The Law of Large Demand for Information

Giuseppe Moscarini and Lones Smith

Outline

- Theory of value *and* pricing of "information", defined as non-sequential sample size of i.i.d. signals.
- Motivation: the emergence of markets for large numbers of cheap units of information (Internet databases).
- Question: If two experiments are not Blackwell-comparable, is it still true that all bayesian Decision Makers rank unanimously the values of their n-replicas? And the marginal values of the n-th observation?
- Answer: Yes, and the ordering is *complete*, for n *large enough* ($n > N$). The minimum $N < \infty$ depends on the Decision Maker, but is always finite. Every experiment has a unique efficiency index ruling asymptotic value and marginal value.
- Implications: The Law of Demand holds for small prices, demand elasticity can be computed.

The (Standard) Model

A) The Decision Problem

- Finite actions a_j, $j = 1, 2..K$
  Finite states of Nature θ_i, $i = 1, 2..M$.
- Decision Maker DM$= (\vec{q}, u)$: Non-degenerate prior beliefs $\vec{q} = \{q_1, q_2, ..q_M\}$ and Payoffs $u(a_j, \theta_i)$.
  Action a_i^* is best in state i.

B) The Experiment

- Family of M probability measures $\mathcal{E} = \{P_\theta\}$ on $\{\Omega, \mathcal{S}\}$.
- Before choosing a_j DM draws from $\mathcal{E}^n$, i.e. observes non-sequentially n i.i.d. realizations $X^n = \{X_1, X_2, ..X_n\}$ of a random variable X on $\{\Omega, \mathcal{S}, P_\theta\}$ with distribution $F(.|\theta)$ and density $f(.|\theta)$.
- De Finetti's Theorem: exchangeability suffices to justify the i.i.d. hypothesis

The Value of n Observations

- Bayesian Updating:

$$q_i\left(X^n\right)=\frac{q_i\Pi_{t=1}^n f\left(X_t|\theta_i\right)}{\Sigma_r q_r\Pi_{t=1}^n f\left(X_t|\theta_r\right)}$$

- Bayesian decision:

$$a\left(X^n\right)=\arg\max_{a_j}\Sigma_i q_i\left(X^n\right)u\left(a_j,\theta_i\right)$$

- Value of $\mathcal{E}^n$: $V_{\mathcal{E}}\left(n\right)-V_{\mathcal{E}}\left(0\right)$, where:

$$V_{\mathcal{E}}\left(n\right)=\int_{X^n}\Sigma_i q_i u\left(a\left(X^n\right)|\theta_i\right)\left[\Sigma_r q_r\Pi_{t=1}^n f\left(X_t|\theta_r\right)\right]dX^n$$

- Full Information Value: $V^*-V\left(0\right)$, where $V^*=\Sigma_i q_i u\left(a_i^*,\theta_i\right)$. Independent of the experiment.

- Full Information Gap (**FIG**): $V^*-V_{\mathcal{E}}\left(n\right)\geq 0$.

Double Dichotomy

Two States, Two Actions

- $a \in \{A, B\}$ and $\theta \in \{L, H\}$. A is best in L.
- Take A iff $f(X^n|L)/f(X^n|H)$ large enough, or:

$$\ln \frac{f(X^n|L)}{f(X^n|H)} = \sum_{t=1}^{n} \ln \frac{f(X_t|L)}{f(X_t|H)} \equiv S_n^L > \xi(u, \vec{q}).$$

- Probabilities of error:

$$\alpha_n = \Pr\left(\frac{S_n^L}{n} \leq \frac{\xi}{n}|L\right) \text{ and } \beta_n = \Pr\left(\frac{S_n^H}{n} \leq \frac{\xi}{n}|H\right).$$

  Clearly $\xi/n \to 0$. But by SLLN for $\theta \neq \bar{\theta}$:

$$\frac{S_n^\theta}{n} \xrightarrow{a.s.} \mathbb{E}\left[\ln f(X|\theta) - \ln f\left(X|\bar{\theta}\right)|\theta\right] > 0.$$

  Each error is a *Large Deviation* in state θ of mean log-LR S_n^θ/n from SLLN limit. α_n and β_n vanish.

- FIG is a linear combination of Large Deviation chances with positive coefficients:

$$\begin{aligned} V^* - V(n) &= q_L[u(A, L) - u(B, L)]\alpha_n + q_H[u(B, H) - u(A, H)]\beta_n \\ &\equiv y_L(\vec{q}, u)\alpha_n + y_H(\vec{q}, u)\beta_n. \end{aligned}$$

Large Deviations in the SLLN

Cramér's Theorem

- Y_t i.i.d. and non-lattice, $\mathbb{E}[Y] > 0$, $\mathbb{V}[Y] = \sigma^2$, $\Pr(Y < 0) > 0$.

  Let $S_n = \sum_{t=1}^{n} Y_t$, $M(t) = \mathbb{E}[\exp\{tY\}]$.

- **Cramér Condition:** $\mathbb{E}[\exp\{\bar{t}|Y|\}] < \infty$, $\exists \bar{t} \neq 0$.

- **Theorem (Cramér 1938).** For every $c < \mathbb{E}[Y]$ let:

$$\rho_c = \inf_t \exp\{-tc\} M(t) = \inf_t \mathbb{E}[\exp\{t(Y-c)\}]$$

  Then:

  1. $\rho_c = M(\tau_c)$ for $\tau_c < 0$.
  2. For a given sequence $\{b_r\}$:

$$\Pr\left(\frac{S_n}{n} \leq c\right) = \frac{\rho_c^n}{\sigma\sqrt{2\pi n \tau_c^2}}\left(1 + \sum_{r=1}^{\infty} \frac{b_r}{n^r}\right)$$

Application to the logLR of the Experiment

- Let $Y_t^L = \ln f(X_t|L) - \ln f(X_t|H)$.

$$\text{Our error} \ : \ \Pr\left(\frac{S_n^L}{n} \leq \frac{\xi}{n}|L\right) = \alpha_n$$

$$\text{Cramér} \ : \ \Pr\left(\frac{S_n^L}{n} \leq 0|L\right) = \frac{\rho_0^n}{\sigma\sqrt{2\pi n \tau_0^2}}\left(1 + \sum_{r=1}^{\infty} \frac{b_r}{n^r}\right)$$

- Apply Cramér to $c = 0 = \lim \xi/n$ with τ_0 and $\rho_0 = \min$ m.g.f., and then correct for $\xi/n > 0$.

- **Lemma.** Assume the logLR of the experiment is non-lattice and satisfies Cramér's condition. Then

$$\frac{\Pr\left(S_n^\theta \leq \xi|\theta\right)}{\Pr\left(S_n^\theta \leq 0|\theta\right)} \to \exp\left\{-\tau_0 \xi\right\}.$$

- **Corollary.**

$$\begin{aligned}
\alpha_n &= \frac{\rho_0^n}{\sigma\sqrt{2\pi n \tau_0^2}}\left(1 + \sum_{r=1}^{\infty} \frac{b_r}{n^r}\right)(1 + o(1)) \exp\left\{-\tau_0 \xi\right\} \\
&\propto \frac{\rho_0^n}{\sqrt{n}}(1 + o(1))
\end{aligned}$$

The Hellinger Transform

1. KULLBACK-LEIBLER RELATIVE ENTROPY, mean logLR Y^θ and drift of S_n^θ in state θ: e.g. for $\theta = L$

$$\lambda^L = \mathbb{E}\left[Y^L|L\right] = \int_X \ln \frac{f(X|L)}{f(X|H)} f(X|L)\, dX \geq 0.$$

2. HELLINGER TRANSFORM: m.g.f. of Y^θ in state θ:

$$\begin{aligned} M_L(t) &= \mathbb{E}\left[\exp\left\{t \ln \frac{f(X|L)}{f(X|H)}\right\}|L\right] \\ &= \int_X f(X|H)^{-t} f(X|L)^{1+t}\, dX \\ &\equiv \mathcal{H}_L(t) = \mathcal{H}_H(-t-1) = \mathcal{H}(t) \end{aligned}$$

Properties: $\mathcal{H}(-1) = \mathcal{H}(0) = 1$, $\mathcal{H}'(-1) = -\lambda^H < 0$ and $\mathcal{H}'(0) = \lambda^L > 0$, $\mathcal{H}''(.) > 0$.

$\mathcal{E}$ sufficient for $\mathcal{F} \;\Rightarrow\; \mathcal{H}_{\mathcal{E}}(.) \geq \mathcal{H}_{\mathcal{F}}(.)$

$$\begin{aligned} \mathcal{H}_{\mathcal{E}\times\mathcal{F}}(t) &= \mathcal{H}_{\mathcal{E}}(t)\,\mathcal{H}_{\mathcal{F}}(t)\,;\; \mathcal{H}_{\mathcal{E}^n}(t) = \left[\mathcal{H}_{\mathcal{E}}(t)\right]^n \\ \mathcal{H}(s) &= \int_X f(X|H)^s f(X|L)^{1-s}\, dX \end{aligned}$$

3. MINIMUM HELLINGER TRANSFORM:

$$\begin{aligned} \inf_t \mathcal{H}(t) &= \mathcal{H}(\tau) \equiv \rho \text{ for } \tau \in (-1, 0) \\ \inf_t \mathcal{H}_L(t) &= \inf_t \mathcal{H}_H(-t-1) = \rho \\ \inf_t \mathcal{H}_{\mathcal{E}^n}(t) &= \mathcal{H}_{\mathcal{E}^n}(\tau) = \left[\mathcal{H}_{\mathcal{E}}(\tau)\right]^n = \rho^n. \end{aligned}$$

Value Ordering

- We have obtained, for ρ the index of the dichotomy:

$$V^* - V(n) \propto \frac{\rho^n}{\sqrt{n}} (1 + o(1))$$

- **Theorem.** For all DM $(\vec{q}, u)$, $\delta \in (0, \rho)$

$$\frac{V^* - V(n)}{(\rho - \delta)^n} \to \infty; \ \frac{V^* - V(n)}{(\rho + \delta)^n} \to 0.$$

- **Corollary.** Given two experiments $\mathcal{E}_1, \mathcal{E}_2$ with indices $\rho_1 < \rho_2$, for every DM $(\vec{q}, u)$ there exists $N(\vec{q}, u) < \infty$ such that DM prefers n observations of $\mathcal{E}_1$ over n of $\mathcal{E}_2$ for all $n > N(\vec{q}, u)$. Extends Chernoff theorem to general decision rules.

- **Remark.** If $\mathcal{E}_1$ is sufficient for $\mathcal{E}_2$ then $N(\cdot, \cdot) \equiv 0$. Else, $N(\vec{q}, u)$ finite but NOT uniformly bounded. But for any finite set of DMs a unique cutoff N suffices for unanimous ordering.

Marginal Value Ordering

- **Lemma.** For all DM and $n > N(\vec{q}, u)$ the marginal value is strictly decreasing:

$$V(n+1) - V(n) > V(n+2) - V(n+1).$$

- **Theorem.** For all DM $(\vec{q}, u)$, $\delta \in (0, \rho)$

$$\frac{V(n+1) - V(n)}{(\rho - \delta)^n} \to \infty; \quad \frac{V(n+1) - V(n)}{(\rho + \delta)^n} \to 0.$$

- **Corollary.** Given two experiments $\mathcal{E}_1, \mathcal{E}_2$ with indices $\rho_1 < \rho_2$, for every DM $(\vec{q}, u)$ there exists $N(\vec{q}, u) < \infty$ such that the marginal value of the $n-th$ observation is larger in experiment 2 for all $n > N(\vec{q}, u)$.

Example: Bernoulli

- Two experiments, no garbling:

$$\begin{array}{ll} \psi_r & 1-\psi_r \\ 1-\phi_r & \phi_r \end{array} = \begin{array}{ll} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{3}{4} \end{array} \text{ and } \begin{array}{ll} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{array}$$

- Hellinger transform for experiment $r = 1, 2$:

$$\mathcal{H}_r\left(t\right) = \psi_r^t \left(1-\phi_r\right)^{1-t} + \left(1-\psi_r\right)^t \phi_r^{1-t}$$

- Hellinger indices: $\rho_2 = 0.912 > 0.908 = \rho_1$. Experiment 1 is superior, because the extra spread is larger:

$$\frac{2}{3} - \frac{3}{5} = \frac{1}{15} > \frac{1}{4} - \frac{1}{5} = \frac{1}{20}$$

Example: Gaussian

- Gaussian. $X \sim N\left(\mu_\theta, \sigma_\theta^2\right)$. After rescaling observations, $X \sim N\left(0, \sigma_L^2\right)$ and $X \sim N\left(1, \sigma_H^2\right)$.

$$
\begin{aligned}
\mathcal{H}_r(t) &= \int_{-\infty}^{+\infty} \left[f_N(X|L)\right]^t \left[f_N(X|H)\right]^{1-t} dX \\
&\propto \exp\left\{\frac{-t(1-t)}{2\left(t\sigma_H^2 + (1-t)\sigma_L^2\right)}\right\} \frac{\sigma_L^{1-t}\sigma_H^t}{\sqrt{t\sigma_H^2 + (1-t)\sigma_L^2}}
\end{aligned}
$$

- If $\sigma_L = \sigma_H = \sigma_r$ for both experiments $r = 1, 2$ and only the means differ across states, $\rho = \exp\left\{-1/8\sigma^2\right\}/\sqrt{2\pi}$ and Blackwell ordering applies.

- If $\mathcal{E}_1 = \left\{\sigma_{L1}, \sigma_{H1}\right\} = \{1, 3\}$ and $\mathcal{E}_2 = \{2, 1\}$ then no garbling. But $\rho_1 = 0.73 < \rho_2 = 0.84$, so the first experiment *with larger variances* eventually dominates.

- The distance between variances is larger, state detection is easier.

The Law of Large Demand for Information

- Fix a DM $(u, \vec{q})$ and an experiment $\mathcal{E}$. Let DM purchase samples at unit price p. Obtain a demand curve $n(p)$ from:

$$\max_{n=0,1,2..} V(n) - np.$$

- **Proposition.** There exists $\bar{p} > 0$ such that:

  1. the unique maximizer $n(p)$ is weakly decreasing in $p \in (0, \bar{p})$, and $n(0) = \infty$.
  2. For any given $p \in (0, \bar{p})$ the demand curve rises in $\rho_{\mathcal{E}}$. Worse information is demanded in larger amounts.

- **Proposition. The limit semielasticity of demand.** There exists a smooth function $z(p)$ such that $\sup_{p \in (0,\bar{p})} |n(p) - z(p)| < 1$ and

$$\lim_{p \downarrow 0} \left[-\frac{dz(p)}{dp} p \right] = -\frac{1}{\ln \rho_{\mathcal{E}}}.$$

- Hence for small prices.

$$n(p) \simeq \frac{\ln p}{\ln \rho_{\mathcal{E}}}.$$

- **Entry Fee Resolves Both Inefficiencies.**
- Firm solves:

$$\max_{p,F} \; F + x(p)(p - c)$$
$$s.t.\; F + x(p)\,p \leq V(x(p)) - V(0)$$

- Solution: $p^{NL} = c$ and

$$F = V(x(c)) - V(0) - x(c)\,c = \pi^{NL}.$$

- Envelope:

$$\frac{d\pi^{NL}}{d\rho} = \frac{\partial V(x(c))}{\partial \rho} < 0.$$

No incentive to reduce the quality of information. The firm prices and produces efficiently, maximizes consumer surplus and extracts it with access fee F.