Consumer theory for cheap information

Gary Baker

October 3, 2020

Agenda

Question

Consider a constrainted decision-maker who has to make a decision under uncertainty. Before acting, she has access to multiple, costly sources of information about the state of the world.

She must now decide from which sources to buy information and how much information to buy from each.

How should she allocate her resources among these different information sources?

Question

For example:

- A researcher trying to determine the true prevalence of some health condition (say COVID-19):
  - State: true prevalence
  - Action: estimate of prevalence
  - Info sources: available tests for the condition
  - Amount of info: how many trial participants
  - Constraint: Grant budget

- Voter trying to decide on a party:
  - State: true optimal policy
  - Action: for which party to vote
  - Info sources: different newspapers
  - Amount of info: how many articles to read
  - Constraint: Limited time to read the news

Towards consumer theory for information

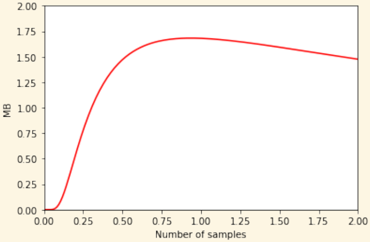
To answer these questions we need an (ordinal) theory of tradeoffs between different information sources.

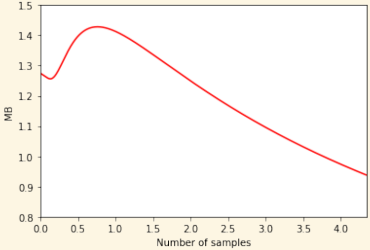
That is, the marginal rate of substitution.

Problem: in general, info value is nasty

For example: In a quasilinear setting, information values are typically nonconcave
[Radner and Stiglitz, 1984]

⇒ MB can be upward sloping at small samples





Exact simple formulae for info values are typically only possible in very specific cases.

“Solution”: large deviations

Find an approximation valid for large information purchases where mistake probabilities are well approximated by **large deviations theory**:

So when we have

- Cheap information,
- Large budgets, or
- Some combination

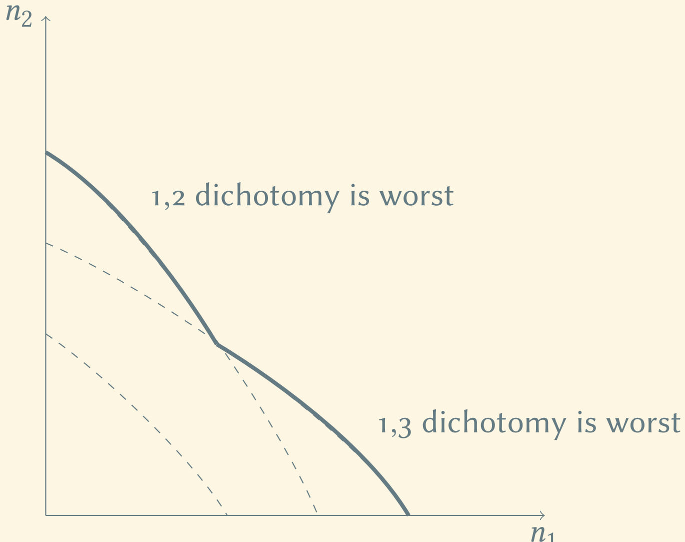
Preview of results

- Define a notion of precision for an information source, $\beta_{\omega i}(\theta\theta')$ – the Chernoff precision
  - Describes how well a source can distinguish state θ from θ'
  - Only a function of characteristics of the information source, and the proportion of each source in the bundle, ω
  - Roughly additive, for a bundle with fractions ω, total precision for a state pair is $\sum n_i \beta_{\omega i}(\theta\theta')$

- "Indifference curves" are approximately loci of bundles of equal **least** precision
  - When defined, MRS is the ratio of precisions

- Low cost optimal proportions satisfy a maxi-min rule: maximize the precision (per dollar) for the worst case state-pair.

Preview of results



Preview of results

- Chernoff precision does <u>not</u> depend on decision-maker characteristics.
- All decision-makers agree on the asymptotically optimal bundle!

Outline

Quick literature overview

Statistics:

- Chernoff [1952] asymptotic relative efficiency
  - How many samples from one statistical test are necessary to perform as well as n from another
  - Comparison of extremes: all one or all the other

- Contribution: I extend this to local (interior) comparisons useful for arbitrary finite-action/finite-state decision problems.

Quick literature overview

Economics:

- Moscarini and Smith [2002]
  - Apply methods similar to Chernoff to write an asymptotic approximation of value, and thus demand for information in the single source case

- Economic contribution: I extend this to an environment with multiple sources and explore implications for tradeoffs between them.
- Technical contribution: I use an alternative approximation approach (saddlepoint method) to get a tighter bound on the convergence rate.

Agenda

Model – Environment

- Finitely many possible underlying states, $\theta \in \Theta$
  - DM has prior $p \in \Delta\Theta$ (no degenerate beliefs)
- Finitely many possible actions $a \in A$.
  - For the presentation, assume $|A| = |\Theta|$
- DM has state-dependent utility function $u(a, \theta)$
  - Chooses action to maximize $\sum_\theta p_\theta u(a, \theta)$

- Prior to acting, the DM can purchase information about the state

Model – Information Sources

- Two information sources, $\mathscr{E}_1, \mathscr{E}_2$ (signals or experiments)
  - $\mathscr{E}_i \equiv \langle \mathrm{F}_i(r \mid \theta) \rangle$ ($r \in \mathrm{R}$ realizations)
  - No signal realization perfectly rules in or rules out any strict subset of states
  - Assume each has conditional density $f_i(r \mid \theta)$ (only for exposition)
- DM can purchase an arbitrary number of *conditionally independent* samples from each source at cost c_i per sample.
- DM has a large, but finite, budget to spend on information.
- After choosing a bundle of information (n_1, n_2), DM observes the vector of realizations, and updates via Bayes Rule

Model – Value with information

Expected value of acting after observing signal realizations

$$v(n_1, n_2) = \sum_{\theta} p_{\theta} \left[\int_{r \in \mathrm{R}} \underbrace{\max_{a} \left\{ \sum_{\theta} u(a, \theta) \mathbb{P}(\theta \mid r) \right\}}_{\text{Payoff to acting after updating}} f_{n_1, n_2}(r \mid \theta) \right]$$

Goal: Maximize subject to budget constraint

$$c_1 n_1 + c_2 n_2 \leq \mathrm{Y}$$

Agenda

Two states – setup

- States:
  - Null hypothesis – H_0
  - Alternative hypothesis – H_1
  - Prior that the alternative is true – p
- Actions:
  - Accept the null – $\mathscr{A}$
  - Reject the null – $\mathscr{R}$

Two states – setup

$$
\begin{aligned}
v(n_1, n_2) = \ & (1-p)\big(\alpha_{\mathrm{I}}(n_1,n_2)u(\mathscr{R},\mathrm{H}_0) + (1-\alpha_{\mathrm{I}}(n_1,n_2))u(\mathscr{A},\mathrm{H}_0)\big) \\
& + p\,\big(\alpha_{\mathrm{II}}(n_1,n_2)u(\mathscr{A},\mathrm{H}_1) + (1-\alpha_{\mathrm{II}}(n_1,n_2))u(\mathscr{R},\mathrm{H}_1)\big)
\end{aligned}
$$

- α_{I} – Type I error probability
- α_{II} – Type II error probability

Full-info gap

We get a bit of simplification by considering the full info-gap instead of value:

$$\mathrm{FIG}(n_1, n_2) \equiv \overbrace{(1-p)u(\mathscr{A}, \mathrm{H}_0) + pu(\mathscr{R}, \mathrm{H}_1)}^{\text{payoff from perfect info}} - v(n_1, n_2)$$

$$\begin{aligned} &= (1-p)\alpha_{\mathrm{I}}(n_1, n_2)\overbrace{(u(\mathscr{A}, \mathrm{H}_0) - u(\mathscr{R}, \mathrm{H}_0))}^{\text{loss from Type-I}} \\ &\quad + p\,\alpha_{\mathrm{II}}(n_1, n_2)\overbrace{(u(\mathscr{R}, \mathrm{H}_1) - u(\mathscr{A}, \mathrm{H}_1))}^{\text{loss from Type-II}} \end{aligned}$$

(Minimizing the FIG is equivalent to maximizing the value)

Roadmap

Goal: Find a nice ordinally-equivalent expression for value

Method:

1. Approximate error probability
2. Simplify with a monotone transformation of value

Error probabilities

Consider the one info source case [Moscarini and Smith, 2002]

$$\alpha_{\mathrm{I}}(n) = \mathbb{P}\left(\frac{p \prod_{i=1}^{n} f(r_i \,|\, \mathrm{H}_1)}{p \prod_{i=1}^{n} f(r_i \,|\, \mathrm{H}_1) + (1-p) \prod_{i=1}^{n} f(r_i \,|\, \mathrm{H}_0)} > \bar{p} \;\middle|\; \mathrm{H}_0 \right)$$

Error probabilities

Change to log-likelihood ratios:

$$\alpha_{\mathrm{I}}(n) = \mathbb{P}\left(\log\left(\frac{p}{1-p}\right) + \sum_{i=1}^{n} \log\left(\frac{f(r_i\,|\,\mathrm{H}_1)}{f(r_i\,|\,\mathrm{H}_0)}\right) > \log\left(\frac{\bar{p}}{1-\bar{p}}\right) \,\middle|\, \mathrm{H}_0\right)$$

$$\equiv \mathbb{P}\left(\sum_{i=1}^{n} s_i > \bar{l} - l \,\middle|\, \mathrm{H}_0\right)$$

$\mathbb{E}(s_i|\mathrm{H}_0) < 0$, so at large sample size, a mistake only occurs when the sample average LLR is far from its mean.

This is a **large deviation**.

Log-likelihood ratio distribution

Has a nice form for its Moment Generating Function (MGF):

$$
\begin{aligned}
M(t) &\equiv \mathbb{E}(e^{ts} \mid H_0) \\
&= \int e^{t \log\left(\frac{f(r \mid H_1)}{f(r \mid H_0)}\right)} f(r \mid H_0) dr \\
&= \int f(r \mid H_1)^t f(r \mid H_0)^{1-t} dr
\end{aligned}
$$

Large deviations – Chernoff Index

Large deviations probabilities often depend on the minimized moment generating function:

$$\begin{aligned}\rho &\equiv \min_t \mathrm{M}(t) \\ &= \min_t \int f(r \mid \mathrm{H}_1)^t f(r \mid \mathrm{H}_0)^{1-t} \mathrm{d}r\end{aligned}$$

Call ρ the Chernoff index of the info source

Properties:

- $\rho \in (0, 1)$
- Blackwell more informative $\Rightarrow$ lower Chernoff index
- n i.i.d. samples has index ρ^n

Large deviations – Chernoff Precision

The Chernoff index is pretty abstract: consider instead

$$\beta \equiv -\log(\rho)$$

Call β the Chernoff Precision.

Properties:

- $\beta > 0$
- Blackwell more informative $\Rightarrow$ higher precision
- n i.i.d. samples has precision $n\beta$

Chernoff precision – example

Gaussian noise: $r \sim \mathcal{N}(0, \sigma^2)$ in state H_0
$r \sim \mathcal{N}(\mu, \sigma^2)$ in state H_1

- Chernoff precision is $\beta = \frac{1}{8}\frac{\mu^2}{\sigma^2}$
  - Proportional to the signal-to-noise ratio
  - Proportional to the classical notion of precision ($1/\sigma^2$)

Approximating the error probability

Lemma (MS02, improved)

The probability of a mistake is falling exponentially in the number of samples according to the Chernoff index.

In particular, both error probabilities—and thus the FIG as well—are proportional to

$$\alpha(n) \propto \frac{\rho^n}{\sqrt{n}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

Proof sketch

Corollary (Chernoff ARE)

n_1 samples from source 1 asymptotically performs as well as $n_1\beta_1/\beta_2$ samples from source 2.

Composite experiments

- Consider a composite composed of n_1 from $\mathscr{E}_1$ and n_2 from $\mathscr{E}_2$
  - Define the composite factor $\omega = n_1/(n_1 + n_2)$

- The LLR distribution is the distribution of the sum of LLRs
  - $M_\omega(t) = M_1(t)^\omega M_2(t)^{1-\omega}$
  - So composite has MGF $M_\omega(t)^{n_1+n_2}$

- Define $\rho_\omega \equiv \min_t M_\omega(t)$
  $\tau_\omega \equiv \arg\min_t M_\omega(t)$

- $\rho_{\omega i} \equiv M_i(\tau_\omega)$

So we have

$$\begin{aligned}\rho_\omega &= \rho_{\omega 1}^{\omega} \rho_{\omega 2}^{1-\omega} \\ &\geq \rho_1^{\omega} \rho_2^{1-\omega}\end{aligned}$$

Composite experiments

$$\begin{aligned}\beta_\omega &= \omega\beta_{\omega 1} + (1-\omega)\beta_{\omega 2} \\ &\leq \omega\beta_1 + (1-\omega)\beta_2\end{aligned}$$

Composite experiments are worse than the sum of their parts.

Intuition

The minimizer, τ, is heuristically a measure of **slant**. Consider 2 news sources reporting about 2 candidates (R and L):

	Source 1 (R leaning)		Source 2 (L leaning)	
Truth	favors R	favors L	favors R	favors L
R actually better	0.99	0.01	0.02	0.98
L actually better	0.98	0.02	0.01	0.99

Precision of both is the same, but minimizers are far apart.

In this case, a most decision makers will prefer 2 samples from one or the other over 1 from each because 97% of the time, the two sources will send contradictory signals.

Approximating the error probability

Proposition

The probability of a mistake is falling exponentially in the number of samples from each experiment according to their respective component Chernoff indices for the given composite factor.

In particular the mistake probabilities—and thus the FIG as well—are proportional to

$$\alpha(n_1, n_2) \propto \frac{\rho_{\omega 1}^{n_1} \rho_{\omega 2}^{n_2}}{\sqrt{n_1 + n_2}} \left(1 + \mathcal{O}\left(\frac{1}{n_1 + n_2}\right)\right)$$

Roadmap

Goal: Find a nice ordinally-equivalent expression for value

Method:

1. Approximate error probability DONE
2. Simplify with a monotone transformation of value

Full-info gap

Plugging in our expression for error probabilities, we have that the FIG is falling exponentially in the Chernoff indices

$$\mathrm{FIG}(n_1, n_2) \propto \frac{\rho_{\omega 1}^{n_1} \rho_{\omega 2}^{n_2}}{\sqrt{n_1 + n_2}} \left(1 + \mathcal{O}\left(\frac{1}{n_1 + n_2}\right)\right)$$

Ordinal value

Take a transformation to get an ordinally-equivalent form for maximization:

$$-\log(\mathrm{FIG}(n_1, n_2)) = (n_1\beta_{\omega 1} + n_2\beta_{\omega 2})\left(1 + \mathcal{O}\left(\frac{\log(n_1 + n_2)}{n_1 + n_2}\right)\right)$$

Maximizing value is roughly equivalent to maximizing total precision!

Heuristically, "indifference curves" are close to iso-precision curves.

Approximately optimal bundles

In a budget-constrained environment it suffices to maximize precision per dollar:

$$\omega^*(c) = \arg\max_{\omega} \left\{ \frac{\omega\beta_{\omega 1} + (1-\omega)\beta_{\omega 2}}{\omega c_1 + (1-\omega)c_2} \right\}$$

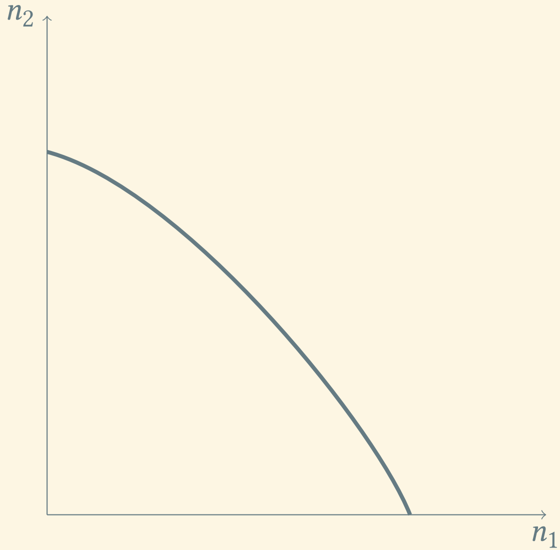
<u>Note:</u> Precisions don't depend on DM characteristics. Everyone (facing the same costs) agrees on the optimal at large samples!

But this is easy with only two states: composites are less precise than the sum of their parts.

⇒ Corners are always (near) optimal at low prices or large budgets.
(We could have just used Chernoff's original result.)

Graphical intuition

Iso-precision lines bow out



Implications for optimization

- Optimal bundles are eventually corners
- Best bundle has the highest Chernoff precision per dollar (β_i / c_i)

Summary so far

- Error probabilities fall exponentially fast with rate $\rho_{1\omega}^{n_1}\rho_{2\omega}^{n_2}$
- Constrained maximization of info value is asymptotically equivalent to constrained maximization of $n_1\beta_{1\omega} + n_2\beta_{2\omega}$
- Precision of composites are less than the sum of their parts so corners are always optimal.

Agenda

Many states

Ok, so two states wasn't too interesting.

What happens when we go to the general finite-state case?

Chernoff precision with many states

- With multiple states, we now have many log-likelihood ratio distributions:
  - e.g. With three states, we have 1 vs 2, 1 vs 3, and 2 vs 3 LLRs.
- So for $\mathscr{E}_i$ we can define a Chernoff index for each pair of states

$$\rho_i(\theta\theta') \equiv \min_t \int f(r \mid \theta)^t f(r \mid \theta')^{1-t} \mathrm{d}r$$

- And thus a precision for each pair

$$\beta_i(\theta\theta') = -\log \rho_i(\theta\theta')$$

Full-info gap

$$\mathrm{FIG}(n_1, n_2) = \sum_{\theta} p_\theta \sum_{\theta' \neq \theta} \overbrace{\alpha(n_1, n_2\,;\, \theta', \theta)}^{\text{mistake prob.}} \underbrace{(u(\theta, \theta) - u(\theta', \theta))}_{\text{gain from correct}}$$

Full-info gap

$$\mathrm{FIG}(n_1, n_2) =$$

$$\sum_{\theta} \sum_{\theta' \neq \theta} (p_\theta + p_{\theta'}) \Bigg(\underbrace{\frac{p_\theta}{p_\theta + p_{\theta'}} \alpha(n_1, n_2\,;\, \theta', \theta)(u(\theta, \theta) - u(\theta', \theta)) + \frac{p_{\theta'}}{p_\theta + p_{\theta'}} \alpha(n_1, n_2\,;\, \theta, \theta')(u(\theta', \theta') - u(\theta, \theta'))}_{\equiv \mathrm{FIG}_{\theta\theta'}(n_1, n_2)} \Bigg)$$

Worst case scenario

Intuition:

- Total FIG is a sum of pairwise FIGs, which are exponentially falling.
  - Only the biggest one matters!
  - i.e. only the most likely mistake matters

Lemma (MS02)

Let θ, θ' *be the dichotomy with the lowest precision, then the FIG is proportional to*

$$\mathrm{FIG}(n_1, n_2) \propto \mathrm{FIG}^*_{\theta\theta'}(n_1, n_2)(1 + \mathcal{O}(\bar{\rho}^n))$$

where $\mathrm{FIG}^*_{\theta\theta'}$ *is the when the state is known to be either* θ *or* θ' *and* $\bar{\rho} < 1$.

Proof sketch

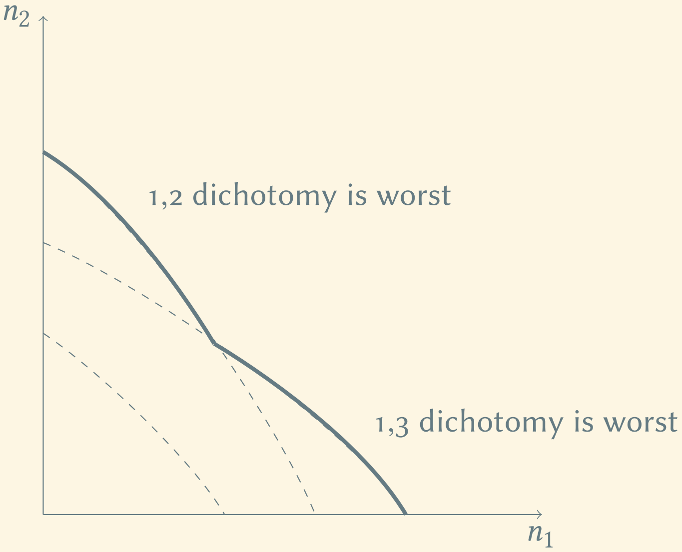
Ordinal value

Writing an ordinally equivalent form like before we have

$$-\log(\mathrm{FIG}(n_1, n_2)) \approx \min_{\{\theta\theta'\}}\{n_1\beta_{\omega 1}(\theta\theta') + n_2\beta_{\omega 2}(\theta\theta')\}$$

- So a composite experiment is worse than the sum of its parts for any single dichotomy
- But because only the worst case matters, experiments can complement each other by covering for each other's weaknesses.
- "Indifference curves" are now iso-*least*-precision curves.

Graphical intuition



Approximately optimal bundles

So for the general finite-state case, the optimal proportions satisfy a maxi-min rule:

Maximize the minimum precision per dollar

$$\omega^*(c) = \arg\max_{\omega} \left\{ \frac{\min_{\theta\theta'}\{\omega\beta_{\omega 1}(\theta\theta') + (1-\omega)\beta_{\omega 2}(\theta\theta')\}}{\omega c_1 + (1-\omega)c_2} \right\}$$

Agenda

Agenda

Defining marginal rate of substitution

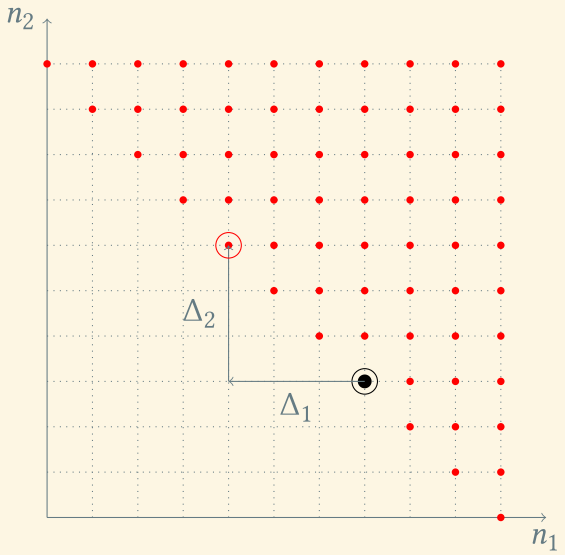
Samples are a *discrete* choice variable. Need to define a notion of *marginal*:
Define the *minimum compensating substitution* as

$$\Delta_2 \equiv \min\{\Delta : v(n_1 - \Delta_1, n_2 + \Delta) \geq v(n_1, n_2)\}$$

And the **discrete rate of substitution** as

$$\mathrm{DRS}_{\Delta_1}(n_1, n_2) = \frac{\Delta_2}{\Delta_1}$$

Defining marginal rate of substitution



Defining marginal rate of substitution

Define the **asymptotic marginal rate of substitution** as

$$\mathrm{AMRS}(\omega) \equiv \lim_{N\to\infty} \mathrm{DRS}_{\Delta_1(N)}(\omega N, (1-\omega)N)$$

where $\Delta_1(N) \to \infty$ as $N \to \infty$, but $\Delta_1(N) = o(N)$.

So we allow the size of the substitution to grow with sample size, just at a much smaller rate.

So *marginal* in this context is a substitution small relative to total sample size.

Marginal rate of substitution

Proposition

The asymptotic marginal rate of substitution is given the ratio of Chernoff precisions:

$$\mathrm{AMRS}(\omega) = \frac{\beta_{\omega 1}}{\beta_{\omega 2}}$$

Intuition.

For large samples, for a much smaller substitution, the change in ω is negligible. Then solve for Δ_1/Δ_2

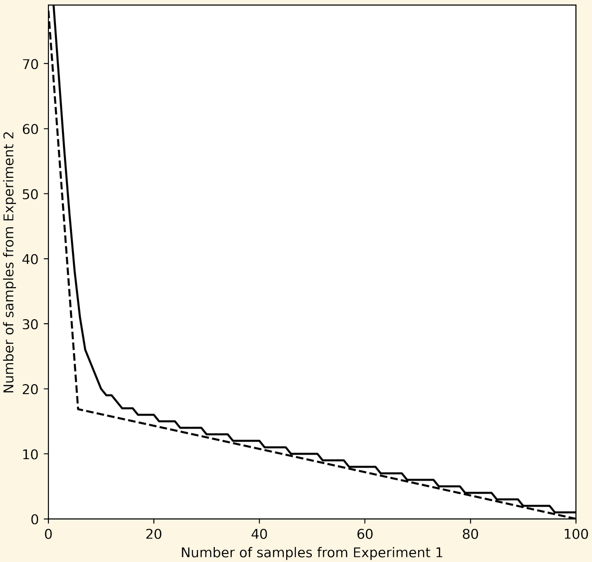
$$n_1\beta_{\omega 1} + n_2\beta_{\omega 2} = (n_1 - \Delta_1)\beta_{\omega 1} + (n_2 + \Delta_2)\beta_{\omega 2}$$

Marginal rate of substitution

Put another way, for a large enough initial sample, any substitution of Δ_1 samples from $\mathscr{E}_1$ for Δ_2 from $\mathscr{E}_2$ will lead to higher payoffs whenever

$$\frac{\Delta_2}{\Delta_1} \geq \frac{\beta_{1\omega}}{\beta_{2\omega}}$$

Indifference curves



Agenda

When do interior solutions occur?

Proposition (Interior solutions: 2 sources)

If there exist generic prices such that $\omega^*(c_1, c_2) \in (0, 1)$ *then two sources differ in their worst case dichotomy.*

⇒Sources can only be complements if they have differing weaknesses.

In practice the converse tends to hold as well.

Income elasticity

Component precisions depend only on the relative proportions, ω, of each source in the bundle.
$\Rightarrow$ Info values are approaching **homothetic**
(Indifference curves are just scalings of each other)

Proposition (Income elasticities)

All information sources are eventually normal goods, and thus all income elasticities are approaching 1.

Price elasticities

Optimal bundles lie near kinks or corners.
⇒ Small price changes don't change relative proportions. (Changes in demand from a price change is pure **income effect**. Hicksian substitution effects are zero at almost all prices.)

Proposition

Holding c_2 fixed, the demand elasticities (both own price for $\mathscr{E}_1$ and cross-price for $\mathscr{E}_2$) are approaching

$$\frac{-\omega^*(c_1, c_2)c_1}{\omega^*(c_1, c_2)c_1 + (1 - \omega^*(c_1, c_2))c_2}$$

except for finitely many values of c_1 where ω^* jumps.

Implications for competition

Fixed price monopolistic competition between two sellers of distinct information sources doesn't work at for large budgets.

At generic prices demand is inelastic, so at least one firm can improve profits by raising prices, unless at a corner.

Agenda

Is the approximation useful in practice?

- Asymptotically, error probabilities are very close to zero no matter what we do.

$$\alpha_1(n) \propto D \frac{\rho_1^n}{\sqrt{n}}$$

Is the approximation useful?

Proposition

Let $n_1^*(\mathrm{Y}, \mathrm{c}), n_2^*(\mathrm{Y}, \mathrm{c})$ be the feasible bundle under budget Y with composite factor as close as possible to the one that maximizes the minimum precision per-dollar under cost vector c, and let $n_1(\mathrm{Y}, \mathrm{c}), n_2(\mathrm{Y}, \mathrm{c})$ be a feasible sampling strategy with fixed (non-optimal) composite factor. Then we have

$$\frac{\mathrm{FIG}(n_1^*(\mathrm{Y}, \mathrm{c}), n_2^*(\mathrm{Y}, \mathrm{c}))}{\mathrm{FIG}(n_1(\mathrm{Y}, \mathrm{c}), n_2(\mathrm{Y}, \mathrm{c}))} \to 0$$

The optimal corner eventually performs <u>much</u> better!

Is the approximation useful in practice?

Put another way, the budget required to achieve a target performance is **much** smaller when following the maximin precision rule.

Required budgets are very sensitive to the sampling strategy as the target FIG gets small.

How good is the approximation?

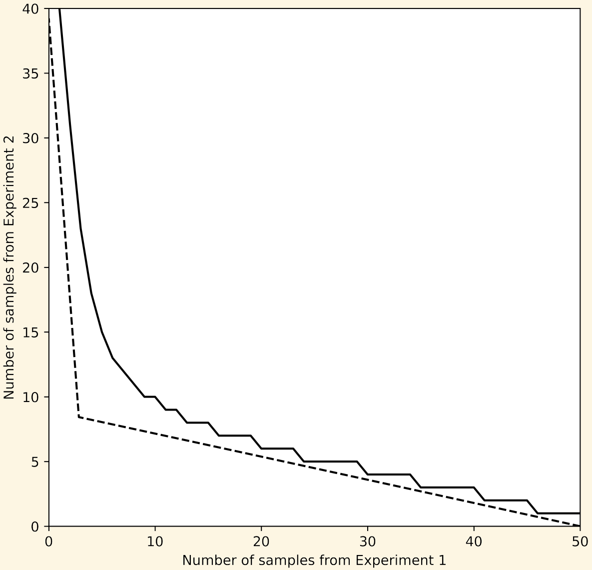


FIG is ~5% of Full-info value.

How good is the approximation?

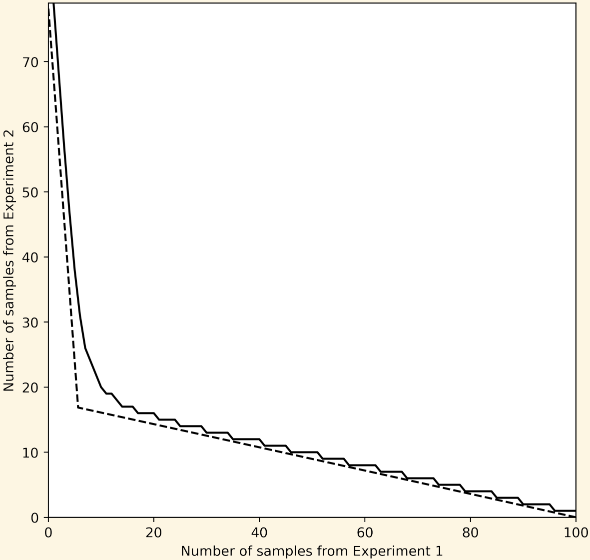


FIG is ~1% of Full-info value.

The approximation in practice

- So long as the number of possible states is small, the approximation works reasonably well.
- Gives fairly accurate predictions for corners vs interior solutions.

The approximation in practice

- ICs are smoothly rounded around kinks.
- With lots of states, the approximation performs relatively poorly.
- ICs are closer to the consumer theory stereotype (but still with corner solutions).
- For a continuous state decision-problem (e.g. estimation), might expect smooth ICs.

Agenda

Future work: continuous states

- The natural next step is to generalize to a continuous state environment.
- Would imply a criterion for optimal experiment design applicable in a general class of estimation problems.

Future work: continuous states

Problem: With continous states, there is no worst-case dichotomy.

For any state, θ, we have $\rho(\theta\theta') \to 1$ and $\beta(\theta\theta') \to 0$ as $\theta' \to \theta$.

Future work: continuous states

Heuristically, the state hardest to distinguish from θ is the one "adjacent" to it, $\theta + d\theta$
Roughly

$$\rho(\theta(\theta + d\theta)) = 1 + \frac{1}{2}\frac{d^2}{d\theta'^2}\rho(\theta\theta')\Big|_{\theta'=\theta}(d\theta)^2$$

Taking a negative log, we get

$$\beta(\theta(\theta + d\theta)) = -\frac{1}{2}\frac{d^2}{d\theta'^2}\rho(\theta\theta')\Big|_{\theta'=\theta}(d\theta)^2$$

So call

$$\hat{\beta}(\theta) \equiv -\frac{1}{2}\frac{d^2}{d\theta'^2}\rho(\theta\theta')$$

Future work: continuous states

It turns out (after a lot of algebra) that

$$\hat{\beta}(\theta) = \int \frac{d^2}{d\theta^2} \log(f(r|\theta))dr$$

This has another name: the Fisher information

Future work: continuous states

- Suggests that the generalization is a maximin Fisher info per dollar rule.
- Fisher informations are additive across info sources, so the asymptotic marginal rate of substitution is likely to be the ratio of Fisher informations for the worst case parameter value.

Thank you!

Proof sketch (Error Approximation)

Define a new distribution: the exponential tilting:

$$dG(s) \equiv \frac{e^{\tau s} dF(s \mid H_0)}{\rho}$$

Properties:

- Moment generating function is $M(t + \tau)/\rho$
  - Thus has mean zero (FOC)
- Variance $\varsigma^2 = M''(\tau)/\rho$

Proof sketch (Error Approximation)

[Bahadur and Rao, 1960]

$$\begin{aligned}\alpha_{\mathrm{I}}(n) &= \int \cdots \int_{\sum_k^{n_1} s_k > \bar{l} - l} \mathrm{dF}(s_1 \mid \mathrm{H}_0) \ldots \mathrm{dF}(s_k \mid \mathrm{H}_0) \\ &= \rho^n \int \cdots \int_{\sum_k^{n_1} s_k > \bar{l} - l} e^{-\tau \sum_k^{n_1} s_k} \mathrm{dG}(s_1) \ldots \mathrm{dG}(s_k) \\ &= \rho^n \int_{\xi_n}^{\infty} e^{-\tau \varsigma \sqrt{n}\, u} \mathrm{dH}_n(u)\end{aligned}$$

Where H_n is the distribution of $\sum s_i / \sqrt{\varsigma^2 n}$ under G. H_n converges to $\mathcal{N}(0, 1)$

Back to main

Proof sketch (many state approximation)

Part 1:

$\mathrm{FIG}^*_{\theta\theta'}(n_1, n_2)$ is the FIG after additionally observing a signal that perfectly reveals the state *unless* the state is either θ or θ' so

$$\mathrm{FIG}(n_1, n_2) \geq \mathrm{FIG}^*_{\theta\theta'}(n_1, n_2) \propto \frac{\rho(\theta\theta')^n}{\sqrt{n}}(1 + \mathcal{O}(n^{-1}))$$

Part 2:

Show that when the state is θ, the probability of a mistake is $\mathcal{O}(\rho(\theta h(\theta))^n)$ where $h(\theta)$ is the state hardest to distinguish from θ.

Part 3:

Squeeze

Back to main

Bahadur, R. R. and R. R. Rao (1960). On deviations of the sample mean. *Annals of Mathematical Statistics 31*(4), 1015–1027.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics 23*(4), 493–507.

Moscarini, G. and L. Smith (2002). The law of large demand for information. *Econometrica 70*(6), 2351–2366.

Pomatto, L., P. Strack, and O. Tamuz (2018). The cost of information.

Radner, R. and J. E. Stiglitz (1984). A nonconcavity in the value of information.