CONSUMER THEORY FOR CHEAP INFORMATION

GARY BAKER

17 September 2021 UW–Madison Theory Seminar

QUESTION

Consider a constrained 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.

QUESTION

Consider a constrained 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

- **how much** information to buy

QUESTION

Consider a constrained 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

- **how much** information to buy
- **from which** sources to get it from.

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

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

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

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

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

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 studying a vaccine:
  - **State:** whether effective or not
  - **Action:** whether to introduce the vaccine or not
  - **Info sources:** different trial protocols
  - **Amount of info:** how many trial participants
  - **Constraint:** grant budget

GOAL

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

- Demand for information in constrained settings
  - Elasticities

GOAL

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

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

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

Going back to Blackwell [1951]:

- Information from different sources can't easily be compared

PROBLEMS WITH INFORMATION

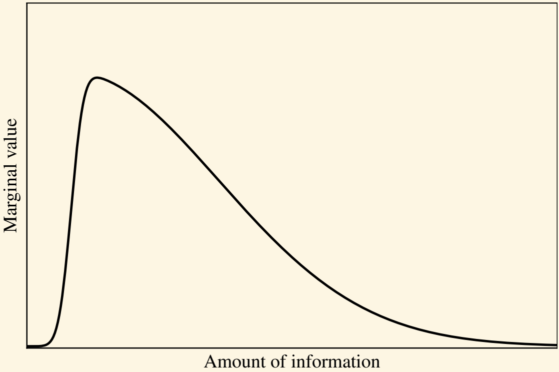
Going back to Blackwell [1951]:

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

PROBLEMS WITH INFORMATION

Another example: Marginal values of information can slope up at small samples.

- FOC analysis doesn't easily work



PROBLEMS WITH INFORMATION

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

WHAT I DO

To answer these questions I develop an **approximate** consumer theory for information.

That is, I will

- Define a generalized notion of **precision**
    - Demand approximately follows a maximin rule

WHAT I DO

To answer these questions I develop an **approximate** consumer theory for information.

That is, I will

- Define a generalized notion of **precision**
  - Demand approximately follows a maximin rule
- Explore implications for consumer theory

WHAT I DO

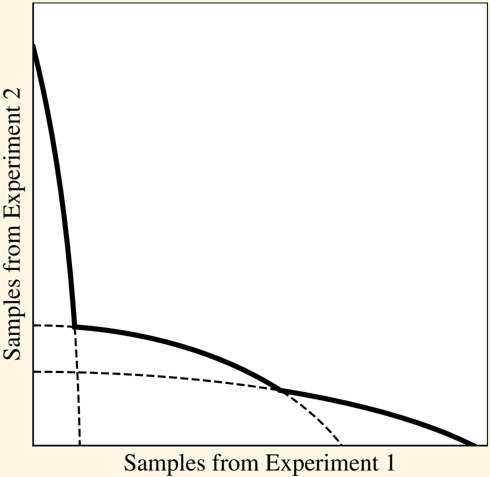
To answer these questions I develop an **approximate** consumer theory for information.

That is, I will

- Define a generalized notion of **precision**
    - Demand approximately follows a maximin rule
- Explore implications for consumer theory

Information is **not** described by the convex-preference benchmark.

WHAT I DO



WHAT I DO

- This approximation will **not** depend on decision-maker characteristics (prior, utility function).

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 at large samples.

AGENDA

AGENDA

LITERATURE

Statistics:

Chernoff [1952] (Asymptotic relative efficiency)

- How many samples from one test needed to do as well as n from another
- Comparison of extremes: all one or the other
- Only covers simple hypothesis tests (2 states)

LITERATURE

Statistics:

Chernoff [1952] (Asymptotic relative efficiency)

- How many samples from one test needed to do as well as n from another
- Comparison of extremes: all one or the other
- Only covers simple hypothesis tests (2 states)

Contribution:

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

LITERATURE

Economics:

Moscarini and Smith [2002]

- Apply similar methods to approximate info value and demand for information in the single source case

Contribution:

LITERATURE

Economics:

Moscarini and Smith [2002]

- Apply similar methods to approximate info value and demand for information in the single source case

Contribution:

- Economic: extend this to environment with multiple sources.
- Technical: tighten the bounds on the convergence rate.

OTHER RELATED LITERATURE

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

Rational inattention:
Sims [2003], &c.

Optimal experiment design:
Elfving [1952], Chernoff [1953], Dette et al. [2007], &c.

AGENDA

MODEL — ENVIRONMENT

- Finitely many possible states of the world, $\theta \in \Theta$
  - DM prior, $p \in \Delta\Theta$ (no degenerate beliefs)

MODEL — ENVIRONMENT

- Finitely many possible states of the world, $\theta \in \Theta$
  - DM prior, $p \in \Delta\Theta$ (no degenerate beliefs)
- Finitely many possible actions, $a \in A$

MODEL — ENVIRONMENT

- Finitely many possible states of the world, $\theta \in \Theta$
  - DM prior, $p \in \Delta\Theta$ (no degenerate beliefs)
- Finitely many possible actions, $a \in A$
- DM state-dependent utility, $u(a, \theta)$
  - Assume each state has a unique and distinct optimal action

MODEL — ENVIRONMENT

- Finitely many possible states of the world, $\theta \in \Theta$
  - DM prior, $p \in \Delta\Theta$ (no degenerate beliefs)
- Finitely many possible actions, $a \in A$
- DM state-dependent utility, $u(a, \theta)$
  - Assume each state has a unique and distinct optimal action
- DM chooses action to maximize expected payoff

MODEL — ENVIRONMENT

- Finitely many possible states of the world, $\theta \in \Theta$
  - DM prior, $p \in \Delta\Theta$ (no degenerate beliefs)
- Finitely many possible actions, $a \in A$
- DM state-dependent utility, $u(a, \theta)$
  - Assume each state has a unique and distinct optimal action
- DM chooses action to maximize expected payoff

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

MODEL — INFORMATION SOURCES

Information sources $\mathscr{E}_1, \mathscr{E}_2$

AKA: tests, signals, (Blackwell) experiments

- $\mathscr{E}_j \equiv \langle \mathrm{F}_j(x \mid \theta) \rangle$ ($x \in \mathrm{X}$ realizations)

MODEL – INFORMATION SOURCES

Information sources $\mathscr{E}_1, \mathscr{E}_2$

AKA: tests, signals, (Blackwell) experiments

- $\mathscr{E}_j \equiv \langle \mathrm{F}_j(x \mid \theta) \rangle$ ($x \in \mathrm{X}$ realizations)
- Assume: "thin tails"

MODEL – INFORMATION SOURCES

Information sources $\mathscr{E}_1, \mathscr{E}_2$

AKA: tests, signals, (Blackwell) experiments

- $\mathscr{E}_j \equiv \langle \mathrm{F}_j(x \mid \theta) \rangle$ ($x \in \mathrm{X}$ realizations)
- Assume: "thin tails"
    - $\Rightarrow$ no realization perfectly reveals or rules out any state.

MODEL – INFORMATION SOURCES

- DM can purchase an arbitrary number of *conditionally independent* samples, n_i, from each source at cost εc_i per sample
  - ε small

- For exposition, assume sources are **infinitely-divisible**, so fractional "samples" are allowed.

MODEL - INFORMATION SOURCES

- DM can purchase an arbitrary number of *conditionally independent* samples, n_i, from each source at cost εc_i per sample
  - ε small

- For exposition, assume sources are **infinitely-divisible**, so fractional "samples" are allowed.

- DM has budget Y to spend on info.

MODEL - INFORMATION SOURCES

- DM can purchase an arbitrary number of *conditionally independent* samples, n_i, from each source at cost εc_i per sample
  - ε small

- For exposition, assume sources are **infinitely-divisible**, so fractional "samples" are allowed.

- DM has budget Y to spend on info.

- After choosing a bundle of information (n_1, n_2), DM observes the vector of realizations, and updates via Bayes Rule.

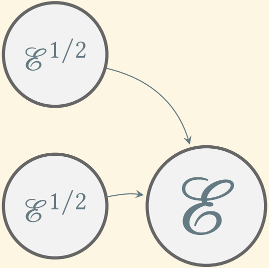
INFINITE-DIVISIBILITY

Definition

Say an information source, $\mathscr{E}$ is **infinitely-divisible** if for any k there exists an information source $\mathscr{E}^{1/k}$ such that k samples conditionally i.i.d. from $\mathscr{E}^{1/k}$ is equivalent to 1 sample from $\mathscr{E}$.

INFINITE-DIVISIBILITY

1/2 "samples" from $\mathscr{E}$ means 1 sample from $\mathscr{E}^{1/2}$



INFINITE DIVISIBILITY - EXAMPLE

An infinitely-divisible analog of any experiment can be acheived by **Poissonizing** it.

Instead of choosing samples directly, DM chooses **expected samples** and then receives an appropriate Poisson draw of samples.

INFINITE DIVISIBILITY - EXAMPLE

An infinitely-divisible analog of any experiment can be acheived by **Poissonizing** it.

Instead of choosing samples directly, DM chooses **expected samples** and then receives an appropriate Poisson draw of samples.

Sum of Poissons is Poisson:

1 sample from the Poissonization with 1 expected sample

$\Leftrightarrow$

2 samples from the Poissonization with 0.5 expected samples.

MODEL — EXPECTED LOSS

Expected difference in value of acting correctly and acting with information

$$\mathrm{L}(n_1, n_2) = \sum_{\theta} p_{\theta} \left[\int_x \underbrace{(u(a^*(\theta), \theta) - u(a(\mathrm{x}), \theta))}_{\text{cost of choosing } a(x) \text{ when } a(\theta) \text{ optimal}} f_{n_1,n_2}(x \mid \theta) \mathrm{d}x \right]$$

MODEL — EXPECTED LOSS

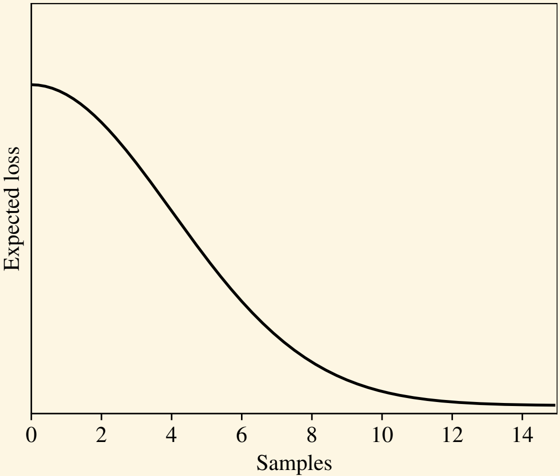
Expected difference in value of acting correctly and acting with information

$$L(n_1, n_2) = \sum_{\theta} p_\theta \left[\int_x \underbrace{\left(u(a^*(\theta), \theta) - u(a(\mathrm{x}), \theta)\right)}_{\text{cost of choosing } a(x) \text{ when } a(\theta) \text{ optimal}} f_{n_1,n_2}(x \mid \theta) \mathrm{d}x \right]$$

Goal: *Minimize subject to budget constraint*

(equivalent to maximizing value of information)

MODEL — EXPECTED LOSS



APPROACH

1. Review the relevant large-deviations approximations
   - Generalize to the multi-source model

APPROACH

1. Review the relevant large-deviations approximations
    - Generalize to the multi-source model
2. Transform into a “utility function”
    - Define **precision**

APPROACH

1. Review the relevant large-deviations approximations
    - Generalize to the multi-source model
2. Transform into a “utility function”
    - Define **precision**
3. Examine properties of the approximation and implications for demand

AGENDA

TWO STATES – SETUP

- States:
  - Null hypothesis – H_0
  - Alternative hypothesis – H_1
  - Prior that alternative is true – p

TWO STATES – SETUP

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

TWO STATES – EXPECTED LOSS

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

- α_{I} – Probability of Type-I error
- α_{II} – Probability of Type-II error

TWO STATES – EXPECTED LOSS

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

- α_{I} – Probability of Type-I error
- α_{II} – Probability of Type-II error

TWO STATES – EXPECTED LOSS

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

- α_{I} – Probability of Type-I error
- α_{II} – Probability of Type-II error

GOAL: Approximate error probabilities

ERROR PROBABILITIES

Start with the one-source case

$$\alpha_{\mathrm{I}}(n) =$$

$$\mathbb{P}\left(\frac{p \prod_{k=1}^{n} f(x_k \,|\, \mathrm{H}_1)}{p\Pi_{k=1}^{n} f(x_k \,|\, \mathrm{H}_1) + (1-p)\Pi_{k=1}^{n} f(x_k \,|\, \mathrm{H}_0)} > \bar{p} \,\middle|\, \mathrm{H}_0 \right)$$

ERROR PROBABILITIES

Change to log-likelihood ratios:

$$\alpha_{\mathrm{I}}(n)$$

$$= \mathbb{P}\left(\log\left(\frac{p}{1-p}\right) + \sum_{k=1}^{n} \log\left(\frac{f(x_k \mid \mathrm{H}_1)}{f(x_k \mid \mathrm{H}_0)}\right) > \log\left(\frac{\bar{p}}{1-\bar{p}}\right) \;\middle|\; \mathrm{H}_0\right)$$

ERROR PROBABILITIES

Change to log-likelihood ratios:

$$\alpha_{\mathrm{I}}(n)$$

$$= \mathbb{P}\left(\log\left(\frac{p}{1-p}\right) + \sum_{k=1}^{n} \log\left(\frac{f(x_k \mid \mathrm{H}_1)}{f(x_k \mid \mathrm{H}_0)}\right) > \log\left(\frac{\bar{p}}{1-\bar{p}}\right) \;\middle|\; \mathrm{H}_0\right)$$

$$= \mathbb{P}\left(\sum_{k=1}^{n} s_k > \bar{l} - l \;\middle|\; \mathrm{H}_0\right)$$

ERROR PROBABILITIES

$$\alpha_{\mathrm{I}}(n) = \mathbb{P}\left(\sum_{k=1}^{n} s_k > \bar{l} - l\right)$$

If we divide by n, we have a statement about sample averages.

ERROR PROBABILITIES

$$\alpha_{\mathrm{I}}(n) = \mathbb{P}\left(\sum_{k=1}^{n} s_k > \bar{l} - l\right)$$

If we divide by n, we have a statement about sample averages.

But can't use CLT: $\mathbb{E}(s_i) < 0$ errors happen far from the mean.
This is a **large** deviation.

MORE INFO

EFFICIENCY INDEX

Large deviations approximations often depend on a minimized moment generating function:

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

EFFICIENCY INDEX

$$\rho = \min_t \int f(x \mid \mathrm{H}_1)^t f(x \mid \mathrm{H}_0)^{1-t} \mathrm{d}x$$

Call ρ the (Chernoff) **efficiency index** of the information source.

Properties:

- $\rho \in (0, 1)$

EFFICIENCY INDEX

$$\rho = \min_t \int f(x \mid \mathrm{H}_1)^t f(x \mid \mathrm{H}_0)^{1-t} \mathrm{d}x$$

Call ρ the (Chernoff) **efficiency index** of the information source.

Properties:

- $\rho \in (0, 1)$
- Blackwell more informative ⇒ **lower** index

EFFICIENCY INDEX

$$\rho = \min_t \int f(x \mid \mathrm{H}_1)^t f(x \mid \mathrm{H}_0)^{1-t} \mathrm{d}x$$

Call ρ the (Chernoff) **efficiency index** of the information source.

Properties:

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

EFFICIENCY INDEX

$$\rho = \min_t \int f(x \mid \mathrm{H}_1)^t f(x \mid \mathrm{H}_0)^{1-t} \mathrm{d}x$$

Call ρ the (Chernoff) **efficiency index** of the information source.

Properties:

- $\rho \in (0, 1)$
- Blackwell more informative $\Rightarrow$ **lower** index
- n i.i.d. samples has index ρ^n
- Doesn't depend on DM characteristics

LARGE-DEVIATION APPROXIMATION

Large-deviations probabilities fall **exponentially**: error probabilities are roughly proportional to

$$\frac{\rho^n}{\sqrt{n}}$$

LARGE-DEVIATION APPROXIMATION

Large-deviations probabilities fall **exponentially**: error probabilities are roughly proportional to

$$\frac{\rho^n}{\sqrt{n}}$$

Lemma (MS02)

The expected loss from n samples from an information source with efficiency index ρ is

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

LARGE-DEVIATION APPROXIMATION

Large-deviations probabilities fall **exponentially**: error probabilities are roughly proportional to

$$\frac{\rho^n}{\sqrt{n}}$$

Lemma (MS02)*

The expected loss from n samples from an information source with efficiency index ρ is

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

TO MULTIPLE SOURCES

$N = n_1 + n_2$: **total sample size**

$r = n_1/N$: **composite factor**

TO MULTIPLE SOURCES

$N = n_1 + n_2$: **total sample size**

$r = n_1/N$: **composite factor**

MGF of a sum is the product of MGFs

$$M_1(t)^{n_1} M_2(t)^{n_2} = \left(M_1(t)^r M_2(t)^{1-r}\right)^N \equiv M_r(t)^N$$

TO MULTIPLE SOURCES

$\mathrm{N} = n_1 + n_2$: **total sample size**

$r = n_1/\mathrm{N}$: **composite factor**

MGF of a sum is the product of MGFs

$$\mathrm{M}_1(t)^{n_1}\mathrm{M}_2(t)^{n_2} = \left(\mathrm{M}_1(t)^r\mathrm{M}_2(t)^{1-r}\right)^{\mathrm{N}} \equiv \mathrm{M}_r(t)^{\mathrm{N}}$$

Define the r-**composite** efficiency index

$$\rho_r \equiv \min_t \mathrm{M}_r(t)$$

TO MULTIPLE SOURCES

$\mathrm{N} = n_1 + n_2$: **total sample size**

$r = n_1/\mathrm{N}$: **composite factor**

MGF of a sum is the product of MGFs

$$\mathrm{M}_1(t)^{n_1}\mathrm{M}_2(t)^{n_2} = \left(\mathrm{M}_1(t)^r\mathrm{M}_2(t)^{1-r}\right)^{\mathrm{N}} \equiv \mathrm{M}_r(t)^{\mathrm{N}}$$

Define the r-**composite** efficiency index

$$\rho_r \equiv \min_t \mathrm{M}_r(t)$$

and similarly the r-**marginal** efficiency indices

$$\rho_{ri} = \mathrm{M}_i(\tau_r)$$

so $\rho_r = \rho_{r1}^r\rho_{r2}^{1-r}$

LOSS WITH MULTIPLE SOURCES

Plugging things in, we have

$$L(n_1, n_2) = A(r)\frac{\rho_r^N}{\sqrt{N}}\left(1 + O\left(\frac{1}{N}\right)\right)$$

$$= A(r)\frac{\rho_{r1}^{n_1}\rho_{r2}^{n2}}{\sqrt{n_1 + n_2}}\left(1 + O\left(\frac{1}{n_1 + n_2}\right)\right)$$

where A depends only on the relative sample proportions r.

FORESHADOWING

The marginal index is the MGF evaluated at the minimizer for the composite.
So we have:

$$\rho_r = \rho_{r1}^{r}\rho_{r2}^{1-r} \geq \rho_1^{r}\rho_2^{1-r}$$

FORESHADOWING

The marginal index is the MGF evaluated at the minimizer for the composite.
So we have:

$$\rho_r = \rho_{r1}^r \rho_{r2}^{1-r} \geq \rho_1^r \rho_2^{1-r}$$

Composite experiments perform **worse** than the sum of their parts.

INTUITION

TO REVIEW

- Defined the **efficiency index**, ρ
- Losses fall exponentially fast in ρ with sample
- Introduced the **marginal** efficiency index, ρ_{rj}
  - loss is reduced by roughly a factor of ρ_{r1} consuming a sample from $\mathscr{E}_1$ in addition to a bundle with sample proportions r

AGENDA

- Preview of results
- Literature
- Model
- Large-deviations approximations
  - Two states
  - Many states
- Consumer theory
  - Demand for samples
  - Substitutability of samples
- How good is the approximation?
- Conclusion and future work

GENERALIZING TO MULTIPLE STATES

- With multiple states, we now have many log-likelihood ratio distributions:

GENERALIZING TO MULTIPLE STATES

- With multiple states, we now have many log-likelihood ratio distributions:
  - e.g. With three states, we have θ_1 vs θ_2, θ_1 vs θ_3, and θ_2 vs θ_3 LLRs.

GENERALIZING TO MULTIPLE STATES

- With multiple states, we now have many log-likelihood ratio distributions:
  - e.g. With three states, we have θ_1 vs θ_2, θ_1 vs θ_3, and θ_2 vs θ_3 LLRs.
- So for $\mathscr{E}_i$ we can define an efficiency index for each pair of states

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

EXPECTED LOSS

$$L(n_1, n_2)$$

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

EXPECTED LOSS APPROXIMATION

Intuition:

- Expected loss is a sum of mistake probabilities

EXPECTED LOSS APPROXIMATION

Intuition:

- Expected loss is a sum of mistake probabilities
- Mistake probabilities fall **exponentially**

EXPECTED LOSS APPROXIMATION

Intuition:

- Expected loss is a sum of mistake probabilities
- Mistake probabilities fall **exponentially**
- Sum of exponentials ⇒ **biggest** term eventually dominates

EXPECTED LOSS APPROXIMATION

Applying a lemma of MS02, we have

$$L(n_1, n_2) = A(r)\frac{\max_{\theta,\theta'}\{\rho_r(\theta,\theta')^N\}}{\sqrt{N}}\left(1 + O\left(\frac{1}{N}\right)\right)$$

$$= A(r)\frac{\max_{\theta,\theta'}\{\rho_{r1}(\theta,\theta')^{n_1}\rho_{r2}(\theta,\theta')^{n_2}\}}{\sqrt{n_1+n_2}}\left(1 + O\left(\frac{1}{N}\right)\right)$$

where A depends only the relative sample proportions.

TO REVIEW

- Efficiency index for each pair of states
  - How well does an experiment distinguish between a pair of states

TO REVIEW

- Efficiency index for each pair of states
  - How well does an experiment distinguish between a pair of states
- Loss dominated by most likely mistake, i.e. **highest** index

APPROACH

1. Review the relevant large-deviations approximations ☑
    - Generalize to the multi-source model ☑
2. Transform into a “utility function”
    - Define **precision**
3. Examine properties of the approximation and implications for demand

PRECISION

The multiplicative form of the approximation screams to have logs taken:

In the one-source, two-state world:

$$-\log(\mathrm{L}(n)) = n\beta\left(1 + \mathrm{O}\left(\log(n)n^{-1}\right)\right)$$

Call $\beta \equiv -\log(\rho)$ the **precision** of the experiment

PRECISION

The multiplicative form of the approximation screams to have logs taken:

In the one-source, two-state world:

$$-\log(\mathrm{L}(n)) = n\beta\left(1 + \mathrm{O}\left(\log(n)n^{-1}\right)\right)$$

Call $\beta \equiv -\log(\rho)$ the **precision** of the experiment

Properties

- $\beta > 0$

PRECISION

The multiplicative form of the approximation screams to have logs taken:

In the one-source, two-state world:

$$-\log(\mathrm{L}(n)) = n\beta\left(1 + \mathrm{O}\left(\log(n)n^{-1}\right)\right)$$

Call $\beta \equiv -\log(\rho)$ the **precision** of the experiment

Properties

- $\beta > 0$
- Blackwell more informative $\Longrightarrow$ **higher** precision

PRECISION

The multiplicative form of the approximation screams to have logs taken:

In the one-source, two-state world:

$$-\log(\mathrm{L}(n)) = n\beta\left(1 + \mathrm{O}\left(\log(n)n^{-1}\right)\right)$$

Call $\beta \equiv -\log(\rho)$ the **precision** of the experiment

Properties

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

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

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

- Precision is $\beta = \frac{1}{8}\frac{\mu^2}{\sigma^2}$

PRECISION — EXAMPLE

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

- 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 $(1/\sigma^2)$

MARGINAL PRECISION

Similarly, define the r-**composite** precision and r-**marginal** precisions.

$$\begin{aligned} -\log(\rho_r) \equiv \beta_r &= r\beta_{r1} + (1-r)\beta_{r2} \\ &\equiv -r\log(\rho_{r1}) - (1-r)\log(\rho_{r2}) \end{aligned}$$

PRECISION AND UTILITY

Putting it all together we have

$$-\log(L(n_1, n_2)) = \min_{\theta,\theta'} \{N\beta_r(\theta, \theta')\} \left(1 + O\left(\frac{\log(N)}{N}\right)\right)$$

$$= \min_{\theta,\theta'} \{n_1\beta_{r1}(\theta, \theta') + n_2\beta_{r2}(\theta, \theta')\} \left(1 + O\left(\frac{\log(N)}{N}\right)\right)$$

PRECISION AND UTILITY

Putting it all together we have

$$-\log(\mathrm{L}(n_1, n_2)) = \min_{\theta,\theta'} \{\mathrm{N}\beta_r(\theta, \theta')\} \left(1 + \mathrm{O}\left(\frac{\log(\mathrm{N})}{\mathrm{N}}\right)\right)$$

$$= \min_{\theta,\theta'} \{n_1\beta_{r1}(\theta, \theta') + n_2\beta_{r2}(\theta, \theta')\} \left(1 + \mathrm{O}\left(\frac{\log(\mathrm{N})}{\mathrm{N}}\right)\right)$$

at high enough total samples, prefer bundles with higher total **minimum** (worst-case) precision!

TO REVIEW

- Defined a generalized notion of **precision** and **marginal** precision
  - Approximate utility and marginal utility for information

TO REVIEW

- Defined a generalized notion of **precision** and **marginal** precision
  - Approximate utility and marginal utility for information
- **Remember:** Precision independent of DM characteristics
  - Everyone agrees on ranking of bundles at large samples

APPROACH

1. Review the relevant large-deviations approximations ☑
   - Generalize to the multi-source model ☑
2. Transform into a “utility function” ☑
   - Define **precision** ☑
3. Examine properties of the approximation and implications for demand

AGENDA

- Preview of results
- Literature
- Model
- Large-deviations approximations
  - Two states
  - Many states
- Consumer theory
  - Demand for samples
  - Substitutability of samples
- How good is the approximation?
- Conclusion and future work

Proposition (Maximin precision)

For budget Υ *and per sample costs* εc_1 *and* εc_2 *optimal sample demand is*

$$\begin{aligned}&(n_1^*, n_2^*)\\ &= \left(\arg\max_{n_1,n_2} \min_{\theta,\theta'} \{n_1\beta_{r1}(\theta,\theta') + n_2\beta_{r2}(\theta,\theta')\}\right)\\ &\qquad \times (1 + O(\varepsilon))\end{aligned}$$

subject to $\varepsilon(n_1 c_1 + n_2 c_2) \leq \Upsilon$.

DEMAND FOR CHEAP INFORMATION

Proposition (Maximin precision)

For budget Y *and per sample costs* εc_1 *and* εc_2 *optimal sample demand is*

$$(n_1^*, n_2^*) = \left(\arg\max_{n_1,n_2} \min_{\theta,\theta'} \{n_1\beta_{r1}(\theta,\theta') + n_2\beta_{r2}(\theta,\theta')\}\right) \times (1 + O(\varepsilon))$$

subject to $\varepsilon(n_1c_1 + n_2c_2) \leq Y$.

Can treat total worst-case precision **as if** utility.

PRECISION PER DOLLAR

Since precision is homothetic, we can equivalently say the optimal sample proportions maximize **precision per dollar**

$$r^* = \arg\max_r \left\{ \min_{\theta,\theta'} \left\{ \frac{\beta_r(\theta,\theta')}{rc_1 + (1-r)c_2} \right\} \right\}$$

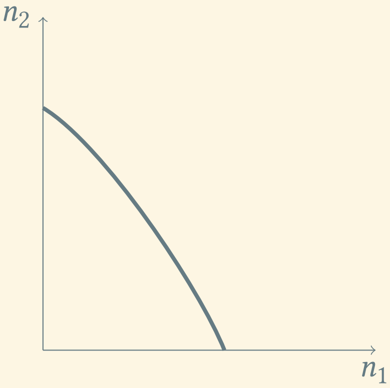
CORNERS?

Recall composites are worse than the sum of their parts for a fixed dichotomy:

$$\rho_{r1}^{n_1}\rho_{r2}^{n_2} \geq \rho_1^{n_1}\rho_2^{n_2} \quad \Leftrightarrow \quad n_1\beta_{r1} + n_2\beta_{r2} \leq n_1\beta_1 + n_2\beta_2$$

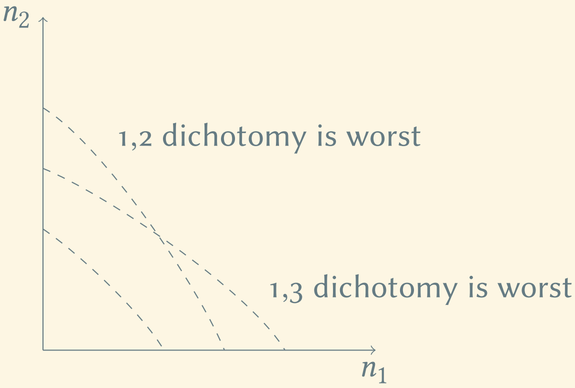
Corners always optimal?

ISO-LEAST-PRECISION CURVES



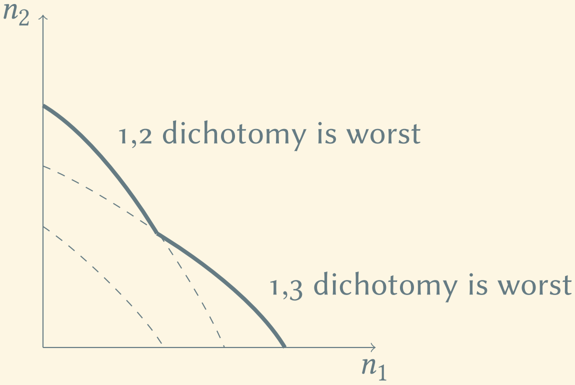
For a single dichotomy, iso-precision curves **bow out**

ISO-LEAST-PRECISION CURVES



For a single dichotomy, iso-precision curves **bow out**

ISO-LEAST-PRECISION CURVES



But the outer contour has inward pointing kinks

PROPERTIES OF PRECISION

Lemma

For a fixed dichotomy, total precision is homothetic and (quasi)convex.
Worst-case precision is thus homothetic and locally quasiconvex at almost all sample proportions r.

MORE INFO

INCOME ELASTICITY

Information sources are always normal goods at large samples.

Corollary (Income elasticity)

The (arc) income elasticity of demand given a fixed change in budget is $1 + O(\varepsilon)$.

DEMAND AT KINKS

Proposition (corners or kinks)

The set of sample proportions that maximize worst-case precision for some cost vector and budget is finite.

$$\left|\left\{r^* \ : \ \exists c_1, c_2, \Upsilon \ s.t. \ r^* \in \arg\max_r \{\min_{\theta,\theta'} \beta_r(\theta,\theta')/(rc_1 + (1-r)c_2)\}\right\}\right| < \infty$$

LOCALLY PERFECT COMPLEMENTS

Info behaves (locally) like **perfect complements**:

Corollary (Price elasticity)

At almost all costs, the (arc) price elasticity of demand for samples from all sources given a small percent change, δ, of c_1 is

$$\eta_1 = \frac{rc_1}{rc_1 + (1-r)c_2}\left(1 + O\left(\varepsilon + \delta\right)\right)$$

LOCALLY PERFECT COMPLEMENTS

Info behaves (locally) like **perfect complements**:

Corollary (Price elasticity)

At almost all costs, the (arc) price elasticity of demand for samples from all sources given a small percent change, δ, of c_1 is

$$\eta_1 = \frac{rc_1}{rc_1 + (1-r)c_2} \left(1 + O\left(\varepsilon + \delta\right)\right)$$

But the sample bundle changes drastically around costs that jump between kinks!

BUNDLE COMPLEXITY

In a two-state environment (one dichotomy) only corners were possible.

Hints that “sophisticated” sample demand requires complicated environments.

BUNDLE COMPLEXITY

In a two-state environment (one dichotomy) only corners were possible.

Hints that “sophisticated” sample demand requires complicated environments.

Proposition

The maximin precision sample bundle has support on at most distinct information sources as there are dichotomies—i.e. $|\Theta|(|\Theta - 1|)/2$.

AGENDA

MARGINAL RATE OF SUBSTITUTION

Solutions are at kinks so there isn't much use for MRS in a classic budget-constrained environment.

MARGINAL RATE OF SUBSTITUTION

Solutions are at kinks so there isn't much use for MRS in a classic budget-constrained environment.

But information is weird:

- Often non-rival, non-excludable

MARGINAL RATE OF SUBSTITUTION

Solutions are at kinks so there isn't much use for MRS in a classic budget-constrained environment.

But information is weird:

- Often non-rival, non-excludable
- Often have additional constraints (finite-sample datasets)

Might want to understand what happens away from the kinks.

MARGINAL RATE OF SUBSTITUTION

Proposition

If the worst-case dichotomy, D, *is unique at sample proportion* r, *then the marginal rate of substitution is*

$$\frac{\partial L/\partial n_1}{\partial L/\partial n_2} = \frac{\beta_{r1}(D)}{\beta_{r2}(D)}\left(1 + O\left(\frac{1}{n_1 + n_2}\right)\right)$$

MARGINAL RATE OF SUBSTITUTION

Proposition

If the worst-case dichotomy, D, *is unique at sample proportion* r, *then the marginal rate of substitution is*

$$\frac{\partial L/\partial n_1}{\partial L/\partial n_2} = \frac{\beta_{r1}(D)}{\beta_{r2}(D)}\left(1 + O\left(\frac{1}{n_1 + n_2}\right)\right)$$

Samples are substitutable in proportion to their marginal precisions.

DISCRETE VERSION

AGENDA

- Preview of results
- Literature
- Model
- Large-deviations approximations
  - Two states
  - Many states
- Consumer theory
  - Demand for samples
  - Substitutability of samples
- How good is the approximation?
- Conclusion and future work

SOURCES OF ERROR

Two sources of approximation error:

1. Large-deviations approximation for single mistake probability

SOURCES OF ERROR

Two sources of approximation error:

1. Large-deviations approximation for single mistake probability
2. Throwing out all but the most likely mistake

SOURCES OF ERROR

Large-deviations errors are small by the standards of large-sample approximations (CLT is $O(n^{-1/2})$)

SOURCES OF ERROR

Large-deviations errors are small by the standards of large-sample approximations (CLT is $O(n^{-1/2})$)

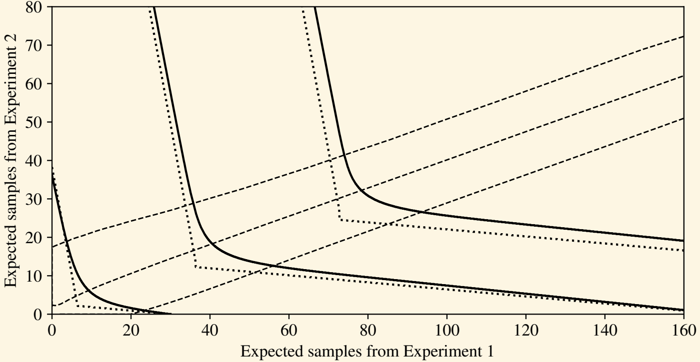
Ignoring less likely mistakes is fine so long as next most likely mistake isn't particularly close

SOURCES OF ERROR

Large-deviations errors are small by the standards of large-sample approximations (CLT is $O(n^{-1/2})$)

Ignoring less likely mistakes is fine so long as next most likely mistake isn't particularly close

Approximation works best when the total number of possible states is small.


Expected samples from Experiment 2
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
Expected samples from Experiment 1

AGENDA

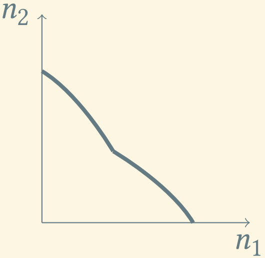
- Preview of results
- Literature
- Model
- Large-deviations approximations
  - Two states
  - Many states
- Consumer theory
  - Demand for samples
  - Substitutability of samples
- How good is the approximation?
- Conclusion and future work

CONCLUSION

- Defined a general notion of **precision**
- Showed demand can be approximately analyzed treating precision **as if** it were a utility function

CONCLUSION

- Information demand behaves as though indifference curves were locally bowed out, kinked, and homothetic
- Locally, sources are perfect complements



IMPLICATIONS

- Suggests a form for information demand for applied work
  - Treat information as a good with care (preferences are **not** convex)

IMPLICATIONS

- Suggests a form for information demand for applied work
  - Treat information as a good with care (preferences are **not** convex)
- Suggests a Bayesian approach to optimal experiment design
  - Interior solutions matter

FUTURE WORK: INFINITE STATES

Most statistical problems are ones of **estimation**.

Unknown state is a real-valued parameter.

FUTURE WORK: INFINITE STATES

Most statistical problems are ones of **estimation**.

Unknown state is a real-valued parameter.

PROBLEM: As two states get close, the precision goes to zero.

NAÏVE APPROACH

THANK YOU!

EMAIL: gary.baker@wisc.edu
WEBSITE: garygbaker.com

LARGE VS SMALL DEVIATIONS

- Could we just use a CLT?
  - No: 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

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 \ 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.

BACK

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 \ 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.

BACK

LOCAL (QUASI)CONVEXITY

Definition

Say a function, f, is locally (quasi)convex around a point x if for ε small enough f is (quasi)convex on $B(x, \varepsilon)$.

BACK

DISCRETE ANALOG TO MRS

Proposition

Suppose $\mathscr{E}_1, \mathscr{E}_2$ not infinitely divisible. If there is a unique worst-case dichotomy at sample proportions, r, then the minimum number of samples from $\mathscr{E}_2$, k_2, required to minimally compensate for a loss of k_1 samples from $\mathscr{E}_1$ is exactly

$$k_2 = \left\lceil k_1 \frac{\beta_{r1}(\mathrm{D})}{\beta_{r2}(\mathrm{D})} \right\rceil$$

for $n_1 + n_2$ high enough.

BACK

INFINITE STATES: NAÏVE APPROACH

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

With some work, it happens to be the case that

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

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

INFINITE STATES: NAÏVE APPROACH

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

With some work, it happens to be the case that

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

Roughly, $\hat{\beta}(\theta)$ measures how well a source can distinguish θ from nearby states.
But you might know $\hat{\beta}(\theta)$ by another name: Fisher information

INFINITE STATES: NAÏVE APPROACH

Suggests a link between Chernoff's efficiency notion and Pitman's efficiency notion.

Names of Contributors	Behavior of Type I Error Probability α_n	Behavior of Type II Error Probability β_n	Behavior of Alternatives
Pitman	$\alpha_n \to \alpha > 0$	$\beta_n \to \beta > 0$	$F^{(n)} \to \mathscr{F}_0$
Chernoff	$\alpha_n \to 0$	$\beta_n \to 0$	$F^{(n)} = F$ fixed
Bahadur	$\alpha_n \to 0$	$\beta_n \to \beta > 0$	$F^{(n)} = F$ fixed
Hodges & Lehmann	$\alpha_n \to \alpha > 0$	$\beta_n \to 0$	$F^{(n)} = F$ fixed
Hoeffding	$\alpha_n \to 0$	$\beta_n \to 0$	$F^{(n)} = F$ fixed
Rubin & Sethuraman	$\alpha_n \to 0$	$\beta_n \to 0$	$F^{(n)} \to \mathscr{F}_0$

BACK

Susan Athey and Jonathan Levin. The value of information in monotone decision problems. *Research in Economics*, 2018. doi: 10.1016/j.rie.2017.01.001.

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

Tilman Börgers, Angel Hernando-Veciana, and Daniel Krähmer. When are signals complements or substitutes? *Journal of Economic Theory*, 148 (1):165–195, jan 2013. doi: 10.1016/j.jet.2012.12.012.

Herman Chernoff. 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, 1952.

Herman Chernoff. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24(4):586–602, dec 1953. doi: 10.1214/aoms/1177728915.

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

G. Elfving. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, 23

(2):255–262, jun 1952. doi: 10.1214/aoms/1177729442.

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

Christopher A Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50: 665–690, 2003. doi: 10.1016/S0304-3932(03)00029-1.