A Random Attention Model

Matias D. Cattaneo
U. Michigan

Xinwei Ma
U. Michigan

Yusufcan Masatlioglu
U. Maryland

Elchin Suleymanov
U. Michigan

Limited Attention

▷ So many options

▷ People do not pay attention to *all* products

Consideration Set

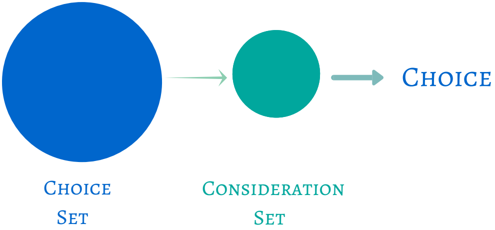
Marketing, Finance and Psychology Literatures: Howard and Sheth 1969, Wright and Barbour 1977, Hauser and Wernerfelt 1990, Nedungadi 1990, Alba et al. 1991, Roberts and Lattin 1991, Shocker et al. 1991, Roberts and Nedungadi 1995, Chiang et al 1999, Punj and Brookes 2001, Swait et al. 2002, Goeree 2008, and many more.

Consideration Set

The consideration set is made up of products that are taken seriously by the consumer in his or her purchase decision.

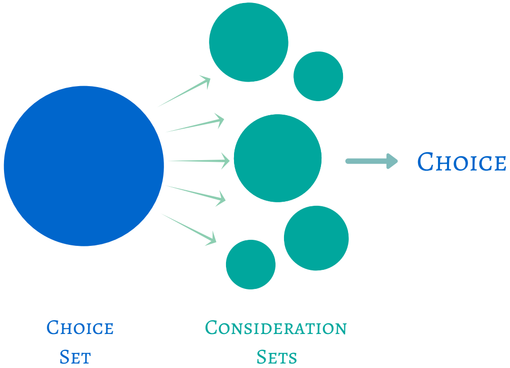
Limited Attention

▷ Two-stage Choice

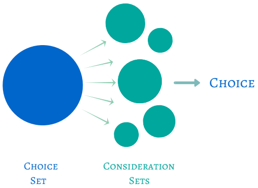


- Masatlioglu, Nakajima, and Ozbay (2012), and many...

Random Attention



Random Attention



Model:

- Consumer is able to rank options,
- Consumer chooses the best option in his consideration set,
- Randomness in choices comes from random attention,

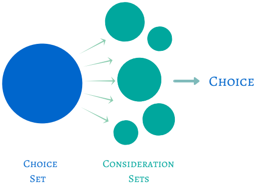
Domain and Data

- DOMAIN
  - $X = \{a_1, a_2, \ldots, a_K\}$: finite and abstract,

- DATA
  - $\pi(a_k|S)$: the probability that a_k is chosen from $S \subset X$,

$$\begin{aligned} \pi(a_k|S) &\geq 0 \text{ for all } a_k \in S \\ \sum_{a_k \in S} \pi(a_k|S) &= 1 \end{aligned}$$

Random Attention



$\mu(T|S)$: frequency of considering T when the feasible set is S

$$\begin{aligned} \mu(T|S) &\geq 0 \text{ for all } T \subset S \\ \sum_{T \subset S} \mu(T|S) &= 1 \end{aligned}$$

Random Attention: Examples

- Full Attention

		$T : Consideration\ \ Sets$						
		$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
	$\{a_1, a_2, a_3\}$	1	0	0	0	0	0	0
S	$\{a_1, a_2\}$		1			0	0	
	$\{a_1, a_3\}$			1		0		0
	$\{a_2, a_3\}$				1		0	0

Random Attention: Examples

- Deterministic Attention

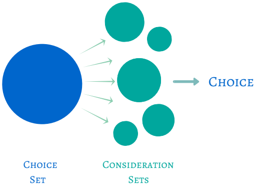
		$T: Consideration\ \ Sets$						
		$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
	$\{a_1, a_2, a_3\}$	0	1	0	0	0	0	0
S	$\{a_1, a_2\}$		1			0	0	
	$\{a_1, a_3\}$			0		0		1
	$\{a_2, a_3\}$				0		1	0

Random Attention: Examples

- Non-Deterministic Attention

		$T: Consideration\ \ Sets$						
		$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
S	$\{a_1, a_2, a_3\}$	0	0.5	0	0	0.5	0	0
	$\{a_1, a_2\}$		1			0	0	
	$\{a_1, a_3\}$			0		0.6		0.4
	$\{a_2, a_3\}$				0.4		0.3	0.3

Stochastic Choice



$$\pi(a_k|S) = \sum_{\substack{T \subset S, \\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S)$$

- Unobservables
  - $\succ$ - complete and transitive
  - μ - ???????

Attention Filter

- Masatlioglu, et al (2012): Deterministic Attention

- Consideration set is unchanged when an alternative to which consumer does not pay attention becomes unavailable.

$$a_k \notin \Gamma(S) \;\Rightarrow\; \Gamma(S) = \Gamma(S - a_k)$$

- What is the corresponding condition for μ?

Attention Filter

$$a_k \notin \Gamma(S) \;\Rightarrow\; \Gamma(S) = \Gamma(S - a_k)$$

- Assume $\Gamma(\{a_1, a_2, a_3\}) = \{a_1, a_2\}$

$\mu(T\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
$\{a_1, a_2, a_3\}$	0	1	0	0	0	0	0

Attention Filter

$$a_k \notin \Gamma(S) \;\Rightarrow\; \Gamma(S) = \Gamma(S - a_k)$$

- Assume $\Gamma(\{a_1, a_2, a_3\}) = \{a_1, a_2\}$
- By the property, $\Gamma(\{a_1, a_2\}) = \{a_1, a_2\}$ (since $a_3 \notin \Gamma(\{a_1, a_2, a_3\})$)

$\mu(T\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
$\{a_1, a_2, a_3\}$	0	1	0	0	0	0	0
$\{a_1, a_2\}$		1			0	0	

Monotonic Attention

Monotonic Attention:

$$\mu(T|S) \leq \mu(T|S - a_k)$$

Monotonic Attention

Monotonic Attention:

$$\mu(T|S) \leq \mu(T|S - a_k)$$

$$a_k \notin \Gamma(S) \;\Rightarrow\; \Gamma(S) = \Gamma(S - a_k)$$

Monotonic Attention

Monotonic Attention: $\mu(T|S) \leq \mu(T|S - a_k)$

For example, "Uniform Attention"

$\mu(T\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
$\{a_1, a_2, a_3\}$	1/7	1/7	1/7	1/7	1/7	1/7	1/7
$\{a_1, a_2\}$		1/3			1/3	1/3	
$\{a_1, a_3\}$			1/3		1/3		1/3
$\{a_2, a_3\}$				1/3		1/3	1/3

- Does μ satisfy monotonicity?

Monotonic Attention

Monotonic Attention: $\mu(T|S) \leq \mu(T|S - a_k)$

$\mu(T\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
$\{a_1, a_2, a_3\}$	1/7	1/7	1/7	1/7	1/7	1/7	1/7
$\{a_1, a_2\}$		1/3			1/3	1/3	
$\{a_1, a_3\}$			1/3		1/3		1/3
$\{a_2, a_3\}$				1/3		1/3	1/3

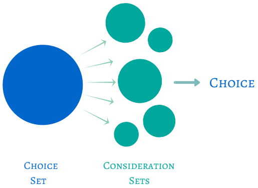
Monotonic Attention

Monotonic Attention: $\mu(T|S) \leq \mu(T|S - a_k)$

$\mu(T\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$
$\{a_1, a_2, a_3\}$	1/7	1/7	1/7	1/7	1/7	1/7	1/7
$\{a_1, a_2\}$		1/3			1/3	1/3	
$\{a_1, a_3\}$			1/3		1/3		1/3
$\{a_2, a_3\}$				1/3		1/3	1/3

- YES, μ is monotonic.
- More examples coming...

Random Attention Model (RAM)



$$\pi(a_k|S) = \sum_{\substack{T \subset S, \\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S)$$

- $\succ$ - complete and transitive
- μ - monotonic

Random Attention Model (RAM)

RAM accommodates well-documented and seemingly anomalous behaviors.

Attraction Effect

- Probabilistic Attraction Effect
  - a_1 and a_2 are equally chosen in a binary comparison,
  - a_3 is a decoy for a_1,

$\pi(a\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$
a_1	1	1/2	1	
a_2	0	1/2		1
a_3	0		0	0

$$\pi(a_1|X) > \pi(a_1|X - a_3)$$

Violation of Regularity

- Random Attention Model allows

$$\pi(a_k|S) > \pi(a_k|S - a_l)$$

  - Removing an alternative can decrease the choice probability

Prediction Power

- Is the model too general?
- The random attention model can be falsified.
  - For example, the following π is outside of the model.

$\pi(a\|S)$	$\{a_1, a_2, a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$
a_1	$1/3$	1	0	
a_2	$1/3$	0		1
a_3	$1/3$		1	0

RUM vs RAM

- Random Utility satisfies Regularity
- RAM is more general

Richness of Model

- To illustrate the richness of model, we provide some examples which satisfy monotonicity.

Examples

▷ **Fixed Independent Consideration (MM, 2014):** Consider a decision maker who pays attention to each alternative with a fixed probability γ.

Examples

▷ **Logit Attention:** (BR, 2017) Consider a decision maker who assigns a positive weight, w_D, for each non-empty subset of X.

$$\mu(T|S) = \frac{w_T}{\sum\limits_{T' \subset S} w_{T'}}$$

Psychologically w_T measures the strength associated with the subset T.

Revealed Preference

- How can we deduce preferences under random attention?
- However, richness does not help us much
  - More degree of freedom
  - Allowing many possibilities
  - Less revelations

Multiple Representations

- Multiple Representations

$$(\succ_1, \mu_1), (\succ_2, \mu_2), (\succ_3, \mu_3), \ldots, (\succ_n, \mu_n)$$

Definition

a_k is **revealed to be preferred** to a_l if $\succ_i$ ranks a_k above a_l for all i.

Revealed Preference

- Finding all possible representations
  - hard and cumbersome

- Need to look for a short-cut

Revealed Preference

- Observation: $\pi(a_k|S) > \pi(a_k|S - a_l)$ implies "a_k is better than a_l"

Revealed Preference

PROOF:

$$\begin{aligned}
\pi(a_k|S) &= \sum_{\substack{T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S) \\
&= \sum_{\substack{a_l\in T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S) + \sum_{\substack{a_l\notin T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S) \\
&\leq \sum_{\substack{a_l\in T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S) + \sum_{\substack{a_l\notin T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S-a_l) \quad \text{(by monotonicity)} \\
&\leq \sum_{\substack{a_l\in T\subset S,\\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S) + \pi(a_k|S-a_l)
\end{aligned}$$

Revealed Preference

PROOF continues...

$$\pi(a_k|S) - \pi(a_k|S - a_l) \quad \leq \quad \sum_{\substack{a_l \in T \subset S, \\ a_k \text{ is } \succ\text{-best in } T}} \mu(T|S)$$

If $\pi(a_k|S) - \pi(a_k|S - a_l) > 0$ then there exists at least one T such that

- $a_l \in T$
- a_k is $\succ$-best in T
- $\mu(T|S) \neq 0$

Hence, a_k is revealed to be preferred to a_l. DONE

Revealed Preference

- Assume $\pi(a_k|S) > \pi(a_k|S - a_l)$ and $\pi(a_l|S') > p(a_l|S' - a_m)$
  - Directly, "a_k is better than a_l" and "a_l is better than a_m"
  - Indirectly, "a_k is better than a_m"

Revealed Preference

- $a_k \mathcal{P} a_l$ if $\pi(a_k|S) > \pi(a_k|S - a_l)$
- Let $\bar{\mathcal{P}}$ be the transitive closure of $\mathcal{P}$
- $a_k \bar{\mathcal{P}} a_l$ implies "a_k is better than a_l"

Revealed Preference

- While $\bar{\mathcal{P}}$ informs us about preference, do we miss some revelation?

Theorem (Revealed Preference)

Let π have a RAM representation. Then a_k is **revealed to be preferred** to a_l if and only if $a_k \bar{\mathcal{P}} a_l$.

- $\bar{\mathcal{P}}$ provides all the information we need to know.

Characterization

Characterization

A stochastic choice π has a RAM representation
iff
$\bar{\mathcal{P}}$ has no cycle.

Related Literature

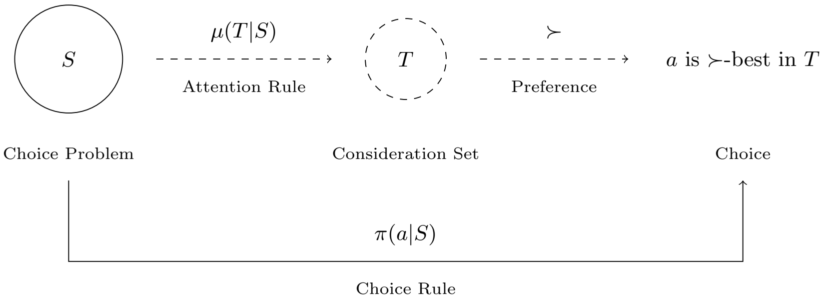
- Gul, Natenzon, and Pesendorfer (2014),
- Manzini and Mariotti (2014),
- Echenique, Saito, and Tserenjigmid (2014),
- Brady and Rehbeck (2016),
- Yildiz (2016),
- Li and Tang (2016).

Warp-Up

- An rich framework,
- Alternative to RUM,
- Revealed Preference is a powerful tool,
- It can be applied
  - both rational and boundedly rational behavior,
  - both deterministic and stochastic choice.

Inference and Estimation

Data Generating Process



Observable: choice problem and choice (solid line).
Unobservable: attention rule, consideration set and preference (dashed line).

Facts

- Choice rule $\pi(\cdot|\cdot)$ is identified (consistently estimable): we observe the choice problems and choices.
- Attention rule $\mu(\cdot|\cdot)$ is not identified: we do not observe the consideration sets.
  - Dimension of $\pi(\cdot|S)$: $|S|$.
  - Dimension of $\mu(\cdot|S)$: $2^{|S|} - 1 \gg |S|$.
- In realistic cases, preference $\succ$ is partially identified.

Definition (Identified Set Θ_π)

Θ_π is the collection of preferences such that

$$\pi(x|S) = \sum_{T \subset S} \mathbb{1}(x \text{ is } \succ\text{-best in } T) \cdot \mu(T|S),$$

for some μ satisfying monotonicity.

Outline

Identification: Equivalent Form 1

- $\pi(x\,|S) > \pi(x\,|S-y) \Rightarrow x \succ y$.
- Hard to implement in practice.
- π being estimated makes it more difficult.

Example (Is $\mathbf{x} \succ \mathbf{y}$?)

(1) Check $\pi(\mathbf{x}\,|S) \leq \pi(\mathbf{x}\,|S-\mathbf{y})$ for all S.

(2) Check $\pi(\mathbf{x}\,|S) \leq \pi(\mathbf{x}\,|S-\mathbf{z})$ and $\pi(\mathbf{z}\,|S) \leq \pi(\mathbf{z}\,|S-\mathbf{y})$ for all $\mathbf{z} \neq \mathbf{x}, \mathbf{y}$ and S.

(3) $\cdots$ (Check for even longer "chains")

Identification: Equivalent Form 2

- $\mathbb{U}$: feasible attention rules, defined by $\mathbf{R}\mu \leq 0$.
- Given $\succ$, a mapping from attention rules to choice rules:

$$\mathbf{C}_{\succ}: \sum_{T \subset S} \mathbb{1}(k \text{ is } \succ\text{-best in } T) \cdot \mu(T|S).$$

- Hence $\succ \in \Theta_\pi$ if and only if $\pi \in \mathbf{C}_{\succ}\mathbb{U} = \{\mathbf{C}_{\succ}\mu : \ \mathbf{R}\mu \leq 0\}$.

$$\mathbb{U} \quad \xrightarrow{\mathbf{C}_{\succ}} \quad \underbrace{\mathbf{C}_{\succ}\mathbb{U} \quad \overset{?}{\ni} \quad \pi}_{\text{Testing Problem}}$$

- A J-test. Alternatively, can show: $\mathbf{C}_{\succ}\mathbb{U} = \{\pi : \ \mathbf{R}_{\succ}\pi \leq 0\}$.
  - Unclear how to construct $\mathbf{R}_{\succ}$ from $\mathbf{C}_{\succ}$ and $\mathbb{U}$.

Identification: Equivalent Form 3

- We cannot identify μ since
  - Consideration sets are not observed.
  - μ has much higher dimension than π.
- That is, the mapping:

$$\mathbf{C}_{\succ}: \sum_{T \subset S} \mathbb{1}(k \text{ is } \succ\text{-best in } T) \cdot \mu(T|S).$$

  is many-to-one.
- Restrict the class of attention rules.
  - Forward engineering: test if $\pi \in \mathbf{C}_{\succ}\mathbb{U}$.
  - Backward engineering: (1) construct a attention rule (in the restricted class) from π and $\succ$; and (2) test if the monotonicity assumption holds.

Identification: Equivalent Form 3 (Cont'd)

Example (Triangular Attention Rule)

Take a preference $b \succ c \succ e \succ d \succ a$. μ is triangular with respect to $\succ$ if, for any menu S, it puts weights only on *lower contour sets*:

$$\Big\{S \cap \{b,c,e,d,a\},\ S \cap \{c,e,d,a\},\ S \cap \{e,d,a\},\ S \cap \{d,a\},\ S \cap \{a\}\Big\}.$$

Construct a unique μ from π. For the previous example,

$$\pi(b|S) = \mu(\{b,c,e,d,a\}|S), \qquad \pi(c|S) = \mu(\{c,e,d,a\}|S), \qquad \text{etc.}$$

Theorem (Identification)

The following are equivalent:

(1) $\succ \in \Theta_\pi$;

(2) The triangular attention rule constructed from π and $\succ$ satisfies the monotonicity assumption.

Main Identification Result

- For each preference $\succ$, we can construct a matrix $\mathbf{R}_{\succ}$, such that $\succ \in \Theta_\pi$ if and only if $\mathbf{R}_{\succ}\pi \leq 0$.
- $\mathbf{R}_{\succ} = \mathbf{R}\mathbf{C}_{\succ}^{-1}$:
  - $\mathbf{C}_{\succ}^{-1}$: maps choice rules back to triangular attention rules.
  - $\mathbf{R}$: restrictions imposed by the monotonicity assumption.

Example

Four elements in the grand set $X = \{a, b, c, d\}$.

(1) $\mathcal{S} = \{\{a, b, c, d\},\ \{b, c, d\}\}$. Then $b \succ d \succ c \succ a$ gives no constraint. $b \succ d \succ a \succ c$ gives one constraint.

(2) $\mathcal{S} = \{\{a, b, c, d\},\ \{b, c, d\},\ \{a, b, c\},\ \{b, c\}\}$. Then $b \succ d \succ a \succ c$ gives ? constraints.

How Many Constraints?

- Five elements: $X = \{a, b, c, d, e\}$.

$\mathcal{S}$	Dim. of π	Restrictions in $\mathbf{R}_{\succ}$
$\{X\}$	5	0
All subsets of size 5 and 4	25	10
All subsets of size 5, 4 and 3	55	50
All subsets of size 5, 4, 3 and 2	75	105
$2^X \backslash \{\emptyset\}$	80	131

Outline

1. Identification
2. Inference
3. Estimation

Testing $\mathsf{H}_0 : \succ \in \Theta_\pi$

- Construct $\mathbf{R}_\succ = \mathbf{M}\mathbf{C}_\succ^{-1}$
- Data: $\{(y_i, Y_i),\ 1 \leq i \leq N\}$
  - i's choice $= y_i \ \in \ Y_i =$ choice problem offered to i.
- For each S and each $a \in S$, estimate the choice rule by

$$\hat{\pi}(a|S) = \frac{\sum_{i=1}^{N} \mathbb{1}(Y_i = S,\ y_i = a)}{\sum_{i=1}^{N} \mathbb{1}(Y_i = S)}$$

- Reject H_0 if $\mathbf{R}_\succ \hat{\pi} \leq 0$ is violated.
- Test statistic: maximum of elements in $\mathbf{R}_\succ \hat{\pi}$ or 0

$$\mathscr{T} = \sqrt{N} \max \left\{ \mathbf{R}_\succ \hat{\pi},\ 0 \right\}.$$

Critical Values

- We need to approximate the distribution of $\mathscr{T}$:

$$\begin{aligned}\mathscr{T} &= \max\left\{\sqrt{N}\mathbf{R}_{\succ}\hat{\pi},\ 0\right\}\\ &= \max\Bigg\{\underbrace{\sqrt{N}\mathbf{R}_{\succ}(\hat{\pi}-\pi)}_{\text{Asy. Normal}} + \underbrace{\sqrt{N}\mathbf{R}_{\succ}\pi}_{\text{Unknown, } \leq 0 \text{ under } \mathsf{H}_0},\ 0\Bigg\}.\\ \mathscr{T}^{\star} &= \max\Bigg\{\underbrace{\mathsf{Normal}(0,\ \hat{\mathbf{V}})}_{\text{Simulate}} + \underbrace{\sqrt{N}\hat{\mathbf{u}}}_{\text{Plug-in Something}},\ 0\Bigg\}.\end{aligned}$$

- Then reject $\mathsf{H}_0 : \succ \in \Theta_\pi$ if $\mathscr{T} > c_\alpha$.

$$c_\alpha = \inf\left\{t :\ \mathbb{P}^{\star}[\mathscr{T}^{\star} > t] \leq \alpha\right\}.$$

Least Favorable Method: $\hat{\mathbf{u}} = 0$

- Motivation: $\mathsf{H}_0 : \succ \in \Theta_\pi \;\Rightarrow\; \mathbf{R}_\succ \pi \leq 0$, hence replace by the conservative upper bound 0.

$$\mathscr{T} = \max\Big\{\underbrace{\sqrt{N}\mathbf{R}_\succ(\hat{\pi} - \pi)}_{\text{Asy. Normal}} + \underbrace{\sqrt{N}\mathbf{R}_\succ \pi}_{\text{Unknown, } \leq 0 \text{ under } \mathsf{H}_0},\ 0\Big\}.$$

$$\mathscr{T}^\star = \max\Big\{\underbrace{\mathsf{Normal}(0,\ \hat{\mathbf{V}})}_{\text{Simulate}} + \underbrace{0}_{\text{Least Favorable}},\ 0\Big\}.$$

- $\succ \in \Theta_\pi \;\Leftrightarrow\; \mathbf{R}_\succ \pi \leq 0$: $\mathscr{T}^\star$ first order stochastically dominates $\mathscr{T}$

$$\lim_N \mathbb{P}[\mathscr{T} > c_\alpha] \leq \lim_N \mathbb{P}^\star[\mathscr{T}^\star > c_\alpha] = \alpha \qquad \text{(size control)}.$$

- $\succ \notin \Theta_\pi \;\Leftrightarrow\; \mathbf{R}_\succ \pi \not\leq 0$: $\mathscr{T} \xrightarrow{\text{p}} +\infty$ while $\mathscr{T}^\star$ remains bounded in probability

$$\lim_N \mathbb{P}[\mathscr{T} > c_\alpha] = 1 \qquad \text{(consistency)}.$$

(Don't!) Plug-in Method: $\hat{\mathbf{u}} = (\mathbf{R}_{\succ}\hat{\pi})_{-}$

- (Bad) Motivation: since π is unknown, replace by its estimate $\hat{\pi}$.

$$\mathscr{T} = \max\Big\{\underbrace{\sqrt{N}\mathbf{R}_{\succ}(\hat{\pi}-\pi)}_{\text{Asy. Normal}} + \underbrace{\sqrt{N}\mathbf{R}_{\succ}\pi}_{\text{Unknown, } \leq 0 \text{ under } \mathsf{H}_0},\ 0\Big\}.$$

$$\mathscr{T}^{\star} = \max\Big\{\underbrace{\mathsf{Normal}(0,\ \hat{\mathbf{V}})}_{\text{Simulate}} + \underbrace{\sqrt{N}(\mathbf{R}_{\succ}\hat{\pi})_{-}}_{\text{Plug-in consistent estimate}},\ 0\Big\}.$$

- Fails if $\mathbf{R}_{\succ}\pi = 0$
  - $\mathbb{P}[\sqrt{N}(\mathbf{R}_{\succ}\hat{\pi})_{-} \lneqq 0] \to c > 0$.
  - With positive probability, $\mathscr{T}$ dominates $\mathscr{T}^{\star}$, hence quantiles computed from $\mathscr{T}^{\star}$ is too small for $\mathscr{T}$.
- Also fails if $\mathbf{R}_{\succ}\pi < 0$ but close to 0 (relative to the sample size).
  - Lack of uniformity: for each sample size N, it is possible to find some DGP $\mathbf{R}_{\succ}\pi < 0$ such that the rejection probability is strictly larger than α.

Plug-in Method with Shrinkage: $\hat{\mathbf{u}} = \frac{1}{\kappa_N}(\mathbf{R}_{\succ}\hat{\pi})_-$

- Motivation: shrink the naive plug-in method by a factor $\kappa_N \overset{\text{slowly}}{\to} \infty$.

$$\mathscr{T} = \max\Big\{\underbrace{\sqrt{N}\mathbf{R}_{\succ}(\hat{\pi}-\pi)}_{\text{Asy. Normal}} + \underbrace{\sqrt{N}\mathbf{R}_{\succ}\pi}_{\text{Unknown, } \leq 0 \text{ under } \mathsf{H}_0},\ 0\Big\}.$$

$$\mathscr{T}^\star = \max\Big\{\underbrace{\mathsf{Normal}(0,\ \hat{\mathbf{V}})}_{\text{Simulate}} + \underbrace{\frac{\sqrt{N}}{\kappa_N}(\mathbf{R}_{\succ}\hat{\pi})_-}_{\text{Plug-in consistent estimate}},\ 0\Big\}.$$

Comparison of Three Methods

$$\mathscr{T} = \max\Big\{\underbrace{\sqrt{N}\mathbf{R}_{\succ}(\hat{\pi}-\pi)}_{\text{Asy. Normal}} + \underbrace{\sqrt{N}\mathbf{R}_{\succ}\pi}_{\text{Unknown, } \le 0 \text{ under } \mathsf{H}_0},\ 0\Big\}.$$

$$\mathscr{T}^{\star} = \max\Big\{\underbrace{\mathsf{Normal}(0,\ \hat{\mathbf{V}})}_{\text{Simulate}} + \underbrace{\sqrt{N}\hat{\mathbf{u}}}_{\text{Plug-in Something}},\ 0\Big\}.$$

- $\succ \in \Theta_\pi \ \Leftrightarrow\ \mathbf{R}_{\succ}\pi \le 0$

Least Favorable	$\hat{\mathbf{u}} = 0$	$\mathbb{P}[\mathscr{T} > c_\alpha] \le \alpha + O\left(\frac{1}{\sqrt{N}}\right)$
Plug-in	$\hat{\mathbf{u}} = (\mathbf{R}_{\succ}\hat{\pi})_-$	$\mathbb{P}[\mathscr{T} > c_\alpha] \le \alpha + O\left(\frac{1}{\sqrt{N}} + 1\right)$
Shrinkage	$\hat{\mathbf{u}} = \frac{1}{\kappa_N}(\mathbf{R}_{\succ}\hat{\pi})_-$	$\mathbb{P}[\mathscr{T} > c_\alpha] \le \alpha + O\left(\frac{1}{\sqrt{N}} + \frac{1}{\kappa_N}\right).$

- $\succ \notin \Theta_\pi \ \Leftrightarrow\ \mathbf{R}_{\succ}\pi \not\le 0$
  - $(\mathbf{R}_{\succ}\hat{\pi})_- \ \le\ \frac{1}{\kappa_N}(\mathbf{R}_{\succ}\hat{\pi})_- \ \le\ 0$
  - Power (ability to rule out improbable preferences): PI > S > LF.

Other Methods for Critical Values

- Different choices of $\hat{\mathbf{u}}$ are proposed in order to
  - Improve power (rule out improbable preferences) relative to the LF method.
  - Maintain size control (false rejection when $\succ \in \Theta_\pi$).
- Review on testing moment inequalities: **?** and **?**.

Outline

Confidence Set

- $\mathcal{C}(\alpha)$ is constructed by inverting the test:

$$\mathcal{C}(\alpha) = \Big\{ \succ :\ \mathscr{T}_{\succ} \leq c_{\alpha,\succ} \Big\}.$$

- Covers preferences in the identified set Θ_π:

$$\lim_{N} \min_{\succ \in \Theta_\pi} \mathbb{P}\Big[\succ \in \mathcal{C}(\alpha) \Big] \geq 1 - \alpha.$$

This Paper

- Generalizes Masatlioglu Nakajima Ozbay (2012) to accommodate random attention scenario.
- Provides conditions under which the preference is partially identified from choice data, without observing consideration sets.
- Constructs test statistics facilitating estimation and inference:
  - Reformulates identification as testing moment inequalities.

    There is a large literature on testing moment inequalities and inference in partially identified models.

    Other test statistics and methods for critical values can be easily adapted.
  - Provides uniformly valid distributional approximations and critical values.
  - Implements in R and Matlab.