Information Orders

Topics in Economic Theory: Part 2

Duarte Gonçalves

University College London

24 February 2022

Introductions

Me: Duarte Gonçalves (“Doo-art”)

- Fields: Information, Game Theory, Behavioral, Experimental, Political Economy
- Topics: info acquisition, belief formation, strategic uncertainty, learning, conflict
- Contact: duarte.goncalves@ucl.ac.uk; Drayton House, rm 112

This course:

- Information: value, cost, acquisition
- Applications: pricing, inflation, voting, bargaining, drug approval, job search, policy exploration, media and political polarization

Reading Group?

- Lectures: groundwork; dry. Limited time
- Reading group with emphasis on applications?

Motivation

Information acquisition is crucial in understanding behavior

- Which computer to buy: learn specs, check reviews, known issues
- Who to vote for: check platform, read interviews to understand personality, watch debates to understand possible post-election coalition
- How price product: learn competitors pricing strategy, study demand
- Conflict: military espionage, covert operations
- Developing a new product: tech may or may not prove successful

Motivation

Our focus in this course: **Acquiring Info**

- Offline Learning (Sampling): learning by existing data
  - e.g. query to database, procuring data, search/expand consideration set (learning product exists), R&D
- Choosing Sources
  - e.g. news outlets, financial advisors

Next year

- Online Learning (Experimentation): learning by doing
  - e.g. experience goods (movies), trying out different investment strategies
- Decision Time: how much data, time reveals preference intensity,
  - timed stochastic choice data
- Information Acquisition in Strategic Settings

Motivation

Related Issues

Design: choosing info structure: e.g. which macro variables to measure, designing tests (education, certification, credit ratings), how to bundle data

Processing: info acquisition s.t. restrictions (can only learn what is available to be learned): e.g. learning specs of computer

Belief Updating: how do people *use* info to revise beliefs: e.g. memory, info about info, motivated reasoning

Pricing: how to price info: e.g. advisory services, databases

Transmission: garbling available info: e.g. advertising, financial reports, speeches, strategic disclosure, statistical office announcements

Encriptation: secure transmission of info

Diffusion: how does info spread: e.g. social networks, misinfo, inflation

Today

‘More information is better’: what is the sense of something being more informative?

Today: acquire basic technical tools on information orders (concepts, results)

Acquiring information *is* costly, but will pay off!

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
   - Payoff Richness
   - Posterior Distribution
   - Infinite Signal Space
   - Examples

2. Lehmann Order
   - Monotonote Likelihood Ratio Property
   - Lehmann Order
   - Equivalence Result

3. Blackwell in Large Samples
   - Setup
   - Large Sample Order
   - Rényi t-Divergence
   - Equivalence Result
   - Dominance Ratio

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
     - Setup
     - Value of Information
     - Sufficiency
     - Equivalence Result
     - Proof by Crémer (’82 JET)
     - Proof by Oliveira (’18 GEB)
   - Payoff Richness
   - Posterior Distribution
   - Infinite Signal Space
   - Examples

2. Lehmann Order

3. Blackwell in Large Samples

Modelling Information

Unknown state of the world $\theta \in \Theta = \{\theta_i\}_{i \in [N]}$

Prior prob distribution $\mu \in \Delta(\Theta)$

Decision problem: $u : A \times \Theta \to \mathbb{R}$; vNM utility, A compact

For simplicity, identify action $a \in A$ with vector of state-contingent payoffs $a_i := u(a, \theta_i)$, $A \subset \mathbb{R}^N$

Statistical experiment or information structure $\pi := (\pi_{ij})_{i \in [N], j \in [K]}$

Prob $x|\theta = \pi(x|\theta)$: column-stochastic matrix ($\pi_{ij} \in [0, 1]$, $\sum_{j \in [N]} \pi_{ij} = 1$)

Finite signal space X, $|X| = K$; signal labels have no content

WLOG $\pi(x) := \sum_\theta \pi(x|\theta)\mu(\theta) > 0 \ \forall x$

Modelling Information

$$\pi = \begin{pmatrix} \pi(x_1 \mid \theta_1) & \cdots & \pi(x_1 \mid \theta_N) \\ \vdots & \pi(x \mid \theta) & \vdots \\ \pi(x_K \mid \theta_1) & \cdots & \pi(x_K \mid \theta_N) \end{pmatrix}$$

Value of Information

Posterior prob $\theta|x$: $\mu(\theta|x) = \mu(\theta)\pi(x|\theta)/\pi(x)$

EU max: $U^A(\mu|x) := \max_{a \in A} \sum_\theta u(a, \theta)\mu(\theta|x)$

Gross value of experiment π in decision problem A given prior μ:

$$G^A(\pi, \mu) := \sum_x U^A(\mu|x)\pi(x)$$

Definition

Given μ and A, π **is more valuable than** π' if $G^A(\pi, \mu) \geq G^A(\pi', \mu)$.

Given μ, π **is more informative than** π' if it is more valuable for all decision problems.

Statistical Sufficiency

Experiment: adding noise to direct observation of state θ

Instead of observing θ, get signal $x \sim \pi(x)$

Fully informative if π is identity matrix (up to permutation rows/cols)

Fully uninformative if $\pi(x|\theta) = \pi(x|\theta')$, $\forall x \in X$, $\forall \theta, \theta' \in \Theta$

(each row is constant vector)

Adding independent noise to π

π' adds state-independent noise to $\pi \equiv \pi'$ produces signal y about signal x

$\mathcal{B}(K', K)$: set of all $K' \times K$ col-stochastic matrices; b is independent from θ

$b \in \mathcal{B}(K', K)$: stochastic transformation matrix

Statistical Sufficiency

Definition

An experiment $\pi_{K\times N}$ is **sufficient** for $\pi'_{K'\times N}$ if $\exists b \in \mathcal{B}(K', K)$ such that $b\pi = \pi'$.

$$
\begin{aligned}
b\pi &= \begin{pmatrix} b(y_1 \mid x_1) & \cdots & b(y_1 \mid x_K) \\ \vdots & b(y \mid x) & \vdots \\ b(y_{K'} \mid x_1) & \cdots & b(y_{K'} \mid x_K) \end{pmatrix} \begin{pmatrix} \pi(x_1 \mid \theta_1) & \cdots & \pi(x_1 \mid \theta_N) \\ \vdots & \pi(x \mid \theta) & \vdots \\ \pi(x_K \mid \theta_1) & \cdots & \pi(x_K \mid \theta_N) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{x'} b(y_1 \mid x')\pi(x' \mid \theta_1) & \cdots & \sum_{x'} b(y_1 \mid x')\pi(x' \mid \theta_N) \\ \vdots & \sum_{x'} b(y \mid x')\pi(x' \mid \theta) & \vdots \\ \sum_{x'} b(y_{K'} \mid x')\pi(x' \mid \theta_1) & \cdots & \sum_{x'} b(y_{K'} \mid x')\pi(x' \mid \theta_N) \end{pmatrix} \\
&= \begin{pmatrix} \pi'(y_1 \mid \theta_1) & \cdots & \pi'(y_1 \mid \theta_N) \\ \vdots & \pi'(y \mid \theta) & \vdots \\ \pi'(y_K \mid \theta_1) & \cdots & \pi'(y_K \mid \theta_N) \end{pmatrix} = \pi'
\end{aligned}
$$

Equivalence Theorem

Theorem (Blackwell '51, '53)

If experiment π is sufficient for experiment π', then π is more informative than π' for any prior $\mu \in \Delta(\Theta)$.

If π is more informative than π' for some $\mu \in \text{int}(\Delta(\Theta))$, then π is sufficient for π'.

Sufficient $\Longrightarrow$ + informative is unsurprising

if agent has π, can always mimic π' by randomizing themselves

Sufficient $\Longleftarrow$ + informative is surprising:

result does not depend on prior

Proof by Crémer ('82 JET)

Proof by contrapositive: not sufficient $\iff$ find problem A & prior μ s.t. $G^A(\pi', \mu) > G^A(\pi, \mu)$

(i) $\pi' \notin \mathcal{B}\pi := \{b\pi \mid b \in \mathcal{B}\}$ (i.e. $\nexists b \in \mathcal{B}$ s.t. $\pi' = b\pi$)

$\iff$ (ii) $\exists q_{y,\theta} \in \mathbb{R}^{K',N}$ s.t. $\forall b \in \mathcal{B}$,

$$\sum_{y,\theta} q_{y,\theta}\pi'(y|\theta) > \sum_{y,\theta} q_{y,\theta}(b\pi)(y|\theta) = \sum_{y,\theta} q_{y,\theta} \sum_{x} b(y|x)\pi(x|\theta) = \sum_{x}\sum_{y} b(y|x) \sum_{\theta} q_{y,\theta}\pi(x|\theta)$$

Proof: $\mathcal{B}\pi$: closed and convex; by separating hyperplane thm, $\pi' \notin \mathcal{B}\pi \iff \exists$ such q

$\iff$ (iii) $\exists q_{y,\theta} \in \mathbb{R}^{K',N}$ s.t. $\forall b \in \mathcal{B}$,

$$\sum_{y,\theta} q_{y,\theta}\pi'(y|\theta) > \sum_{x} \max_{y} \left\{ \sum_{\theta} q_{y,\theta}\pi(x|\theta) \right\}$$

Proof: $\sum_x \max_y \sum_\theta q_{y,\theta}\pi(x|\theta) \geq \sum_x \sum_y b(y|x) \sum_\theta q_{y,\theta}\pi(x|\theta)$ and $\exists b \in \mathcal{B} : b(y|x) = 1_{y=y^*(x)}$, where $y^*(x)$ selection of $\arg\max_y \sum_\theta q_{y,\theta}\pi(x|\theta)$

Proof by Crémer ('82 JET)

(iii) $\exists q_{y,\theta} \in \mathbb{R}^{K',N}$ s.t. $\forall b \in \mathcal{B}$,

$$\sum_{y,\theta} q_{y,\theta}\pi'(y|\theta) > \sum_{x} \max_{y} \left\{ \sum_{\theta} q_{y,\theta}\pi(x|\theta) \right\}$$

$\iff$ (iv) $\forall \mu \gg 0$, $\exists u_{a^y,\theta} \in \mathbb{R}^{K',N}$ s.t.

$$\sum_{y} \pi'(y) \sum_{\theta} u(a^y,\theta)\mu(\theta|y) > \sum_{x} \pi(x) \max_{y} \left\{ \sum_{\theta} u(a^y,\theta)\mu(\theta|x) \right\}$$

Proof: $u(a^y,\theta) := q_{y,\theta}/\mu(\theta)$; then $q_{y,\theta}\pi'(y|\theta) = q_{y,\theta}\frac{\pi'(y)\mu(\theta|y)}{\mu(\theta)} = \pi'(y)u(a^y,\theta)\mu(\theta|y)$

$\iff$ (v) $\exists A, \mu$ such that $G^A(\pi',\mu) > G^A(\pi,\mu)$

Proof: $A = \{a^y\}_y$, $u : A \times \Theta$ s.t. $u(a^y,\theta)$

$G^A(\pi',\mu) \geq \sum_y \pi'(y) \sum_\theta u(a^y,\theta)\mu(\theta|y) > \sum_x \pi(x) \max_y \left\{ \sum_\theta u(a^y,\theta)\mu(\theta|x) \right\} = G^A(\pi,\mu)$

Proof by Oliveira ('18 GEB)

Category:

(i) collection of objects ($X, Y, Z, ...$),

(ii) arrows between objects ($\alpha, \beta, \gamma, ...$),

(iii) binary operation $\circ$ s.t. if $\alpha : X \to Y$ and $\beta : Y \to Z$, then $\beta \circ \alpha : X \to Z$

satisfying

(1) Every object Y has identity arrow $id_Y : Y \to Y$ s.t. for any $\alpha : X \to Y$ and $\beta : Y \to Z$,

$id_Y \circ \alpha = \alpha$ and $\beta \circ id_Y = \beta$

(2) $\circ$ is associative

Proof by Oliveira ('18 GEB)

Category:

(i) objects $(X, Y, Z, ...)$, (ii) arrows between objects $(\alpha, \beta, \gamma, ...)$, (iii) $\circ$ s.t. if $\alpha : X \to Y$ and $\beta : Y \to Z$, then $\beta \circ \alpha : X \to Z$ sat. (1) $\exists id_Y : Y \to Y$ s.t. $\alpha : X \to Y$ and $\beta : Y \to Z$; (2) $\circ$ associative

Re-write definitions

Information structure: $\pi : \Theta \to \Delta(X)$; feasible signals $S(\pi) := \cup_\theta \text{supp}(\pi(\theta))$

Stochastic map: $\alpha : X \to \Delta(Y)$

Extension of a stochastic map: $\bar{\alpha} : \Delta(X) \to \Delta(Y)$, $\bar{\alpha}(q) := \sum_X \mu(x)\alpha(x)$

Composition $\beta \circ \alpha : X \to \Delta(Z)$ of two stochastic maps $\alpha : X \to \Delta(Y)$ and $\beta : Y \to \Delta(Z)$

$(\beta \circ \alpha)(z|x) = \sum_y \beta(z|y)\alpha(y|x)$; composition is commutative, $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$

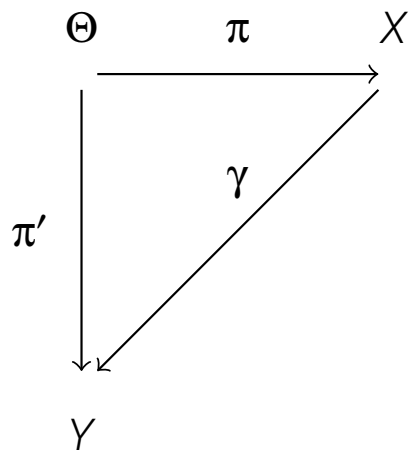
Proof by Oliveira ('18 GEB)

Category:

(i) objects $(X, Y, Z, ...)$, (ii) arrows between objects $(\alpha, \beta, \gamma, ...)$, (iii) $\circ$ s.t. if $\alpha : X \to Y$ and $\beta : Y \to Z$, then $\beta \circ \alpha : X \to Z$ sat. (1) $\exists id_Y : Y \to Y$ s.t. $\alpha : X \to Y$ and $\beta : Y \to Z$; (2) $\circ$ associative

Re-write definitions

Garbling: π' is a garbling of π if $\exists$ stochastic map $\gamma : S(\pi) \to \Delta(S(\pi'))$ s.t. $\pi' = \gamma \circ \pi$



Proof by Oliveira ('18 GEB)

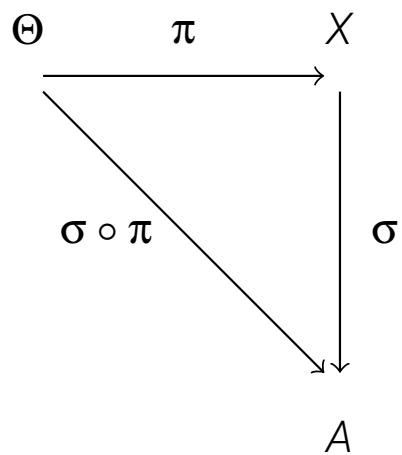
Information allows agent to correlate actions with state

Feasible strategies given π: $\Sigma(\pi) := \{\sigma : S(\pi) \to \Delta(A)\}$

Feasible conditional distribution over actions given π:

$P(\pi) := \{\sigma \circ \pi : \Theta \to \Delta(A) \mid \sigma \in \Sigma(\pi)\}$

$\implies G^A(\pi, \mu) = \max_{q \in P(\pi)} \sum_\theta \left(\sum_a q(a|\theta) u(a, \theta)\right) \mu(\theta)$



Proof by Oliveira ('18 GEB)

Theorem (Blackwell '51, '53)

The following are equivalent:

(1) π' is a garbling of π

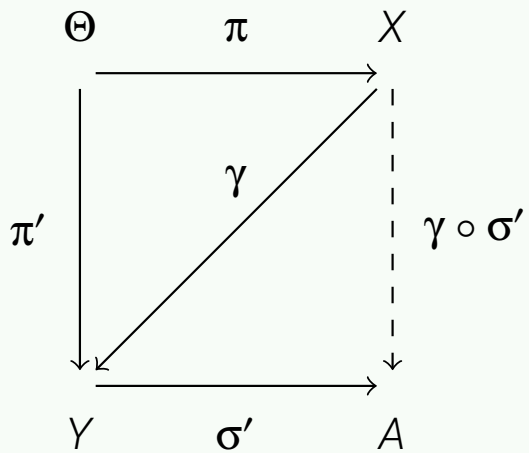
(2) $\forall A, P(\pi) \supseteq P(\pi')$

(3) $\forall A, \mu, G^A(\pi, \mu) \geq G^A(\pi', \mu)$

Proof

(1) $\Longrightarrow$ (2)

(1) $\equiv \exists \gamma : \gamma \circ \pi = \pi'$



Proof by Oliveira ('18 GEB)

Theorem (Blackwell '51, '53)

The following are equivalent:

(1) π' is a garbling of π

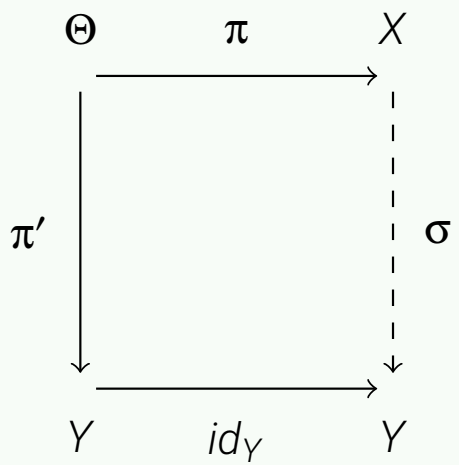
(2) $\forall A, P(\pi) \supseteq P(\pi')$

(3) $\forall A, \mu, G^A(\pi, \mu) \geq G^A(\pi', \mu)$

Proof

(1) $\Longleftarrow$ (2) $\equiv \forall A, (q \in P(\pi') \Longrightarrow q \in P(\pi)) \Longrightarrow q = \sigma' \circ \pi' = \sigma \circ \pi)$

Set $A = Y, \sigma' = id_Y \Longrightarrow \pi' = id_Y \circ \pi' = \sigma \circ \pi$



Proof by Oliveira ('18 GEB)

Theorem (Blackwell '51, '53)

The following are equivalent:

(1) π' is a garbling of π

(2) $\forall A, P(\pi) \supseteq P(\pi')$

(3) $\forall A, \mu, G^A(\pi, \mu) \geq G^A(\pi', \mu)$

Proof

(2) $\Longrightarrow$ (3) immediate

(2) $\Longleftarrow$ (3)

$\neg$(2) implies $\exists q \in P(\pi') \setminus P(\pi) \subset \mathbb{R}^{\Theta \times A}$

$P(\pi)$ convex and compact $\rightarrow$ use sep. hyperplane thm

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
   - Payoff Richness
   - Posterior Distribution
   - Infinite Signal Space
   - Examples

2. Lehmann Order

3. Blackwell in Large Samples

Payoff Richness

Original proof: experiment is payoff richer than another if set state-contigent feasible payoffs is a superset; implied by $P(\pi) \supseteq P(\pi')$

$L(\pi, A) := \{(\sum_a q(a|\theta)u(a,\theta))_\theta,\ q \in P(\pi)\}$

$P(\pi) \supseteq P(\pi') \implies L(\pi, A) \supseteq L(\pi', A),\ \forall A$

What about $\impliedby$?

Payoff Richness

Theorem 4 (Blackwell '51, '53)

Let $|S(\pi)| = K$, $|S(\pi')| = K'$, $|\Theta| = N$. If π is payoff richer than π', then $\forall D'_{N\times K'} \subset \mathbb{R}^{N,K'}$, $\exists b \in \mathcal{B}(K', K)$ s.t. $\text{tr}(D'b\pi) = \text{tr}(D'\pi')$

Interpreting D': let $A = Y$ and view each col as payoff vector $(u(a,\theta))_{a\in Y}$

Proof

$(D'\pi')_{N\times N} = (\sum_a u(a,\theta)\pi'(a|\theta'))_{\theta,\theta'}$

$\text{diag}(D'\pi') = (\sum_a u(a,\theta)\pi'(a|\theta))_\theta \in L(\pi', Y) \subseteq L(\pi, Y)$

Hence, by payoff richness, $\exists D$ such that $\text{diag}(D'\pi') = \text{diag}(D\pi) = (\sum_\theta D_{\theta,x}\pi(x|\theta))_\theta$

Every col in D is in convex hull D' cols

$\implies \exists b \in \mathcal{B}(K', K)$ s.t. $\text{tr}(D'b\pi) = \text{tr}(D'\pi')$

Payoff Richness

Want to show $P(\pi) \supseteq P(\pi') \impliedby L(\pi, A) \supseteq L(\pi', A)$, $\forall A$

Proof

Let $\mathcal{D} := N \times K'$ matrices such that $D_{ij} \in [0, 1]$

Set auxillary 2-player zero-sum game: P1 chooses $D \in \mathcal{D}$, P2 chooses $b \in \mathcal{B}(K', K)$

Payoffs to player 1 $\text{tr}(D(b\pi - \pi'))$

Trace: N expected payoff under uniform prior

Strategy space: convex, compact; Payoffs: continuous in s, linear in s_i

$\implies \exists$ PSNE

Theorem 4 $\implies$ P2 can secure payoff ≤ 0; i.e. $\exists b : \text{tr}(D(b\pi - \pi')) \leq 0$

$\implies b\pi - \pi' \leq 0$ (otherwise P1 could deviate and get strictly positive payoff)

As $b\pi$ and π' are stochastic matrices, $\sum_y (\pi'(y|\theta) - \sum_x b(y|x)\pi(x|\theta)) = 0 \; \forall y \in Y$

Hence, it is not possible that $b\pi - \pi' \leq 0$ and $b\pi - \pi' \neq 0$

$\implies b\pi = \pi'$

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
   - Payoff Richness
   - Posterior Distribution
     - Stochastic Orders
     - MPS of Posterior Beliefs
   - Infinite Signal Space
   - Examples

2. Lehmann Order

3. Blackwell in Large Samples

Stochastic Orders

Definition

Let X, Y be r.v. taking values in $\mathbb{R}^n$.

P_X **first-order stochastically dominates** P_Y ($P_X \geq_{FOSD} P_Y$) if $\mathbb{E}_{P_X}[u(X)] \geq \mathbb{E}_{P_Y}[u(Y)]$ for all nondecreasing $u : \mathbb{R}^n \to \mathbb{R}$ whenever both expectations exist.

P_X **second-order stochastically dominates** P_Y ($P_X \geq_{SOSD} P_Y$) if $\mathbb{E}_{P_X}[u(X)] \geq \mathbb{E}_{P_Y}[u(Y)]$ for all nondecreasing concave $u : \mathbb{R}^n \to \mathbb{R}$ whenever both expectations exist.

Stochastic Orders

Definition

$P_X \geq_{FOSD} P_Y$ if $\mathbb{E}_{P_X}[u(X)] \geq \mathbb{E}_{P_Y}[u(Y)]$ $\forall$ nondec $u : \mathbb{R}^n \to \mathbb{R}$

$P_X \geq_{SOSD} P_Y$ if $\mathbb{E}_{P_X}[u(X)] \geq \mathbb{E}_{P_Y}[u(Y)]$ for all nondec concave $u : \mathbb{R}^n \to \mathbb{R}$

Proposition

$P_X \geq_{FOSD} P_Y$ iff $\mathbb{P}_{P_X}(X \in M) \geq \mathbb{P}_{P_Y}(Y \in M)$ for all upper sets $M \subseteq \mathbb{R}^n$ s.t. M is closed and sat. $x \in M \implies y \in M \, \forall y \geq x$

$P_X \geq_{FOSD} P_Y$ iff $\mathbb{P}_{P_X}(X \in L) \leq \mathbb{P}_{P_Y}(Y \in L)$ for all lower sets $L \subseteq \mathbb{R}^n$ s.t. M is closed and sat. $x \in L \implies y \in L \, \forall y \leq x$

$P_X \geq_{SOSD} P_Y$ iff $\exists X \sim F$ and $Y \sim G$ s.t. $X \geq \mathbb{E}[Y|X]$ a.s. (Strassen's ('65) Theorem)

References: Mü & Stoyan ('02), Sriboonchitta, Wong, Dhompongsa, & Nguyen ('09)

Great "tools" paper on comparative statics using stochastic orders: Jensen ('02, REStud)

Stochastic Orders

Definition

Let X, Y be r.v. taking values in $\mathbb{R}^n$.

A P_X is a **mean-preserving spread** of P_Y ($P_Y \geq_{MPS} P_X$ if $\mathbb{E}[X|Y] = Y$. P_Y is said a mean-preserving contraction of P_X.

Proposition

P_X MPS P_Y iff $X \overset{d}{=} Y + \boldsymbol{\varepsilon}$ where $\mathbb{E}[\boldsymbol{\varepsilon}|Y] = \mathbf{0}$ $(\boldsymbol{\varepsilon} := \mathbb{E}[X|Y] - X)$

P_X MPS $P_Y \Longrightarrow P_Y \geq_{SOSD} P_X$

P_X MPS P_Y iff $\mathbb{E}_{P_X}[u(X)] \geq (\leq) \mathbb{E}_{P_Y}[u(Y)]$ for all ~~nondecreasing~~ convex (concave) u : $\mathbb{R}^n \to \mathbb{R}$ whenever both expectations exist

$\geq_{MPS}, \geq_{FOSD} \subseteq \geq_{SOSD}$

MPS of Posterior Beliefs

MPS: why are we talking about this?

$\mu \in \Delta(\Theta)$, $x \sim \pi(\cdot|\theta)$

$U^A : \Delta(\Theta) \to \mathbb{R}$, $U^A(\mu) := \max_{a \in A} \mathbb{E}_\mu[u(a, \theta)]$

U^A is convex in beliefs

Posterior beliefs are a mean preserving spread of the prior: $\mathbb{E}[\mu(\theta|x)] = \mu(\theta)\ \forall \theta$

i.e. information generates MPS of beliefs

$\implies G^A(\pi, \mu) \geq G^A(\pi', \mu) = U^A(\mu)$ for any fully uninformative π' ($\equiv$ no info)

Idea: + informative $\iff$ MPS of posteriors

MPS of Posterior Beliefs

Proposition

π is sufficient for π' iff $\mu(\theta|x)$ MPS of $\mu(\theta|y)$ $\forall \mu$, where $x \sim \pi(\cdot|\theta)$, $y \sim \pi'(\cdot|\theta)$

Proof

Let $\mu \gg 0$ wlog, $|\Theta| = N$, $|S(\pi)| = |X| = K$, $|S(\pi')| = |Y| = K'$, $\beta, b \in \mathcal{B}(K, K')$

$\mu_{1\times N}$, $\pi_{N\times K}$, $\pi'_{N\times K'}$

(i) $\mu(\cdot|x)$ MPS $\mu(\cdot|y)$ if $\exists \beta \in \mathcal{B}(K, K')$ s.t. $\sum_x \beta_{xy}\mu(\cdot|x) = \mu(\cdot|y)$

(ii) π sufficient for π' if $\exists b \in \mathcal{B}(K, K')$ s.t. $\pi b = \pi'$

(i) $\Longrightarrow$ (ii): $G^A(\pi, \mu) = \mathbb{E}[U^A(\mu(\cdot|x))] \geq \mathbb{E}[U^A(\mu(\cdot|y))] = G^A(\pi', \mu)$ and use Blackwell's thm

Focus on (i) $\Longleftarrow$ (ii)

MPS of Posterior Beliefs

Proposition

π is sufficient for π' iff $\mu(\theta|x)$ MPS of $\mu(\theta|y)$ $\forall \mu$, where $x \sim \pi(\cdot|\theta)$, $y \sim \pi'(\cdot|\theta)$

Proof

Let $\mu \gg 0$ wlog, $|\Theta| = N$, $|S(\pi)| = |X| = K$, $|S(\pi')| = |Y| = K'$, $\beta, b \in \mathcal{B}(K, K')$

$\mu_{1\times N}$, $\pi_{N\times K}$, $\pi'_{N\times K'}$

(i) $\mu(\cdot|x)$ MPS $\mu(\cdot|y)$ if $\exists \beta \in \mathcal{B}(K, K')$ s.t. $\sum_x \beta_{xy}\mu(\cdot|x) = \mu(\cdot|y)$

(ii) π sufficient for π' if $\exists b \in \mathcal{B}(K, K')$ s.t. $\pi b = \pi'$

Prelim: $\text{diag}(\mu)\pi$ is $N \times K$, $(\text{diag}(\mu)\pi)_{\theta,x} = \mu(\theta)\pi(x|\theta)$

$\mu\pi = \sum_\theta \mu(\theta)\pi(x|\theta)$, $\text{diag}(\mu\pi)_{K\times K}$

Posteriors: $\mu(\theta|x)_{\theta,x} = \text{diag}(\mu)\pi(\text{diag}(\mu\pi))^{-1}$ is $N \times K$

From b, π, μ construct β s.t. $\text{diag}(\mu)\pi(\text{diag}(\mu\pi))^{-1}\beta = \text{diag}(\mu)\pi b(\text{diag}(\mu\pi b))^{-1}$

MPS of Posterior Beliefs

Proposition

π is sufficient for π' iff $\mu(\theta|x)$ MPS of $\mu(\theta|y)$ $\forall \mu$, where $x \sim \pi(\cdot|\theta)$, $y \sim \pi'(\cdot|\theta)$

Proof

$$\begin{aligned}
& \operatorname{diag}(\mu)\pi(\operatorname{diag}(\mu\pi))^{-1}\beta = \operatorname{diag}(\mu)\pi b(\operatorname{diag}(\mu\pi b))^{-1} \\
\iff \quad & \pi(\operatorname{diag}(\mu\pi))^{-1}\beta = \pi b(\operatorname{diag}(\mu\pi b))^{-1} \\
\iff \quad & 0 = \pi\left[(\operatorname{diag}(\mu\pi))^{-1}\beta - b(\operatorname{diag}(\mu\pi b))^{-1}\right] \\
\iff \quad & 0 = \pi(\operatorname{diag}(\mu\pi))^{-1}\left[\beta - \operatorname{diag}(\mu\pi)b(\operatorname{diag}(\mu\pi b))^{-1}\right]
\end{aligned}$$

$$\beta = \operatorname{diag}(\mu\pi)b(\operatorname{diag}(\mu\pi b))^{-1} = \left(\frac{b(y|x)\sum_\theta \mu(\theta)\pi(x|\theta)}{\sum_{x'} b(y|x')\sum_\theta \mu(\theta)\pi(x'|\theta)}\right)_{x,y} = \left(\frac{b(y|x)\pi(x)}{\sum_{x'} b(y|x')\pi(x')}\right)_{x,y} \in \mathcal{B}(K, K')$$

Infinite Signal Space

Results generalize to infinite signal spaces, but need to be careful

(Euclidean spaces, Borel σ-algebra, measurability, etc; see Le Cam '64, '96)

(X, π): experiment; $X \sim \pi(\cdot|\theta)$

Finite Θ: $\exists b : S(\pi) \to \Delta(S(\pi'))$ s.t. $\pi'(\tilde{S}|\theta) = \int b(s) d\pi(s|\theta)$

Straightforward general definition:

Definition

Experiment (X, π) is sufficient for (Y, π') if $\exists$ r.v. $Z \sim F$ independent of θ and $h(X, Z)$ s.t. $h(X, Z) \sim \pi'(\cdot|\theta)$, $\forall \theta \in \Theta$

$h(X, Z)$ is a garbling of X: adds noise to $X \longrightarrow$ conveys less info about θ

e.g. $Y \overset{d}{=} h(X, Z) = X + Z$

Examples

Experiments satisfying sufficiency

$\Theta = S = \{0, 1\}$; $X \sim \mathbb{P}(s = \theta|\theta) = p$ and $Y \sim \mathbb{P}(s' = \theta|\theta) = p' \leq p$

$X \sim N(\theta, \sigma)$ and $Y \sim N(\theta, \sigma')$, $\sigma' \geq \sigma$ and known

$X \sim U(\theta - 1, \theta + 1)$ and $X \sim U(\theta - k, \theta + k)$, $k \in \mathbb{N}$

Experiments that do not satisfy sufficiency

$\Theta = \mathbb{R}$; $X = \theta + \varepsilon$, $Y = \theta + \delta$, ε normally distrib and $\delta \perp\!\!\!\perp \varepsilon$, δ **not** normally distrib

$\mathbb{V}(\varepsilon), \mathbb{V}(\delta) > 0$; not ranked no matter how much noisier one is relative to the other

$\Theta \subseteq \mathbb{R}^2$, $X \sim N(\theta_1, \theta_2)$ and $Y \sim N(\theta_1, \theta_2')$, $\sigma' \geq \sigma$ unknown

$X \sim U(\theta - 1, \theta + 1)$ and $X \sim U(\theta - k, \theta + k)$, $k \notin \mathbb{N}$

Blackwell ordering of information is very incomplete

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
   - Payoff Richness
   - Posterior Distribution
   - Infinite Signal Space
   - Examples

2. Lehmann Order
   - Monotonote Likelihood Ratio Property
   - Lehmann Order
   - Equivalence Result

3. Blackwell in Large Samples
   - Setup
   - Large Sample Order
   - Rényi t-Divergence
   - Equivalence Result
   - Dominance Ratio

Monotonote Likelihood Ratio Property

Setup

State: $\Theta \subset \mathbb{R}$; Signal space: $S \subseteq \mathbb{R}$; Prior $\mu \in \Delta(\Theta)$

Experiment: family of cdfs over signals, conditional on state

$P : S \times \Theta \to [0, 1]$; $P(s|\theta)$; admit pdf (or pmf) $p(s|\theta)$

Posterior cdf given s: $P(\cdot|s; \mu)$

Definition

P has MLRP if $\forall s_H > s_L$ and $\theta_H > \theta_L$, $p(s_L|\theta_L)p(s_H|\theta_H) \geq p(s_L|\theta_H)p(s_L|\theta_H)$

If $f > 0$, $\dfrac{p(s_H|\theta_H)}{p(s_H|\theta_L)} \geq \dfrac{p(s_L|\theta_H)}{p(s_L|\theta_L)}$

Higher signals more likely in higher states in likelihood-ratio sense

With two states (or signals), can relabel signals (resp. states) so that MLRP is satisfied

Monotonote Likelihood Ratio Property

Theorem (Milgrom '81 Bell)

P has MLRP if and only if it generates FOSD-ranked posteriors: $P(\theta|s_H;\mu) \leq P(\theta|s_L;\mu)$, $\forall\theta, s_H \geq s_L$ for any $\mu \in \Delta(\Theta)$.

Proof

For simplicity, $p > 0$.

$\Longrightarrow$: $\dfrac{p(\theta_H|s;\mu)}{p(\theta_L|s;\mu)} = \dfrac{\mu(\theta_H)}{\mu(\theta_L)}\dfrac{p(s|\theta_H)}{p(s|\theta_L)}$ inc. in s; LR dominance implies FOSD

$\Longleftarrow$: Take uniform prior. $\dfrac{p(\theta_H|s;\mu)}{p(\theta_L|s;\mu)} = \dfrac{p(s|\theta_H)}{p(s|\theta_L)}$

Lehmann Order

References: Lehmann ('88 AnnStats); Persico ('00 Ecta)

Setup

State: $\Theta \subset \mathbb{R}$; Signal space: $S \subseteq \mathbb{R}$; Prior $\mu \in \Delta(\Theta)$

Experiment: family of cdfs over signals, conditional on state

$P : S \times \Theta \to [0, 1]$; write $P_\theta(s) \equiv P(s|\theta)$; admit pdf $p_\theta(s) \equiv p(s|\theta)$

Posterior cdf given s: $P(\cdot|s; \mu)$

Action space: $A \subseteq \mathbb{R}$; $u : A \times \Theta \to \mathbb{R}$

Definition

P is more accurate than Q if $P_\theta^{-1} \circ Q_\theta$ is nondecreasing in θ

Lehmann Order

Definition

P is more accurate than Q if $P_\theta^{-1} \circ Q_\theta$ is nondecreasing in θ

$\theta_H > \theta_L \iff (P_{\theta_H}^{-1} \circ Q_{\theta_H})(s) \geq (P_{\theta_L}^{-1} \circ Q_{\theta_L})(s) \iff (Q_{\theta_H} \circ Q_{\theta_L}^{-1})(\tau) \geq (P_{\theta_H} \circ P_{\theta_L}^{-1})(\tau)$

(replace $s = Q_{\theta_L}^{-1}(\tau)$ and apply P_{θ_H} to both sides)

Interpretation: $\Theta = \{\theta_L, \theta_H\}$; $p_\theta, q_\theta > 0$

Suppose $S = [0, 1]$ and $P_{\theta_L}, Q_{\theta_L}$ are uniform

$(Q_{\theta_H} \circ Q_{\theta_L}^{-1})(\tau) = \int_0^q q_{\theta_H}(t)dt \geq \int_0^q p_{\theta_H}(t)dt = (P_{\theta_H} \circ P_{\theta_L}^{-1})(\tau)$:

If state is low, equally likely to see high signals under P or Q (both uniform)

If state is high, more likely to see high signals under P than Q (FOSD)

i.e. higher signals are more informative of higher state under P than under Q

Lehmann Order

Definition

P is more accurate than Q if $P_\theta^{-1} \circ Q_\theta$ is nondecreasing in θ

$\theta_H > \theta_L \iff (P_{\theta_H}^{-1} \circ Q_{\theta_H})(s) \geq (P_{\theta_L}^{-1} \circ Q_{\theta_L})(s) \iff (Q_{\theta_H} \circ Q_{\theta_L}^{-1})(\tau) \geq (P_{\theta_H} \circ P_{\theta_L}^{-1})(\tau)$

(replace $s = Q_{\theta_L}^{-1}(\tau)$ and apply P_{θ_H} to both sides)

Keep same intuition for arbitrary P_{θ_L}

Higher signals are more informative of higher state under P than under Q

(1) P_{θ_H} relative to dP_{θ_L} (instead of relative to Lebesgue): Radon-Nikodym derivative

$$P_{\theta_H}(s) = \int_{-\infty}^{s} \frac{dP_{\theta_H}}{dP_{\theta_L}}(t) dP_{\theta_L}(t)$$

$\frac{dP_{\theta_H}}{dP_{\theta_L}}(t)$: "density of P_{θ_H} relative to dP_{θ_L}"

(2) Ensure comparability by 'normalizing' using inverse cdf (change of variable)

$$(P_{\theta_H} \circ P_{\theta_L}^{-1})(\tau) = \int_0^q 1 d(P_{\theta_H} \circ P_{\theta_L}^{-1})(t) = \int_0^q \left(\frac{dP_{\theta_H}}{dP_{\theta_L}} \circ P_{\theta_L}^{-1} \right)(t) dt$$

Lehmann Order

Definition

P is more accurate than Q if $P_\theta^{-1} \circ Q_\theta$ is nondecreasing in θ

$\theta_H > \theta_L \iff (P_{\theta_H}^{-1} \circ Q_{\theta_H})(s) \geq (P_{\theta_L}^{-1} \circ Q_{\theta_L})(s) \iff (Q_{\theta_H} \circ Q_{\theta_L}^{-1})(\tau) \geq (P_{\theta_H} \circ P_{\theta_L}^{-1})(\tau)$

(replace $s = Q_{\theta_L}^{-1}(\tau)$ and apply P_{θ_H} to both sides)

Graphically: $(P_{\theta_H} \circ P_{\theta_L}^{-1}) : [0,1] \to [0,1]$

If P sat. MLRP, $(P_{\theta_H} \circ P_{\theta_L}^{-1})$ increasing and convex

Lehmann Order

Definition

$u : A \times \Theta \to \mathbb{R}$ satisfies the weak single-crossing property (wSCP) if $\forall a' > a, \theta' > \theta$,

$u(a', \theta) - u(a, \theta) \geq 0 \implies u(a', \theta') - u(a, \theta') \geq 0$

If higher action preferred to lower in low state then also the case in high state

"Weak" moniker: std single-crossing prop has $\geq$ (>) instead of just $\geq$

$$G^A(P, \mu) := \mathbb{E}_{\theta \sim \mu}\left[\mathbb{E}_{S \sim P(\cdot|\theta)}\left[\max_{a \in A} \mathbb{E}_{\theta \sim \mu|S}[u(a, \theta)|s]\right]\right]$$

Theorem (Lehmann, '88; Persico, '00)

Let P and Q be experiments satisfying MLRP. P is more accurate than Q if and only if $G^A(P, \mu) \geq G^A(Q, \mu)$ for all priors μ and all utility functions $u : A \times \Theta \to \mathbb{R}$ satisfying wSCP

Sufficiency implies accuracy but the converse is not true in general

if state is binary, Sufficiency/Blackwell order = Accuracy/Lehmann order

Applications

Persico ('00 Ecta): common value auction: bidders want to acquire + info in 1PA than 2PA

Idea: 1PA more risk-sensitive

Jensen ('18 REStud): distributional effects of income/productivity/trade shocks in workhorse macro models

Mekonnen & Vizcaíno ('20 WP): how much info should financial investment adviser give; when should firms make info acq public

Info in general: selling data; disclosure of platform by politicians, macro info by central banks, asset holdings by CEOs; R&D; industrial espionage; etc

Overview

1. Blackwell Order
   - Value of Information and Sufficiency
   - Payoff Richness
   - Posterior Distribution
   - Infinite Signal Space
   - Examples

2. Lehmann Order
   - Monotonote Likelihood Ratio Property
   - Lehmann Order
   - Equivalence Result

3. Blackwell in Large Samples
   - Setup
   - Large Sample Order
   - Rényi t-Divergence
   - Equivalence Result
   - Dominance Ratio

Blackwell in Large Samples

Motivation

2 experiments: cheap and expensive; both informative

Keep repeating cheap, eventually learn truth

Will *n*-repeats of cheap be more informative than expensive? How many?

Applications: How many antigen tests = pcr test? Which/how much data to acquire?

Setting

Mu, Pomatto, Strack, and Tamuz ('21 Ecta)

Binary state: $\Theta = \{0, 1\}$; Signal space: $S \subseteq \mathbb{R}^n$; Prior $\mu \in \Delta(\Theta)$

Experiment: family of cdfs over signals, conditional on state

$P : S \times \Theta \to [0, 1]$; $P_\theta(s) \equiv P(s|\theta)$

Joint experiment: $P \otimes Q$

n iid repetitions of P: joint experiment $P^{\otimes n}$

Restrictions on experiments:

P_0, P_1 mutually absolutely continuous ('no signal perfectly reveals the state')

dP_1/d_0 bounded if bounded above and bounded away from 0

(P, Q) generic pair if essential max and min of $\log dP_1/dP_0$ and $\log dQ_1/dQ_0$ are different

ess max$f$:= $\inf\{a \in \mathbb{R} \mid \mu(\{f(x) \geq a\}) = 0\}$

Note: (P, P) is not a generic pair

Large Sample Order

Definition

P dominates Q in large samples if $\exists n_0 \in \mathbb{N}$ such that $P^{\otimes n}$ Blackwell dominates ($\equiv$ is sufficient for) $Q^{\otimes n}$ for all $n \geq n_0$

Blackwell ('51, Thm 12): P, P' suff for Q, Q', then $P \otimes P'$ suff for $Q \otimes Q'$ (finite Θ)

$\Longrightarrow$ Large sample order extends Blackwell order

WT characterize large sample order in terms of P and Q

Rényi t-Divergence

Definition

For $\mu, \nu \in \Delta(\Theta)$, $t > 0$,

$$R_t(\mu||\nu) := \begin{cases} \frac{1}{t-1} \log \int_\Omega \left(\frac{d\mu}{d\nu}(\omega)\right)^{t-1} d\mu(\omega) & , \text{ if } t \neq 1 \\ \int_\Omega \log\left(\frac{d\mu}{d\nu}(\omega)\right) d\mu(\omega) & , \text{ if } t = 1 \end{cases}$$

$R_1(\mu||\nu)$: KL divergence; R_t increases in t and is continuous whenever finite

$t \to \infty$, $R_\infty(\mu||\nu)$ ess max of $\log(d\mu/d\nu)$

Definition

For experiment P, $t > 0$, $R_P^\theta(t) := R_t(P_\theta||P_{1-\theta})$

Properties of $R_P^\theta(t)$

Higher $R_P^\theta(2) \implies$ experiment in expectation produces more strongly evidence in favor of θ when it is the true state

Rényi t-Divergence

Definition

For $\mu, \nu \in \Delta(\Theta)$, $t > 0$,

$$R_t(\mu||\nu) := \begin{cases} \frac{1}{t-1} \log \int_\Omega \left(\frac{d\mu}{d\nu}(\omega)\right)^{t-1} d\mu(\omega) & , \text{ if } t \neq 1 \\ \int_\Omega \log\left(\frac{d\mu}{d\nu}(\omega)\right) d\mu(\omega) & , \text{ if } t = 1 \end{cases}$$

Definition

For experiment P, $t > 0$, $R_P^\theta(t) := R_t(P_\theta||P_{1-\theta})$

Properties of $R_P^\theta(t)$

$R_P^1(t) = \frac{t}{1-t} R_P^0(1-t)$

Additivity: $P = B \otimes C \implies R_P^\theta = R_B^\theta + R_C^\theta$ i.e. $R_{P^{\otimes n}}^\theta = n R_P^\theta$

Rényi Order

Definition

P dominates Q in the Rényi order if $\forall \theta \in \{0, 1\}$, $t > 0$, $R_P^\theta(t) > R_Q^\theta(t)$

Additivity implies P dominates Q in the Rényi order if and only if $P^{\otimes n}$ dominates $Q^{\otimes n}$ in the Rényi order

Equivalence Result

Theorem

Let P, Q be a generic pair of bounded experiments.

P dominates Q in large samples if and only if P dominates Q in the Rényi order

Genericity condition and boundedness of Q necessary; boundedness of P is not

Note: Rényi order extends strict Blackwell order

Proof

Proof to $\Longrightarrow$:

P induces conditional posterior distributions $P_{\mu,\theta}$

Write $P_\mu = 1/2P_{\mu,0} + 1/2P_{\mu,1}$

Blackwell order: $P \geq_B Q$ if and only if

$$\int_0^1 v(p)dP_\mu(p) \geq \int_0^1 v(p)dQ_\mu(p)$$

for every convex function $v : (0,1) \to \mathbb{R}$

Using: P_μ is MPS of Q_μ

binary state implies beliefs in $(0,1)$ ($p = \mu(\theta = 1)$)

inequality strict (>) if v strictly convex and $P_\mu \neq Q_\mu \Longleftrightarrow$ strict Blackwell order ($>_B$)

Idea: construct $v(p)$ convex s.t. monotone in Rényi t-divergence

Equivalence Result

Proof

(i) $dP_\mu := 1/2(dP_{\mu,0} + dP_{\mu,1})$

(ii) $p = \frac{dP_{\mu,1}(p)}{dP_{\mu,1}+dP_{\mu,0}} \iff dP_{\mu,1}(p) = \frac{p}{1-p} dP_{\mu,0}(p) \implies dP_\mu = \frac{1}{2p} dP_{\mu,1}(p)$

Posterior $p = \frac{dP_1(s)}{dP_1(s)+dP_0(s)} = \frac{dP_1(s)/dP_0(s)}{dP_1(s)/dP_0(s)+1}$

for $t > 1$, $v_1(p) := 2p^t(1-p)^{1-t}$; strictly convex

$$\int_0^1 v_1(p) dP_\mu(p) = \int_0^1 2p^t(1-p)^{1-t} \frac{1}{2p} dP_{\mu,1}(p) = \int_0^1 \left(\frac{1-p}{p}\right)^{1-t} dP_{\mu,1}(p)$$

$$= \int_S \left(\frac{dP_1(s)}{dP_0(s)}\right)^{1-t} dP_1(s) = \exp^{(t-1)R_P^1(t)}$$

last line: change of variables

$P >_B Q \implies R_P^1(t) > R_Q^1(t)$

Equivalence Result

Proof

for $t \in (0,1)$, define $v_2(p) := -v_1(p)$, strictly convex, and

$$\int_0^1 v_2(p) dP_\mu(p) = -\exp^{(t-1)R_P^1(t)}$$

$$P >_B Q \Longrightarrow R_P^1(t) > R_Q^1(t)$$

for $t = 1$, define $v_3(p) := 2p\log(p/(1-p))$, strictly convex

$$\int_0^1 v_3(p) dP_\mu(p) = \int_0^1 \log\left(\frac{p}{1-p}\right) dP_{\mu,1}(p) = \int_S \log\left(\frac{dP_1(s)}{dP_0(s)}\right) dP_1(s) = R_P^1(1)$$

$$P >_B Q \Longrightarrow R_P^1(1) > R_Q^1(1)$$

Symmetric argument gives $R_P^0(t) > R_Q^0(t)$

Equivalence Result

Proof

$\Longrightarrow$ Large sample order requires $nR_P^{\theta}(t) > nR_Q^{\theta}(t) \iff R_P^{\theta}(t) > R_Q^{\theta}(t)\ \forall t > 0, \theta \in \{0, 1\}$

Converse: more complicated...

(i) Blackwell order equivalent to FOSD of log LR of experiments + carefully crafted independent rv

(ii) Use large deviations to relate distribution of log LR of experiments

Dominance Ratio

Definition

For P and Q nontrivial experiments, let

$$P/Q := \sup\left\{m/n \in \mathbb{Q}_{++} \mid P^{\otimes n} \text{ Blackwell dominates } Q^{\otimes m}\right\}$$

$$= \inf\left\{n/m \in \mathbb{Q}_{++} \mid P^{\otimes n} \text{ Blackwell dominates } Q^{\otimes m}\right\}$$

$$= \sup\left\{a > \mathbf{0} \mid P^{\otimes n} \text{ Blackwell dominates } Q^{\otimes \lceil an \rceil} \text{ for all large enough } n\right\}$$

Interpretation: "P is (asymptotically) P/Q times as informative as Q"

Theorem

Let P, Q be nontrivial bounded experiments. Then $P/Q > \mathbf{0}$ and given by

$$P/Q = \inf_{\theta \in \{0,1\},\, t > 0} \frac{R_P^\theta(t)}{R_Q^\theta(t)}$$

Immediate from equivalence result

How Many Samples More?

What this is **not** saying: if $a \leq P/Q$, then $P^{\otimes n} \geq_B Q^{\otimes \lceil an \rceil}$ $\forall n$

Only holds true for *large n*

Would be interesting to get smallest $a > 0$ such that $\forall m$, $P^{\otimes \lceil am \rceil} \geq_B Q^{\otimes m}$

Useful for applications: pricing data, allocation of tests

Note that it is necessary and sufficient: $a \in \mathbb{N}$ such that $P^{\otimes a} \geq_B Q$

$P^{\otimes a} \geq_B Q \implies P^{\otimes a} \otimes P^{\otimes a} \geq_B Q \otimes Q$

Lemma 2 in paper provides (implicitly defined) bound