Solutions to Problem Set 03: Choice Functions, WARP and Stochastic Choice

Goes with Lectures 2 and 3. You test whether choices from menus can be explained by a utility function, adapt WARP to indifference, connect it with Sen's conditions α and β, and then move to choice probabilities: random utility, regularity, the logit model, and how to infer a choice correspondence from choice frequencies. **Solutions only; see the problem set for the statements.**

Notation (as in the lectures). X is a finite set of alternatives and $\mathcal{A}$ the set of nonempty menus $A \subseteq X$. A choice function C picks one option $C(A) \in A$; a choice correspondence picks a nonempty set $C(A) \subseteq A$. A stochastic choice function gives probabilities $p(x \mid A)$ with $\sum_{x \in A} p(x \mid A) = 1$; for pairs, $p(x, y) = p(x \mid \{x, y\})$. Logit with utility u and noise $\lambda > 0$: $p(x \mid A) = e^{u(x)/\lambda} / \sum_{y \in A} e^{u(y)/\lambda}$.

Problem 1 Can these choices be explained? Warm-up

Solution. (a) D_1: no. A rationalizing u would need $u(a) > u(b)$, $u(b) > u(c)$ and $u(c) > u(a)$, hence $u(a) > u(a)$.

D_2: no. From $\{a, b\}$, $u(a) > u(b)$; from $\{a, b, c\}$, $u(b) > u(a)$. This is the lecture's "day 4" pattern: the third option changes the choice between the first two.

D_3: yes. The choices require $u(b) > u(a), u(c), u(d)$; $u(c) > u(a)$; $u(a) > u(d)$; $u(c) > u(d)$. So $u(b) > u(c) > u(a) > u(d)$ works, e.g. $u(a, b, c, d) = (1, 3, 2, 0)$.

D_4: yes, e.g. $u(a, b, c, d) = (3, 2, 1, 0)$, but also $(1, 0, 3, 2)$.

(b) D_3: the inequalities above force $u(b) > u(c) > u(a) > u(d)$ for *every* rationalizing u. So every such u predicts b from $\{a, b\}$ and c from $\{a, c, d\}$. D_4: the two utilities in (a) predict a and c from $\{a, c\}$. The data say nothing about this menu: a dataset can be rationalizable and still leave many predictions open.

(c)(i) If $x \; R \; y$, then $x = C(A)$ for an observed A with $y \in A$, $y \neq x$, so $u(x) > u(y)$. A cycle would give $u(x_1) > u(x_2) > \cdots > u(x_k) > u(x_1)$, which is impossible.

(ii) Suppose every x is revealed preferred to some other element. Start anywhere and follow $x_1 \; R \; x_2 \; R \; x_3 \; R \cdots$. Since X is finite, some element appears twice, and the piece of the path between its two appearances is a cycle. So, without a cycle, some x is revealed preferred to nothing. Removing elements cannot create a cycle, so the argument applies again to the remaining set, and the procedure gives every element a utility, all different. Now let $x \; R \; y$. When x was removed, it was revealed preferred to no remaining element; so y had been removed earlier, and $u(x) > u(y)$. On an observed menu A, $C(A) \; R \; y$ for all $y \in A \setminus \{C(A)\}$, so $u(C(A)) > u(y)$: u rationalizes the data. (This acyclicity condition is the *Strong* Axiom of Revealed Preference, SARP.)

(d) WARP requires: if $x \neq y$ are both in $A \cap B$ and $C(A) = x$, then $C(B) \neq y$. In D_1, two different observed menus share only one element, and from a single menu only one option is chosen. So there is nothing to check, and WARP holds. Yet $a \; R \; b \; R \; c \; R \; a$ is a cycle: by (c), no utility rationalizes D_1. The lecture's theorem assumes *richness*: choices from *all* menus. If $\{a, b, c\}$ were observed, every choice from it would violate WARP (this is the lecture's proof of transitivity). With partial data, WARP checks only pairs of menus; one needs the whole chain, i.e. SARP.

Problem 2 Choice correspondences: adapting WARP to indifference Core

Solution. (a) Take $u(x) = u(y) = 1, u(z) = 0$. The maximizers are $\{x, y\}$ in $\{x, y\}$ and in $\{x, y, z\}$, $\{x\}$ in $\{x, z\}$, $\{y\}$ in $\{y, z\}$: exactly C_1. Naive WARP fails with $A = B = \{x, y\}$: $x \in C_1(A)$ and $y \in C_1(B)$. In fact the naive axiom rules out *every* indifference: as soon as two options are chosen from the same menu, it is violated. With indifference, "x chosen when y is available" only reveals $x \succsim y$, not $x \succ y$; so choosing y when x is available is not a contradiction. What *is* a contradiction is choosing y and rejecting x, since that reveals $y \succ x$. This is what the lecture's version forbids.

(b) Let $x \in C(A)$, $y \in C(B)$, $x, y \in A \cap B$. Since x maximizes u on $A \ni y$, $u(x) \geq u(y)$. Since y maximizes u on B, $u(y) \geq u(v)$ for all $v \in B$; so $u(x) \geq u(y) \geq u(v)$ for all $v \in B$, and $x \in B$. Hence $x \in C(B)$.

(c) No. Take $A = \{x, y\}$, $B = \{x, y, z\}$: $y \in C_2(A)$ and $x \in C_2(B)$, so WARP requires $y \in C_2(B) = \{x\}$, which fails. So by (b), no utility generates C_2. Directly: $\{x, y\}$ gives $u(x) = u(y)$, $\{y, z\}$ gives $u(y) = u(z)$, so $u(x) = u(z)$; but $\{x, z\}$ gives $u(x) > u(z)$. Indifference is not transitive in C_2.

(d) Write $C(A) = \{c(A)\}$. Suppose the correspondence version holds, and let $x \neq y$ in $A \cap B$ with $c(A) = x$. If $c(B) = y$, the axiom gives $x \in C(B) = \{y\}$, so $x = y$: contradiction. So $c(B) \neq y$: the function version holds. Conversely, suppose the function version holds, and let $x \in C(A)$, $y \in C(B)$, $x, y \in A \cap B$. If $x \neq y$, the function version (with $c(A) = x$) says $c(B) \neq y$, a contradiction. So $x = y \in C(B)$.

(e)(i) $C(\{x, y\})$ is nonempty, so it contains x or y: $x \succsim^\star y$ or $y \succsim^\star x$.

(ii) ($\subseteq$) Let $x \in C(A)$ and $y \in A$. If $x \in C(\{x, y\})$, done. Otherwise $y \in C(\{x, y\})$; apply WARP to A and $\{x, y\}$: $x \in C(A)$ and $y \in C(\{x, y\})$ give $x \in C(\{x, y\})$. Either way $x \succsim^\star y$. ($\supseteq$) Let $x \in A$ with $x \succsim^\star y$ for all $y \in A$, and take any $z \in C(A)$. Since $x \in C(\{x, z\})$ and $z \in C(A)$, WARP (with the menus $\{x, z\}$ and A) gives $x \in C(A)$.

(iii) Let $x \succsim^\star y$ and $y \succsim^\star z$, and $T = \{x, y, z\}$. If $x \in C(T)$, then $x \succsim^\star z$ by (ii). If $y \in C(T)$: since $x \in C(\{x, y\})$, WARP gives $x \in C(T)$, and we are back to the first case. If $z \in C(T)$: since $y \in C(\{y, z\})$, WARP gives $y \in C(T)$, and we are back to the second case. As $C(T) \neq \emptyset$, $x \succsim^\star z$.

So $\succsim^\star$ is complete and transitive. The weak-order theorem gives u with $x \succsim^\star y \iff u(x) \geq u(y)$, and by (ii), $C(A)$ is the set of maximizers of u on A.

Problem 3 Sen's conditions α and β Core

Solution. (a) α: if the world champion is Pakistani, he is also the champion of Pakistan. β: if two Pakistanis are (joint) champions of Pakistan, and one of them is a world champion, then so is the other. Both sound obvious: a ranking of athletes does not depend on who else competes.

(b) α: let $A \subseteq B$, $x \in A$, $x \in C(B)$. Take any $y \in C(A)$ (it is nonempty). Then $x, y \in A \cap B$, $x \in C(B)$ and $y \in C(A)$; WARP (with B in the role of the first menu and A in the role of the second) gives $x \in C(A)$. β: let $A \subseteq B$, $x, y \in C(A)$, and $x \in C(B)$. Then $x, y \in A \cap B$, $y \in C(A)$ and $x \in C(B)$, so WARP gives $y \in C(B)$. The same argument with x and y swapped gives the other direction.

(c) Let $x, y \in A \cap B$, $x \in C(A)$, $y \in C(B)$. Since $\{x, y\} \subseteq A$ and $x \in C(A)$, α gives $x \in C(\{x, y\})$. Since $\{x, y\} \subseteq B$ and $y \in C(B)$, α gives $y \in C(\{x, y\})$. So $x, y \in C(\{x, y\})$, and β (with $\{x, y\} \subseteq B$) gives: $y \in C(B)$ implies $x \in C(B)$. WARP holds.

(d) C_2 and α: when A is a singleton, α holds trivially. So the only case to check is $B = X$ and A a pair containing x, the only element of $C_2(X)$: indeed $x \in C_2(\{x, y\})$ and $x \in C_2(\{x, z\})$. So α holds. β fails: $x, y \in C_2(\{x, y\})$, $x \in C_2(X)$ but $y \notin C_2(X)$. And by Problem 2(c), WARP fails.

β without α: take the choice function $C(\{x,y\}) = x$, $C(\{x,y,z\}) = y$ (other menus arbitrary, e.g. $C(\{x,z\}) = x$, $C(\{y,z\}) = y$). With one option per menu, the premise "$x, y \in C(A)$" forces $x = y$, so β holds trivially. α fails: $y \in C(X)$ and $y \in \{x,y\}$, but $y \notin C(\{x,y\})$.

(e) For a choice function, β always holds (see (d)), so by (b) and (c), WARP $\iff \alpha \wedge \beta \iff \alpha$. For functions, α reads: if $A \subseteq B$ and $C(B) \in A$, then $C(A) = C(B)$. This is the "contraction" property; its stochastic version is *regularity* (Problem 4).

(f) Menus $A = \{a, o\} \subseteq B = \{a, a', o\}$. She takes a from B (say), $a \in A$, but $C(A) = \{o\}$: she violates α. If we describe the options only as fruits, she is not maximizing any utility. But she cares about something else: taking the *last* apple. The option "take a" is not the same option in both baskets: in A it means leaving nothing for the host. Redescribe the options with this feature, and her choices are consistent. Whether a violation of α shows irrationality depends on how the options are described, a theme that returns with framing and description invariance (Lecture 4).

Problem 4 Random utility and regularity

Warm-up

Solution. (a) Write π_{xyz} etc. The data give six equations:

$$p(x \mid X) = \pi_{xyz} + \pi_{xzy} = 0.5, \quad p(x,y) = \pi_{xyz} + \pi_{xzy} + \pi_{zxy} = 0.6,$$
$$p(y \mid X) = \pi_{yxz} + \pi_{yzx} = 0.3, \quad p(x,z) = \pi_{xyz} + \pi_{xzy} + \pi_{yxz} = 0.7,$$
$$p(z \mid X) = \pi_{zxy} + \pi_{zyx} = 0.2, \quad p(y,z) = \pi_{xyz} + \pi_{yxz} + \pi_{yzx} = 0.6.$$

Subtracting: $\pi_{zxy} = 0.1$, so $\pi_{zyx} = 0.1$; $\pi_{yxz} = 0.2$, so $\pi_{yzx} = 0.1$; then $\pi_{xyz} = 0.6 - 0.2 - 0.1 = 0.3$ and $\pi_{xzy} = 0.2$. All are nonnegative and they add up to 1: p is a RUM. The solution is unique: each step was forced. (With three options, the shares are the Block–Marschak polynomials of Lecture 3; with more options, different distributions can give the same p.)

(b) x: $p(x \mid X) = 0.5 \leq 0.6$ and ≤ 0.7. y: $0.3 \leq p(y,x) = 0.4$ and $\leq p(y,z) = 0.6$. z: $0.2 \leq p(z,x) = 0.3$ and $\leq p(z,y) = 0.4$. Regularity holds, as it must for a RUM.

(c) Suppose some $\mathbb{P}$ over rankings generates the data. Every ranking under which x is the best of $\{x,y,d\}$ ranks x above y. So $p(x \mid \{x,y,d\}) = \mathbb{P}(x \text{ best in } \{x,y,d\}) \leq \mathbb{P}(x \succ y) = p(x,y)$, i.e. $0.6 \leq 0.5$: impossible. This is the decoy effect of Lecture 4.

(d) Let $A \subseteq B$ and $x \in A$. Regularity says $p(x \mid B) \leq p(x \mid A)$. If $p(x \mid B) = 0$ it holds trivially; if $p(x \mid B) = 1$, i.e. $C(B) = x$, it says $p(x \mid A) = 1$, i.e. $C(A) = x$. So regularity is exactly: $C(B) = x \in A \Rightarrow C(A) = x$, which is condition α for a choice function. By Problem 3(e), $\alpha \iff$ WARP. If C satisfies WARP, the lecture's theorem gives a ranking $\succ$ whose best element is chosen from every menu; the RUM that puts probability 1 on $\succ$ generates p. Conversely, a RUM satisfies regularity, hence α, hence WARP.

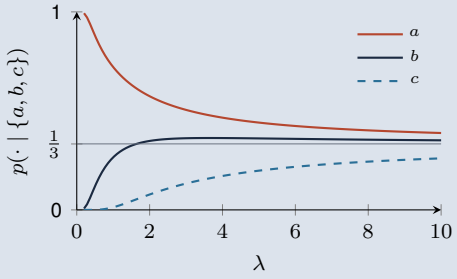
(e) If p_1 comes from $\mathbb{P}_1$ and p_2 from $\mathbb{P}_2$, then $t\,p_1 + (1-t)\,p_2$ comes from $t\,\mathbb{P}_1 + (1-t)\,\mathbb{P}_2$, which is again a distribution over rankings. So a population of consumers who each maximize a fixed ranking (each satisfies WARP) produces market shares that are a RUM, and satisfy regularity. Contrapositive: if adding d raises the share of x, then not every consumer is a random utility maximizer; at least some consumers' choices depend on the menu. Population data can reveal that *someone* violates WARP, but not who.

Problem 5 Logit: what the noise parameter λ does

Core

Solution. (a) With weights $e^{u/\lambda}$:

λ	$p(a)$	$p(b)$	$p(c)$
0.5	0.881	0.119	< 0.001
1	0.721	0.265	0.013
2	0.574	0.348	0.078
10	0.388	0.351	0.260



For $\lambda = 1$: $e^4 = 54.60$, $e^3 = 20.09$, $e^0 = 1$, sum 75.68.

(b) Let $u^* = \max_A u$. Divide numerator and denominator by $e^{u^*/\lambda}$:

$$p(x \mid A) = \frac{e^{-(u^* - u(x))/\lambda}}{\sum_{y \in A} e^{-(u^* - u(y))/\lambda}}.$$

As $\lambda \to 0$, each term with $u(y) < u^*$ goes to 0 and each term with $u(y) = u^*$ equals 1. So the denominator tends to k, the number of maximizers, and $p(x \mid A) \to 1/k$ for a maximizer and $\to 0$ for the others; with a unique maximizer, the limit is 1. Small noise gives utility maximization, with ties broken uniformly: the choice correspondence of Problem 2, plus a fair coin. As $\lambda \to \infty$, every $e^{u(y)/\lambda} \to 1$, so $p(x \mid A) \to 1/|A|$: choices become uniform and reveal nothing.

(c) For the best option x^*, $p(x^* \mid A) = 1/\sum_{y \in A} e^{-(u^* - u(y))/\lambda}$. Each term $e^{-d/\lambda}$ with $d \geq 0$ increases in λ, so $p(x^* \mid A)$ decreases. For the worst option x_* with utility u_*, $p(x_* \mid A) = 1/\sum_{y \in A} e^{(u(y) - u_*)/\lambda}$, and each term $e^{d/\lambda}$ with $d \geq 0$ decreases in λ, so $p(x_* \mid A)$ increases. The middle option b is *not* monotone: $p(b) \to 0$ as $\lambda \to 0$, $p(b) \to \frac{1}{3}$ as $\lambda \to \infty$, but $p(b) = 0.351 > \frac{1}{3}$ at $\lambda = 10$ (the maximum is about 0.363, near $\lambda \approx 3.6$). As noise rises, b first gains from a, then loses to c.

(d) Let $\theta = 1/\lambda$ and $Z = \sum_{y \in A} e^{\theta u(y)}$, so $p(x) = e^{\theta u(x)}/Z$. Then $\frac{dp(x)}{d\theta} = p(x)(u(x) - \bar{u})$, since $\frac{d \ln Z}{d\theta} = \bar{u}$. Hence

$$\frac{d\bar{u}}{d\theta} = \sum_x u(x)\, p(x)(u(x) - \bar{u}) = \sum_x p(x)(u(x) - \bar{u})^2 = \mathrm{Var}_p(u) \geq 0,$$

and $\bar{u}'(\lambda) = \frac{d\bar{u}}{d\theta} \cdot \frac{d\theta}{d\lambda} = -\mathrm{Var}_p(u)/\lambda^2$. More noise means worse choices on average, measured by Robinson's own utility. In (a): $\bar{u} = 3.88, 3.68, 3.34, 2.61$ for $\lambda = 0.5, 1, 2, 10$. If λ is read as mistakes (not as taste shocks), this is the welfare cost of noisy choice.

(e) $(\kappa u(x) + m)/(\kappa\lambda) = u(x)/\lambda + m/(\kappa\lambda)$: every weight is multiplied by the same factor $e^{m/(\kappa\lambda)}$, which cancels. What the data reveal is the log-odds

$$\ln \frac{p(x \mid A)}{p(y \mid A)} = \frac{u(x) - u(y)}{\lambda},$$

i.e. u/λ up to an additive constant. So utility *differences* are revealed, relative to the noise: in logit, the statement "going from b to a is a smaller improvement than going from c to b" has a meaning (it is $1 < 3$, in units of λ), and it is testable. With deterministic choice, only the ranking is revealed. The cardinal information comes from the frequency of mistakes, and therefore from the assumed model of noise: with another noise distribution, the same frequencies give other differences.

(f) $\ln \frac{0.8}{0.2} = \ln 4$ and $\ln \frac{0.95}{0.05} = \ln 19$, so with $w = e^{u/\lambda}$ normalized by $w(c) = 1$: $w(b) = 19$, $w(a) = 76$. Hence $p(a, c) = 76/77 \approx 0.987$, and $p(\cdot \mid \{a, b, c\}) = (76, 19, 1)/96 \approx (0.792,\ 0.198,\ 0.010)$. Logit predicts all menus from two binary choices; any deviation in the data is a test of IIA.

Problem 6 Why Gumbel errors give logit

Challenge

Solution. (a) As $t \to -\infty$, $e^{-t} \to \infty$ and $F(t) \to e^{-\infty} = 0$; as $t \to \infty$, $e^{-t} \to 0$ and $F(t) \to e^0 = 1$. F is continuous and strictly increasing ($t \mapsto -e^{-t}$ is increasing, and so is $\exp$). Density: $f(t) = F'(t) = e^{-t}e^{-e^{-t}}$.

(b) For any t, by independence,

$$\mathbb{P}\Big(\max_i (v_i + \varepsilon_i) \le t\Big) = \prod_i F(t - v_i) = \exp\Big(-e^{-t}\sum_i e^{v_i}\Big) = \exp\big(-e^{-(t-\ln S)}\big) = F(t - \ln S),$$

with $S = \sum_i e^{v_i}$. This is the distribution function of $\ln S + \varepsilon$. The maximum of shifted Gumbels is a shifted Gumbel: this property is what makes the model tractable.

(c) Condition on $\varepsilon_x = t$: x wins iff $\varepsilon_y < t + v_x - v_y$, which has probability $F(t + d)$ with $d = v_x - v_y$. So

$$\mathbb{P}(x \text{ wins}) = \int_{-\infty}^{\infty} f(t)\, F(t + d)\, dt = \int_{-\infty}^{\infty} e^{-t} \exp\big(-e^{-t}\big) \exp\big(-e^{-t}e^{-d}\big)\, dt.$$

With $s = e^{-t}$, $ds = -e^{-t}dt$, and t from $-\infty$ to ∞ means s from ∞ to 0:

$$\mathbb{P}(x \text{ wins}) = \int_0^{\infty} e^{-s(1+e^{-d})}\, ds = \frac{1}{1 + e^{-d}} = \frac{e^{v_x}}{e^{v_x} + e^{v_y}}.$$

(d) x is chosen iff $v_x + \varepsilon_x > M$, where $M = \max_{y \in A\setminus\{x\}}(v_y + \varepsilon_y)$. M is independent of ε_x, and by (b) it has the distribution of $v_M + \varepsilon$ with $v_M = \ln \sum_{y \in A\setminus\{x\}} e^{v_y}$. By (c),

$$\mathbb{P}(x \text{ chosen}) = \frac{e^{v_x}}{e^{v_x} + e^{v_M}} = \frac{e^{v_x}}{e^{v_x} + \sum_{y \in A\setminus\{x\}} e^{v_y}} = \frac{e^{v_x}}{\sum_{y \in A} e^{v_y}}.$$

Ties have probability 0, because the ε have a density: $\mathbb{P}(v_x + \varepsilon_x = v_y + \varepsilon_y) = 0$.

(e) $U_x = u(x) + \lambda\varepsilon_x = \lambda(u(x)/\lambda + \varepsilon_x)$. Multiplying all utilities by $\lambda > 0$ does not change which one is largest, so we can apply (d) with $v_x = u(x)/\lambda$: $p(x \mid A) = e^{u(x)/\lambda} / \sum_{y \in A} e^{u(y)/\lambda}$. The noise has standard deviation $\lambda\pi/\sqrt{6} \approx 1.28\,\lambda$: λ is the scale of the noise, in units of utility.

(f) $\max_x U_x = \lambda \max_x \left(u(x)/\lambda + \varepsilon_x\right)$, which by (b) has the distribution of $\lambda(\ln \sum_x e^{u(x)/\lambda} + \varepsilon)$. Taking expectations gives $\lambda \ln \sum_{x \in A} e^{u(x)/\lambda} + \lambda\gamma$. Adding an option adds a positive term inside the logarithm, so the expected maximum rises: more options never hurt a random utility maximizer, even a bad option, because its random utility is sometimes the highest. As $\lambda \to 0$: with $u^* = \max_A u$ and k maximizers, $\lambda \ln \sum_x e^{u(x)/\lambda} = u^* + \lambda \ln \sum_x e^{-(u^* - u(x))/\lambda}$, and the last sum tends to k, so the expression tends to u^*; and $\lambda\gamma \to 0$. (This "log-sum" is the consumer-surplus formula used with logit demand.)

(g) Independence is used twice: in (b), to write the distribution function of the maximum as a product, and in (d), to treat M and ε_x as independent. Two identical buses should have the *same* random utility: $\varepsilon_{\text{red}} = \varepsilon_{\text{blue}}$. Independent errors treat them instead as two separate draws; adding the red bus then adds a fresh chance that "some bus" beats the car, and the bus share rises from $\frac{1}{2}$ to $\frac{2}{3}$. With perfectly correlated bus errors, the two buses act as one option, and the car keeps $\frac{1}{2}$. Models with correlated errors (probit, nested logit) drop independence, and with it IIA.

Problem 7 From choice frequencies to a choice correspondence

Core

Solution. (a) The thresholds are δ times the largest probability in the menu.

	$\{x,y\}$	$\{x,z\}$	$\{y,z\}$	$\{x,y,z\}$
p	$(0.6, 0.4)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.3, 0.2)$
C_1	$\{x\}$	$\{x\}$	$\{y\}$	$\{x\}$
$C_{1/2}$	$\{x,y\}$	$\{x\}$	$\{y,z\}$	$\{x,y\}$
C_0	$\{x,y\}$	$\{x,z\}$	$\{y,z\}$	$\{x,y,z\}$

For $\delta = \frac{1}{2}$: in $\{x,z\}$ the threshold is $0.35 > 0.3$, so z is dropped; in $\{x,y,z\}$ it is 0.25, so y (0.3) stays and z (0.2) is dropped. C_1 is the choice function of the ranking $x \succ y \succ z$: WARP holds. $C_{1/2}$ violates WARP: $z \in C(\{y,z\})$ and $y \in C(\{x,y,z\})$, so WARP requires $z \in C(\{x,y,z\})$, which fails. On pairs it coincides with C_2 of Problem 2: $x \sim y$, $y \sim z$, but $x \succ z$. $C_0(A) = A$ for every menu: WARP holds (a constant utility generates it), but this says nothing. The same data are "rational" or not depending on how we read them. The choice of δ is the analyst's, not the data's.

(b) If $p(x,y) \geq \frac{1}{2}$, then x is the most chosen option and belongs to C_δ; and $\frac{1}{2} \geq \delta/(1+\delta)$. If $p(x,y) < \frac{1}{2}$, the largest probability is $1 - p(x,y)$, and $x \in C_\delta$ iff $p(x,y) \geq \delta(1 - p(x,y))$, i.e. $p(x,y) \geq \delta/(1+\delta)$. On pairs, the rule is a fixed threshold: $\frac{1}{2}$ for $\delta = 1$, $\frac{1}{3}$ for $\delta = \frac{1}{2}$.

(c) With logit, $p(x \mid A)/\max_y p(y \mid A) = e^{(u(x)-u^*)/\lambda}$, with $u^* = \max_A u$. This is $\geq \delta$ iff $u^* - u(x) \leq \lambda \ln(1/\delta)$. So C_δ keeps the options within a utility distance $\lambda \ln(1/\delta)$ of the best. In the example the distance is $\ln 4 \approx 1.39$. Hence $C(\{a,b\}) = \{a,b\}$, $C(\{b,c\}) = \{b,c\}$, $C(\{a,c\}) = \{a\}$ (as $2 > 1.39$), and $C(\{a,b,c\}) = \{a,b\}$. WARP fails, as in (a): $c \in C(\{b,c\})$, $b \in C(\{a,b,c\})$, but $c \notin C(\{a,b,c\})$. This is the "just noticeable difference" of Problem Set 2: indifference within a fixed distance is not transitive. So even a logit chooser, the textbook model of a utility maximizer with noise, has an inferred correspondence that violates WARP, for every $\delta < 1$ and some utilities.

(d) $\mathbb{P}(K \leq 3) \approx 0.060 > 0.05$: we cannot reject, so x stays, although its frequency 15% is below $\frac{1}{3}$. $\mathbb{P}(K \leq 2) \approx 0.018 \leq 0.05$: x is dropped. $\mathbb{P}(K = 0) = (\frac{2}{3})^5 \approx 0.132$: x stays, although it was *never* chosen. The procedure keeps an option unless the data are strong evidence against it. With few observations, it is inclusive: we cannot tell an option chosen with probability $\frac{1}{3}$ from one never chosen after only five trials.

(e) By (b), $x \succ y$ in the inferred correspondence (i.e. $C(\{x,y\}) = \{x\}$) requires $p(x,y) > 1/(1+\delta)$, which is $\frac{2}{3}$ for $\delta = \frac{1}{2}$. A cycle of majorities with, say, $p(x,y) = p(y,z) = p(z,x) = 0.6$ gives $x \sim y \sim z \sim x$: all indifferences, which are transitive. Only a cycle of *strong* majorities gives a strict cycle. In finite samples, (d) adds more indifferences. Applying their method to Tversky's data, Balakrishnan et al. (2022) find that with $\delta = \frac{1}{2}$ and a 5% level most subjects (about 62%) have transitive inferred preferences. Cycles in frequencies (violations of weak stochastic transitivity, Lecture 3) are not the same as cycles in preferences.

References

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, 26(102):121–127.

Balakrishnan, N., Ok, E. A., and Ortoleva, P. (2022). Inference of choice correspondences. Working paper, version of December 7, 2022.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Sen, A. K. (1971). Choice functions and revealed preference. *Review of Economic Studies*, 38(3):307–317.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.