16. An Introduction to Information

Duarte Gonçalves

University College London

MRes Microconomics

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 today: **Modelling Info**.

Next year: **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.
- 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. Modelling Information

2. Value of Information and Sufficiency
   - Setup
   - Value of Information
   - Sufficiency

3. Equivalence Result
   - Proof by Crémer ('82 JET)

4. Posterior Distribution

5. More
   - Infinite Signal Space
   - Examples
   - Recap

Overview

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 $b\pi = \pi'$.

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

Example: Clinical test + diligent vs sloppy nurse.

Quality control with better/worse technology.

Overview

1. Modelling Information
2. Value of Information and Sufficiency
3. Equivalence Result
   - Proof by Crémer ('82 JET)
4. Posterior Distribution
5. More

Equivalence Theorem

Re-write definitions

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

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

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

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 associative, $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$.

Equivalence Theorem

Theorem (Blackwell '51, '53)

The following are equivalent:

(1) π' is a garbling of π;

(2) π is more informative than π', i.e., $\forall A, \mu$, $G^A(\pi, \mu) \geq G^A(\pi', \mu)$;

(3) $\forall A$, $P(\pi) \supseteq P(\pi')$.

We'll focus on showing (1) $\iff$ (2).

Sufficient $\implies$ + informative is unsurprising

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

Sufficient $\impliedby$ + 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
$\exists b \in \mathcal{B} : b(y|x) = \mathbf{1}_{y=y^*(x)}$,
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) $\forall \mu \gg \mathbf{0}$, $\exists u_{a^y,\theta} \in \mathbb{R}^{K',N}$ 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. $u(a^y,\theta)$

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

Overview

Mean-Preserving Spreads

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 then said a mean-preserving contraction of P_X.

Proposition

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

P_X MPS $P_Y \implies P_Y \geq_{SOSD} P_X$.

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 also called convex order and written as $\geq_{CX}$.

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 $\mu \gg \mathbf{0}$ 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) $\implies$ (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) $\impliedby$ (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 $\mu \gg \mathbf{0}$ 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: $\mathrm{diag}(\mu)\pi$ is $N \times K$, $\quad (\mathrm{diag}(\mu)\pi)_{\theta,x} = \mu(\theta)\pi(x|\theta)$

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

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

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

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

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

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

$$= \left(\frac{b(y|x)\pi(x)}{\sum_{x'} b(y|x')\pi(x')}\right)_{x,y} \in \mathcal{B}(K, K')$$

Overview

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 $X \implies$ 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 $Y \sim N(\theta_1, \theta_2')$, $\theta_2' \geq \theta_2$ but 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.

Recap

Framework is the backbone of information economics (theory, macro, finance, IO, etc.). And you'll see it again in term 2.

Blackwell Order: $\pi \geq_B \pi' \iff \forall A, \mu, G^A(\pi, \mu) \geq G^A(\pi', \mu)$ $\iff \exists b : \pi' = b\pi \iff$ posterior beliefs induced by π are MPS of those induced by π'.

- Less informative = Garbling. Tight characterisation.
- More informative = more valuable
  - $\implies$ Costs to information should be monotone in the Blackwell order.
- Blackwell order very incomplete. Lots of ways to complete it; e.g., mutual information (the Cobb-Douglas of information costs).
- Recent interest in trying to go beyond Bayesian DM framework and incorporate belief updating biases.
- Providing genuinely private information? Private private information.
- Comparing many weak diagnostic tests vs expensive clinical test: Blackwell in large samples.
- Measuring effectiveness of information campaign?