Lecture 1: Models of Bounded Rationality

Joel P. Flynn
Yale

May 22, 2024
NBER Behavioral Macroeconomics Bootcamp

Why Model Bounded Rationality?

Elementary observation: people make mistakes

Economic hypothesis: smaller mistakes when they are more "costly"

Goals:

1. Overview existing models of making mistakes
2. Explore their *equilibrium* implications

Outline

Psychometrics: Stylized Facts About Choices

A General Model of Stochastic Choice

The Mutual Information Model

Back to Macro: Rational Inattention in Markets, Games

Takeaways

Facts to Motivate Our Search for a Model of Mistakes

- Our search for a good model of bounded rationality should be guided by our psychological evidence about how people make mistakes

- Five stylized facts

  1. Choice is random
  2. Choice responds to incentives
  3. Choice depends on prior beliefs
  4. Choice depends on the decision context
  5. Choice can depend on decision-irrelevant context

Fact 1: Choice is Random

Psychometric Function: The probability of choosing the correct answer as a function of properties of stimuli

Examples:

1. Which light is brighter?
2. Which object is heavier?
3. **How many objects are there?**

Full rationality benchmark: get the correct answer

Fact 1: Choice is Random

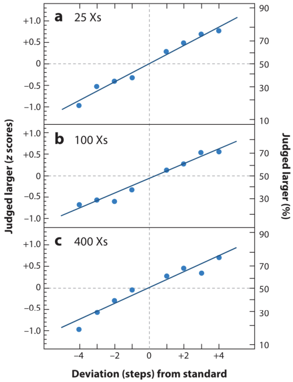


Figure 1

Psychometric functions for comparisons of numerosity. The number of items (Xs) in the reference array is (*a*) 25, (*b*) 100, and (*c*) 400. The right scale indicates the probability of judging the second numerosity to be larger (in percent), while the left scale indicates the corresponding *z* score. Figure adapted with permission from Krueger (1984).

Figure from Woodford (2020): Modeling Imprecision in Perception, Valuation, and Choice (Ann. Rev. Econ.)

Facts 2/3: Choice Responds to Incentives and Prior Beliefs

Key to economics being relevant (as distinct from relevant to economics!): do choices respond to payoffs (incentives) and beliefs (priors)?

An example experiment (Green and Swets, 1966)

1. Two auditory stimuli s and n
2. Subject played audio and can answer S or N
3. Hit rate: $P(S|s)$. False alarm rate: $P(S|n)$.

Vary both the prior probabilities of s and n (0.1,0.3.0.5,0.7,0.9) and the incentive for correct identification

Facts 2/3: Choice Responds to Incentives and Prior Beliefs

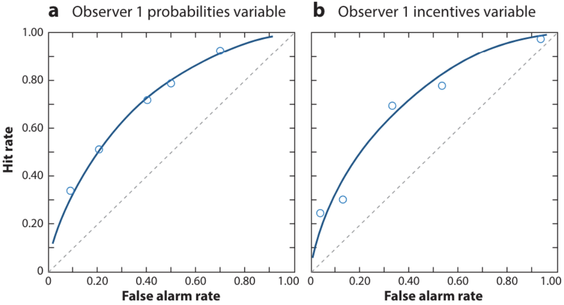


Figure 2

Conditional response probabilities in a signal detection task. Each of the circles in the figure plots the subject's conditional response probabilities for one block of trials. The trade-off is shown between the hit rate (vertical axis) and the false alarm rate (horizontal axis), as one varies the prior probability of occurrence of the two stimuli (*a*) or the relative rewards for correct identification of the two stimuli (*b*). In each case, the efficient frontier (ROC curve) is shown by the bowed solid curve. Abbreviation: ROC, receiver operating characteristic. Figure adapted from Green & Swets (1966) with permission of Peninsula Publishing.

Figure from Woodford (2020): Modeling Imprecision in Perception, Valuation, and Choice (Ann. Rev. Econ.)

Fact 4: Choice Depends on the Decision Context

The accuracy and precision of choice varies with the "context of decision problems," such as the action space and the state space

Dean and Neligh (2023): Experimental Tests of Rational Inattention (JPE)

1. The action space matters:
   1.1 Pick between two options
   1.2 Introduce a third option (expand action space)
   1.3 **Probability that one of the actions is taken increases**

2. The state space matters:
   2.1 **Probabilities are more inaccurate when people are asked to distinguish states that look more similar**

Fact 5: Choice Can Depend on Decision-Irrelevant Context

The accuracy and precision of choice can also depend on context that is *not* decision-relevant

Mani et al. (2013): Poverty Impedes Cognitive Function (Science)

Performance in cognitive task declines when

1. People are reminded if the difficulty of making financial decisions
2. Have higher or lower income from a seasonal cycle

See also Gorodnichenko et al. (2024): The Economics of Financial Stress (WP)

Outline

Primitives

- Actions $x \in X \subset \mathbb{R}$
- State $\theta \in \Theta \subset \mathbb{R}$, where Θ is finite; (subjective) probability distribution $\pi \in \Delta(\Theta)$

Primitives

- Actions $x \in X \subset \mathbb{R}$
- State $\theta \in \Theta \subset \mathbb{R}$, where Θ is finite; (subjective) probability distribution $\pi \in \Delta(\Theta)$
- Choice variable ("plan") is $(p(\cdot \mid \theta))_{\theta \in \Theta} \in (\Delta(X))^{|\Theta|} =: \mathcal{P}$, a probability distribution over actions in each state of the world

Primitives

- Actions $x \in X \subset \mathbb{R}$
- State $\theta \in \Theta \subset \mathbb{R}$, where Θ is finite; (subjective) probability distribution $\pi \in \Delta(\Theta)$
- Choice variable ("plan") is $(p(\cdot \mid \theta))_{\theta \in \Theta} \in (\Delta(X))^{|\Theta|} =: \mathcal{P}$, a probability distribution over actions in each state of the world
- Payoffs come from two source:
  - ex post felicity, $u : X \times \Theta \to \mathbb{R}$
  - ex ante cost of implementing plan, $K : (\Delta(X))^{|\Theta|} \times \Delta(\Theta) \to \mathbb{R}$

Primitives

- Actions $x \in X \subset \mathbb{R}$
- State $\theta \in \Theta \subset \mathbb{R}$, where Θ is finite; (subjective) probability distribution $\pi \in \Delta(\Theta)$
- Choice variable ("plan") is $(p(\cdot \mid \theta))_{\theta \in \Theta} \in (\Delta(X))^{|\Theta|} =: \mathcal{P}$, a probability distribution over actions in each state of the world
- Payoffs come from two source:
    - ex post felicity, $u : X \times \Theta \to \mathbb{R}$
    - ex ante cost of implementing plan, $K : (\Delta(X))^{|\Theta|} \times \Delta(\Theta) \to \mathbb{R}$
- So in summary, the agent solves

$$\max_{p \in \mathcal{P}} \left\{ \sum_{\theta \in \Theta} \sum_{x \in X} u(x, \theta) p(x \mid \theta) \pi(\theta) - K[p, \pi] \right\} \tag{1}$$

Defining "Mistakes"

$$\max_{p \in \mathcal{P}} \left\{ \sum_{\theta \in \Theta} \sum_{x \in X} u(x, \theta) p(x \mid \theta) \pi(\theta) - K[p, \pi] \right\} \tag{2}$$

- Assume there is a unique $x^* : \Theta \to X$ that defines the *ex post* optimum: that is, for each $\theta \in \Theta$, $\{x^*(\theta)\} = \arg\max_{x \in X} u(x, \theta)$
- It is possible to put full probability on x^*. Let's call *any* deviation of $p(\cdot \mid \theta)$ from $x^*(\theta)$ a "mistake" in that state.

Why Make Mistakes?

$$\max_{p \in \mathcal{P}} \left\{ \sum_{\theta \in \Theta} \sum_{x \in X} u(x, \theta) p(x \mid \theta) \pi(\theta) - K[p, \pi] \right\} \tag{3}$$

Why Make Mistakes?

$$\max_{p\in\mathcal{P}} \left\{ \sum_{\theta\in\Theta} \sum_{x\in X} u(x,\theta) p(x \mid \theta) \pi(\theta) - K[p,\pi] \right\} \tag{3}$$

What could K embody?

- Trembling hands: decision making (especially precision) requires effort, and people don't put in 100%
- Information acquisition: it's hard to learn perfectly about θ
- Complexity: it is hard to think through the "logic" of a problem, even form the right expected utility representation. $\Rightarrow$ huge open questions, that for now are mostly confined to experimental economics
- Ambiguity aversion: I might randomize to hedge against uncertainty about the state or probability distribution. see, e.g., Fudenberg, Iijima, Strzalecki (ECMA), 2015, Section 5

Trembling Hands (I)

- The simplest, most classic models of trembling hands are "state-separable":

$$K[p, \phi] = \sum_{\theta} \lambda(\theta) \sum_{x} \phi(p(x \mid \theta)) \tag{4}$$

  where ϕ is increasing and convex, and $\lambda > 0$ are state-specific weights

- Interpretation: *precision* is difficult to achieve, weights measure how easy or hard different states are to plan for
  1. Ex post planning: $\lambda = \pi$
  2. Ex ante planiing: $\lambda(\theta) \equiv \text{cons}$

Trembling Hands (II)

- Logit model (Harsanyi, 1973; McKelvey and Palfrey, 1995), up to constant $\lambda > 0$

$$\phi(x) = \lambda x \log x \tag{5}$$

- Key property: Inada condition

Trembling Hands (II)

- Logit model (Harsanyi, 1973; McKelvey and Palfrey, 1995), up to constant $\lambda > 0$

$$\phi(x) = \lambda x \log x \tag{5}$$

- Key property: Inada condition

- Choice probabilities available in closed form

$$p(x \mid \theta) = \frac{\exp(\lambda^{-1} u(x, \theta))}{\sum_{x' \in X} \exp(\lambda^{-1} u(x', \theta))} \tag{6}$$

  - Resembles McFadden's model with Type-I Extreme Value shocks
  - Generates choice that is random, responds to incentives, and gets more precise when utility differences are large.

Trembling Hands (III)

- Quadratic model (Rosenthal, 1989)

$$\phi(x) = \lambda \frac{x^2}{2} \tag{7}$$

- Key property: no Inada condition!

Trembling Hands (III)

- Quadratic model (Rosenthal, 1989)

$$\phi(x) = \lambda \frac{x^2}{2} \tag{7}$$

- Key property: no Inada condition!

- Choice probabilities available in closed form

$$p(x \mid \theta) = \frac{1}{\lambda}(u(x, \theta) - \bar{u}(\theta)) 1_{u(x,\theta) \geq \bar{u}(\theta)} \tag{8}$$

  where constants $\bar{u}(\theta)$ chosen so each distribution integrates to unity
  - Endogenous "consideration sets": possible that only a strict subset of actions is played with positive probability

Outline

Information Theory (I)

- One important approach in economics, due to Sims (2003): model costly information acquisition by appealing to established theory in mathematics about how to "optimally" encode information

Information Theory (I)

- One important approach in economics, due to Sims (2003): model costly information acquisition by appealing to established theory in mathematics about how to "optimally" encode information

- Claude Shannon introduced an important definition of information content in a random variable: the *entropy* for a random variable θ, with distribution function π

$$H[\pi] = -\sum_{\theta} p(\theta) \log p(\theta) = \mathbb{E}_{\pi}[-\log p(\theta)] \tag{9}$$

Information Theory (II)

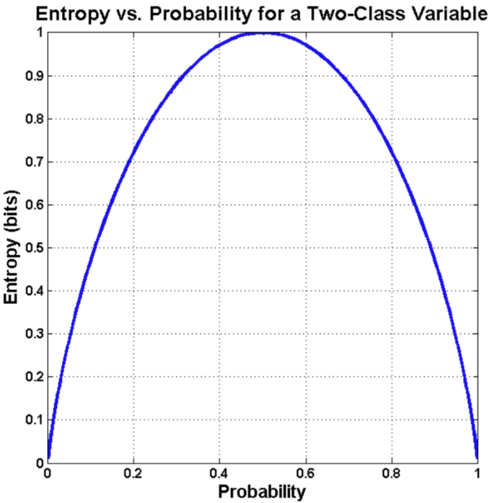


Figure: Entropy of a Bernoulli R.V.

Information Theory (III)

Justification 1: the encoding theorem

- If we have a random variable X with entropy $H(X)$, then (i) it can be compressed into $H(X)$ bits with negligible probability of information loss (ii) compressing it into less than $H(X)$ bits implies almost certain loss of information
- This is something you would learn to understand how computers do lossless compression (e.g., RAW and TIFF are lossless, JPG is lossy)

Information Theory (III)

Justification 1: the encoding theorem

- If we have a random variable X with entropy $H(X)$, then (i) it can be compressed into $H(X)$ bits with negligible probability of information loss (ii) compressing it into less than $H(X)$ bits implies almost certain loss of information
- This is something you would learn to understand how computers do lossless compression (e.g., RAW and TIFF are lossless, JPG is lossy)

Justification 2: axioms. see the "Characterization" section on the Wikipedia article about Entropy (information theory).

Information Theory (IV)

Further facts about entropy that will help with intuition later:

- Entropy doesn't depend on the "labels" of the states
- It is zero for a deterministic random variable (under usual continuity convention for $0 \log 0$)
- It is not affected by adding zero probability states
- If we transform a random variable X by applying some function f, entropy can only go down
  - and it goes down only if f coarsens the states

The Rational Inattention Cost: Information Formulation

Proposed in articles by Sims (1998, 2003)

$$K[\psi, \pi] = \sum_{\omega \in \Omega} (H[\pi] - H[\psi(\cdot \mid \omega)])\psi(\omega) = \mathrm{MI}[\psi] \tag{10}$$

- Information cost is *expected reduction in entropy after observing* ω
- which is also the definition of "Mutual Information," another information-theoretic notion

The Rational Inattention Cost: Information Formulation

Proposed in articles by Sims (1998, 2003)

$$K[\psi, \pi] = \sum_{\omega \in \Omega} (H[\pi] - H[\psi(\cdot \mid \omega)])\psi(\omega) = \mathrm{MI}[\psi] \tag{10}$$

- Information cost is *expected reduction in entropy after observing* ω
- which is also the definition of "Mutual Information," another information-theoretic notion
- n.b., there are other cost functionals which are "expected difference in informativeness between prior and posterior": the *posterior separable* class studied by Denti (2022), Caplin, Dean and Leahy (2022), etc.

The Rational Inattention Cost: Information Formulation

Proposed in articles by Sims (1998, 2003)

$$K[\psi, \pi] = \sum_{\omega \in \Omega} (H[\pi] - H[\psi(\cdot \mid \omega)])\psi(\omega) = \mathrm{MI}[\psi] \tag{10}$$

- Information cost is *expected reduction in entropy after observing* ω
- which is also the definition of "Mutual Information," another information-theoretic notion
- n.b., there are other cost functionals which are "expected difference in informativeness between prior and posterior": the *posterior separable* class studied by Denti (2022), Caplin, Dean and Leahy (2022), etc.
- Note that this has a natural extension to cases in which θ is continuous, using *differential entropy* (previous definition with an integral rather than sum). But this does not follow all of the axioms!

The Rational Inattention Cost: Stochastic Choice Formulation

$$K[p,\pi] = \underbrace{\sum_{x,\theta} p(x|\theta)\log p(x|\theta)\pi(\theta)}_{\text{Entropy stochastic choice}} - \underbrace{\sum_{x} p(x)\log p(x)}_{\text{Cross-State Interactions}} \tag{11}$$

The Rational Inattention Cost: Stochastic Choice Formulation

$$K[p,\pi] = \underbrace{\sum_{x,\theta} p(x|\theta)\log p(x|\theta)\pi(\theta)}_{\text{Entropy stochastic choice}} - \underbrace{\sum_{x} p(x)\log p(x)}_{\text{Cross-State Interactions}} \tag{11}$$

- Mutual information model likes random choice, just like our first example of stochastic choice
- But it also likes the distribution of actions (unconditional) to have *low* entropy, or be concentrated on a few points

Result 1: Logit Choice Formula

Theorem (Theorem 1, Matejka and McKay (2015))

Choice probabilities are implicitly defined by

$$\begin{aligned} p(x \mid \theta) &= \frac{p(x)\exp(\lambda^{-1}u(x,\theta))}{\sum_{x' \in X} p(x')\exp(\lambda^{-1}u(x',\theta))} \\ p(x) &= \sum_{\theta} p(x \mid \theta) \end{aligned} \tag{12}$$

- "Logit plus anchoring"
- Defines a fixed-point equation

Result 2: Endogenous Consideration Sets (I)

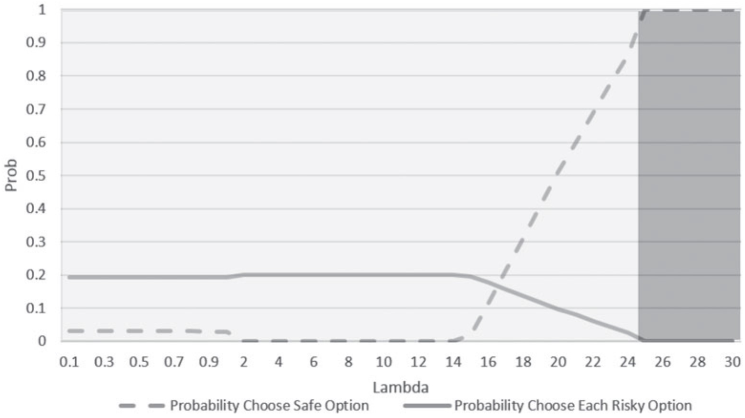
Corollary (Corollary 2, Matejka and Mckay (2015))

For all $x \in X$ such that $p(x) > 0$, we have that:

$$\mathbb{E}\left[\frac{\exp(\lambda^{-1}u(x,\theta))}{\sum_{\tilde{x}\in X} p(\tilde{x})\exp(\lambda^{-1}u(\tilde{x},\theta))}\right] = 1 \tag{13}$$

- Example 2 from Caplin, Dean, and Leahy (ReStud, 2019)
- Safe choice $x = 0$ always give payoff 5.5
- Risky choices $x \in \{1, 2, 3, 4, 5\}$ are all (independent) lotteries that give payoff 10 with probability $1/2$, 0 with probability $1/2$
- Problem is to maximize expected utility minus $\lambda > 0$ times MI cost

Result 2: Endogenous Consideration Sets (II)



- Why? In a certain way, commonly played actions are "self-reinforcing" because marginal costs decrease

Result 3: Discrete Solutions (I)

- Jung, Kim, Matejka, Sims in ReStud (2020) show that mutual information often implies *discrete support of actions* in problems that allow a continuous action space

- That is: even though a continuum of prices (e.g., $p \in [\$0, \$10]$) are possible, a seller facing mutual information constraints may want only to charge a few prices (e.g., \$1 or \$5 or \$7)

- Economically, this is shown most clearly for "tracking problems with bounded support": that is, the agent's objective is $u(x, \theta) = V(x - \theta)$ where V is (uniquely) maximized at 0, and Θ is a compact subset of $\mathbb{R}^N$

- Mathematically, the result is quite interesting as well

Result 4: Endogenous Gaussian-Gaussian Structure

- Finally, the most celebrated result of them all...
- Let $u(x, \theta) = -(x - \theta)^2$ and let $\lambda > 0$ be the scaling on the cost
- Then the optimal action x is normally distributed with variance $\max\{\sigma^2 - 2\lambda, 0\}$ and the optimal signal is Gaussian

Result 4: Endogenous Gaussian-Gaussian Structure

- Finally, the most celebrated result of them all...
- Let $u(x, \theta) = -(x - \theta)^2$ and let $\lambda > 0$ be the scaling on the cost
- Then the optimal action x is normally distributed with variance $\max\{\sigma^2 - 2\lambda, 0\}$ and the optimal signal is Gaussian
- A microfoundation for all the Gaussian-fundamentals, Gaussian-signals, quadratic problems that we solved earlier in the course

Result 4: Endogenous Gaussian-Gaussian Structure

- Finally, the most celebrated result of them all...
- Let $u(x,\theta) = -(x-\theta)^2$ and let $\lambda > 0$ be the scaling on the cost
- Then the optimal action x is normally distributed with variance $\max\{\sigma^2 - 2\lambda, 0\}$ and the optimal signal is Gaussian
- A microfoundation for all the Gaussian-fundamentals, Gaussian-signals, quadratic problems that we solved earlier in the course
- Widely used result but also not robust (as we have seen)

There are Lots of Other Information Costs...

- Pomatto, Strack, and Tamuz, *AER*: "The Cost of Information: The Case of Constant Marginal Costs"
- Hebert and Woodford, *JET*: "Rational Inattention When Decisions Take Time"
- Hebert and Woodford, *AER*: "Neighborhood-Based Information Costs"
- Denti, Marinacci, and Rustichini, *AER*: "Experimental Costs of Information"
- Bloedel and Zhong, WP: "The Cost of Optimally Acquired Information"
- Morris and Strack, WP: "The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs"

Outline

Equilibrium Predictions with Stochastic Choice (I)

- In a paper with Karthik ("Strategic Mistakes," *JET*) we study general aggregative games with state-separable costs
- Continuum of agents care about own actions $x \in \mathcal{X}$, the state $\theta \in \Theta$ and an aggregate of actions of others $X : \Delta(\mathcal{X}) \to \mathbb{R}$. Payoff function: $u(x, X, \theta)$

Equilibrium Predictions with Stochastic Choice (I)

- In a paper with Karthik ("Strategic Mistakes," *JET*) we study general aggregative games with state-separable costs
- Continuum of agents care about own actions $x \in \mathcal{X}$, the state $\theta \in \Theta$ and an aggregate of actions of others $X : \Delta(\mathcal{X}) \to \mathbb{R}$. Payoff function: $u(x, X, \theta)$
- Agents choose a non-parametric *stochastic choice* rule $p : \Theta \to \Delta(\mathcal{X})$, $p(x|\theta)$ PDF of actions $x \in \mathcal{X}$ in state $\theta \in \Theta$

Equilibrium Predictions with Stochastic Choice (I)

- In a paper with Karthik ("Strategic Mistakes," *JET*) we study general aggregative games with state-separable costs
- Continuum of agents care about own actions $x \in \mathcal{X}$, the state $\theta \in \Theta$ and an aggregate of actions of others $X : \Delta(\mathcal{X}) \to \mathbb{R}$. Payoff function: $u(x, X, \theta)$
- Agents choose a non-parametric *stochastic choice* rule $p : \Theta \to \Delta(\mathcal{X})$, $p(x|\theta)$ PDF of actions $x \in \mathcal{X}$ in state $\theta \in \Theta$
- Cost of playing more precise actions mediated by a cost functional ($\phi : \mathbb{R}_+ \to \mathbb{R}$ strictly convex, *e.g.* entropy $\phi(x) = x \log x$)

$$c(P; \hat{X}) = \sum_{\Theta} \int_{\mathcal{X}} \phi(p(x|\theta)) \, \mathrm{d}x \, \lambda(\theta, \hat{X}(\theta))$$

where $\lambda : \Theta \times \mathbb{X} \to \mathbb{R}_+$ is a weighting function and $\hat{X}$ is the conjecture of the equilibrium law of motion

Conditions for Equilibrium Uniqueness: Payoffs and Costs (I)

Assumption (Supermodularity and Sufficient Concavity)

The payoff function u and weighting function λ are such that the following holds for all $x' \geq x, X' \geq X$, and θ:

$$\frac{u(x', X', \theta) - u(x, X', \theta)}{\lambda(X', \theta)} \geq \frac{u(x', X, \theta) - u(x, X, \theta)}{\lambda(X, \theta)}$$

Moreover, for all $\alpha \in \mathbb{R}_+$, $x' \geq x$, X, and θ, the following holds:

$$\frac{u(x' - \alpha, X, \theta) - u(x - \alpha, X, \theta)}{\lambda(X, \theta)} \geq \frac{u(x', X + \alpha, \theta) - u(x, X + \alpha, \theta)}{\lambda(X + \alpha, \theta)}$$

Conditions for Equilibrium Uniqueness: Payoffs and Costs (II)

Lemma

If $u(\cdot, \theta)$ and $\lambda(\cdot, \theta)$ are twice continuously differentiable, then this assumption is equivalent to:

$$0 \leq u_{xX}(x, X, \theta) - u_x(x, X, \theta)\frac{\lambda_X(X, \theta)}{\lambda(X, \theta)} \leq -u_{xx}(x, X, \theta)$$

Conditions for Equilibrium Uniqueness: Aggregation

Assumption (Monotone and Discounted Aggregator)

For all $g, g' \in \Delta(\mathcal{X})$*:*

$$g' \succeq_{FOSD} g \implies X(g') \geq X(g)$$

Moreover, there exists $\beta \in (0,1)$ *such that for any distribution* $g \in \Delta(\mathcal{X})$ *and any* $\alpha \in \mathbb{R}_+$*:*

$$X(\{g(x-\alpha)\}_{x\in\mathcal{X}}) \leq X(\{g(x)\}_{x\in\mathcal{X}}) + \beta\alpha$$

1. Linear aggregators:

$$X(g) = \beta \int_{\mathcal{X}} f(x)g(x)dx$$

   where $\beta \in [0,1)$ and $f : X \to \mathbb{R}$ is such that $f' \in [0,1]$.
2. Quantile aggregators $X(g) = \beta G^{-1}(l)$ where $\beta \in [0,1)$ and $l \in (0,1)$.

Conditions for Equilibrium Uniqueness: Costs

Definition (Quasi-MLRP)

A function $f : \mathbb{R}_+ \to \mathbb{R}$ satisfies quasi-MLRP if for any two distributions $g', g \in \Delta(\mathcal{X})$:

$$\left(f(g'(x')) - f(g'(x)) \geq f(g(x')) - f(g(x)) \; \forall x' \geq x \right) \implies g' \succeq_{FOSD} g$$

Assumption (Quasi-MLRP Kernel)

Costs have a differentiable kernel ϕ such that ϕ' satisfies quasi-MLRP.

Lemma

The entropy kernel $\phi(p) = p \log p$ and the quadratic kernel $\phi(p) = \frac{1}{2}p^2$ satisfy this assumption.

Equilibrium Uniqueness

Theorem (Existence, Uniqueness, and Symmetry)

Under the assumptions above, there exists a unique equilibrium. The unique equilibrium is symmetric.

- Define the operator $T : \mathcal{B} \to \mathcal{B}$ and show it's a contraction:

$$T\hat{X} = X \circ p^*(\hat{X})$$

- Use variational arguments to show that (for any two points in support of optimal stochastic choice rule):

$$u(x', \hat{X}(\theta), \theta) - \lambda(\hat{X}(\theta), \theta)\phi'(p(x'|\theta; \hat{X})) = u(x, \hat{X}(\theta), \theta) - \lambda(\hat{X}(\theta), \theta)\phi'(p(x|\theta; \hat{X}))$$

- Under the assumptions, if $\hat{X}' \geq \hat{X}$ then $p^*(\theta; \hat{X}') \succeq_{FOSD} p^*(\theta; \hat{X})$. Moreover, if $\hat{X}$ and $\hat{X}' = \hat{X} + \alpha$ for $\alpha \in \mathbb{R}_+$, then we have that $p^*_{-\alpha}(\theta; \hat{X}) \succeq_{FOSD} p^*(\theta; \hat{X}')$

Robust Predictions with Stochastic Choice (III)

Theoretical results: find conditions on (u, ϕ, X) such that:

1. Equilibria exist and are unique
2. Equilibrium action distribution $p(x|\theta)$ is FOSD-monotone in θ (and action support is monotone in strong set order), aggregate $X(\theta)$ is monotone in θ
3. Equilibrium action distribution features dispersion or extent of mistakes that is monotone in the state
4. Equilibria are efficient

Technique: contraction-mapping arguments, which are essentially impossible under unrestricted information acquisition

Robust Predictions with Stochastic Choice (III)

Theoretical results: find conditions on (u, ϕ, X) such that:

1. Equilibria exist and are unique
2. Equilibrium action distribution $p(x|\theta)$ is FOSD-monotone in θ (and action support is monotone in strong set order), aggregate $X(\theta)$ is monotone in θ
3. Equilibrium action distribution features dispersion or extent of mistakes that is monotone in the state
4. Equilibria are efficient

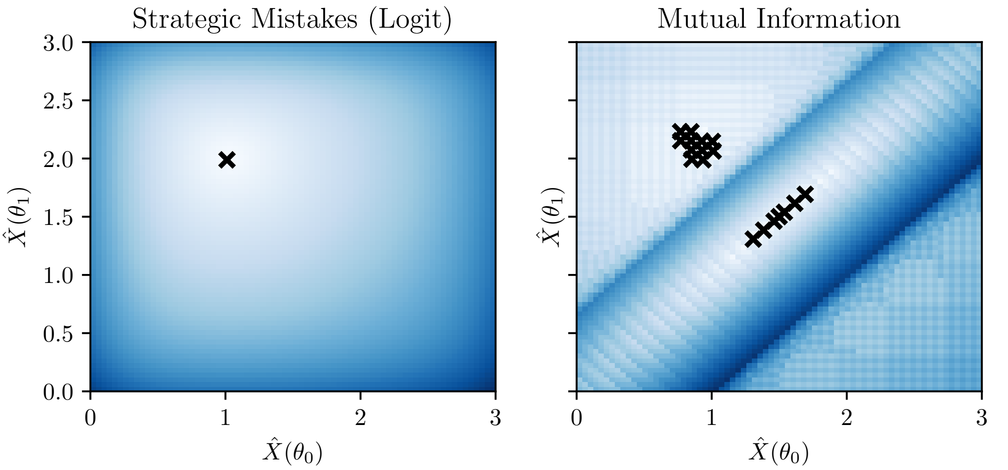
Technique: contraction-mapping arguments, which are essentially impossible under unrestricted information acquisition

What goes wrong with mutual information, in particular?

Robust Predictions with Stochastic Choice (IV)

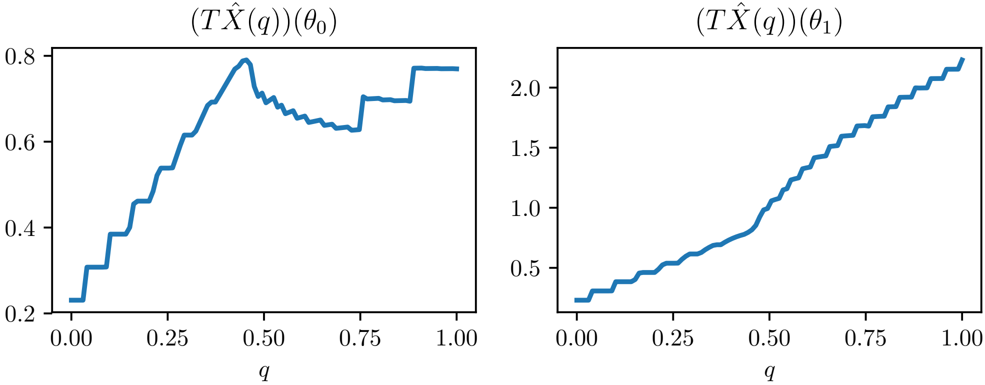
- Payoffs
$$u(x, X, \theta) = -(x - (0.15\theta + 0.85X))^2 \tag{14}$$
- Aggregator $X = \int_0^1 x_i \,\mathrm{d}i$
- State $\Theta = \{1, 2\}$, with $50/50$ prior
- Action space is 40-piont grid between 0 and 3
- Costs: either logit stochastic choice or mutual information

Robust Predictions with Stochastic Choice (V)



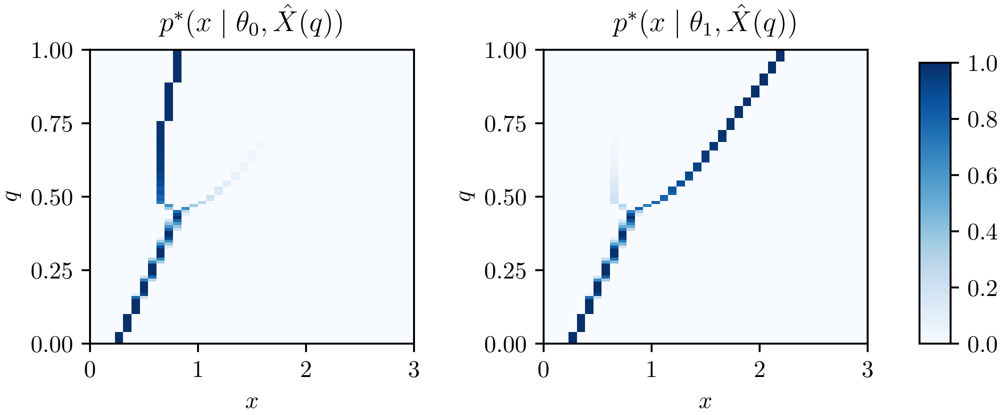
Robust Predictions with Stochastic Choice (VI)

Parameterize, relative to some equilibrium: $\hat{X}(q) = (q \cdot X^*_{MI}(\theta_0), q \cdot X^*_{MI}(\theta_1))$



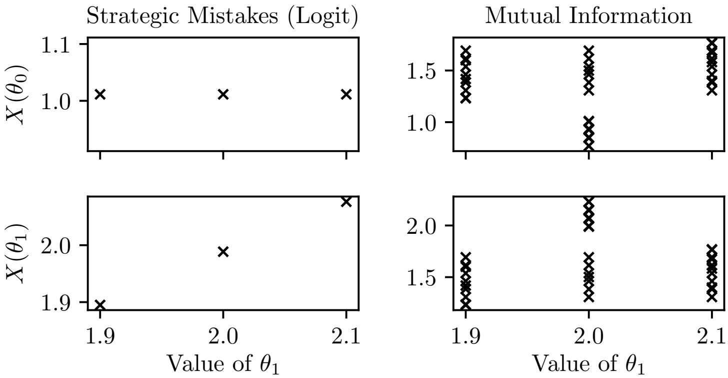
Robust Predictions with Stochastic Choice (VI)

Parameterize, relative to some equilibrium: $\hat{X}(q) = (q \cdot X^*_{MI}(\theta_0), q \cdot X^*_{MI}(\theta_1))$



Robust Predictions with Stochastic Choice (VI)

Parameterize, relative to some equilibrium: $\hat{X}(q) = (q \cdot X^*_{MI}(\theta_0), q \cdot X^*_{MI}(\theta_1))$



Outline

Takeaways for Modelling Bounded Rationality

- No simple answer of what method is best (depends upon what you want to capture)
- Different models have different pros and cons
- My stylized recommendation:
    1. If an informational interpretation is not important, use a state-separable stochastic choice formulation
    2. If an informational interpretation is important, use a cost function that allows for the important properties of the problem that you wish to study
- Beware of blindly using mutual information: it has poor properties in games (general equilibrium) outside of the very special Gaussian-Gaussian case