# Consumer theory for cheap information

Gary Baker
University of Wisconsin–Madison

16 October 2020
UW–Madison, Theory Seminar

# 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 decide not only

- <u>from which</u> sources to buy information, but also
- <u>how much</u> information to buy from each.

# Question

For example:

- 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

- A researcher trying to determine the effectiveness of some vaccine (say, for COVID-19):
  - State: true effectiveness
  - Action: whether to introduce the vaccine or not
  - Info sources: available tests for the condition
  - Amount of info: how many trial participants
  - Constraint: Grant budget

# Goal

We'd like to have a consumer theory for information.

- Tradeoffs between different sources
  - marginal rate of substitution
- Demand for information in constrained settings
  - Elasticities

# Potential Applications

- Media and rational inattention: how people allocate their resources (e.g. time) between different news/info sources
- Research design and optimal treatment allocation

# Problems with information as a good

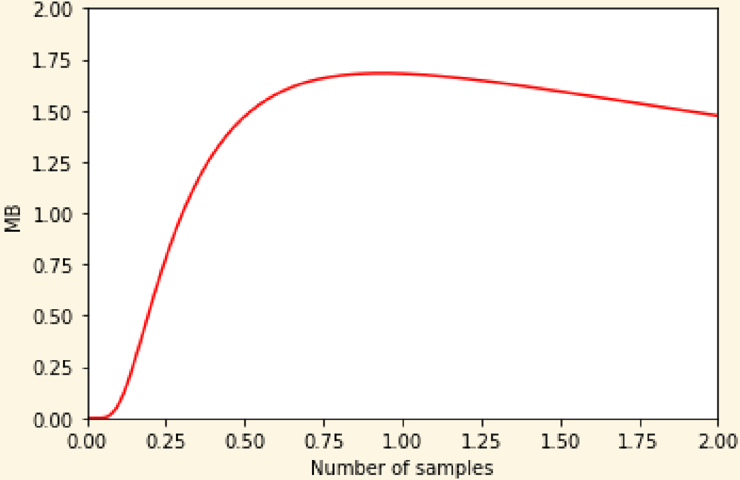
Going back to Blackwell [1951]:

- Information from different sources can't easily be compared
- In the broadest sense, information sources can only be ordered by garbling.

# Problems with information as a good

Another example: In a quasilinear setting, marginal values of information is typically upward sloping at small samples.

- First-order condition analysis doesn't easily work



# Problems with information as a good

In general, information value doesn't have a nice, closed-form expression.

# What I do

To answer these questions I develop an (approximate) ordinal theory of tradeoffs between information source.

That is, I will

- Find an approximate ordinal expression for information values, valid at large samples (when info is cheap)
- Characterize the marginal rate of substitution between information sources
- Explore implications for information demand in a budget constrained setting

# What I do

- This approximation will not depend on decision-maker characteristics (prior, utility function)
- Everyone facing the same costs will agree on the optimal bundle

# Method: large deviations

Information is valuable insofar as it prevents you from taking a suboptimal action.

- We can characterize info values by the probability that it misleads the decision maker.
- That is, when the decision maker takes the wrong action after seeing the info.

With a lot of information (large samples), the probability of being misled is very small (a tail event).

Approximating this is the realm of large deviations theory.

# Method: large deviations

So my approximations will be valid when the DM purchases a lot of info.

So a scenario with

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

# Agenda

# Literature

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:

- extend to local (interior) comparisons (MRS), and
- to arbitrary finite-action/finite-state decision problems.

# Literature

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: extend this to environment with multiple sources and explore implications for tradeoffs between them.
  - Technical contribution: Proof approach for the approximation gives tighter bounds on convergence rate and implies a full, asymptotic expansion.

# Other related literature

Value of and comparisons between information sources:
Börgers et al. [2013], Athey and Levin [2018]

Rational inattention:
Sims [2003], and many, many others

Optimal experiment design / treatment assignment:
Elfving [1952], Chernoff [1953], Dette et al. [2007]
(another huge literature)

# 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 each state has a unique optimal action.
- DM has state-dependent utility function $u(a, \theta)$
  - Chooses action, $a$, 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, $\mathscr{E}_1, \mathscr{E}_2$
  (AKA: tests, signals, or experiments)
  - $\mathscr{E}_i \equiv \langle \mathrm{F}_i(r \mid \theta) \rangle$ ($r \in \mathrm{R}$ realizations)
  - Assume: No signal realization perfectly rules in or rules out any strict subset of states
  - Assume for exposition: each has conditional density $f_i(r \mid \theta)$

# Model – Information sources

- DM can purchase an arbitrary number of *conditionally independent* samples, $n_i$, 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 – $\mathrm{H}_0$
  - Alternative hypothesis – $\mathrm{H}_1$
  - Prior that the alternative is true – $p$
- Actions:
  - Accept the null – $\mathscr{A}$
  - Reject the null – $\mathscr{R}$

# Two states – setup

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

- $\alpha_{\mathrm{I}}$ – Type I error probability
- $\alpha_{\mathrm{II}}$ – Type II error probability

# Full-information gap

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

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

$$= (1-p)\alpha_{\text{I}}(n_1, n_2)\overbrace{(u(\mathscr{A}, \text{H}_0) - u(\mathscr{R}, \text{H}_0))}^{\text{loss from Type-I}}$$

$$+ p\,\alpha_{\text{II}}(n_1, n_2)\overbrace{(u(\mathscr{R}, \text{H}_1) - u(\mathscr{A}, \text{H}_1))}^{\text{loss from Type-II}}$$

(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 from MS02

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

# 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**. More info

# Large deviations – Chernoff Index

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

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

Call $\rho$ the Chernoff index of the info source

Properties:

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

# Large deviations – Chernoff Precision

The Chernoff index is pretty abstract: consider instead

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

Call $\beta$ the Chernoff Precision.

Properties:

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

This will be the key object for my results.

# Chernoff precision – example

Gaussian noise: $r \sim \mathcal{N}(0, \sigma^2)$ in state $\mathrm{H}_0$
$r \sim \mathcal{N}(\mu, \sigma^2)$ in state $\mathrm{H}_1$

- 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

# 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
  - $\mathrm{M}_\omega(t) = \mathrm{M}_1(t)^\omega \mathrm{M}_2(t)^{1-\omega}$
  - So composite has MGF given by $\mathrm{M}_\omega(t)^{n_1+n_2}$
- Define $\rho_\omega \equiv \min_t \mathrm{M}_\omega(t)$
  $\tau_\omega \equiv \arg\min_t \mathrm{M}_\omega(t)$

# Composite experiments

Define the component index for each source as

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

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, $\tau$, 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 \ Report | 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, 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:

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

All DMs agree: 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:

*Proposition*

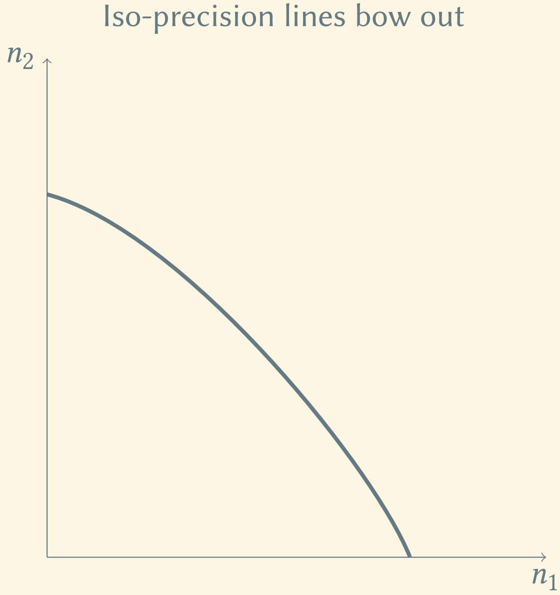
*As the budget, Υ, gets large, for (generically) any DM the proportion of samples from source 1 in the optimal bundle approaches*

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

Note: 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

# Graphical intuition



# Implications for optimization

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

# Summary: Two states

- 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

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 $\mathcal{E}_i$ 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

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

# Full-info gap

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

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

# 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* $\theta, \theta'$ *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* $\theta$ *or* $\theta'$ *and* $\bar{\rho} < 1$.

Proof sketch

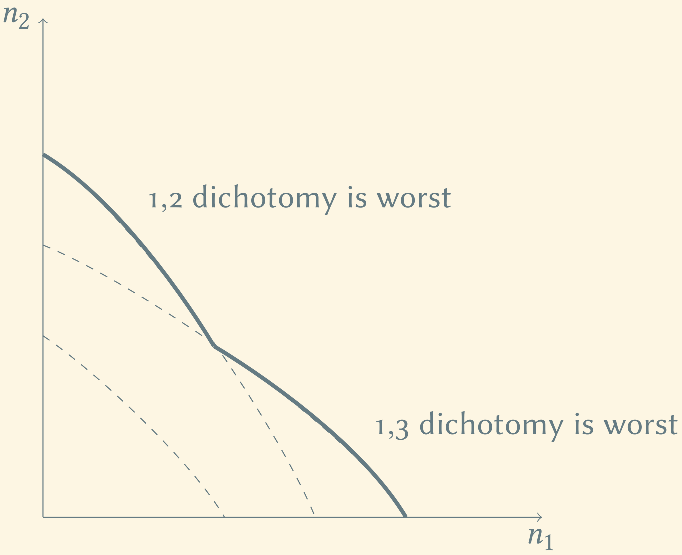
# Ordinal value

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

*Proposition (Maxi-min precision per dollar)*

*As the budget, Y, gets large, for (generically) any DM the proportion of samples from source 1 in the optimal bundle approaches*

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

# Agenda

- Preview of results
- Literature
- Model
- Large deviations approximations
  - The two-state case
  - The many-state case
- Consumer theory
  - “Marginal” rate of substitution
  - Implications for information demand
  - Is the approximation useful?
- Future work

# Agenda

Preview of results

Literature

Model

Large deviations approximations
- The two-state case
- The many-state case

Consumer theory
- “Marginal” rate of substitution
- Implications for information demand
- Is the approximation useful?

Future work

# Defining marginal rate of substitution

Samples are a *discrete* choice variable. Need to define a notion of *marginal*:

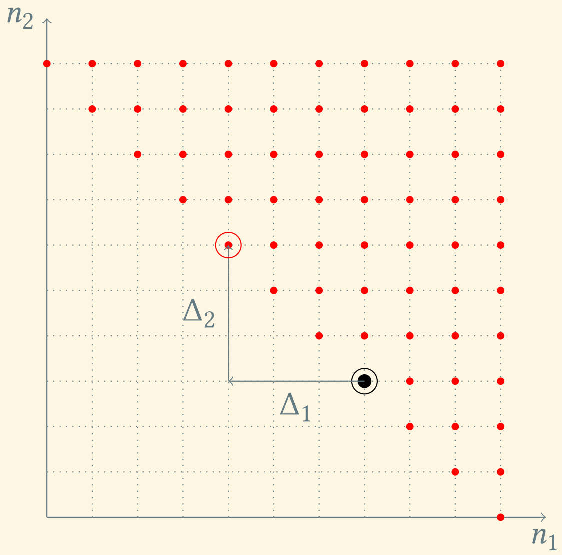
Define the *minimum compensating substitution* as

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

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 $\omega$ is negligible. Then solve for $\Delta_1/\Delta_2$

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

# Marginal rate of substitution

Put another way:

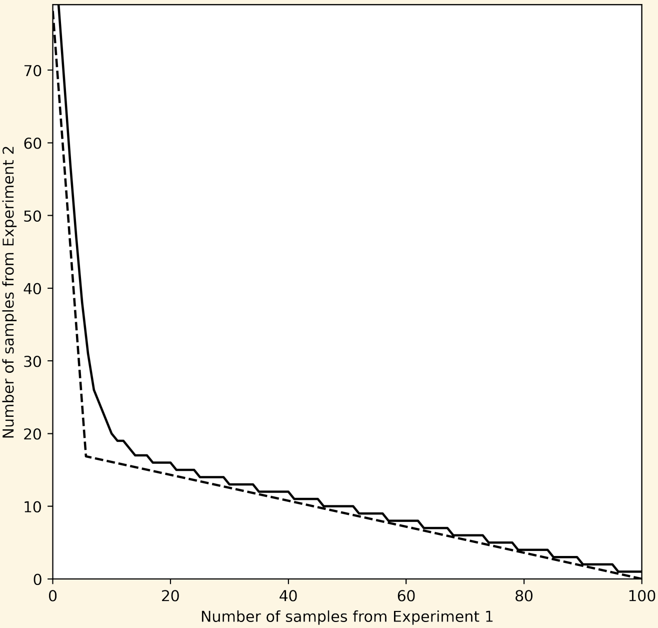
*Corollary*

*Fix a subsitution $\Delta_2, \Delta_1$. Then for (generically) any DM, there exists $n_1 + n_2$ high enough such that if*

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

*then the DM prefers the bundle $(n_1 - \Delta_1, n_2 + \Delta_2)$ to $(n_1, n_2)$*

# Indifference curves



# Agenda

Preview of results

Literature

Model

Large deviations approximations

- The two-state case
- The many-state case

Consumer theory

- “Marginal” rate of substitution
- Implications for information demand
- Is the approximation useful?

Future work

# 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.

# Income elasticity

Component precisions depend only on the relative proportions, $\omega$, 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.

$\Rightarrow$ 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 $\omega^*$ jumps.*

# Implications for competition

Fixed price monopolistic competition between two sellers of distinct information sources doesn't work 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?

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

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

# Is the approximation useful?

*Proposition*

*Let* $n_1^*(\mathrm{Y}, \mathrm{c}), n_2^*(\mathrm{Y}, \mathrm{c})$ *be the feasible bundle under budget* $\mathrm{Y}$ *with composite factor as close as possible to the one that maximizes the minimum precision per-dollar under cost vector* $\mathrm{c}$, *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*

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

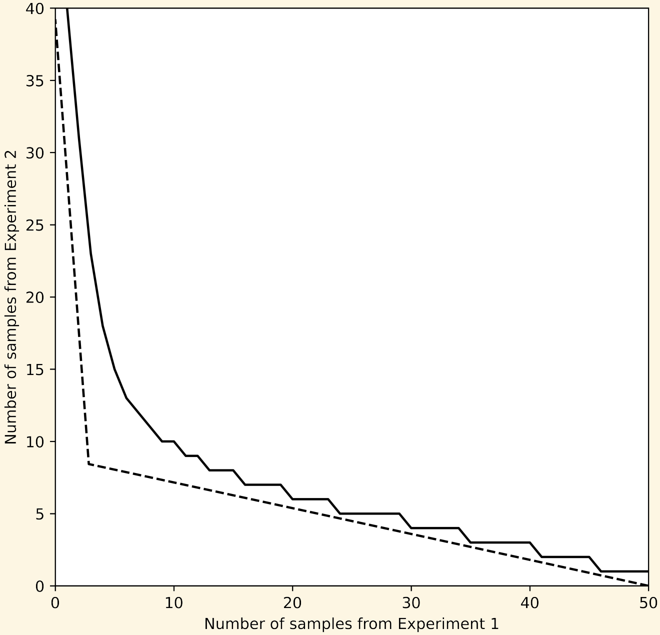
The optimal bundle eventually performs much better!

# Is the approximation useful?

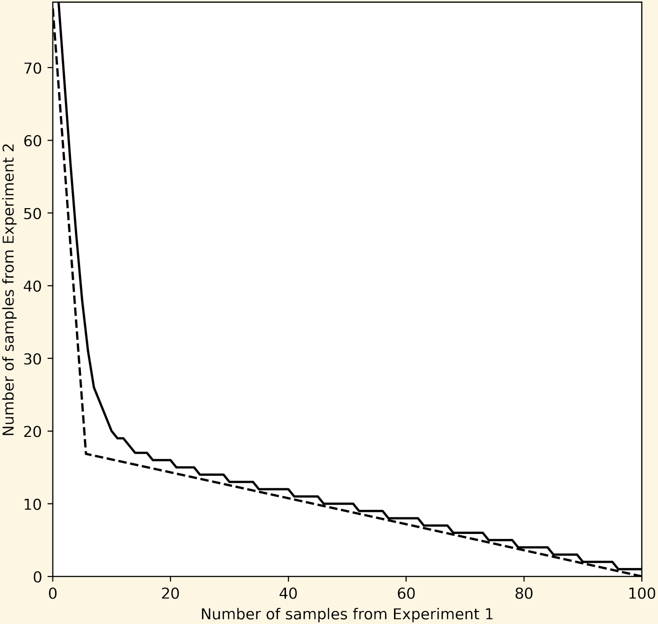
Put another way, the budget required to achieve a target performance is <u>much</u> smaller when following the maximin precision rule.

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

# How good is the approximation?



# How good is the approximation?



# 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.

# Summary

When information is cheap and/or budgets large:

- Maximizing information value under a constraint is equivalent to maximizing the precision per dollar of the worst case state pair.
- Optimal bundles always at a corner or one of finitely many interior kink points where multiple state pairs have equal precision.
- All DMs agree on the optimal composition!

# Summary

When information is cheap and/or budgets large:

- Information sources are all unit income elastic.
  - No inferior or luxury information goods
- For simple environments (relatively few possible states) demand is inelastic at interior solutions except for finitely many prices where the optimal composition jumps.

# 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/treatment assignment 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, $\theta$, 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 $\theta$ is the one "adjacent" to it, $\theta + d\theta$

With some work, it happens to be the case that

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

Roughly, $\hat{\beta}(\theta)$ measures how well a source can distinguish $\theta$ from nearby states.

But you might know $\hat{\beta}(\theta)$ by another name: 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!

email: gary.baker@wisc.edu
website: garygbaker.com

# Large vs Small Deviations

- Could we just use a CLT?
  - Problem: CLT approximates $\mathbb{P}(\bar{x}_n - \mu > \epsilon/\sqrt{n})$
  - Pr. that the deviation from the true mean is bigger than some shrinking cutoff
  - i.e. that the deviation is small.

- We have $\mathbb{P}(\bar{x}_n - \mu > -\mu + L/n)$
  - Pr. that the deviation from the true mean is more than a fixed amount
  - This is a large deviation.

Back to main

# Proof sketch (Error Approximation)

Define a new distribution: the exponential tilting:

$$\mathrm{dG}(s) \equiv \frac{e^{\tau s}\mathrm{dF}(s \mid \mathrm{H}_0)}{\rho}$$

Properties:

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

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

Back to main

# Proof sketch (many state approximation)

**Part 1:**

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

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

**Part 2:**

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

**Part 3:**

Squeeze

Back to main

Athey, S. and J. Levin (2018). The value of information in monotone decision problems. *Research in Economics*.

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

Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 93–102. University of California Press.

Börgers, T., A. Hernando-Veciana, and D. Krähmer (2013, jan). When are signals complements or substitutes? *Journal of Economic Theory 148*(1), 165–195.

Chade, H. and E. E. Schlee (2002). Another look at the Radner-Stiglitz nonconcavity in the value of information. *Journal of Economic Theory 107*(2), 421–452.

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.

Chernoff, H. (1953, dec). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics 24*(4), 586–602.

Dette, H., L. M. Haines, and L. A. Imhof (2007). Maximin and Bayesian optimal designs for regression models. *Statistica Sinica 17*(2), 463–480.

Elfving, G. (1952, jun). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics 23*(2), 255–262.

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

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

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*, 665–690.