The Cost of Information

Luciano Pomatto[•] Philipp Strack[◦] Omer Tamuz[•]

[•]Caltech and [◦]Yale

Information as a Commodity

Much of contemporary economic theory is built on the idea that information is scarce and valuable.

> *"In a world of uncertainty it was no great leap to realize that information is valuable in an economic sense. Nevertheless, it has proved difficult to frame a general theory of information as an economic commodity..." —Arrow (1984)*

Key questions studied in the literature:

- How to represent information?
- How to measure its value?
- How to model its cost?

Information as a Commodity

Much of contemporary economic theory is built on the idea that information is scarce and valuable.

> *"In a world of uncertainty it was no great leap to realize that information is valuable in an economic sense. Nevertheless, it has proved difficult to frame a general theory of information as an economic commodity..." —Arrow (1984)*

Key questions studied in the literature:

- How to represent information?
- How to measure its value?
- How to model its cost?

This paper: a model for the cost of generating information.

Representing Information

Following Bohnenblust-Shapley-Sherman (1949) and Blackwell (1951), we represent information as a Blackwell experiment: a signal whose distribution depends on an unknown state.

Representing Information

Following Bohnenblust-Shapley-Sherman (1949) and Blackwell (1951), we represent information as a Blackwell experiment: a signal whose distribution depends on an unknown state.

Formally:

- $\Theta = \{1, \ldots, n\}$: unknown **state** of the world.
- S: set of possible signal realization.
- An **experiment** is a collection

$$\mu = (S, (\mu_i)_{i \in \Theta})$$

associating to each $i \in \Theta$ a distribution (probability measure) μ_i defined on S.

Representing Information

Following Bohnenblust-Shapley-Sherman (1949) and Blackwell (1951), we represent information as a Blackwell experiment: a signal whose distribution depends on an unknown state.

Formally:

- $\Theta = \{1, \ldots, n\}$: unknown **state** of the world.
- S: set of possible signal realization.
- An **experiment** is a collection

$$\mu = (S, (\mu_i)_{i \in \Theta})$$

  associating to each $i \in \Theta$ a distribution (probability measure) μ_i defined on S.

- Example: $\Theta = \{1, 2\}$, $S = \{R, B\}$, $\mu_1(R) = \mu_2(B) = 0.6$

Representing Information

Following Bohnenblust-Shapley-Sherman (1949) and Blackwell (1951), we represent information as a Blackwell experiment: a signal whose distribution depends on an unknown state.

Formally:

- $\Theta = \{1, \ldots, n\}$: unknown **state** of the world.
- S: set of possible signal realization.
- An **experiment** is a collection

$$\mu = (S, (\mu_i)_{i \in \Theta})$$

  associating to each $i \in \Theta$ a distribution (probability measure) μ_i defined on S.

- Example: $\Theta = \{1, 2\}$, $S = \{R, B\}$, $\mu_1(R) = \mu_2(B) = 0.6$
- Example: $\Theta = \{1, 2\}$, $S = \mathbb{R}$, $\mu_1 = N(+1, \sigma^2)$, $\mu_2 = N(-1, \sigma^2)$.

Approaches to modeling the cost of information

- Parametric family of experiments:

  ...the agent acquires a signal $S \sim N(i, \sigma^2)$ *at cost* $c(\sigma)$*...*

- Wald's sequential sampling: there is a single experiment μ. Agent chooses how many independent copies to carry out, pays linear cost.

- A non-parametric information acquisition model where the cost of information is described by a functional over all experiments:

$$C : \mathcal{E} \longrightarrow \mathbb{R}_+$$

Approaches to modeling the cost of information

- Parametric family of experiments:

  ...the agent acquires a signal $S \sim N(i, \sigma^2)$ *at cost* $c(\sigma)$*...*

- Wald's sequential sampling: there is a single experiment μ. Agent chooses how many independent copies to carry out, pays linear cost.

- A non-parametric information acquisition model where the cost of information is described by a functional over all experiments:

$$C : \mathcal{E} \longrightarrow \mathbb{R}_+$$

- Following Sims (2003), most of the literature has taken C to be the mutual information between the signal and state, or equivalently the expected reduction in entropy.

This paper

- We propose an axiomatic approach: state axioms that describe properties of the cost function, and explore their implications.
- We study axioms that capture constant marginal costs.
- Leads to testably different predictions from mutual information cost.

Literature

- **Measures of Informativeness and the Value of Information:** Kullback and Leibler (1951), Marschak (1959), Arrow (1971), Moscarini and Smith (2001, 2002), Cabrales, Gossner, Serrano (2017), Frankel and Kamenica (2018).
- **Rational Inattention:** Sims (2003, 2010), Caplin and Dean (2013, 2015) Caplin (2016), Mackowiak, Matejka, and Wiederholt (2018), etc.
- **Information Theory:** Survey by Csiszar (2008).
- **Decision Theory and Revealed Preferences:** Caplin and Dean (2015), Caplin Deah Leahy (2016), Oliveira et. al (2017), Denti (2018), Mensch (2018).
- **Alternatives to Mutual Information Cost:** Hebert and Woodford (2017), Morris and Strack (2018).

Model

- $\Theta = \{1, \ldots, n\}$: unknown **state** of the world.

- An **experiment** is a measurable space (S, Σ) and a collection of mutually absolutely continuous probability measures

$$\mu = (S, (\mu_i)_{i \in \Theta})$$

- $\ell_{ij}(s) = \log \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)$ log-likelihood ratio

- $\mathcal{E}$: class of experiments s.t. all moments of ℓ_{ij} are finite.

- $C : \mathcal{E} \longrightarrow \mathbb{R}_+$: **cost function**.

Characterization

Axiom 1 If μ dominates ν in the Blackwell order, then $C(\mu) \geq C(\nu)$.

Characterization

Axiom 1 If μ dominates ν in the Blackwell order, then $C(\mu) \geq C(\nu)$.

- If μ dominates ν in the Blackwell order then μ "contains more information about the state".
- ν can be generated from μ by adding noise (Blackwell's Theorem).
- "Free disposal".
- There are many physical ways to generate a particular experiment μ. $C(\mu)$ is the cost of the cheapest one.

Axiom 2: Independent Experiments

Performing two independent experiments

$$\mu = (S, (\mu_i)_{i \in \Theta}) \text{ and } \nu = (T, (\nu_i)_{i \in \Theta})$$

is represented as acquiring information by means of the single experiment

$$\mu \otimes \nu = (S \times T, (\mu_i \times \nu_i)_{i \in \Theta})$$

Conditional on the state, the observations are independent.

- E.g. Drawing independent samples from two populations.

Axiom 2: Independent Experiments

Additivity with respect to independent experiments:

Axiom 2 $\quad C(\mu \otimes \nu) = C(\mu) + C(\nu)$

Axiom 2: Independent Experiments

Additivity with respect to independent experiments:

Axiom 2 $\quad C(\mu \otimes \nu) = C(\mu) + C(\nu)$

- Constant marginal cost.
- The cost of surveying $2n$ people is twice the cost of surveying n.
- Non-parametric generalization of Wald's sequential sampling model.
- Natural baseline assumption to study.
- Under mutual information $C(\mu \otimes \nu) < C(\mu) + C(\nu)$.

Axiom 3: Diluted Experiments

We study experiments that generate information with probability α.

Fix $\mu = (S, (\mu_i))$. In the diluted experiment $\alpha \cdot \mu$:

- with probability α, μ is performed
- with probability $1 - \alpha$, an uninformative signal $o \notin S$ is observed.

We posit that the cost of such an experiment is linear in the probability of success α.

Axiom 3: Diluted Experiments

Linearity in the probability of success for diluted experiments:

Axiom 3 $\quad C(\alpha \cdot \mu) = \alpha C(\mu)$

Axiom 3: Diluted Experiments

Linearity in the probability of success for diluted experiments:

Axiom 3 $\quad C(\alpha \cdot \mu) = \alpha C(\mu)$

- Constant marginal cost.
- $\leq$: randomization is free.
- $\geq$: cost of repeating is linear.
- Also holds under mutual information cost, and furthermore under any posterior separable cost.

Axiom 4: Continuity

We define increasingly fine topologies d_N on the set of experiments.

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

Axiom 4: Continuity

We define increasingly fine topologies d_N on the set of experiments.

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

- P_i^μ : distribution of LLRs conditional on state i.
- **Moments:** For every $\alpha \in \mathbb{N}^n$

$$M_i^\mu(\alpha) = \int \ell_{1i}^{\alpha_1} \cdots \ell_{ni}^{\alpha_n} \, \mathrm{d}\mu_i$$

- **Distance over experiments:**

$$d_N(\mu,\nu) = \max_{i \in \Theta} d_{tv}(P_i^\mu, P_i^\nu) + \max_{i \in \Theta} \max_{\alpha \in \{0,\ldots,N\}^n} \left| M_i^\mu(\alpha) - M_i^\nu(\alpha) \right|$$

- Closeness in total-variation and similar LLRs moments.
- Redundant axiom when $|\Theta| = 2$.

Characterization

Theorem

C satisfies Axioms 1-4 if and only if there exist unique (β_{ij}) in $\mathbb{R}_+$ s.t.

$$C(\mu) = \sum_{i,j \in \Theta} \beta_{ij} \int_S \log \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s) \, \mathrm{d}\mu_i(s).$$

Model: LLR cost

$$C(\mu) = \sum_{i,j\in\Theta} \beta_{ij} \underbrace{\int \log \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)\,\mathrm{d}\mu_i(s)}_{\text{KL-divergence } D(\mu_i\|\mu_j)}$$

- LLR Cost.
- $\beta_{ij} \geq 0$: one parameter for each ordered pair (i, j) of distinct states.
- $D(\mu_i\|\mu_j)$.
  1. Kullback-Leibler (KL) divergence.
  2. Expected LLR between i and j conditional on i.
  3. Proposed as measure of informativeness by Kullback and Leibler (1951).
- The higher β_{ij} the more costly it is to distinguish between i and j.

Bayesian Representation

Consider a Bayesian with prior q who observes an experiment μ. Denote the distribution of her posterior by π_μ.

Theorem

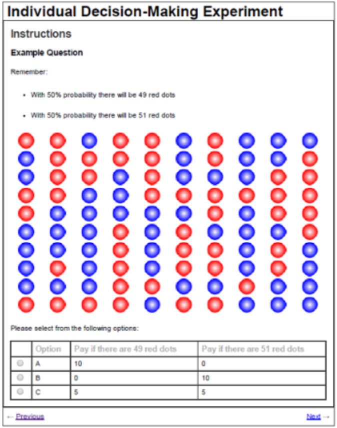
$$C(\mu) = \int_{\Delta(\Theta)} F(p) - F(q) \, \mathrm{d}\pi_\mu(p)$$

where

$$F(p) = \sum_{i,j \in \Theta} \gamma_{ij} p_i \log \frac{p_i}{p_j} \qquad \gamma_{ij} = \beta_{ij}/q_i$$

Perception Tasks

Example: Bet on Red or Blue



Individual Decision-Making Experiment

Instructions

Example Question

Remember:

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

Please select from the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	10	0
○	B	0	10
○	C	5	5

← Previous Next →

Figure: Experiment by Dean and Neligh (2017)

Example: Bet on Red or Blue

- A DM sees red and blue dots on a screen.

- The state is the number of blue dots $i \in \{1, \dots, 99\}$.
  Prior is uniform.

- The DM must bet on red or blue.

Example: Bet on Red or Blue

- A DM sees red and blue dots on a screen.
- The state is the number of blue dots $i \in \{1, \ldots, 99\}$.
  Prior is uniform.
- The DM must bet on red or blue.
- Assume she solves the information acquisition problem:

$$\max_{\mu} \int_{\Delta(\Theta)} \left(\max_{a \in A} u_a \cdot p \right) \mathrm{d}\pi_{\mu}(p) - C(\mu).$$

Example: Bet on Red or Blue

- A DM sees red and blue dots on a screen.
- The state is the number of blue dots $i \in \{1, \ldots, 99\}$.
  Prior is uniform.
- The DM must bet on red or blue.
- Assume she solves the information acquisition problem:

$$\max_{\mu} \int_{\Delta(\Theta)} \left(\max_{a \in A} u_a \cdot p \right) \mathrm{d}\pi_{\mu}(p) - C(\mu).$$

Compare two predictions:

1. C : mutual information cost.
2. C : LLR cost

Example: Bet on Red or Blue

- A DM sees red and blue dots on a screen.

- The state is the number of blue dots $i \in \{1, \dots, 99\}$.
  Prior is uniform.

- The DM must bet on red or blue.

- Assume she solves the information acquisition problem:

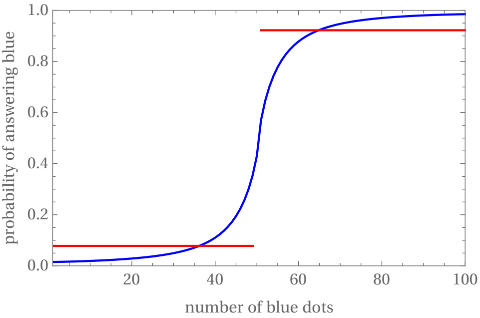
$$\max_{\mu} \int_{\Delta(\Theta)} \left(\max_{a \in A} u_a \cdot p \right) \mathrm{d}\pi_{\mu}(p) - C(\mu).$$

Compare two predictions:

1. C : mutual information cost.
2. C : LLR cost, setting $\beta_{ij} = 1/(i-j)^2$.

Example: Bet on Red or Blue

- **Mutual information cost**
  The DM will, in every state $i \in \{0, \ldots, 99\}$, have the same probability of being correct.
- **LLR cost**
  The subject will be correct with probability that is higher when the difference between red and blue is higher.



Binary Choice

Example: Bet on Red or Blue

- There are are two states and two actios: $\Theta = A = \{1, 2\}$.
- The DM has a uniform prior over the state.
- If action matches state payoff is v, otherwise 0.

Example: Bet on Red or Blue

- There are are two states and two actios: $\Theta = A = \{1, 2\}$.
- The DM has a uniform prior over the state.
- If action matches state payoff is v, otherwise 0.
- Assume she solves the information acquisition problem:

$$\max_{\mu} \int_{\Delta(\Theta)} \left(\max_{a \in A} u_a \cdot p \right) \mathrm{d}\pi_{\mu}(p) - C(\mu).$$

Example: Bet on Red or Blue

- There are are two states and two actios: $\Theta = A = \{1, 2\}$.
- The DM has a uniform prior over the state.
- If action matches state payoff is v, otherwise 0.
- Assume she solves the information acquisition problem:

$$\max_{\mu} \int_{\Delta(\Theta)} \left(\max_{a \in A} u_a \cdot p \right) \mathrm{d}\pi_{\mu}(p) - C(\mu).$$

Compare two predictions:

1. C : mutual information cost.
2. C : LLR cost, setting $\beta_{12} = \beta_{21}$.

Example: Bet on Red or Blue

- Denote $p_m(v)$ the probability of mismatching when the payoff for matching is v.
- Fix cost so that $p_m(\$1) = 20\%$.
- What is $p_m(\$10)$?

Example: Bet on Red or Blue

- Denote $p_m(v)$ the probability of mismatching when the payoff for matching is v.
- Fix cost so that $p_m(\$1) = 20\%$.
- What is $p_m(\$10)$?

Compare two predictions:

1. Under mutual information cost:
    1. $p_m(v) \sim \mathrm{e}^{-v}$ for large v.
    2. $p_m(10v_0) \approx 0.000001$.
2. Under LLR cost
    1. $p_m(v) \sim 1/v$ for large v.
    2. $p_m(10v_0) \approx 1/60$

Conclusion

- We propose a new cost of information: LLR cost.
  1. Allows flexible choice of (almost) any experiment.
  2. Based on simple economic assumptions that formalize constant marginal cost.
  3. Has simple, tractable form.
  4. Yields testable predictions that differ from entropy cost.
- Follow-up questions.
  1. Increasing / decreasing marginal costs.
  2. Continuum of states.
  3. Applications.

Thank you!