Blackwell-Monotone Information Costs

Xiaoyu Cheng

Florida State University

Yonggyun (YG) Kim

Florida State University

Introduction

Introduction

- Agenda: integration of costly information across various fields
- Question: Which information cost function *should* or *could* be used
- Examples
  - Entropy Costs: Sims (2003); Matějka, McKay (2015)
  - Posterior Separable Costs: Caplin, Dean, Leahy (2022); Denti (2022)
  - Log-Likelihood Ratio Costs: Pomatto, Strack, Tamuz (2023)
- Common Principle: Blackwell Monotonicity
  - Higher rank in Blackwell's order $\Rightarrow$ higher cost
  - Minimum requirement for plausible information costs

Introduction

- Agenda: integration of costly information across various fields
- Question: Which information cost function *should* or *could* be used
- Examples
  - Entropy Costs: Sims (2003); Matějka, McKay (2015)
  - Posterior Separable Costs: Caplin, Dean, Leahy (2022); Denti (2022)
  - Log-Likelihood Ratio Costs: Pomatto, Strack, Tamuz (2023)
- Common Principle: Blackwell Monotonicity
  - Higher rank in Blackwell's order $\Rightarrow$ higher cost
  - Minimum requirement for plausible information costs

Introduction

- Agenda: integration of costly information across various fields
- Question: Which information cost function *should* or *could* be used
- Examples
  - Entropy Costs: Sims (2003); Matějka, McKay (2015)
  - Posterior Separable Costs: Caplin, Dean, Leahy (2022); Denti (2022)
  - Log-Likelihood Ratio Costs: Pomatto, Strack, Tamuz (2023)
- Common Principle: Blackwell Monotonicity
  - Higher rank in Blackwell's order $\Rightarrow$ higher cost
  - Minimum requirement for plausible information costs

Introduction

- Agenda: integration of costly information across various fields
- Question: Which information cost function *should* or *could* be used
- Examples
  - Entropy Costs: Sims (2003); Matějka, McKay (2015)
  - Posterior Separable Costs: Caplin, Dean, Leahy (2022); Denti (2022)
  - Log-Likelihood Ratio Costs: Pomatto, Strack, Tamuz (2023)
- Common Principle: Blackwell Monotonicity
  - Higher rank in Blackwell's order $\Rightarrow$ higher cost
  - Minimum requirement for plausible information costs

Introduction

- **Blackwell's Theorem**: the followings are equivalent
  1. For any Bayesian decision problem, the expected payoff under f is greater than or equal to that under g
  2. There exists a stochastic matrix M such that $g = f \cdot M$

- **Goals**
  - identify elementary necessary and sufficient conditions for Blackwell monotonicity
  - characterize a practical and tractable class of information cost functions

▸ Literature

Introduction

- **Blackwell's Theorem**: the followings are equivalent

  1. For any Bayesian decision problem, the expected payoff under f is greater than or equal to that under g
  2. There exists a stochastic matrix M such that $g = f \cdot M$

- **Goals**
  - identify elementary necessary and sufficient conditions for Blackwell monotonicity
  - characterize a practical and tractable class of information cost functions

▸ Literature

Introduction

- **Blackwell's Theorem**: the followings are equivalent

  1. For any Bayesian decision problem, the expected payoff under f is greater than or equal to that under g
  2. There exists a stochastic matrix M such that $g = f \cdot M$

- **Goals**
  - identify elementary necessary and sufficient conditions for Blackwell monotonicity
  - characterize a practical and tractable class of information cost functions

Literature

Roadmap

1. Preliminaries
2. Blackwell Monotonicity under Binary Experiments
   - Examples of Information Costs
3. Blackwell Monotonicity under General Experiments
   - Additively Separable Costs
4. Applications
   - Costly Persuasion
   - Bargaining with Information Acquisition

Preliminaries

Experiments

- $\Omega = \{\omega_1, \cdots, \omega_n\}$: a finite set of states
- $\mathcal{S} = \{s_1, \cdots, s_m\}$: a finite set of signals
- A *statistical experiment* $f : \Omega \to \Delta(\mathcal{S})$ can be represented by an $n \times m$ matrix:

$$f = \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix},$$

  where $f_{ij} = \Pr(s_j|\omega_i)$, thus, $f_{ij} \geq 0$ and $\sum_{j=1}^{m} f_{ij} = 1$

- $\mathcal{E}_m \subset \mathbb{R}^{n\times m}$: the space of all experiments with m possible signals

Blackwell Informativeness

- $f \succeq_B g$: f is *Blackwell more informative* than g
  if there exists a stochastic matrix M such that $g = f \cdot M$
    - M is a stochastic matrix iff $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i
- **Permutation**
    - A stochastic matrix P is called a *permutation matrix* if it has exactly one non-zero entry in each row and each column.
    - If P is a permutation matrix, so is P^{-1}.
    - **Observation**: f and $f \cdot P$ are equally Blackwell informative:

$$f \succeq_B f \cdot P \succeq_B f \cdot P \cdot P^{-1} = f \tag{1}$$

    - **Intuition**: relabeling signals does not change the informativeness

Blackwell Informativeness

- $f \succeq_B g$: f is *Blackwell more informative* than g
  if there exists a stochastic matrix M such that $g = f \cdot M$
  - M is a stochastic matrix iff $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i

- **Permutation**
  - A stochastic matrix P is called a *permutation matrix* if it has exactly one non-zero entry in each row and each column.
  - If P is a permutation matrix, so is P^{-1}.
  - **Observation:** f and $f \cdot P$ are equally Blackwell informative:

$$f \succeq_B f \cdot P \succeq_B f \cdot P \cdot P^{-1} = f \tag{1}$$

  - **Intuition:** relabeling signals does not change the informativeness

Blackwell Informativeness

- $f \succeq_B g$: f is *Blackwell more informative* than g
  if there exists a stochastic matrix M such that $g = f \cdot M$
  - M is a stochastic matrix iff $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i

- **Permutation**
  - A stochastic matrix P is called a *permutation matrix* if it has exactly one non-zero entry in each row and each column.
  - If P is a permutation matrix, so is P^{-1}.
  - **Observation**: f and $f \cdot P$ are equally Blackwell informative:

$$f \succeq_B f \cdot P \succeq_B f \cdot P \cdot P^{-1} = f \tag{1}$$

  - **Intuition**: relabeling signals does not change the informativeness

Blackwell Informativeness

- $f \succeq_B g$: f is *Blackwell more informative* than g
  if there exists a stochastic matrix M such that $g = f \cdot M$
  - M is a stochastic matrix iff $M_{ij} \geq 0$ and $\sum_j M_{ij} = 1$ for all i

- **Permutation**
  - A stochastic matrix P is called a *permutation matrix* if it has exactly one non-zero entry in each row and each column.
  - If P is a permutation matrix, so is P^{-1}.
  - **Observation**: f and $f \cdot P$ are equally Blackwell informative:
  $$f \succeq_B f \cdot P \succeq_B f \cdot P \cdot P^{-1} = f \tag{1}$$
  - **Intuition**: relabeling signals does not change the informativeness

Information Costs and Blackwell Monotonicity

- **Information Costs**
  - $C : \mathcal{E}_m \to \mathbb{R}_+$: an information cost function
  - $\mathcal{C}_m$: the set of all Lipschitz continuous information cost functions defined over $\mathcal{E}_m$
  - Lipschitz continuity ensures that a derivative exists a.e. and is integrable.

- **Blackwell Monotonicity**
  - An information cost function $C \in \mathcal{C}_m$ is **Blackwell monotone** if for all $f, g \in \mathcal{E}_m$, $C(f) \geq C(g)$ whenever $f \succeq_B g$.

- **Permutation Invariance**
  - Any Blackwell-monotone information cost function is **permutation invariant**, i.e., $C(f) = C(f \cdot P)$

$$f \succeq_B f \cdot P \succeq_B f \quad \Rightarrow \quad C(f) \geq C(f \cdot P) \geq C(f).$$

Information Costs and Blackwell Monotonicity

- **Information Costs**
  - $C : \mathcal{E}_m \to \mathbb{R}_+$: an information cost function
  - $\mathcal{C}_m$: the set of all Lipschitz continuous information cost functions defined over $\mathcal{E}_m$
  - Lipschitz continuity ensures that a derivative exists a.e. and is integrable.

- **Blackwell Monotonicity**
  - An information cost function $C \in \mathcal{C}_m$ is **Blackwell monotone** if for all $f, g \in \mathcal{E}_m$, $C(f) \geq C(g)$ whenever $f \succeq_B g$.

- **Permutation Invariance**
  - Any Blackwell-monotone information cost function is **permutation invariant**, i.e., $C(f) = C(f \cdot P)$

$$f \succeq_B f \cdot P \succeq_B f \quad \Rightarrow \quad C(f) \geq C(f \cdot P) \geq C(f).$$

Information Costs and Blackwell Monotonicity

- **Information Costs**
  - $C : \mathcal{E}_m \to \mathbb{R}_+$: an information cost function
  - $\mathcal{C}_m$: the set of all Lipschitz continuous information cost functions defined over $\mathcal{E}_m$
  - Lipschitz continuity ensures that a derivative exists a.e. and is integrable.

- **Blackwell Monotonicity**
  - An information cost function $C \in \mathcal{C}_m$ is **Blackwell monotone** if for all $f, g \in \mathcal{E}_m$, $C(f) \geq C(g)$ whenever $f \succeq_B g$.

- **Permutation Invariance**
  - Any Blackwell-monotone information cost function is **permutation invariant**, i.e., $C(f) = C(f \cdot P)$

$$f \succeq_B f \cdot P \succeq_B f \quad \Rightarrow \quad C(f) \geq C(f \cdot P) \geq C(f).$$

Binary Experiments

Blackwell Informativeness under Binary Experiments

- Focus on the case where $n = m = 2$
- Any experiment can be represented by $f \equiv (f_1, f_2)^\intercal \in [0,1]^2$:

$$[\mathbf{1} - f, f] = \begin{array}{c|cc} & s_L & s_H \\ \hline \omega_L & 1 - f_1 & f_1 \\ \omega_H & 1 - f_1 & f_2 \end{array}$$

- Which of the followings (defined over $f_2 \geq f_1$) are Blackwell-monotone?

1. $C(f_1, f_2) = (f_2 - f_1)^2$
2. $C(f_1, f_2) = f_2 - 2f_1$
3. $C(f_1, f_2) = \dfrac{f_2(1 - f_2)}{f_1(1 - f_1)} - 1$
4. $C(f_1, f_2) = \dfrac{f_2}{f_1} + \dfrac{1 - f_1}{1 - f_2} - 2$

Blackwell Informativeness under Binary Experiments

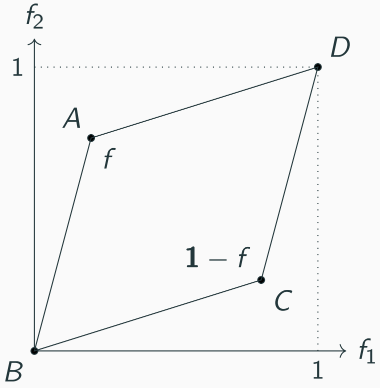
- Focus on the case where $n = m = 2$
- Any experiment can be represented by $f \equiv (f_1, f_2)^\intercal \in [0,1]^2$:

$$[\mathbf{1} - f, f] = \begin{array}{c|cc} & s_L & s_H \\ \hline \omega_L & 1 - f_1 & f_1 \\ \omega_H & 1 - f_1 & f_2 \end{array}$$

- Which of the followings (defined over $f_2 \geq f_1$) are Blackwell-monotone?

1. $C(f_1, f_2) = (f_2 - f_1)^2$
2. $C(f_1, f_2) = f_2 - 2f_1$
3. $C(f_1, f_2) = \dfrac{f_2(1 - f_2)}{f_1(1 - f_1)} - 1$
4. $C(f_1, f_2) = \dfrac{f_2}{f_1} + \dfrac{1 - f_1}{1 - f_2} - 2$

Parallelogram Hull



Any stochastic matrix can also be represented by $(a, b) \in [0, 1]^2$:

$$M = \begin{bmatrix} 1 - a & a \\ 1 - b & b \end{bmatrix}.$$

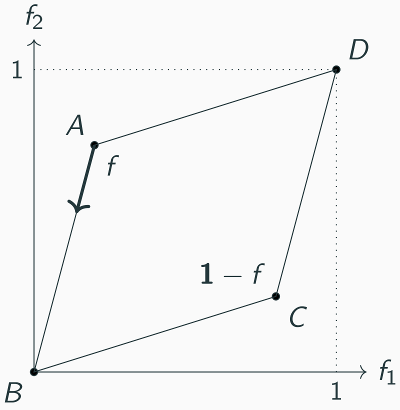
Then, $[\mathbf{1} - g, g] = [\mathbf{1} - f, f] \cdot M$ implies

$$g = a \cdot (\mathbf{1} - f) + b \cdot f.$$

- $f \succeq_B g$ iff g is in the *parallelogram hull* of f and $\mathbf{1} - f$:

$$\mathrm{PARL}(f, \mathbf{1} - f) = \left\{ a \cdot (\mathbf{1} - f) + b \cdot f \in \mathbb{R}^2_+ : a, b \in [0, 1] \right\}.$$

When C is Blackwell monotone,



1. $\langle \nabla C(f), -f \rangle \leq 0$

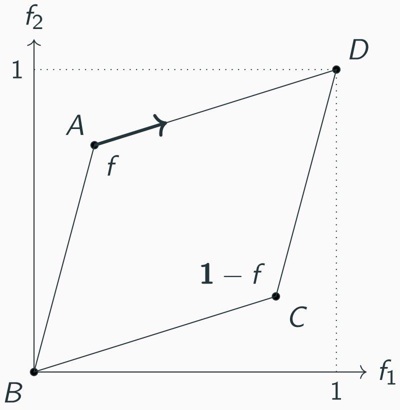
$$\begin{tikzcd}
s_L \arrow[r, "1"] & s'_L \\
s_H \arrow[ur, "\epsilon"'] \arrow[r, "1-\epsilon"'] & s'_H
\end{tikzcd}$$

2. $\langle \nabla C(f), \mathbf{1} - f \rangle \leq 0$

$$\begin{tikzcd}
s_L \arrow[r, "1-\epsilon"] \arrow[dr, "\epsilon"] & s'_L \\
s_H \arrow[r, "1"'] & s'_H
\end{tikzcd}$$

Necessary Conditions for Blackwell Monotonicity

When C is Blackwell monotone,



1. $\langle \nabla C(f), -f \rangle \leq 0$

$$\begin{tikzcd}
s_L \arrow[r, "1"] & s'_L \\
s_H \arrow[ur, "\epsilon"'] \arrow[r, "1-\epsilon"'] & s'_H
\end{tikzcd}$$

2. $\langle \nabla C(f), \mathbf{1} - f \rangle \leq 0$

$$\begin{tikzcd}
s_L \arrow[r, "1-\epsilon"] \arrow[dr, "\epsilon"] & s'_L \\
s_H \arrow[r, "1"'] & s'_H
\end{tikzcd}$$

Theorem for Binary Experiments

Theorem 1

$C \in \mathcal{C}_2$ is Blackwell monotone if and only if it is

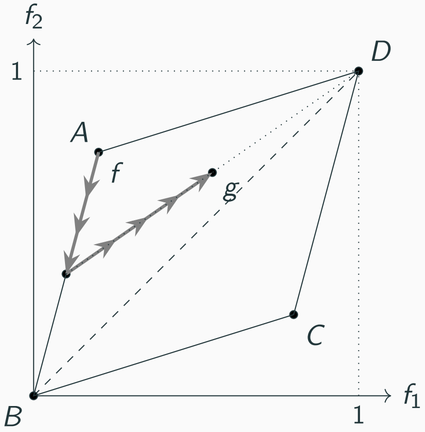
1. permutation invariant;
2. for all $f \in \mathcal{E}_2$,

$$\langle \nabla C(f), f \rangle \geq 0 \geq \langle \nabla C(f), \mathbf{1} - f \rangle. \tag{2}$$

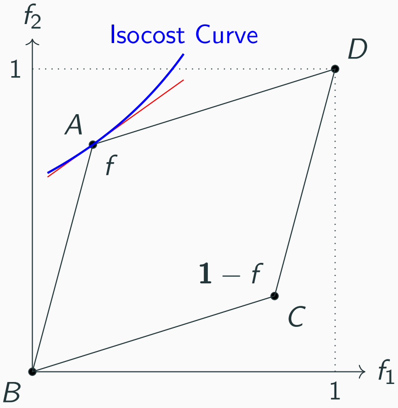
- This theorem holds for the cases with more than two states, but the binary signal assumption is curcial.

Proof for Sufficiency

For any $f \succeq_B g$, we can find a path from f to g (or the permutation of it) along which Blackwell informativeness decreases



Further Characterizations with Binary States



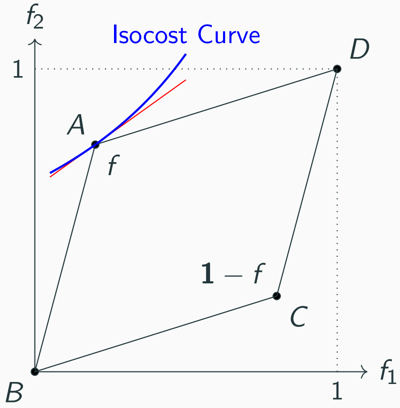
$\langle \nabla C(f), f \rangle \geq 0 \geq \langle \nabla C(f), \mathbf{1} - f \rangle$

is equivalent to:

$$\underbrace{\frac{f_2}{f_1}}_{\substack{\text{the slope} \\ \text{of } \overline{AB}}} \geq \underbrace{-\frac{\partial C/\partial f_1}{\partial C/\partial f_2}}_{\substack{\text{the slope of} \\ \text{the isocost curve}}} \geq \underbrace{\frac{1-f_2}{1-f_1}}_{\substack{\text{the slope} \\ \text{of } \overline{AD}}}$$

- **Interpretation**: a *marignal rate of information transformation* (MRIT) lies between the two likelihood ratios provided by the experiment.

Further Characterizations with Binary States



$\langle \nabla C(f), f \rangle \geq 0 \geq \langle \nabla C(f), \mathbf{1} - f \rangle$
is equivalent to:

$$\underbrace{\frac{f_2}{f_1}}_{\substack{\text{the slope} \\ \text{of } \overline{AB}}} \geq \underbrace{-\frac{\partial C / \partial f_1}{\partial C / \partial f_2}}_{\substack{\text{the slope of} \\ \text{the isocost curve}}} \geq \underbrace{\frac{1 - f_2}{1 - f_1}}_{\substack{\text{the slope} \\ \text{of } \overline{AD}}}$$

- **Interpretation**: a *marignal rate of information transformation* (MRIT) lies between the two likelihood ratios provided by the experiment.

Further Characterizations with Binary States

1. $C(f_1, f_2) = (f_2 - f_1)^2$ with $1 > f_2 > f_1 > 0$

$$\frac{f_2}{f_1} \geq -\frac{\partial C/\partial f_1}{\partial C/\partial f_2} = 1 \geq \frac{1-f_2}{1-f_1}$$

- The above inequalities hold for all $1 > f_2 > f_1 > 0$, thus, it is **Blackwell monotone**.

2. $C(f_1, f_2) = f_2 - 2f_1$ with $1 > f_2 > f_1 > 0$

$$\frac{f_2}{f_1} \geq -\frac{\partial C/\partial f_1}{\partial C/\partial f_2} = 2 \geq \frac{1-f_2}{1-f_1}$$

- The above inequalities does not always hold, e.g., $f_1 = .5$ and $f_2 = .6$, thus, it is not Blackwell monotone.

Further Characterizations with Binary States

1. $C(f_1, f_2) = (f_2 - f_1)^2$ with $1 > f_2 > f_1 > 0$

$$\frac{f_2}{f_1} \geq -\frac{\partial C/\partial f_1}{\partial C/\partial f_2} = 1 \geq \frac{1 - f_2}{1 - f_1}$$

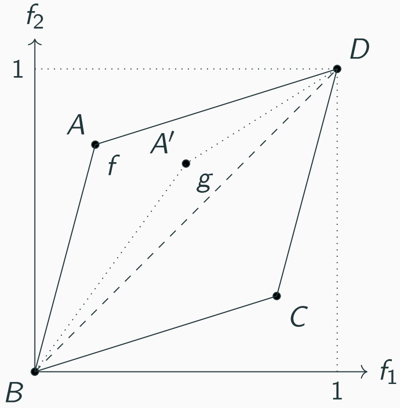
- The above inequalities hold for all $1 > f_2 > f_1 > 0$, thus, it is **Blackwell monotone**.

2. $C(f_1, f_2) = f_2 - 2f_1$ with $1 > f_2 > f_1 > 0$

$$\frac{f_2}{f_1} \geq -\frac{\partial C/\partial f_1}{\partial C/\partial f_2} = 2 \geq \frac{1 - f_2}{1 - f_1}$$

- The above inequalities does not always hold, e.g., $f_1 = .5$ and $f_2 = .6$, thus, it is not Blackwell monotone.

Further Characterizations with Binary States



$f \succeq_B g$ is equivalent to:

1. AB steeper than $A'B$:

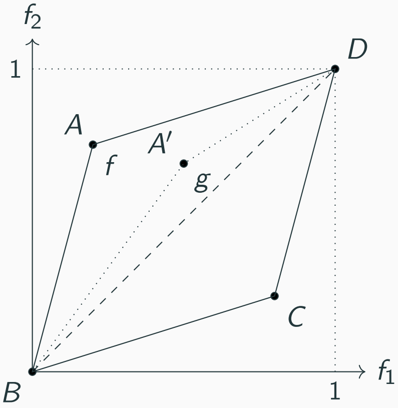
$$\alpha \equiv \frac{f_2}{f_1} \geq \frac{g_2}{g_1} \equiv \alpha'$$

2. AD shallower than $A'D$:

$$\beta \equiv \frac{1-f_1}{1-f_2} \geq \frac{1-g_1}{1-g_2} \equiv \beta'$$

- C is Blackwell monotone iff it is increasing in α and β after reparametrization

Further Characterizations with Binary States



$f \succeq_B g$ is equivalent to:

1. AB steeper than $A'B$:

$$\alpha \equiv \frac{f_2}{f_1} \geq \frac{g_2}{g_1} \equiv \alpha'$$

2. AD shallower than $A'D$:

$$\beta \equiv \frac{1-f_1}{1-f_2} \geq \frac{1-g_1}{1-g_2} \equiv \beta'$$

- C is Blackwell monotone iff it is increasing in α and β after reparametrization

Further Characterizations with Binary States

3. $C(f_1, f_2) = \dfrac{f_2(1-f_2)}{f_1(1-f_1)} - 1$ with $1 > f_2 > f_1 > 0$

$$\tilde{C}(\alpha, \beta) = \frac{\alpha}{\beta} - 1$$

- $\tilde{C}$ is increasing in α but not in β, thus, $\tilde{C}$ is not Blackwell monotone.

4. $C(f_1, f_2) = \dfrac{f_2}{f_1} + \dfrac{1-f_1}{1-f_2} - 2$ with $1 > f_2 > f_1 > 0$

$$\tilde{C}(\alpha, \beta) = \alpha + \beta - 2$$

- $\tilde{C}$ is increasing in both α and β, thus, $\tilde{C}$ is **Blackwell monotone**.

Answer for the Motivating Question

Which of the followings are Blackwell-monotone information cost functions?

1. $C(f_1, f_2) = (f_2 - f_1)^2$
2. $C(f_1, f_2) = f_2 - 2f_1$
3. $C(f_1, f_2) = \dfrac{f_2(1 - f_2)}{f_1(1 - f_1)} - 1$
4. $C(f_1, f_2) = \dfrac{f_2}{f_1} + \dfrac{1 - f_1}{1 - f_2} - 2$

General Experiments

Necessary Conditions for Blackwell Monotonicity

Now assume that there are more than two signals.

- Permutation invariance is still necessary

- For any pair (i, j), the following garbling worsens the informativeness:

$$\begin{tikzcd} s_i \arrow[r, "1-\epsilon"] \arrow[dr, "\epsilon"] & s_i' \\ s_j \arrow[r, "1"'] & s_j' \end{tikzcd}$$

- This gives us $\langle \nabla^j C(f) - \nabla^i C(f), f^i \rangle \leq 0$, where

$$\langle \nabla^j C(f) - \nabla^i C(f), f^i \rangle = \sum_{s=1}^{n} \frac{\partial C}{\partial f_{sj}} \cdot f_{si} - \sum_{s=1}^{n} \frac{\partial C}{\partial f_{si}} \cdot f_{si}$$

Necessary Conditions for Blackwell Monotonicity

Now assume that there are more than two signals.

- Permutation invariance is still necessary

- For any pair (i, j), the following garbling worsens the informativeness:

$$\begin{tikzcd} s_i \arrow[r, "1-\epsilon"] \arrow[dr, "\epsilon"] & s_i' \\ s_j \arrow[r, "1"'] & s_j' \end{tikzcd}$$

- This gives us $\langle \nabla^j C(f) - \nabla^i C(f), f^i \rangle \leq 0$, where

$$\langle \nabla^j C(f) - \nabla^i C(f), f^i \rangle = \sum_{s=1}^{n} \frac{\partial C}{\partial f_{sj}} \cdot f_{si} - \sum_{s=1}^{n} \frac{\partial C}{\partial f_{si}} \cdot f_{si}$$

Sufficient Conditions for Blackwell Monotonicity

When $m \geq 3$, there may not exist a path along which informativeness decreases

Proposition

Let

$$g = \begin{bmatrix} 4/5 & 1/5 & 0 \\ 0 & 4/5 & 1/5 \\ 1/5 & 0 & 4/5 \end{bmatrix} \in \mathcal{E}_3.$$

If $f \succeq_B g$ and $f \in \mathcal{E}_3$, then f is a permutation of I_3 or g.

▸ Illustrations

- I_3 is Blackwell more informative than g, but we cannot find a path from I_3 to g along which Blackwell informativeness decreases

Sufficient Conditions for Blackwell Monotonicity

When $m \geq 3$, there may not exist a path along which informativeness decreases

Proposition

Let

$$g = \begin{bmatrix} 4/5 & 1/5 & 0 \\ 0 & 4/5 & 1/5 \\ 1/5 & 0 & 4/5 \end{bmatrix} \in \mathcal{E}_3.$$

If $f \succeq_B g$ and $f \in \mathcal{E}_3$, then f is a permutation of I_3 or g.

▸ Illustrations

- I_3 is Blackwell more informative than g, but we cannot find a path from I_3 to g along which Blackwell informativeness decreases

Quasiconvexity

- Observe that there is a permutation of I_3 such that

$$g = \frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot (I_3 \cdot P).$$

- If we impose **quasiconvexity**, with permutation invariance, we have

$$C(I_3) = C(I_3 \cdot P) \geq C\left(\frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot I_3 \cdot P\right) = C(g).$$

- Caveat: Quasiconvexity is not a necessary condition for Blackwell monotonicity

▸ Details

Quasiconvexity

- Observe that there is a permutation of I_3 such that

$$g = \frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot (I_3 \cdot P).$$

- If we impose **quasiconvexity**, with permutation invariance, we have

$$C(I_3) = C(I_3 \cdot P) \geq C\left(\frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot I_3 \cdot P\right) = C(g).$$

- Caveat: Quasiconvexity is not a necessary condition for Blackwell monotonicity

Details

Quasiconvexity

- Observe that there is a permutation of I_3 such that

$$g = \frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot (I_3 \cdot P).$$

- If we impose **quasiconvexity**, with permutation invariance, we have

$$C(I_3) = C(I_3 \cdot P) \geq C\left(\frac{4}{5} \cdot I_3 + \frac{1}{5} \cdot I_3 \cdot P\right) = C(g).$$

- Caveat: Quasiconvexity is not a necessary condition for Blackwell monotonicity

▸ Details

Theorem for General Experiments

Theorem 2

Suppose that $C \in \mathcal{C}_m$ is Lipschitz continuous and quasiconvex. Then, C is Blackwell monotone if and only if it is

1. permutation invariant;
2. for all $f \in \mathcal{E}_2$ and $i \neq j$,

$$\langle \nabla^j C(f) - \nabla^i C(f), f \rangle \leq 0. \tag{3}$$

- $S_B(f)$: the set of experiments that are less Blackwell informative than f
- Two conditions ensure that extreme points of $S_B(f)$ are not more costly than f
- Then, we can apply quasiconvexity

Theorem for General Experiments

Theorem 2

Suppose that $C \in \mathcal{C}_m$ is Lipschitz continuous and quasiconvex. Then, C is Blackwell monotone if and only if it is

1. permutation invariant;
2. for all $f \in \mathcal{E}_2$ and $i \neq j$,

$$\langle \nabla^j C(f) - \nabla^i C(f), f \rangle \leq 0. \tag{3}$$

- $S_B(f)$: the set of experiments that are less Blackwell informative than f
- Two conditions ensure that extreme points of $S_B(f)$ are not more costly than f
- Then, we can apply quasiconvexity

Examples: Additively Separable Costs

Additively Separable Costs

C is additively separable if there exists Lipschitz continuous functions $\psi : \mathbb{R}^n_+ \to \mathbb{R}_+$ such that, for all m and $f \in \mathcal{E}_m$,

$$C(f) = \sum_{j=1}^{m} \psi(f^j).$$

Theorem 3

When C is additively separable, C is Blackwell monotone if and only if ψ is sublinear:

1. positive homogeneity: $\psi(\alpha h) = \alpha \psi(h)$;
2. subadditivity: $\psi(k) + \psi(l) \geq \psi(k + l)$

Examples: Additively Separable Costs

Additively Separable Costs

C is additively separable if there exists Lipschitz continuous functions $\psi : \mathbb{R}^n_+ \to \mathbb{R}_+$ such that, for all m and $f \in \mathcal{E}_m$,

$$C(f) = \sum_{j=1}^{m} \psi(f^j).$$

Theorem 3

When C is additively separable, C is Blackwell monotone if and only if ψ is sublinear:

1. positive homogeneity: $\psi(\alpha h) = \alpha \psi(h)$;
2. subadditivity: $\psi(k) + \psi(l) \geq \psi(k + l)$

Proof of Theorem 3

[Sublinearity $\Rightarrow$ Blackwell Monotonicity]

- From sublinearity, we can show that C is convex.
- Consider the garbling of replacing s_j to s_k with prob. ϵ:

$$\Delta C = \psi(f^k + \epsilon \cdot f^j) + \psi((1-\epsilon)f^j) - \left[\psi(f^k) + \psi(f^j)\right]$$
$$= \psi(f^k + \epsilon \cdot f^j) + (1-\epsilon) \cdot \psi(f^j) - \psi(f^k) - \psi(f^j)$$
$$= \psi(f^k + \epsilon \cdot f^j) - \psi(f^k) -$$

Necessary Condition

Proof of Theorem 3

[Sublinearity $\Rightarrow$ Blackwell Monotonicity]

- From sublinearity, we can show that C is convex.
- Consider the garbling of replacing s_j to s_k with prob. ϵ:

$$\begin{aligned}\Delta C &= \psi(f^k + \epsilon \cdot f^j) + \psi((1-\epsilon)f^j) - \left[\psi(f^k) + \psi(f^j)\right] \\ &= \psi(f^k + \epsilon \cdot f^j) + (1-\epsilon)\cdot\psi(f^j) - \psi(f^k) - \psi(f^j) \\ &= \psi(f^k + \epsilon \cdot f^j) - \psi(f^k) -\end{aligned}$$

Necessary Condition

Proof of Theorem 3

[Sublinearity $\Rightarrow$ Blackwell Monotonicity]

- From sublinearity, we can show that C is convex.

- Consider the garbling of replacing s_j to s_k with prob. ϵ:

$$\begin{aligned}\Delta C &= \psi(f^k + \epsilon \cdot f^j) + \psi((1-\epsilon)f^j) - \left[\psi(f^k) + \psi(f^j)\right] \\ &= \psi(f^k + \epsilon \cdot f^j) + (1-\epsilon) \cdot \psi(f^j) - \psi(f^k) - \psi(f^j) \\ &= \psi(f^k + \epsilon \cdot f^j) - \psi(f^k) - \epsilon \cdot \psi(f^j)\end{aligned}$$

▸ Necessary Condition

Proof of Theorem 3

[Sublinearity $\Rightarrow$ Blackwell Monotonicity]

- From sublinearity, we can show that C is convex.
- Consider the garbling of replacing s_j to s_k with prob. ϵ:

$$\begin{aligned}
\Delta C &= \psi(f^k + \epsilon \cdot f^j) + \psi((1-\epsilon)f^j) - \left[\psi(f^k) + \psi(f^j)\right] \\
&= \psi(f^k + \epsilon \cdot f^j) + (1-\epsilon) \cdot \psi(f^j) - \psi(f^k) - \psi(f^j) \\
&= \psi(f^k + \epsilon \cdot f^j) - \psi(f^k) - \psi(\epsilon \cdot f^j) \leq 0
\end{aligned}$$

Necessary Condition

Examples: Additively Separable Costs

1. Supnorm Costs
$$C(f) = \sum_{j=1}^{m} \max_i f_{ij}.$$
2. Linear Costs
$$C(f) = \sum_{j=1}^{m} |\langle a, f^j \rangle| = \sum_{j=1}^{m} \left| \sum_{i=1}^{n} a_i f_{ij} \right|.$$
3. Linear ϕ-divergence Costs (including LLR costs)
$$C(f) = \sum_{j=1}^{m} \sum_{i,i'} \beta_{ii'} f_{i'j} \phi_{ii'} \left(\frac{f_{ij}}{f_{i'j}} \right).$$
4. Posterior Separable Costs (including Entropy costs)
$$C_\mu(f) = H(\mu) - \sum_{j=1}^{m} \tau(f^j) \cdot H\left[\left(\frac{\mu_i f_i^j}{\tau(f^j)} \right)_i \right]$$
where $\tau(f^j)$ is the probability of receiving signal j, i.e., $\tau(f^j) \equiv \sum_{i=1}^{n} \mu_i \cdot f_i^j$.

Application I: Costly Persuasion

Gentzkow, Kamenica (2014) Revisited

- Consider a costly persuasion problem with the standard example
  - State: $\{innocent, guilty\}$
  - Receiver's action: **A**cquit or **C**onvict
  - Sender's payoff: $u_S(C) = 1$, $u_S(A) = 0$
  - Receiver's payoff: $u_R(A, innocent) = u_R(C, guilty) = 1$
    $u_R(C, innocent) = u_R(A, guilty) = 0$
  - Sender commits to an experiment at some cost

- GK focuses on posterior separable costs (e.g., entropy cost) to utilize concavification technique

- Can we solve this problem with any Blackwell-monotone information cost function?

Gentzkow, Kamenica (2014) Revisited

- Consider a costly persuasion problem with the standard example
  - State: $\{innocent, guilty\}$
  - Receiver's action: **A**cquit or **C**onvict
  - Sender's payoff: $u_S(C) = 1$, $u_S(A) = 0$
  - Receiver's payoff: $u_R(A, innocent) = u_R(C, guilty) = 1$
    $u_R(C, innocent) = u_R(A, guilty) = 0$
  - Sender commits to an experiment at some cost
- GK focuses on posterior separable costs (e.g., entropy cost) to utilize concavification technique
- Can we solve this problem with any Blackwell-monotone information cost function?

Gentzkow, Kamenica (2014) Revisited

- Consider a costly persuasion problem with the standard example
  - State: $\{innocent, guilty\}$
  - Receiver's action: **A**cquit or **C**onvict
  - Sender's payoff: $u_S(C) = 1$, $u_S(A) = 0$
  - Receiver's payoff: $u_R(A, innocent) = u_R(C, guilty) = 1$
    $u_R(C, innocent) = u_R(A, guilty) = 0$
  - Sender commits to an experiment at some cost
- GK focuses on posterior separable costs (e.g., entropy cost) to utilize concavification technique
- Can we solve this problem with any Blackwell-monotone information cost function?

Costly Persuasion with Blackwell-Monotone Information Cost

- It is without loss to consider binary experiments since **R**'s action is binary
    - $f_2 = \Pr(C|guilty)$ and $f_1 = \Pr(C|innocent)$

- When the prior is p, the sender's problem is

$$\max_{0 \le f_1 \le f_2 \le 1} pf_2 + (1-p)f_1 - C(f_1, f_2)$$

subject to

$$\frac{pf_2}{pf_2 + (1-p)f_1} \ge \frac{1}{2}.$$

- When $p \ge 1/2$, the solution is $f_1 = f_2 = 1$: always convict costlessly

Costly Persuasion with Blackwell-Monotone Information Cost

- It is without loss to consider binary experiments since **R**'s action is binary
  - $f_2 = \Pr(C|guilty)$ and $f_1 = \Pr(C|innocent)$

- When the prior is p, the sender's problem is

$$\max_{0 \leq f_1 \leq f_2 \leq 1} pf_2 + (1-p)f_1 - C(f_1, f_2)$$

subject to

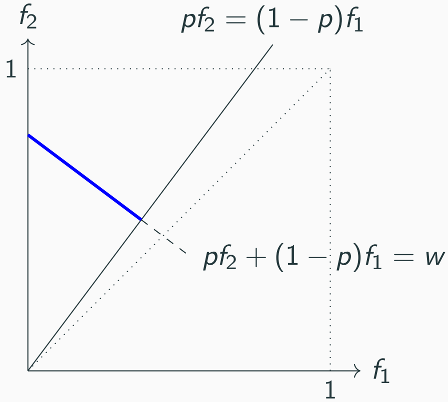
$$\frac{pf_2}{pf_2 + (1-p)f_1} \geq \frac{1}{2}.$$

- When $p \geq 1/2$, the solution is $f_1 = f_2 = 1$: always convict costlessly

Cost Minimization

- Suppose $p < 1/2$.
- Cost minimization problem under $pf_2 + (1-p)f_1 = w$:

$$\min C(f_1, f_2) \quad \text{s.t.} \quad \begin{aligned} pf_2 + (1-p)f_1 &= w, \\ pf_2 &\geq (1-p)f_1 \end{aligned}$$

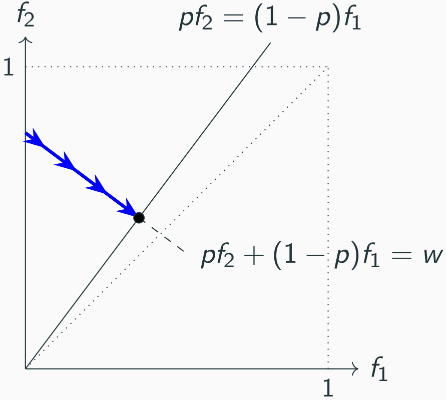


- **Proposition**: for any Blackwell-monotone information cost function, the cost is minimized when $pf_2 = (1-p)f_1$

Cost Minimization

- Suppose $p < 1/2$.
- Cost minimization problem under $pf_2 + (1-p)f_1 = w$:

$$\min C(f_1, f_2) \quad \text{s.t.} \quad \begin{array}{l} pf_2 + (1-p)f_1 = w, \\ pf_2 \geq (1-p)f_1 \end{array}$$



- **Proposition**: for any Blackwell-monotone information cost function, the cost is minimized when $pf_2 = (1-p)f_1$

Sender's Problem

- When $pf_2 + (1-p)f_1 = w$, the cost is minimized at

$$f_2 = \frac{w}{2p} \quad \text{and} \quad f_1 = \frac{w}{2(1-p)}.$$

- Now the sender's problem is

$$\max_{0 \leq w \leq 2p} w - C\left(\frac{w}{2(1-p)}, \frac{w}{2p}\right) \tag{4}$$

- From here on, a specific cost function is needed

Sender's Problem

- When $pf_2 + (1-p)f_1 = w$, the cost is minimized at

$$f_2 = \frac{w}{2p} \quad \text{and} \quad f_1 = \frac{w}{2(1-p)}.$$

- Now the sender's problem is

$$\max_{0 \le w \le 2p} w - C\left(\frac{w}{2(1-p)}, \frac{w}{2p}\right) \tag{4}$$

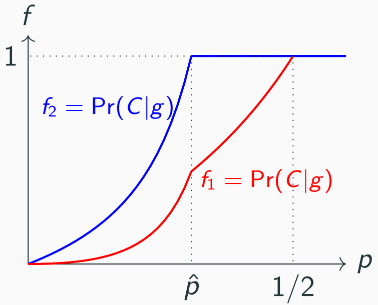
- From here on, a specific cost function is needed

Costly Persuasion with Non-Posterior-Separable Cost

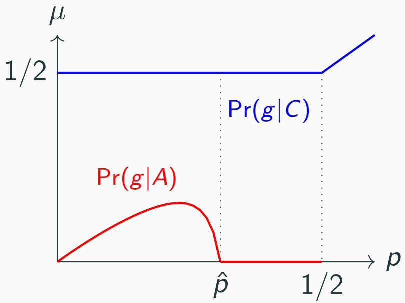
- When $C(f_1, f_2) = (f_2 - f_1)^2$, the solution for $p < 1/2$ is

$$f_2(p) = \min\left\{1, \ \frac{(1-p)^2 p}{(1-2p)^2}\right\} \quad \text{and} \quad f_1(p) = \frac{p}{1-p} \cdot f_2(p).$$

▸ Entropy



Optimal Experiments



Posteriors

Application II: Bargaining and Information Acquisition

Chatterjee, Dong, Hoshino (2023)

- Consider a bargaining problem with information acquisition
  - Players: **S**eller and **B**uyer
  - State (**B**'s valuation): $v \in \{L, H\}$ with $H > L > 0$
    - Prior belief: $\pi \equiv \Pr(v = H) \in (0, 1)$
  - Timing of the game
    1. Nature draws v and **S** observes v
    2. **S** offers p
    3. **B** costly acquires information about v and then accepts or rejects

- Chatterjee et al. focus on specific types of information acquisition

- We extend their analysis by allowing **B** to choose information flexibly

Chatterjee, Dong, Hoshino (2023)

- Consider a bargaining problem with information acquisition
  - Players: **S**eller and **B**uyer
  - State (**B**'s valuation): $v \in \{L, H\}$ with $H > L > 0$
    - Prior belief: $\pi \equiv \Pr(v = H) \in (0, 1)$
  - Timing of the game
    1. Nature draws v and **S** observes v
    2. **S** offers p
    3. **B** costly acquires information about v and then accepts or rejects

- Chatterjee et al. focus on specific types of information acquisition

- We extend their analysis by allowing **B** to choose information flexibly

Chatterjee, Dong, Hoshino (2023): H-focused information

B's cost: $\lambda \cdot c(f_H)$

Result 1: pooling eq'm

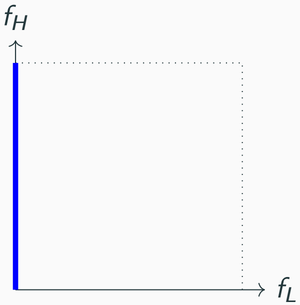
under H-focused signal structure, for any λ, there exists $\epsilon > 0$ such that every equilibrium is a pooling equilibrium where

1. both types of **S** offer $p^* \in [L, L+\epsilon)$;
2. **B** accepts without information acquisition.

Moreover, $\epsilon \to 0$ as $\lambda \to 0$, thus, **B** extracts full surplus as $\lambda \to 0$

H-focused Information

	s_L	s_H
L	1	0
H	$1 - f_H$	f_H



Chatterjee, Dong, Hoshino (2023): L-focused information

B's cost: $\lambda \cdot c(1 - f_L)$

Result 2: almost-separating eq'm

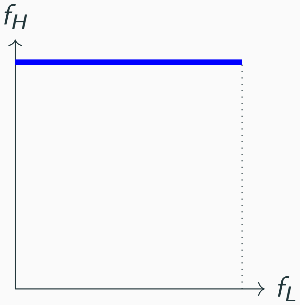
under L-focused signal structure, for any small enough λ, there exists an equilibrium such that

1. type H **S** offers $p^* \approx H$;
2. type L **S** offers L with prob. $1 - \epsilon$, p^* with prob. ϵ;
3. **B** acquires information and conditions her purchase decision on the signal realization

Moreover, **S**'s payoff is close to v and **B**'s payoff is close to zero

L-focused Information

	s_L	s_H
L	$1 - f_L$	f_L
H	0	1



Flexible Information Acquisition

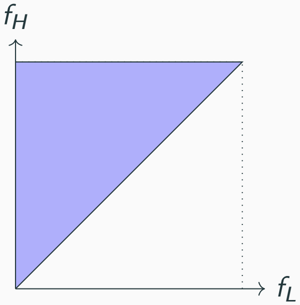
- We extend to the full domain and consider $\lambda|f_2 - f_1|$ and $\lambda(f_2 - f_1)^2$

Result 1': when $C(f_1, f_2) = \lambda|f_2 - f_1|$, the unique equilibrium is the pooling equilibrium, and as $\lambda \to 0$, **B** extracts full surplus

Result 2': when $C(f_1, f_2) = \lambda(f_2 - f_1)^2$, there exists an almost-separating equilibrium, and **S**'s payoff is close to v and **B**'s payoff is close to zero

Flexible Information

	s_L	s_H
L	$1 - f_L$	f_L
H	$1 - f_H$	f_H



Conclusion

Conclusion

- We identify necessary and sufficient conditions for Blackwell Monotonicity.
- Under additive separability, we show that the sublinearity of the primitive function is equivalent to Blackwell Monotonicity.
- Our technique allows us to
  - solve the costly persuasion problem with any Blackwell-monotone information costs
  - solve the bargaining problem with information acquisition in the extended domain
- Future Research: Lehmann-Monotone Information Costs

Conclusion

- We identify necessary and sufficient conditions for Blackwell Monotonicity.
- Under additive separability, we show that the sublinearity of the primitive function is equivalent to Blackwell Monotonicity.
- Our technique allows us to
  - solve the costly persuasion problem with any Blackwell-monotone information costs
  - solve the bargaining problem with information acquisition in the extended domain
- Future Research: Lehmann-Monotone Information Costs

Thank You!

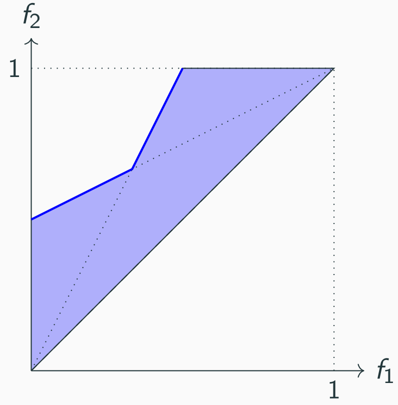
Related Literature

- **Posterior-based information costs**
  - Entropy cost: Sims [2003]; Matějka, Mckay [2015]
  - Decision theory: Caplin, Dean [2015]; Caplin, Dean, Leahy [2022]; Chambers, Liu, Rehbeck [2020]; Denti [2022]
  - Applications: Ravid [2020]; Zhong [2022]; Gentzkow, Kamenica [2014]
- **Experiment-based information costs**
  - LLR cost: Pomatto, Strack, Tamuz [2023];
  - Applications: Denti, Marinacci, Rustichini [2022]; Ramos-Mercado [2023]

Go Back

Quasiconvexity

- The following information cost function for binary experiments is not quasiconvex

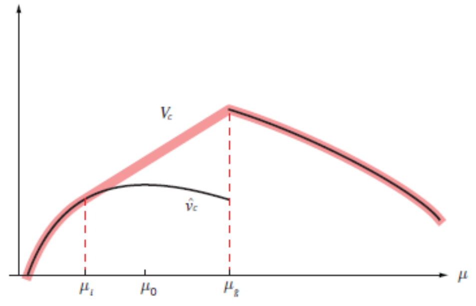


$$C(f_1, f_2) = \min\left\{\frac{f_2}{f_1}, \frac{1-f_1}{1-f_2}\right\}$$
$$= \min\{\alpha, \beta\}$$

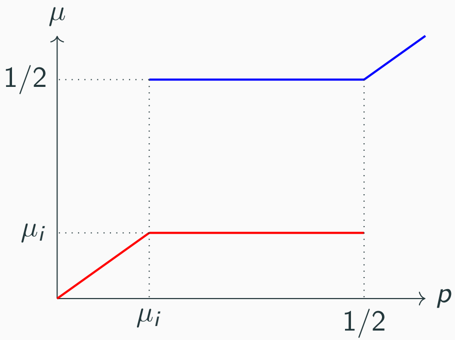
Go Back

Gentzkow, Kamenica (2014) Revisited

- Entropy cost: $k \cdot \mathbb{E}_{\pi|p}[H(p) - H(\mu_s)]$ where $H(\mu) \equiv -\sum_\omega \mu(\omega)\log(\mu(\omega))$
  - p is prior, and μ_i and μ_g are posteriors from an experiment π



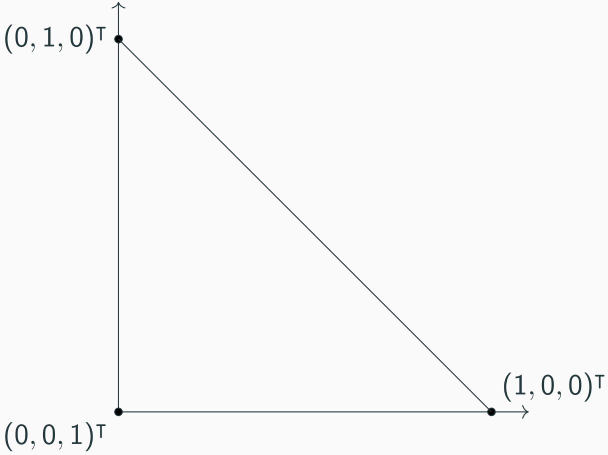
Concavification



Posteriors

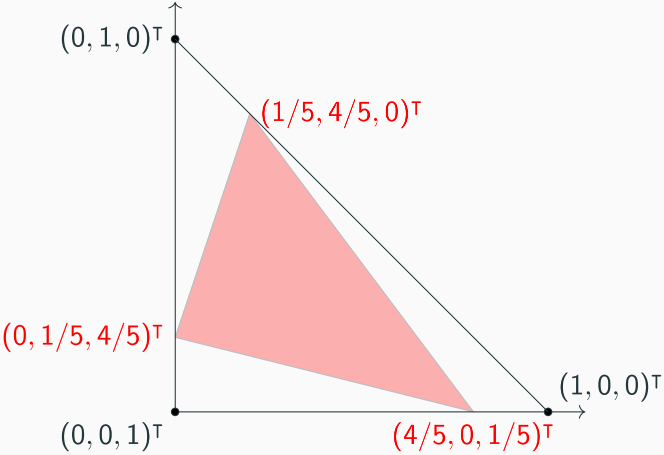
Go Back

- When $n = m = 3$, $f \succeq_B g$ iff the triangle generated by f^1, f^2, f^3 includes the one generated by g^1, g^2, g^3



Go Back

- When $n = m = 3$, $f \succeq_B g$ iff the triangle generated by f^1, f^2, f^3 includes the one generated by g^1, g^2, g^3



1. **Positive homegenity:** Note that $\psi(\mathbf{0}) = 0$. For any $k \in \mathbb{N}$,

$$[\hat{f}, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{1} - \hat{f}] \sim_B [\hat{f}/k, \hat{f}/k, \cdots, \hat{f}/k, \mathbf{1} - \hat{f}] \quad \Rightarrow \quad \psi(\hat{f}) = k\, \psi(\hat{f}/k).$$

   Then, for any $(k, l) \in \mathbb{N}^2$, we also have

$$\frac{l}{k}\, \psi(\hat{f}) = l\, \psi\left(\frac{\hat{f}}{k}\right) = \psi\left(\frac{l}{k}\, \hat{f}\right)$$

   By density of $\mathbb{Q}$ in $\mathbb{R}$ and the continuity of ψ, $\psi(\alpha\, \hat{f}) = \alpha\, \psi(\hat{f})$ for all $\alpha \in \mathbb{R}_+$

2. **Subadditivity**:

$$[\hat{f}, \hat{g}, \mathbf{1} - \hat{f} - \hat{g}] \succeq_B [\hat{f} + \hat{g}, \mathbf{0}, \mathbf{1} - \hat{f} - \hat{g}] \quad \Rightarrow \quad \psi(\hat{f}) + \psi(\hat{g}) \geq \psi(\hat{f} + \hat{g})$$

Go Back