Lecture 6 - Rational Inattention

Rational Inattention

- Fairly recent field of study in microeconomics theory - originally studied in macro.
- Today:
  - Matekja & McKay (2015 AER)
  - Caplin, Dean & Leahy (2019 Restud)
  - Pomatto, Strack & Tamuz (2022 AER)

Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model

Matekja & McKay

Motivation

- Individuals often need to pay attention to evaluate objects when choosing between them.
  - Utility isn't immediate.

Motivation

- Individuals often need to pay attention to evaluate objects when choosing between them.
  - Utility isn't immediate.

- Natural economic assumption: attention is chosen optimally.

Motivation

- Individuals often need to pay attention to evaluate objects when choosing between them.
  - Utility isn't immediate.

- Natural economic assumption: attention is chosen optimally.

- Matekja & McKay impose:
  - No restriction on the kind of information you can acquire.
  - A cost function that is the entropy reduction.

Motivation

- Individuals often need to pay attention to evaluate objects when choosing between them.
  - Utility isn't immediate.
- Natural economic assumption: attention is chosen optimally.
- Matekja & McKay impose:
  - No restriction on the kind of information you can acquire.
  - A cost function that is the entropy reduction.
- They solve this problem and find logit like choice probability rule.

Motivation

- Individuals often need to pay attention to evaluate objects when choosing between them.
  - Utility isn't immediate.

- Natural economic assumption: attention is chosen optimally.

- Matekja & McKay impose:
  - No restriction on the kind of information you can acquire.
  - A cost function that is the entropy reduction.

- They solve this problem and find logit like choice probability rule.

- Side note: you may wonder what kind of choices can be rationalized by more general attention setting: see Caplin & Dean (2015 AER).

Model Setup: Decision Environment

- **Actions:** The decision maker chooses an action from a finite set

$$A = \{1, \ldots, N\}.$$

- **State:** The state of nature is a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)$ representing the payoffs of each action.

- **Prior:** The decision maker holds a prior belief $G(\boldsymbol{v})$ over the state.

Information Acquisition

- **Information Strategy:** Before choosing an action, the decision maker selects an information strategy.

$$F(s, \boldsymbol{v}) \quad \text{with marginal} \int F(\boldsymbol{s}, \boldsymbol{v})\, ds = G(\boldsymbol{v}).$$

- **Signal:** The signal s refines the decision maker's belief about v.
- **Action Rule:** After observing s, the decision maker chooses the action that maximizes expected payoff.

Cost of Information

- The cost of processing information is measured via an entropy-based function:

$$c(F) = \lambda\Big(H(G) - \mathbb{E}_s\big[H(F(\cdot \mid s))\big]\Big),$$

  where $A > 0$ is the unit cost of information.

- $H(F) = -\sum_k P_k \log(P_k)$, if states are finite.

- **Interpretation:** This cost equals the mutual information between the state and the signal.

The Two-Stage Decision Problem

- **Stage 1:** Choose an information structure $F(s \mid \boldsymbol{v})$.
- **Stage 2:** Given the updated (posterior) belief, choose the action $a(s)$ that maximizes expected payoff:

$$V(B) = \max_{a \in A} \mathbb{E}_B[v_a].$$

- **Overall Objective:** Maximize the ex-ante expected payoff minus the cost of information:

$$\max_{F \in \Delta(\mathbb{R}^{2N})} \int_{\boldsymbol{v}} \int_{s} V(F(\cdot|s)) F(ds|\boldsymbol{v}) G(d\boldsymbol{v}) - c(F)$$

$$\text{such that} \int_{s} F(ds, \boldsymbol{v}) = G(\boldsymbol{v}), \ \forall \boldsymbol{v} \in \mathbb{R}^N.$$

Information and Action Strategies

Joint Distribution of Selected Action and State.
The combination of an information strategy and an action strategy together induce a joint distribution between the selected action and the state. For any strategy (F, a), let

$$S_i := \{\, s \in \mathbb{R}^n : a\big(F(\cdot \mid s)\big) = i\}$$

be the set of signals that lead to the action i. Let

$$P_i(\boldsymbol{v}) \;=\; \int_{s\,\in\,S_i} F\big(ds \mid \boldsymbol{v}\big).$$

Let $\mathcal{P}$ denote the collection $\{P_i(\boldsymbol{v})\}_{i=1}^{N}$ and let

$$P_i^0 \;=\; \int P_i(\boldsymbol{v})\, G(d\boldsymbol{v})$$

be the unconditional probability of selecting action i.

Reformulation in Terms of Choice Probabilities

- Instead of focusing on signals, one can work with the induced *state-contingent choice probabilities.*

- An important property of rationally inattentive behavior is that under an optimal strategy each action is selected in at most one posterior. Receiving distinct signals that lead to the same action is inefficient. This behavior follows from the convexity of the entropy-based cost function.

- Then under an optimal F and choice of a given s

$$\int_{\boldsymbol{v}}\int_{s} V\big(F(\cdot \mid s)\big)\, F(\mathrm{d}s \mid \boldsymbol{v})\, G(\mathrm{d}\boldsymbol{v}) = \sum_{i=1}^{N} \int_{\boldsymbol{v}} v_i\, P_i(\boldsymbol{v})\, G(\mathrm{d}\boldsymbol{v}),$$

$$\hat{c}(F) \;=\; c(P, G) = \lambda\left(-\sum_{i=1}^{N} P_i^0 \,\log P_i^0 \;+\; \int \Big(\sum_{i=1}^{N} P_i(\boldsymbol{v})\,\log P_i(\boldsymbol{v})\Big)\, G(\mathrm{d}\boldsymbol{v})\right),$$

- Can rewrite the optimization as:

$$\max_{P=\{P_i(\boldsymbol{v})\}_{i=1}^N} \sum_{i=1}^N \int v_i \, P_i(\boldsymbol{v}) \, G(\mathrm{d}\boldsymbol{v}) \; - \; c\big(P, G\big), \tag{1}$$

$$\text{subject to} \quad P_i(\boldsymbol{v}) \;\geq\; 0 \quad \forall\, i, \; \forall\, \boldsymbol{v} \in \mathbb{R}^N, \tag{2}$$

$$\sum_{i=1}^N P_i(\boldsymbol{v}) \;=\; 1 \quad \forall\, \boldsymbol{v} \in \mathbb{R}^N. \tag{3}$$

Main Result (Theorem 1)

Theorem 1

For $\lambda > 0$, the optimal strategy satisfies the conditional choice probabilities

$$P_i(\boldsymbol{v}) = \frac{P_i^0\, e^{v_i/\lambda}}{\sum_{j=1}^{N} P_j^0\, e^{v_j/\lambda}},$$

where P_i^0 is the unconditional probability of choosing action i.

For $\lambda = 0$, the decision maker selects the action with the highest payoff with probability 1.

Main Result (Theorem 1)

Theorem 1

For $\lambda > 0$, the optimal strategy satisfies the conditional choice probabilities

$$P_i(\boldsymbol{v}) = \frac{P_i^0 \, e^{v_i/\lambda}}{\sum_{j=1}^{N} P_j^0 \, e^{v_j/\lambda}},$$

where P_i^0 is the unconditional probability of choosing action i.
For $\lambda = 0$, the decision maker selects the action with the highest payoff with probability 1.

Interpretation:

- When actions are a priori homogeneous (i.e. P_i equal for all i), the formula reduces to the standard multinomial logit.
- The term P_i captures the influence of prior beliefs and the cost of information.

Proof of Theorem 1

- Write out the Lagrangian:

$$\begin{aligned}\mathcal{L}(\mathcal{P}) = &\sum_{i=1}^{N} \int_{\boldsymbol{v}} v_i \, P_i(\boldsymbol{v}) \, G(\mathrm{d}\boldsymbol{v}) \\ &- \lambda \left(- \sum_{i=1}^{N} P_i^0 \log P_i^0 \; + \; \int_{\boldsymbol{v}} \sum_{i=1}^{N} P_i(\boldsymbol{v}) \; \log P_i(\boldsymbol{v}) \, G(\mathrm{d}\boldsymbol{v}) \right) \\ &+ \int_{\boldsymbol{v}} \xi_i(\boldsymbol{v}) \, P_i(\boldsymbol{v}) \, G(\mathrm{d}\boldsymbol{v}) \; - \; \int_{\boldsymbol{v}} \mu(\boldsymbol{v}) \left(\sum_{i=1}^{N} P_i(\boldsymbol{v}) \; - \; 1 \right) G(\mathrm{d}\boldsymbol{v}).\end{aligned}$$

- $\xi_i(\boldsymbol{v}) \geq 0, \mu$ are the Lagrangian multipliers.

Proof of Theorem 1

- $P_i^0 > 0$ then take the foc with respect to $P_i(\boldsymbol{v})$

$$v_i + \xi(\boldsymbol{v}) - \mu(\boldsymbol{v}) + \lambda(\log P_i^0 + 1 - \log P_i(\boldsymbol{v}) - 1) = 0$$

Proof of Theorem 1

- $P_i^0 > 0$ then take the foc with respect to $P_i(\boldsymbol{v})$

$$v_i + \xi(\boldsymbol{v}) - \mu(\boldsymbol{v}) + \lambda(\log P_i^0 + 1 - \log P_i(\boldsymbol{v}) - 1) = 0$$

- Note if $P_i^0 > 0$ and $v_i < -\infty$ then $P_i(\boldsymbol{v}) > 0$ for all $\boldsymbol{v}$.

Proof of Theorem 1

- $P_i^0 > 0$ then take the foc with respect to $P_i(\boldsymbol{v})$

$$v_i + \xi(\boldsymbol{v}) - \mu(\boldsymbol{v}) + \lambda(\log P_i^0 + 1 - \log P_i(\boldsymbol{v}) - 1) = 0$$

- Note if $P_i^0 > 0$ and $v_i < -\infty$ then $P_i(\boldsymbol{v}) > 0$ for all $\boldsymbol{v}$.
- $P_i(\boldsymbol{v}) > 0$ so one of the condition does not bind, $\xi_i(\boldsymbol{v}) = 0$

$$P_i(\boldsymbol{v}) = P_i^0 \exp(\frac{v_i - \mu(\boldsymbol{v})}{\lambda})$$

Proof of Theorem 1

- $P_i^0 > 0$ then take the foc with respect to $P_i(\boldsymbol{v})$

$$v_i + \xi(\boldsymbol{v}) - \mu(\boldsymbol{v}) + \lambda(\log P_i^0 + 1 - \log P_i(\boldsymbol{v}) - 1) = 0$$

- Note if $P_i^0 > 0$ and $v_i < -\infty$ then $P_i(\boldsymbol{v}) > 0$ for all $\boldsymbol{v}$.
- $P_i(\boldsymbol{v}) > 0$ so one of the condition does not bind, $\xi_i(\boldsymbol{v}) = 0$

$$P_i(\boldsymbol{v}) = P_i^0 \exp(\frac{v_i - \mu(\boldsymbol{v})}{\lambda})$$

- Probabilities have to sum up to 1 by the other constraint so we have:

$$\exp(\frac{\mu(\boldsymbol{v})}{\lambda}) = \sum_i P_i^0 \exp(\frac{v_i}{\lambda})$$

Comparison to Random Utility

- Recall the popular multinomial logit formula:

$$\Pr(x\text{chosen from A}) = \frac{\exp(v(x))}{\sum y \in A \exp(v(y))}$$

Comparison to Random Utility

- Recall the popular multinomial logit formula:

$$\Pr(x\text{chosen from A}) = \frac{\exp(v(x))}{\sum y \in A \exp(v(y))}$$

- Luce shows that it is equivalent to the IIA axiom.

$$\forall A, B, \quad \frac{\Pr(x \text{ chosen from } A)}{\Pr(y \text{ chosen from } A)} = \frac{\Pr(x \text{ chosen from } B)}{\Pr(y \text{ chosen from } B)}$$

Comparison to Random Utility

- Recall the popular multinomial logit formula:

$$\Pr(x \text{chosen from A}) = \frac{\exp(v(x))}{\sum y \in A \exp(v(y))}$$

- Luce shows that it is equivalent to the IIA axiom.

$$\forall A, B, \quad \frac{\Pr(x \text{ chosen from } A)}{\Pr(y \text{ chosen from } A)} = \frac{\Pr(x \text{ chosen from } B)}{\Pr(y \text{ chosen from } B)}$$

- IIA of course has been criticized by many.

- M&M show that they require weaker axioms:
  - IIA in consequence: $\frac{P_i(\boldsymbol{v})}{P_j(\boldsymbol{v})} = \frac{P_i(\boldsymbol{u})}{P_j(\boldsymbol{u})}$ whenever $u_i = v_i$ and $u_j = v_j$.

- M&M show that they require weaker axioms:
  - IIA in consequence: $\frac{P_i(\boldsymbol{v})}{P_j(\boldsymbol{v})} = \frac{P_i(\boldsymbol{u})}{P_j(\boldsymbol{u})}$ whenever $u_i = v_i$ and $u_j = v_j$.
  - IIA in actions: $\frac{P_i(\boldsymbol{v})}{P_j(\boldsymbol{v})} = \frac{P_i(\boldsymbol{u})}{P_j(\boldsymbol{u})}$ whenever $u_i = u_j$ and $v_i = v_j$.

- M&M show that they require weaker axioms:
  - IIA in consequence: $\frac{P_i(\boldsymbol{v})}{P_j(\boldsymbol{v})} = \frac{P_i(\boldsymbol{u})}{P_j(\boldsymbol{u})}$ whenever $u_i = v_i$ and $u_j = v_j$.
  - IIA in actions: $\frac{P_i(\boldsymbol{v})}{P_j(\boldsymbol{v})} = \frac{P_i(\boldsymbol{u})}{P_j(\boldsymbol{u})}$ whenever $u_i = u_j$ and $v_i = v_j$.
- They showed that if there are at least 3 actions, then the above 2 are equivalent to having an RI model.

Rational Inattention, Optimal Consideration Sets, and Stochastic Choice

Caplin, Dean & Leahy

Motivation

- Matekja & McKay 2015 find the conditional choice probabilities but does not say which objects have positive probabilities.

Motivation

- Matekja & McKay 2015 find the conditional choice probabilities but does not say which objects have positive probabilities.
- Consideration sets: related paper is Manzini & Mariotti (2014 ECMA), they show that a random choice rule satisfying weak axioms must be a two-stage process where only a subset of actions have positive probability.
- CDL 2015 finds conditions for consideration sets in rational inattention.

Model

- $\omega \in \Omega$, finitely many state of the world.
- $a \in A$, set of actions
- $u : A \times \Omega \to \mathbb{R}$ utility.
- $\mu \in \Delta(\Omega)$ is a prior.

M&M in this framework

Choose $P \in \mathcal{P}$ in order to maximize:

$$\sum_{\omega\in\Omega} \mu(\omega)\Big(\sum_{a\in A} P(a \mid \omega)\, u(a,\omega)\Big) \; - \; \lambda\Bigg[\sum_{\omega\in\Omega} \mu(\omega)\Big(\sum_{a\in A} P(a \mid \omega)\; \ln P(a \mid \omega)\Big) \; - \; \sum_{a\in A} P(a)\; \ln P(a)\Bigg],$$

where

$$P(a) \; = \; \sum_{\omega\in\Omega} \mu(\omega)\, P(a \mid \omega).$$

Optimal Policy Condition

M & M: For all actions $a \in A$ such that $P(a) > 0$,

$$P(a \mid \omega) \;=\; \frac{P(a)\, z(a,\omega)}{\sum_{b \in A} P(b)\, z(b,\omega)},$$

where

$$z(a,\omega) \;\equiv\; \exp\!\big(u(a,\omega)/\lambda\big).$$

Interpretation: The optimal policy "twists" the choice probabilities toward states in which the payoffs $u(a,\omega)$ are high.

Consideration Sets

Proposition 1. A policy $P \in \mathcal{P}$ is optimal if and only if

$$\sum_{\omega \in \Omega} \frac{z(a,\omega)\,\mu(\omega)}{\sum_{b \in A} P(b)\,z(b,\omega)} \;\leq\; 1,$$

for all $a \in A$, with equality if $a \in B(P)$ ($B(P)$ is the consideration set of all actions chosen with strictly positive probability). Moreover, for all such actions a and states ω, $P(a \mid \omega)$ satisfies last slide's equations.

- The above establishes that we must be looking at consideration sets but not what these necessarily are.

- Note this is still somewhat novel because
  - In M&M it could have been all options were taken. This hints that there may be a lot of non-chosen options.
  - In M&M the logit probabilities arises by assuming all options are chosen with positive probability. It is not clear initially that with considerations sets this holds.

- Now we want to understand a little bit better how these options are evaluated and when one is in a consideration set.

Recall in M&M, the innovation was to forget about posteriors and only focus on choice probabilities.

Recall in M&M, the innovation was to forget about posteriors and only focus on choice probabilities.

Now let's take the choice probabilities, turns out we can back out the posteriors.

$$\gamma^a(\omega) = \frac{P(a|\omega)\mu(\omega)}{P(a)}$$

So from the choice probabilities we can infer the posterior about states.

Proposition 2

Proposition 2. Consider the choice problem (A, μ) and a policy, $\{P(a)\}_{a \in A}$ and $\{\gamma^a\}_{a \in B(P)}$. The policy is optimal if and only if

$$\sum_{a \in A} P(a)\, \gamma^a(\omega) \;=\; \mu(\omega) \quad \text{for all } \omega \in \Omega,$$

and the following conditions hold:

1. **Invariant Likelihood Ratio (ILR) Equations for Chosen Options:**
   For any $a, b \in B(P)$ and $\omega \in \Omega$,
   $$\frac{\gamma^a(\omega)}{z(a,\omega)} \;=\; \frac{\gamma^b(\omega)}{z(b,\omega)}.$$

2. **Likelihood Ratio Inequalities for Unchosen Options:**
   For any $a \in B(P)$ and $c \in A \setminus B(P)$,
   $$\sum_{\omega \in \Omega} \left(\frac{\gamma^a(\omega)}{z(a,\omega)}\right) z(c,\omega) \;\leq\; 1.$$

Example

- M objects, $a_1, .., a_M$.
- $X \subset \mathbb{R}$ are possible utility levels (finite)/
- State-space $\Omega = X^M$.

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_M \end{pmatrix},$$

  where $\omega_i \in X$ for all $a_i \in A$, example

  The utility of a state/action pair is then given by

$$u(a_i, \omega) = \omega_i.$$

- $\mu(\omega) = \prod_{i=1}^{M} \mu_i(\omega_i)$

Example

Theorem 2. Any optimal policy for an independent consumption problem will involve a cut-off $c \in \mathbb{R}$ such that, for any $a_i \in A$, $P(a_i) > 0$ if

$$\mathbb{E}z(a_i, \omega) \;=\; \sum_{\omega \in \Omega} z(a_i, \omega)\, \mu(\omega) \;=\; \sum_{\omega \in X} \exp(\alpha_i/\lambda)\, \mu_i(\omega) \;>\; c,$$

and $P(a_i) = 0$ otherwise.

Example

Theorem 2. Any optimal policy for an independent consumption problem will involve a cut-off $c \in \mathbb{R}$ such that, for any $a_i \in A$, $P(a_i) > 0$ if

$$\mathbb{E}z(a_i, \omega) \;=\; \sum_{\omega \in \Omega} z(a_i, \omega)\, \mu(\omega) \;=\; \sum_{\omega \in X} \exp(\alpha_i / \lambda)\, \mu_i(\omega) \;>\; c,$$

and $P(a_i) = 0$ otherwise.

- Note that conditional on learning, risky options more attractive than safe option of the same value due to convex transformation of payoff.

Example 2: Independent Consumption Problem

Utility Levels:

$$X = \{0,\ 5.5,\ 10\}.$$

Available Actions:

- **Action a_1:** Always yields a value of 5.5 (the "safe" option).
- **Actions $a_2, \ldots, a_6$:** Each has a 50% chance of yielding 10 and a 50% chance of yielding 0 (the "risky" options).

Example 2: Independent Consumption Problem

Utility Levels:

$$X = \{0,\ 5.5,\ 10\}.$$

Available Actions:

- **Action a_1:** Always yields a value of 5.5 (the "safe" option).
- **Actions $a_2, \ldots, a_6$:** Each has a 50% chance of yielding 10 and a 50% chance of yielding 0 (the "risky" options).

- What are possible consideration sets given the entropy cost?

Example 2: Independent Consumption Problem

Utility Levels:

$$X = \{0,\ 5.5,\ 10\}.$$

Available Actions:

- **Action a_1:** Always yields a value of 5.5 (the “safe” option).
- **Actions $a_2, \ldots, a_6$:** Each has a 50% chance of yielding 10 and a 50% chance of yielding 0 (the “risky” options).

- What are possible consideration sets given the entropy cost?
  - Only a_1, only $a_2, .., a_6$, all of them.

Q: When is risky only consideration set used?

Q: When is risky only consideration set used?

Lemma 1. Let (μ, A) be an independent consumption problem and $\{a_1, \ldots, a_N\} = B \subseteq A$ be a set of ex ante identical actions (i.e. $\mu_i(x) = \mu_b(x)$ for all $x \in X$ and $i, j \leq N$). Then a strategy that picks each $a_j \notin B$ with the unconditional probability $\frac{1}{N}$ and assigns conditional probabilities according to (2) is optimal if, for each $i \in B$,

$$\sum_{x \in X} \exp\!\big(x/\lambda\big)\, \mu_j(x) \leq \frac{1}{N} \Big[\sum_{\overline{x} \in X^N} \frac{\prod_{n=1}^{N} \mu_B(\overline{x}_n)}{\sum_{n=1}^{N} \exp(\overline{x}_n/\lambda)} \Big]^{-1}$$

- So how does this work?
  - $\lambda = 30$ then only safe option.
  - $\lambda = 20$ both safe and risky.
  - $\lambda = 2$ only risky option.
  - $\lambda = 1$ both safe and risky.

Cost of Information

Cost of Information

- Entropy reduction is mathematically elegant and offers a clean solution.

Cost of Information

- Entropy reduction is mathematically elegant and offers a clean solution.
- But it has a few drawbacks:
  - Choice probability have some unintuitive properties.
  - Arbitrary functional form without foundations.
  - Experimentally rejected.

Cost of Information

- Entropy reduction is mathematically elegant and offers a clean solution.
- But it has a few drawbacks:
  - Choice probability have some unintuitive properties.
  - Arbitrary functional form without foundations.
  - Experimentally rejected.
- Naturally people have looked for alternatives.

Posterior Separable Cost Functions

Definition. A cost function $C\colon \mathcal{M}_\pi \to (-\infty, \infty]$ is *posterior separable* if there is an upper semicontinuous function $H\colon \Delta(\Theta) \to \mathbb{R}$ such that for all $\mu \in \mathcal{M}_\pi$,

$$C(\mu) \;=\; H(\pi) \;-\; \int_{\Delta(\Theta)} H(p)\,\mathrm{d}\mu(p).$$

We call H a *measure of uncertainty*. A typical choice for H is *entropy* (Matějka and McKay 2015):

$$H(p) \;=\; -\sum_{\theta} p(\theta)\,\ln\big[p(\theta)\big].$$

- Posterior separable cost functions are nice because they preserve the analytical simplicity of RI. They also allow for a connection between Bayesian persuasion and RI.

- Posterior separable cost functions are nice because they preserve the analytical simplicity of RI. They also allow for a connection between Bayesian persuasion and RI.

- There are a few experimental papers that have talked about posterior separability: Denti (2023), Walker-Jones (2024).
    - Denti (2023): Adding a more risky option should increase learning and drop the probability a safe is chosen to 0. Contradicted.
    - Walker-Jones (2024): Denti (2023) only works if you assume constant cost of information. Aggregating RI agents $\neq$ representative RI agent (note Aggregating REU agent = Representative REU agent).

Pomatto, Strack & Tamuz (2023)

- They consider cost functions over Blackwell experiments.

Pomatto, Strack & Tamuz (2023)

- They consider cost functions over Blackwell experiments.
- No more choices for now - primitives are Blackwell experiments, hence the cost of information is *prior-free*.

Pomatto, Strack & Tamuz (2023)

- They consider cost functions over Blackwell experiments.
- No more choices for now - primitives are Blackwell experiments, hence the cost of information is *prior-free*.
- They characterize the set of information costs which are *constant marginal cost*.

Model and Representation

- $i, j \in \Theta$ set of states finite.
- An experiment μ provides a distribution of signal realizations $\mu_i(s)$ in state i.
- μ_i and μ_j are, therefore, two different distributions, and the more different they are..., the more distinct these two states are.
- They characterize the following cost function:

$$C(\mu) \;=\; \sum_{i,j \in \Theta} \beta_{ij} \, D_{\mathrm{KL}}\big(\mu_i \,\|\, \mu_j\big).$$

β_{ij} is the "cost" of distinguishing between j and i when the realized state is i. The larger this is, the more costly it is to reject j when the true state is i.

Axioms

- We say μ and ν are Blackwell equivalent if they yield the same distribution of posteriors under Bayesian updating for any priors.
- Axiom 1: if μ and ν are Blackwell equivalent then $C(\mu) = C(\nu)$.

Axioms

- We say μ and ν are Blackwell equivalent if they yield the same distribution of posteriors under Bayesian updating for any priors.
- Axiom 1: if μ and ν are Blackwell equivalent then $C(\mu) = C(\nu)$.

Implication: Invariant by indexing of signals - only statistical content matters.

Axioms

- Given two experiments $\mu = \left(S, \mu_i\right)$ and $\nu = \left(T, \nu_i\right)$, we define their product

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

where $\mu_i \times \nu_i$ denotes the product of the two measures.

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

Axioms

- Given two experiments $\mu = \big(S, \mu_i\big)$ and $\nu = \big(T, \nu_i\big)$, we define their product

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

  where $\mu_i \times \nu_i$ denotes the product of the two measures.

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

- The cost of observing two independent draws is the sum of observing each separately.

Axioms

- Given an experiment μ, denote by $\alpha \cdot \mu$ for $\alpha \in [0,1]$ the experiment which gives with probability $(1-\alpha)$ an uninformative signal and α probability of μ.

- Axiom 3: $C(\alpha \cdot \mu) = \alpha C(\mu)$

Axioms

- Given an experiment μ, denote by $\alpha \cdot \mu$ for $\alpha \in [0,1]$ the experiment which gives with probability $(1-\alpha)$ an uninformative signal and α probability of μ.
- Axiom 3: $C(\alpha \cdot \mu) = \alpha C(\mu)$
- This combines two assumptions: uninformative signals have 0 cost and dilution has linear cost. Note this is satisfied by posterior-separable costs.

Theorem 1

Theorem. An information cost function C satisfies Axioms 1,2,3 plus continuity* if and only if there exists a collection $\{\beta_{i,j}\}_{i \neq j} \subseteq \mathbb{R}_+$ such that for every experiment $\mu = \big(S, \mu_i\big)$,

$$C(\mu) \;=\; \sum_{i,j} \beta_{i,j} \int_S \ln\!\Big(\frac{d\mu_j}{d\mu_i}(s)\Big)\, d\mu_i(s). \tag{3}$$

Moreover, the collection $\{\beta_{i,j}\}$ is unique given C.

Corollary: C is increasing in Blackwell ordering.