Attention and Perception

Lecture 6: Rational Inattention (Shannon Costs)[1]

Daniel Martin

danielmartin@ucsb.edu

[1] Thanks to Mark Dean, whose slides these are based on

Costs of Information

- Previously we introduced a general model of rational inattention
  - Characterized by NIAS and NIAC
  - Made only limited assumptions about the cost of attention
- But what if we want to make stronger assumptions about costs?
- Several proposals:
  - Costly sequential search (e.g. McCall 1970)
  - Cost to reduce variance of normal signal (e.g. Verrecchia 1982)
  - Shannon mutual information costs (e.g. Matějka & McKay 2015)

Rational Inattention and Shannon Information Costs

- We now consider the option based on Shannon entropy (Shannon 1948)
  - Most common cost function used in the rational inattention literature
  - Consider this like the Cobb-Douglas form of demand
- Long history of research in information theory
  - Quite a lot is known about how Shannon entropy behaves
  - Cover & Thomas (2006) is a great resource

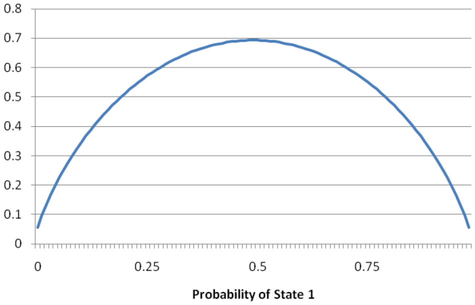
Entropy

- Entropy measures how uncertain you are before learning which outcome actually happened
- It captures how much missing information remains in a probability distribution
- If one outcome is almost certain, entropy is low
- If probability is spread evenly across many outcomes, entropy is high

Shannon Entropy

- For a random variable X that takes the value x_i with probability $p_i = p(x_i)$

$$\begin{aligned} H(X) &= E[-\ln(p_i)] \\ &= -\sum_i p_i \ln(p_i) \end{aligned}$$



Entropy and Information Costs

- Related to the notion of entropy is the notion of mutual information
- It is a measure of how much information one variable tells you about another
- Shannon mutual information takes the following form:

$$I(X, Y) = \sum_x \sum_y p(x, y) \ln \frac{p(x, y)}{p(x)p(y)}$$

- Note that $I(X, Y) = 0$ if X and Y are independent

Entropy and Information Costs

- Note also that Shannon mutual information can be rewritten in the following way

$$
\begin{aligned}
I(X,Y) &= \sum_x \sum_y p(x,y) \ln \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y) \ln \frac{p(x|y)}{p(x)} \\
&= \sum_y \sum_x p(x,y) \ln p(x|y) - \sum_x \sum_y p(x,y) \ln p(x) \\
&= \sum_y p(y) \sum_x p(x|y) \ln p(x|y) - \sum_x p(x) \ln p(x) \\
&= H(X) - E(H(X|Y))
\end{aligned}
$$

- Difference between entropy of X and the expected entropy of X once Y is known

Mutual Information and Information Costs

- Mutual information between states and posteriors is used to model information costs
  - i.e., cost is proportional to how much uncertainty about the state is removed when the agent learns which posterior they have
- Two approaches in economics to putting this into the optimization problem
  1. Place a bound on mutual information (Sims 2003)
     - All π that have mutual information above some limit are infinitely costly
  2. Impose linear costs on mutual information (Matějka & McKay 2015, Caplin, Dean & Leahy 2022)

$$\begin{aligned} K(\mu,\pi) &= \kappa\left(H(\mu) - E_\pi\left[H(\gamma)\right]\right) \\ &= \kappa\left(\begin{array}{c} \sum_{\gamma\in\Gamma}\pi(\gamma)\sum_{\omega\in\Omega}\gamma(\omega)\ln\gamma(\omega) \\ -\sum_{\omega\in\Omega}\mu(\omega)\ln\mu(\omega) \end{array}\right) \end{aligned}$$

Solving Rational Inattention Models

- Solving this model can be difficult analytically
  - Though easier than many other models (logs useful in FOCs!)
- General approach: focus on joint distribution of actions and states
  - As if the agent chooses this joint distribution
  - Can do this because optimal strategy will always be 'well behaved'
  - Each action taken in at most one signal
- We will talk about analytical approaches
  - Also algorithmic approaches, such as the Blahut-Arimoto algorithm (Blahut 1972, Arimoto 1972)
  - See Cover & Thomas (2006, p. 191)

Solving Rational Inattention Models

- $\mathcal{P}$ is the set of all state-contingent stochastic choice functions for some state space Ω and set of acts A
  - Equivalent to SDSC data of Caplin & Martin (2015)
- Remember $P(a|\omega)$ is the probability of choosing a in state ω
- Remember that, for $P \in \mathcal{P}$, the mutual information between choices a and objective state ω is given by

$$I(A, \Omega) = H(A) - H(A|\Omega)$$

Solving Rational Inattention Models

- The agent's decision problem is to choose $P \in \mathcal{P}$ to maximize

$$\sum_{a\in A}\int_{\omega} u(a,\omega)P(a|\omega)\mu(d\omega)$$
$$-\kappa\left[\sum_{a\in A}\int_{\omega} P(a|\omega)\ln P(a|\omega)\mu(d\omega) - \sum_{a\in A} P(a)\ln P(a)\right]$$

- Subject to:

$$\textstyle\sum_{a\in A} P(a|\omega) = 1$$

- Note: $P(a)$ is the unconditional probability of choosing a
- Note there is another constraint which we will ignore for now

$$P(a|\omega) \geq 0 \;\forall\; a,\omega$$

The Lagrangian Function

$$\sum_{a\in A}\int_{\omega} u(a,\omega)P(a|\omega)\mu(d\omega)$$
$$-\kappa\left[\sum_{a\in A}\int_{\omega} P(a|\omega)\ln P(a|\omega)\mu(d\omega)-\sum_{a\in A}P(a)\ln P(a)\right]$$
$$-\int_{\omega}\rho(\omega)\left[\sum_{a\in A}P(a|\omega)-1\right]\mu(d\omega)$$

- $\rho(\omega)$ Lagrangian multiplier on the condition that $\sum_{a\in A}P(a|\omega)=1$
- FOC with respect to $P(a|\omega)$ (assuming >0)

$$u(a,\omega)-\kappa[\ln P(a|\omega)+1-\ln P(a)-1]-\rho(\omega)=0$$

Solution

- FOC with respect to $P(a|\omega)$ (assuming > 0)

$$u(a,\omega) - \kappa[\ln P(a|\omega) + 1 - \ln P(a) - 1] - \rho(\omega) = 0$$

- Which gives

$$P(a|\omega) = P(a) \exp^{\frac{u(a,\omega)-\rho(\omega)}{\kappa}}$$

- Plug this into

$$\begin{aligned} \sum_{a' \in A} P(a'|\omega) &= 1 \\ \Rightarrow \quad \exp^{\frac{\rho(\omega)}{\kappa}} &= \sum_{a' \in A} P(a') \exp^{\frac{u(a',\omega)}{\kappa}} \end{aligned}$$

- Which in turn gives...

$$P(a|\omega) = \frac{P(a)\exp^{\frac{u(a,\omega)}{\kappa}}}{\sum_{c\in A} P(c)\exp^{\frac{u(c,\omega)}{\kappa}}}$$

- Similar in form to logistic random choice
  - Identified by Matějka & McKay (2015) (MM henceforth)
- If alternatives are ex ante identical, this *is* logistic choice
- Otherwise choice probabilities are 'warped' by $P(a)$, which contains information on the prior value of each option
  - Important: note that $P(a)$ is endogenous, **not** a parameter
- Recent: Brown & Jeon (2024) provide an analytically tractable version with EV1 'prior' and preference heterogeneity

Comments

- The MM conditions ignore the constraint

$$P(a|\omega) \geq 0 \ \forall \ a, \omega$$

- Need to know which acts will be chosen with positive probability
- Typically there will be many acts not chosen at the optimum (Jung, Kim, Matějka & Sims 2019)
- So there will be many solutions to the necessary conditions
- Sufficient conditions provided by Caplin, Dean & Leahy (2019)
  - Their application is RI-driven consideration sets
  - Sufficient condition identifies acts that are never chosen (considered)

Necessary and Sufficient Conditions

- Let $z(a,\omega)$ be 'normalized utilities'

$$z(a,\omega) = \exp\left\{\frac{u(a,\omega)}{\kappa}\right\}$$

- Note that the MM conditions are

$$P(a|\omega) = \frac{P(a)z(a,\omega)}{\sum_{c\in A} P(c)z(c,\omega)}$$

Necessary and Sufficient Conditions

Theorem

P is consistent with rational inattention with mutual information costs ***if and only if***

$$\begin{aligned}\sum_{\omega\in\Omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] &\leq 1 \quad \text{for all } a\in A\\ \sum_{\omega\in\Omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] &= 1 \quad \text{for all } a\in A \text{ such that } P(a)>0.\end{aligned}$$

and

$$P(a|\omega)=\frac{P(a)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}$$

Posterior-Based Approach

- Can also take the posterior-based approach used in Caplin & Martin (2015) and Caplin & Dean (2015)
- Assume we are choosing π, a mapping from states to (simple) distributions over posterior beliefs
- General class: uniformly posterior-separable costs based on generic T (Caplin, Dean & Leahy 2022, Bloedel & Zhong 2025)

$$K(\mu, \pi) = T(\mu) - E_{\pi}\left[T(\gamma)\right]$$

- Special class of UPS model: "Shannon model" with costs based on Shannon entropy H

$$K(\mu, \pi) = \kappa\left(H(\mu) - E_{\pi}\left[H(\gamma)\right]\right)$$

- What are the necessary and sufficient conditions for the Shannon model?

Necessary and Sufficient Conditions (Posterior-Based Approach)

$$\boxed{\gamma^a\text{=posterior beliefs for action } a,\ B_i\text{=set of chosen actions, } z(a,\omega) = e^{u(a,\omega)/\kappa}}$$

Theorem

Given decision problem $i \in D$, a set of posteriors is rationally inattentive iff:

1. ***Invariant Likelihood Ratio (ILR) Equations for Chosen Acts**: given $a, b \in B_i$,*

$$\frac{\gamma^a(\omega)}{z(a,\omega)} = \frac{\gamma^b(\omega)}{z(b,\omega)} \text{ for all } \omega \in \Omega$$

2. ***Likelihood Ratio Inequalities for Unchosen Acts**: given $a \in B_i$ and $b \in A_i \setminus B_i$,*

$$\sum_{\omega\in\Omega} \left[\frac{\gamma^a(\omega)}{z(a,\omega)}\right] z(b,\omega) \leq 1$$

More on ILR

- Here is Invariant Likelihood Ratio (ILR) without the z notation:

$$\frac{\gamma^a(\omega)}{e^{u(a,\omega)/\kappa}} = \frac{\gamma^b(\omega)}{e^{u(b,\omega)/\kappa}}$$

- And here are some equivalent ways to write ILR:

$$\frac{\gamma^a(\omega)}{\gamma^b(\omega)} = e^{\frac{u(a,\omega)-u(b,\omega)}{\kappa}}$$

$$\ln\left(\frac{\gamma^a(\omega)}{\gamma^b(\omega)}\right) = \frac{u(a,\omega) - u(b,\omega)}{\kappa}$$

Summary

- Introduced Shannon mutual information as a potential cost function
  - Popular in the literature
  - Cobb-Douglas analog
- Introduced some analytical tools to help solve the Shannon model
  - Necessary conditions (Matějka & McKay 2015)
  - Necessary + sufficient conditions (Caplin, Dean & Leahy 2019)
    - Can be written two ways!

References I

Arimoto, S. (1972), 'An algorithm for computing the capacity of arbitrary discrete memoryless channels', *IEEE Transactions on Information Theory* **18**(1), 14–20.

Blahut, R. E. (1972), 'Computation of channel capacity and rate-distortion functions', *IEEE Transactions on Information Theory* **18**(4), 460–473.

Bloedel, A. W. & Zhong, W. (2025), The cost of optimally acquired information. arXiv:2511.05466, working paper.

Brown, Z. Y. & Jeon, J. (2024), 'Endogenous information and simplifying insurance choice', *Econometrica* **92**(3), 881–911.

Caplin, A. & Dean, M. (2015), 'Revealed preference, rational inattention, and costly information acquisition', *American Economic Review* **105**(7), 2183–2203.

Caplin, A., Dean, M. & Leahy, J. (2019), 'Rational inattention, optimal consideration sets, and stochastic choice', *The Review of Economic Studies* **86**(3), 1061–1094.

Caplin, A., Dean, M. & Leahy, J. (2022), 'Rationally inattentive behavior: Characterizing and generalizing Shannon entropy', *Journal of Political Economy* **130**(6), 1676–1715.

Caplin, A. & Martin, D. (2015), 'A testable theory of imperfect perception', *The Economic Journal* **125**(582), 184–202.

References II

Cover, T. M. & Thomas, J. A. (2006), *Elements of Information Theory*, 2nd edn, Wiley-Interscience, Hoboken, NJ.

Jung, J., Kim, J. H., Matějka, F. & Sims, C. A. (2019), 'Discrete actions in information-constrained decision problems', *The Review of Economic Studies* **86**(6), 2643–2667.

Matějka, F. & McKay, A. (2015), 'Rational inattention to discrete choices: A new foundation for the multinomial logit model', *American Economic Review* **105**(1), 272–298.

McCall, J. J. (1970), 'Economics of information and job search', *The Quarterly Journal of Economics* **84**(1), 113–126.

Shannon, C. E. (1948), 'A mathematical theory of communication', *The Bell System Technical Journal* **27**(3–4), 379–423, 623–656.

Sims, C. A. (2003), 'Implications of rational inattention', *Journal of Monetary Economics* **50**(3), 665–690.

Verrecchia, R. E. (1982), 'Information acquisition in a noisy rational expectations economy', *Econometrica* **50**(6), 1415–1430.