \beta_{ij}^{\Theta} = \beta_{kl}^{\Xi}$$

One-Dimensional States

Let T be a finite subset of W, a nonempty open interval of $\mathbb{R}$.

Axiom A: For all $\Theta, \Xi \in T$ such that $|\Theta| = |\Xi|$ and for all $i, j \in \Theta$ and $k, l \in \Xi$,

$$|i - j| = |k - l| \;\Rightarrow\; \beta_{ij}^{\Theta} = \beta_{kl}^{\Xi}$$

Given Θ, let ζ^{Θ} denote the experiment that produces noisy signals $s = i + \varepsilon$ where $\varepsilon \sim N(0, 1)$.

Axiom B: For all $\Theta, \Xi \in T$, $C^{\Theta}(\zeta^{\Theta}) = C^{\Xi}(\zeta^{\Xi})$.

One-Dimensional States

Let T be a finite subset of W, a nonempty open interval of $\mathbb{R}$.

Axiom A: For all $\Theta, \Xi \in T$ such that $|\Theta| = |\Xi|$ and for all $i, j \in \Theta$ and $k, l \in \Xi$,

$$|i - j| = |k - l| \;\Rightarrow\; \beta_{ij}^{\Theta} = \beta_{kl}^{\Xi}$$

Given Θ, let ζ^{Θ} denote the experiment that produces noisy signals $s = i + \varepsilon$ where $\varepsilon \sim N(0, 1)$.

Axiom B: For all $\Theta, \Xi \in T$, $C^{\Theta}(\zeta^{\Theta}) = C^{\Xi}(\zeta^{\Xi})$.

Proposition: The collection C^{Θ} satisfies Axioms A and B if and only if $\exists \kappa \in \mathbb{R}_{++}$ such that for all $i, j \in \Theta \in T$,

$$\beta_{ij}^{\Theta} = \frac{\kappa}{n(n-1)} \frac{1}{(i-j)^2}, \quad n = |\Theta|$$

Example: LLR Cost for Binary States/Experiments

Consider the set of states $\Theta = \{H, L\}$ and the experiment $\nu^p = (S, (\nu_i))$ also yields binary signals $s \in S = \{0, 1\}$ where $\nu_H = B(p)$ and $\nu_L = B(1-p)$ for some $p > \frac{1}{2}$.

In this case, the cost increases in p and is given by:

$$C(\nu^p) = (\beta_{HL} + \beta_{LH}) \left[p ln \left(\frac{p}{1-p} \right) + (1-p) ln \left(\frac{1-p}{p} \right) \right]$$

Example: The Optimal Experiment with Multiple States

Since C is monotone w.r.t. Blackwell order, WLOG, we can restrict attention to experiments where the set of realizations S equals the set of actions A and the decision maker chooses $a = s$ as the optimal action given signal s from the experiment.

Example: The Optimal Experiment with Multiple States

Since C is monotone w.r.t. Blackwell order, WLOG, we can restrict attention to experiments where the set of realizations S equals the set of actions A and the decision maker chooses $a = s$ as the optimal action given signal s from the experiment.

The optimal experiment is thus

$$\mu^* \in \underset{\mu \in P(A)^n}{argmax} \sum_{i \in \Theta} q_i \left(\sum_{a \in A} \mu_i(a) u(a, i) \right) - C(\mu)$$

$\sum_{i \in \Theta} q_i \left(\sum_{a \in A} \mu_i^*(a) u(a, i) \right) - C(\mu^*)$ is upper-semicontinuous and concave, and it always exists.

Example: The Optimal Experiment with Multiple States

The optimal experiment yields the optimal choice probabilities, meaning the marginal benefit of choosing an action a marginally more often than b must exactly offset its marginally cost.

Example: The Optimal Experiment with Multiple States

The optimal experiment yields the optimal choice probabilities, meaning the marginal benefit of choosing an action a marginally more often than b must exactly offset its marginally cost.

Marginal benefit is:

$$MB_i(a, b) = q_i[u(a, i) - u(b, i)]$$

Example: The Optimal Experiment with Multiple States

The optimal experiment yields the optimal choice probabilities, meaning the marginal benefit of choosing an action a marginally more often than b must exactly offset its marginally cost.

Marginal benefit is:

$$MB_i(a, b) = q_i[u(a, i) - u(b, i)]$$

Marginal cost is:

$$MC_i(a, b) = \sum_{j \in \Theta} \beta_{ij} \left(ln\left(\frac{\mu_i(a)}{\mu_j(a)}\right) - ln\left(\frac{\mu_i(b)}{\mu_j(b)}\right) \right) - \sum_{j \in \Theta} \beta_{ji} \left(\frac{\mu_j(a)}{\mu_i(a)} - \frac{\mu_j(b)}{\mu_i(b)} \right)$$

Example: The Optimal Experiment with Multiple States

The optimal experiment yields the optimal choice probabilities, meaning the marginal benefit of choosing an action a marginally more often than b must exactly offset its marginally cost.

Marginal benefit is:

$$MB_i(a,b) = q_i[u(a,i) - u(b,i)]$$

Marginal cost is:

$$MC_i(a,b) = \sum_{j\in\Theta} \beta_{ij} \left(ln\left(\frac{\mu_i(a)}{\mu_j(a)}\right) - ln\left(\frac{\mu_i(b)}{\mu_j(b)}\right)\right) - \sum_{j\in\Theta} \beta_{ji} \left(\frac{\mu_j(a)}{\mu_i(a)} - \frac{\mu_j(b)}{\mu_i(b)}\right)$$

Proposition: $\forall a, b \in supp(\mu^*),\ MB_i(a,b) = MC_i(a,b)$