Rational Inattention Lecture 2

Mark Dean

Behavioral Economics G6943
Fall 2022

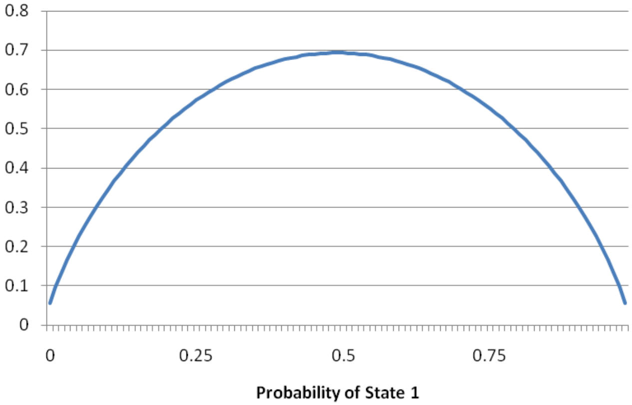
Rational Inattention and Shannon Information Costs

- We have so far considered what we can say when we are agnostic about information costs
- We now move consider behavior under a specific assumed cost for information
- Based on the concept of Shannon Entropy
  - Extremely popular in the applied literature
  - Consider this the 'Cobb Douglas' case to last week's 'revealed preference' treatment
- Long history of research in information theory
  - Quite a lot is known about how these costs behave
  - Cover and Thomas is a great resource

- Shannon Entropy is a measure of how much 'missing information' there is in a probability distribution
- In other words - how much we do not know, or how much we would learn from resolving the uncertainty
- For a random variable X that takes the value x_i with probability $p(x_i)$ for $i = 1...n$, defined as

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

Shannon Entropy



- Can think of it as how much we learn from result of experiment

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
  - $H(X) = H(p)$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
    - $\max_{p \in \Delta^M} H(p) = H\left(\left\{\frac{1}{M}, \frac{1}{M}, ..., \frac{1}{M}\right\}\right)$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
    - $H(\{p_1 .... p_M\}) = H(\{p_1 .... p_M, 0\})$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
- Additive
  - $H(X, Y) = H(X) + \sum_x p(x)H(Y|x)$
  - How much you learn from observing X, plus how much you additionally learn from observing Y
  - Implies that the entropy of two independent variables is just $H(X) + H(Y)$
  - 'Constant returns to scale' assumption

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
- Additive
- Then Entropy must be of the form (Khinchin 1957)

$$H(X) = -\sum_i p(x_i)\ln(p_i)$$

- Note, other entropies are available! e.g. Tsallis

$$\frac{k}{q-1}(1 - \sum_i p(x_i)^q)$$

Entropy and Information Costs

- Related to the notion of entropy is the notion of Mutual Information

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

- Measure of how much information one variable tells you about another
- Note that $I(X, Y) = 0$ if X and Y are independent

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

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

- 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 signals often used to model information constraints
- Sims [2003] focused on a hard constraint on the amount of entropy a DM can use
- We will start by focussing on the case of **costs that are linear in mutual information**

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

- For convenience use γ to refer to the posterior beliefs generated by signal γ

Mutual Information and Information Costs

- Can be justified by information theory
  - Mutual information is related to the rate of information flow
- Say you are going to observe n repetitions of the state Ω (let ω^n be a typical element)
- You are allowed to send a message consisting of nR bits (R is the rate)
- Decoded in order to generate n repetitions of the signal space Γ (let γ^n be a typical element)
- Define $d(\omega, \gamma)$ be the loss associated with receiving signal γ in state ω, and $\hat{d}(\omega^n, \gamma^n) = \frac{1}{n} \sum d(\omega_i^n, \gamma_i^n)$

Mutual Information and Information Costs

- Rate Distortion Theorem: Let $R(D)$ be the minimal rate needed to generate loss D as $n \to \infty$, then

$$R(D) = \min_{\pi \in \Pi} I(\Omega, \Gamma) \text{ s.t. } \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \leq D$$

- Implies (assuming strict monotonicity)

$$\min \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \text{ s.t. } I(\Omega, \Gamma) \leq R(D)$$

- is equivalent to

$$\min \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \text{ s.t. } R \leq R(D)$$

- See Cover and Thomas Chapter 10.

- Key feature: Entropy is strictly *concave*
- So negative of entropy is strictly convex
- Say we choose a signal structure with two posteriors γ and γ'
- It must be that

$$P(\gamma)\gamma + P(\gamma')\gamma' = \mu$$

- so

$$\begin{aligned} P(\gamma)H(\gamma) + P(\gamma')H(\gamma') &< H(P(\gamma)\gamma + p(\gamma')\gamma') \\ &= H(\mu) \end{aligned}$$

- So the cost of ‘learning something’ is always positive

Solving Rational Inattention Models

- Solving the Shannon model can be difficult analytically
  - Though easier than many other models
- General approach - ignore choice of information structure, instead focus on joint distribution of choice variable and state
  - i.e. choose state dependent stochastic choice directly
  - Can do this because optimal strategy will always be 'well behaved'
  - Each action taken in at most one state
- Example (Matejka and McKay 2015) - continuous state space, finite action space
- We will talk about analytical approaches
  - Alternative, algorithmic approaches
  - e.g. Blahut-Arimotio algorithm
  - See Cover and Thomas (page 191)

- $\mathcal{P}$ set of all state contingent stochastic choice functions for some state space Ω and set of acts A
- 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

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

- Decision problem of agent is to choose $P \in \mathcal{P}$ to maximize

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

- Subject to

$$\sum_{a \in A} P(a|\omega) = 1 \text{ Almost surely}$$

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

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

The Lagrangian Function

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

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

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

- Note that this is a convex problem

Solution

- FOC WRT $P(a|\omega)$ (assuming >0)

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

- Which gives

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

- Plug this into

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

- Which in turn gives...

Comments

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

- Similar in form to logistic random choice
- 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
- As costs go to zero, deterministically pick best option in that state
- As costs go to infinity, deterministically pick the best option ex ante

Comments

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

- Importantly, not the same as logistic RUM
  - This has to be true - we know this model violates monotonicity
- There are papers that play up this link
  - It *is* true that there is an equivalence result for the two models **if we look at a single choice problem**
  - But the models make very different predictions across choice problems

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 et al. 2019)
- There will be many solutions to the necessary conditions
- Ideally, would like necessary and sufficient conditions

Necessary and Sufficient Conditions

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

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

- 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**

$$\sum_{\omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] \leq 1 \text{ all } a\in A$$

$$\sum_{\omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] = 1 \text{ all } a \text{ s.t. } P(a)>0$$

and

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

1. Identify correct **unconditional** choice probabilities
   - Equality condition for chosen actions
   - Check inequality condition for unchosen actions
2. Read off **conditional** choice probabilities using MM conditions

Example: Finding the Good Act

- Choose from a set of goods $A = \{a_1, ..., a_N\}$
- Only one of these goods is of high quality
  - u_h utility of the high quality good
  - u_l utility of the low quality good
  - μ_i prior probability that good i is the high quality good
  - WLOG assume $\mu_1 \geq \mu_2 .... \geq \mu_N$
- Common set up in many psychology experiments

Solution

- Cutoff strategy in prior probabilities: Exists c such that
  - $\mu_i > c \Rightarrow i$ chosen with positive probability
  - $\mu_i < c \Rightarrow i$ never chosen and nothing is learned about their quality
- Endogenously form a 'consideration set'
- Let $\delta = \frac{\exp(\frac{u_h}{\lambda})}{\exp(\frac{u_l}{\lambda})} - 1$: 'additional' utility from high act
- Search the best K alternatives, where K solves

$$\mu_K > \frac{\sum_{k=1}^{K} \mu_k}{K + \delta} \geq \mu_{K+1}.$$

Consideration Set Formation

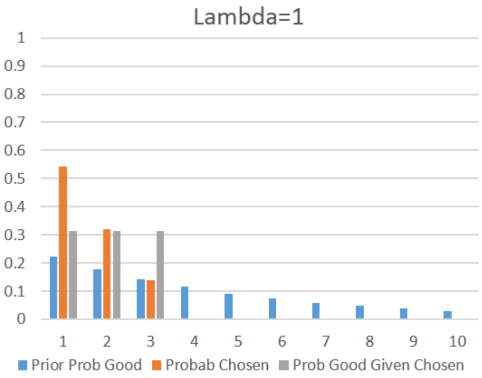
- Can use equality constraints to solve for unconditional choice probabilities

$$P(a_i) = \frac{\mu(\omega_i)(K+\delta) - \sum_{k=1}^{K} \mu(\omega_k)}{\delta \sum_{k=1}^{K} \mu(\omega_k)}$$

- MM conditions to solve for conditional choice probabilities

$$P(b|b = u_h) = \frac{P(b)\delta}{\sum_{c \in A} P(c)}$$

Choice Probabilities - Example

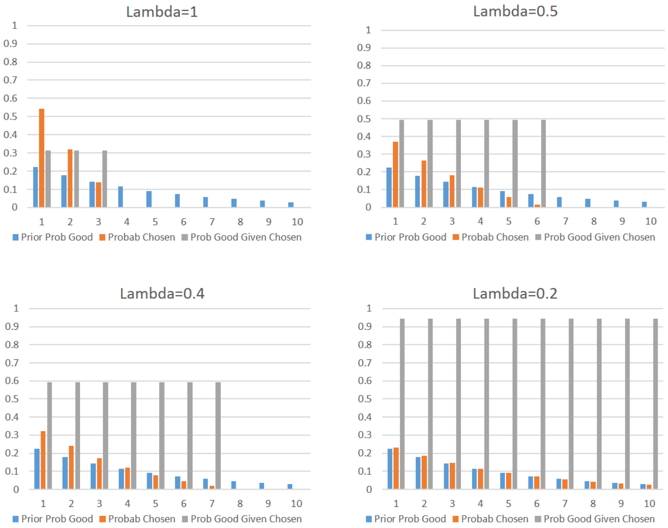


- Exponential priors
- $u_h = 1,\ u_l = 0$

Features of the Solution

- ‘Consideration set’ of alternatives chosen with positive probability
- Mistakes even amongst alternatives in the consideration sets
- Ex ante probability of alternative being good conditional on being chosen is same for all alternatives

Choice Probabilities - Example



Importance of Sufficient Conditions

- The MM necessary conditions could be solved for many possible ‘consideration sets’
  - Choosing any option with probability 1 will solve the necessary conditions
  - For any set C with worst alternative $\mu_{\bar{C}}$ there is a solution to the necessary conditions if

$$\frac{\mu_{\bar{C}}}{\sum_{k \in C} \mu_k} > \frac{1}{|C| + \delta}.$$

- Do no reference unchosen actions
- Do not determine whether higher utility could be obtained with a different consideration sets
- This is the advantage of the sufficient conditions

The Linear Quadratic Gaussian Case

- One case in which this problem becomes more tractable is if the input and output signal are both normal
- The entropy of a normal variable $X \sim N(\mu, \sigma_x^2)$ is given by

$$H(X) = \frac{1}{2} \ln(2\pi e \sigma_x^2)$$

- If Y and X are both normal, then

$$E(H(Y|X)) = \int_x f(x) \int_y f(y|x) \ln f(y|x) d(y) d(x)$$

- As $y|x$ is distributed normally with variance $(1-\rho^2)\sigma_y^2$, this becomes

$$\begin{aligned} E(H(Y|X)) &= \int_x f(x) \frac{1}{2} \ln(2\pi e \sigma_{y|x}^2) d(x) \\ &= \frac{1}{2} \ln(2\pi e (1-\rho^2)\sigma_y^2) \end{aligned}$$

where ρ is the correlation coefficient between X and Y

The Linear Quadratic Gaussian Case

- As mutual information is given by

$$\begin{aligned} & H(Y) - E(H(Y|X)) \\ = \quad & \frac{1}{2}\ln(2\pi e\sigma_y^2) - \frac{1}{2}\ln(2\pi e(1-\rho^2)\sigma_y^2) \end{aligned}$$

- In this case, the mutual information is given by

$$-\frac{1}{2}\ln(1-\rho^2)$$

- So information costs depend only on the covariance of the two signals!
- It turns out that joint normality is optimal if the utility function is quadratic in the relationship between the objective and subjective state
  - Choice of variance on some normally distributed error term
- However, note that some papers *assume* normality (this is bad)

The Linear Quadratic Gaussian Case

- In fact, the LQG case may be our best hope of a workhorse rational inattention model that can be applied to a wide range of problems
  - Because it is so simple to solve
- If there are a vector of states and a vector of actions this framework can be used to approximate a number of situations
  - Tracking problems (e.g. Sims [2003], Fulton [2018])
  - Pricing (e.g. Maćkowiak and Wiederholt [2009], Paciello and Wiederholt [2014])
  - Consumption with many sources of income and many goods (e.g. Koszegi and Matejka 2020)
  - Portfolio selection (e.g. Van Nieuwerburg and Veldkamp [2009], Mondria [2010])
- Some of these paper assume that information has to be gathered on each shock separately
  - Either for analytical tractability or realism

The Linear Quadratic Gaussian Case

- Recent work has provided analytic solutions to the multi state/multi action problem
  - Even when there is prior correlation between states.
- One way to characterize solution [Fulton 2018]
  - DM recombines states α into a set of 'canonical signals'

    $$y_c = S\alpha + \varepsilon$$

    Where S is a matrix derived from the prior covariance matrix and payoff matrix
  - The optimal ε will be distributed normally with the covariance matrix being diagonal.
    - Transforms the original problem into n independent problems
  - The variance of the noise on each canonical shock is decided by a 'water filling' algorithm
    - Some shocks will have no attention paid to them, the others will have attention paid to equalize cost and benefits

- For further information see
  - Fulton, Chad. "Choosing what to pay attention to." Theoretical Economics 17.1 (2022): 153-184.
  - Miao, Jianjun, Jieran Wu, and Eric R. Young. "Multivariate rational inattention." Econometrica 90.2 (2022): 907-945.
  - Dewan, A “Costly Multidimensional Information”, Working paper [2019]
  - Kőszegi, Botond, and Filip Matějka. "Choice simplification: A theory of mental budgeting and naive diversification." The Quarterly Journal of Economics 135.2 (2020): 1153-1207.
- Or ask our very own Hassan Afrouzi!

- There is another way to approach this problem which possibly gives more insight
- Assume we are choosing Q, a (simple) distribution over posterior beliefs, with $Q(\gamma)$ the probability of belief γ
- We can also work with a generalized cost function

$$\sum_{\Gamma} Q(\gamma) T(\gamma) - T(\mu)$$

  where T is some strictly convex function
- For example, we could replace Shannon entropy with other types of entropy.

- One way to gain insight into what is going on is to rewrite the objective function

$$
\begin{aligned}
& \sum_{\Gamma} Q(\gamma)\left[\max_{a\in A}\sum_{\Omega}\gamma(\omega)u(a,\omega)\right]-\left[\sum_{\Gamma} Q(\gamma)T(\gamma)-T(\mu)\right] \\
= & \sum_{\Gamma} Q(\gamma)\left[\max_{a\in A}\sum_{\Omega}\gamma(\omega)u(a,\omega)-T(\gamma)\right]+T(\mu) \\
= & \sum_{\Gamma} Q(\gamma)\max_{a\in A} N_a(\gamma)
\end{aligned}
$$

- Each γ and a has a net utility associated with it

$$N_a(\gamma)=\sum_{\Omega}\gamma(\omega)u(a,\omega)-[T(\gamma)-T(\mu)]$$

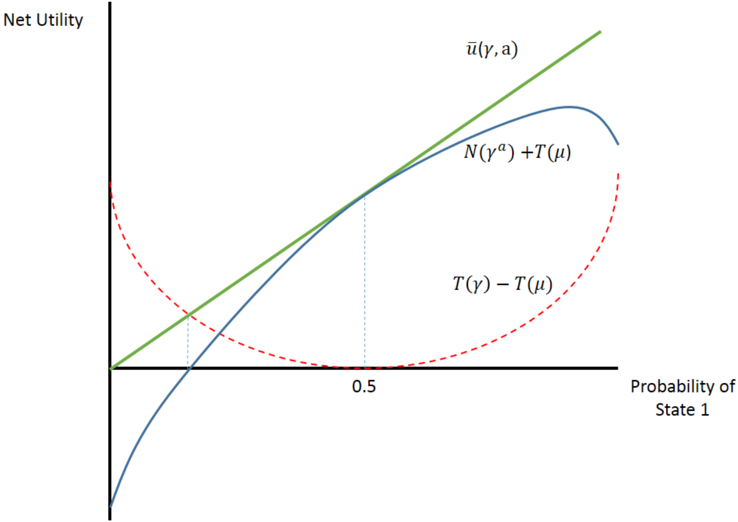
- Aim is to pick distribution of posteriors which maximizes the expected value of net utilities subject to

$$\sum_{\gamma\in\Gamma(\pi)} Q(\gamma)\gamma=\mu$$

- Consider a simple case with two states and two acts

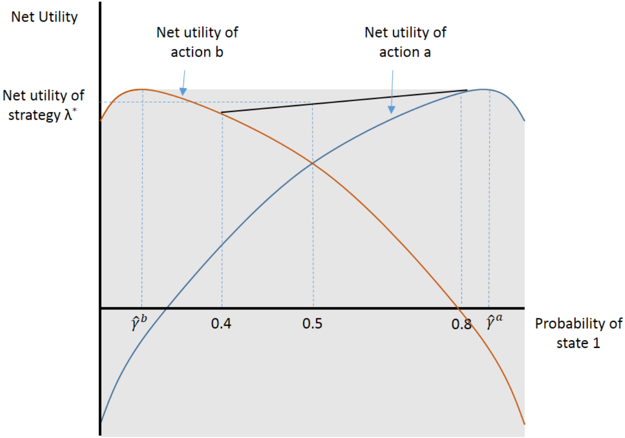
Action	**Payoff in state 1**	**Payoff in state 2**
a	10	0
b	0	10

Net Utility



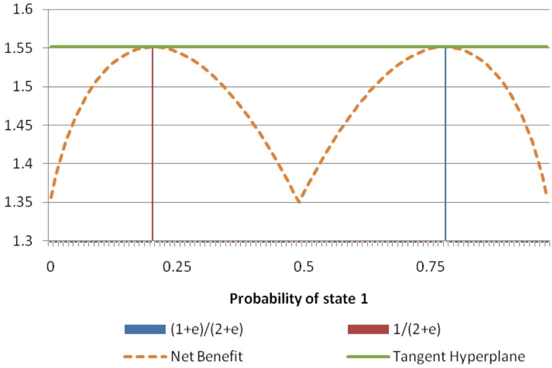
Net Utility
$\bar{u}(\gamma, \mathrm{a})$
$N(\gamma^a) + T(\mu)$
$T(\gamma) - T(\mu)$
0.5
Probability of State 1

Optimal Strategy



- What to find the posteriors which support the highest chord above the prior
- The solution for every possible prior defined by the lower epigraph of the concavified net utility function

Finding the Optimal Strategy



- Optimal posteriors identified by hyperplane that supports the set of feasible net utilities.

Theorem

Given decision problem $(\mu, A) \in \Gamma \times \mathcal{F}$ a set of posteriors are rationally inattentive if and only if:

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

$$\frac{\gamma^a(\omega)}{z(a(\omega))} = \frac{\gamma^b(\omega)}{z(b(\omega))}$$

2. **Likelihood Ratio Inequalities for Unchosen Acts**: *given act a chosen with positive probability and $b \in A$,*

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

Behavioral Properties

- We have necessary and sufficient conditions to characterize the Shannon model
- But these do not necessarily help us understand the behaviors that it predicts
- Also results apply only to the Shannon Model
- Might be helpful to have a more 'behavioral' characterization
  - See Caplin, Dean and Leahy [2022]
- Define two additional classes of model
  - Posterior Separable

$$\sum_{\Gamma} Q(\gamma)\, T_{\mu}(\gamma) - T_{\mu}(\mu)$$

  - Uniformly Posterior Separable

$$\sum_{\Gamma} Q(\gamma)\, T(\gamma) - T(\mu)$$

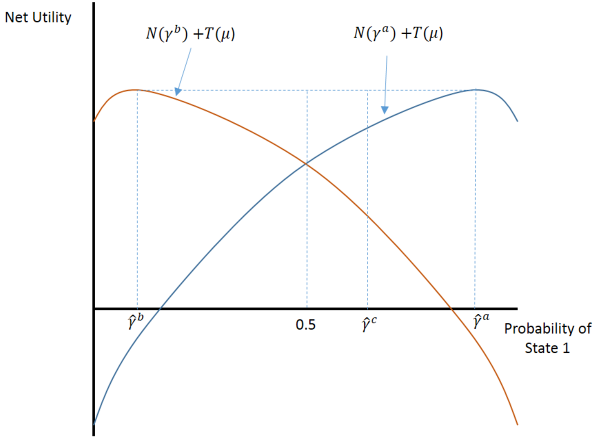
- UPS class (in particular) has proved popular in the literature
  - Can fix some of the behavioral problems of Shannon
  - Still maintains a lot of tractability
- Nice features
  - Plays well with Bayesian Persuation (e.g. Matyskova [2018])
  - Consistent with optimal dynamic information acquisition (e.g. Morris and Strack [2017], Hebert and Woodford [2019])
  - Useful for dynamic programming (e.g. Miao and Xing [2020]

Posterior Separability

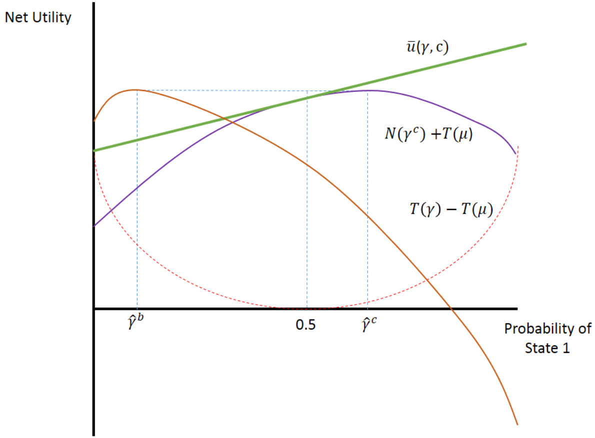
- Turns out that we can characterize using three behavioral axioms
  - Plus some technical ones that we won't bother with

1. Separability
2. Locally Invariant Posteriors
3. Invariance Under Compression

Separability



Separability



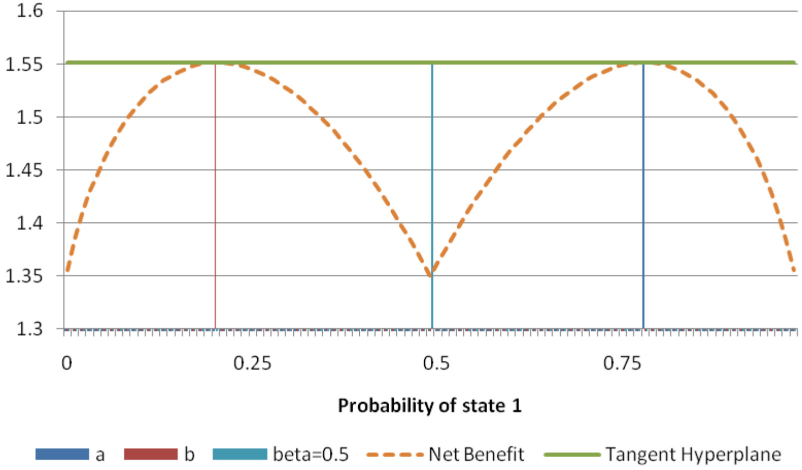
Separability

- Separability states you can always do this
    - For any set of chosen acts and associated posteriors
    - Can switch out one posterior and replace it with another posterior
    - Changing only the associated act.
- This is a property of the Posterior Separable model

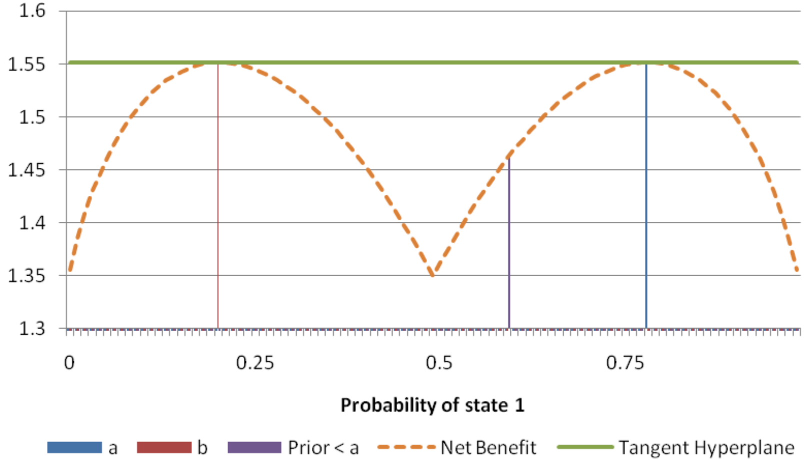
- Example: 2 states, 2 actions

Action	**Payoff in state 1**	**Payoff in state 2**
$\mathbf{f}^1$	x	0
$\mathbf{f}^2$	0	x

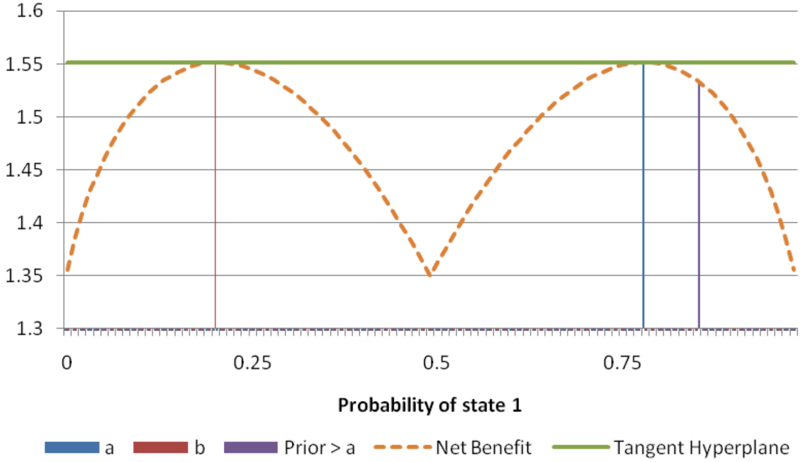
Behavior at 0.5 Prior



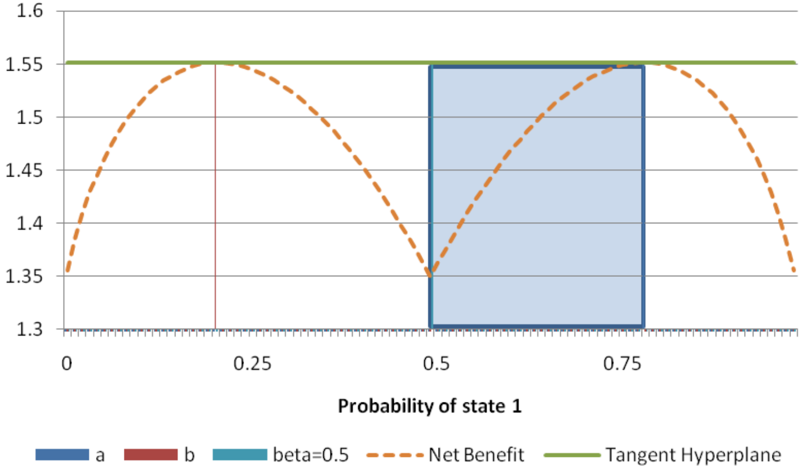
Behavior for prior$<$a



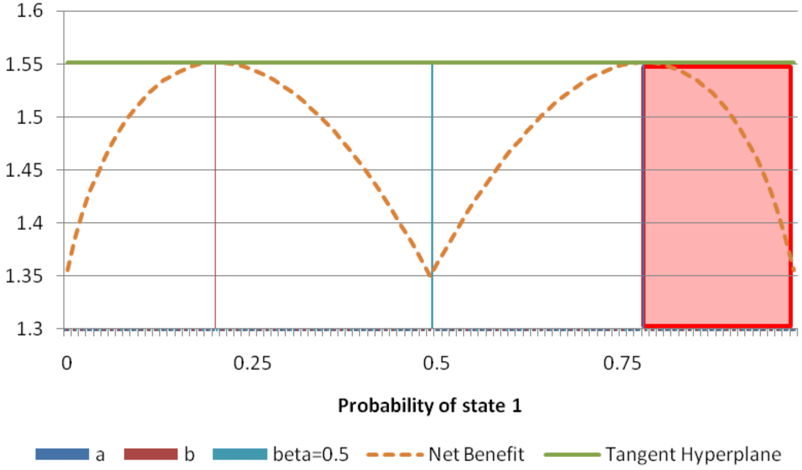
Behavior for prior>a



Same Posteriors as for 0.5 prior



No Information Gathered



Locally Invariant Posteriors

- Locally Invariant posteriors: If a set of posteriors $\{\gamma^a\}_{a \in A}$ are optimal for decision problem $\{\mu, A\}$ and are also feasible for $\{\mu', A\}$ then they are also optimal for that decision problem
- Choice probabilities move ‘mechanically’ with prior to maintain posteriors
- Useful in, for example, models in which consumers are rationally inattentive to quality
    - As the prior distribution of quality changes, posterior beliefs do not
    - See Martin [2014]
- This is a property of the Uniformly Posterior Separable Model
    - See Denti [2022] for an alternative characterization of the UPS model

Invariance Under Compression

- The Shannon model is clearly ‘special’ in many ways in the class of UPS model
- The literature has noted many properties
  - Symmetry
  - Separability of Orthogonal Decisions
  - Lack of Complementarities
- All of these properties can be captured in a single axiom
  - Invariance Under Compression

Invariance Under Compression - An Example

- Consider decision problem (i)

State	ω_1	ω_2
Prior Prob	0.5	0.5
Payoff Action A	10	0
Payoff Action B	0	10

- And now decision problem (ii) which splits ω_2

State	ω_1	ω_2	ω_3
Prior Prob	0.5	0.2	0.3
Payoff Action A	10	0	0
Payoff Action B	0	10	10

Invariance Under Compression - An Example

- How should behavior change between the two decision problems?
- In principal, many things could happen
  - Could be harder to learn about two states that one, so less accurate in (*ii*) than (*i*)
  - Could be easier to learn about two states that one, so more accurate in (*ii*) than (*i*)
- Shannon model says that behavior should not change
  - $P_i(a|\omega_2) = P_{ii}(a|\omega_2) = P_{ii}(a|\omega_3)$

Behavioral Characterization

- Invariance under Compression formalizes this
- Defines the concept of a 'basic' decision problem
  - No two states have the same payoff for all acts
- Every decision problem has associated basic forms
- Choice behavior the same when moving between decision problems and their basic forms
- Corollaries
  - Behavior the same in every state which is payoff equivalent
  - Moving prior probabilities between payoff equivalent states does not change behavior

Summary

- Introduced Shannon Mutual Information as a potential cost function
  - Popular in the literature
  - 'Cobb Douglas' vs 'Revealed Preference'
- Introduced some analytical tools to help solve the Shannon model
  - MM - necessary conditions
  - Necessary + Sufficient Conditions
  - Posterior-based approach
  - Behavioral characterization
- Shown that the Shannon model can give rise to endogenous consideration set formation