Online Appendix: Stochastic choice analysis with mismeasured data *

Henrik Petri

November 16, 2021

Abstract

This online appendix contains the rather lengthy proof of prosposition 8.4 in Petri (2021).

0.1 Proof of proposition 6.3

In the following proofs we will use I as a shorthand for the set $\{1, ..., n\}$.

Lemma 0.1. *Let $A \subseteq X$. If it holds that*

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$$

for all $P \in \mathscr{P}$ then there is a $P \in \mathscr{P}$ with

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P).$$

Proof. Since $\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \leq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$ for all $P \in \mathscr{P}$ it suffices to show that there is a $P \in \mathscr{P}$ with $\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) \geq \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P)$. To show this we note that $\sum_{\{i \in I: a_i = b\}} k_i - \sum_{\{i \in I: b_i = b\}} k_i \geq 0$ for some $b \in \{a_1, ..., a_n\}$. If not, then $\sum_{i \in I} k_i = \sum_{a \in \{a_1, ..., a_n\}} \sum_{\{i \in I: a_i = a\}} k_i < \sum_{a \in \{a_1, ..., a_n\}} \sum_{\{i \in I: b_i = a\}} k_i \leq \sum_{i \in I} k_i$. A contradiction. Let b be such that $\sum_{\{i \in I: a_i = b\}} k_i - \sum_{\{i \in I: b_i = b\}} k_i \geq 0$.Then if $P \in \mathscr{P}(b, X)$ it follows that

$$\sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(a_i, A)}(P) = \sum_{\{i \in I: a_i = b\}} k_i \geq \sum_{\{i \in I: b_i = b\}} k_i = \sum_{i \in I} k_i \mathbf{1}_{\mathscr{P}(b_i, A)}(P).$$

The claim follows. □

*Department of Economics, University of Bath, 3 East, Bath BA2 7AY, UK, henrik@petri.se

Lemma 0.2. *If it holds that*

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P\in\mathscr{P}$ then there is a $P\in\mathscr{P}$ with

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P).$$

Proof. The proof is by induction on the number of distinct sets in $A_1,\dots,A_n$. The base case $A_1=\dots.=A_n=A$ for some $A\subseteq X$ follows directly by lemma 0.1. As induction hypothesis assume that the claim is true for all sequences $A_1,\dots,A_n$ with $k-1$ distinct sets. Assume that $A_1,\dots,A_n$ is a sequence with k distinct sets such that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P\in\mathscr{P}$. Let $b\in\arg\max_{a\in\{a_1,\dots,a_n\}}\left(\sum_{\{i\in I:a_i=a\}}k_i-\sum_{\{i\in I:b_i=a\}}k_i\right)$ then we have that $\sum_{\{i\in I:a_i=b\}}k_i-\sum_{\{i\in I:b_i=b\}}k_i\ge 0$. If not, then $\sum_{i\in I}k_i=\sum_{a\in\{a_1,\dots,a_n\}}\sum_{\{i\in I:a_i=a\}}k_i<\sum_{a\in\{a_1,\dots,a_n\}}\sum_{\{i\in I:b_i=a\}}k_i\le\sum_{i\in I}k_i$. A contradiction.

Thus $\sum_{\{i\in I:a_i=b\}}k_i-\sum_{\{i\in I:b_i=b\}}k_i\ge 0$. Let $I(2)=\{i\in I: b\notin A_i\}$ and note that $\sum_{i\in I\setminus I(2)}k_i\mathbf{1}_{\mathscr{P}(a_i,A_i)}(P)-\sum_{i\in I\setminus I(2)}k_i\mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)\ge 0$ for all $P\in\mathscr{P}(b,X)$. Further, by assumption, it holds that

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P\in\mathscr{P}$. Rearranging the above expression we obtain

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \le \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) - \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le 0$$

for all $P\in\mathscr{P}(b,X)$. This in particular implies that $\sum_{i\in I(2)}k_i\mathbf{1}_{\mathscr{P}(a_i,A_i)}(P)-\sum_{i\in I(2)}k_i\mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)\le 0$ for all $P\in\mathscr{P}$. Since $(A_i)_{i\in I(2)}$ is a sequence with $k-1$ distinct sets it follows by the induction hypothesis that there is a $P\in\mathscr{P}$ with $\sum_{i\in I(2)}k_i\mathbf{1}_{\mathscr{P}(a_i,A_i)}(P)=\sum_{i\in I(2)}k_i\mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$. Since $b\notin A_i$ for any $i\in I(2)$ we w.l.o.g. assume that $P\in\mathscr{P}(b,X)$. Using this we obtain that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \ge$$

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P).$$

The claim follows.

□

Lemma 0.3. *If it holds that*

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$ then it follows that

$$\sum_{\{i\in I: a_i=b\}} k_i - \sum_{\{i\in I: b_i=b\}} k_i = 0$$

for all $b \in X$. This further implies that $\{a_1,...,a_n\} = \{b_1,...,b_n\}$.

Proof. Let

$$b \in \operatorname*{argmax}_{a\in\{a_1,...,a_n\}} \left(\sum_{\{i\in I: a_i=a\}} k_i - \sum_{\{i\in I: b_i=a\}} k_i \right)$$

then we have that $\sum_{\{i\in I: a_i=b\}} k_i - \sum_{\{i\in I: b_i=b\}} k_i \geq 0$.

CASE 1: Assume that $\sum_{\{i\in I: a_i=b\}} k_i - \sum_{\{i\in I: b_i=b\}} k_i > 0$.

Let $I(2) = I \backslash \{i \in I : b \in A_i\}$. Then it follows that $\sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) - \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) < 0$ for all $P \in \mathscr{P}(b, X)$. Further, by assumption, it holds that

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$. Rearranging the above expression we obtain

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \leq \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) - \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) < 0$$

for all $P \in \mathscr{P}(b, X)$. This in particular implies that $\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) - \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) \leq 0$ for all $P \in \mathscr{P}$. It folllows that there is a $P \in \mathscr{P}$ with $\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$. Since $b \notin A_i$ for any $i \in I(2)$ we w.l.o.g. assume that $P \in \mathscr{P}(b, X)$. Using this we obtain that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) + \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) >$$

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) + \sum_{i\in I\backslash I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P).$$

A contradiciton.

CASE 2: If $\sum_{\{i\in I: a_i=b\}} k_i - \sum_{\{i\in I: b_i=b\}} k_i = 0$ then $\sum_{\{i\in I: a_i=a\}} k_i - \sum_{\{i\in I: b_i=a\}} k_i = 0$ for all $a \in \{a_1,...,a_n\}$. If not, then

$$\sum_{i\in I} k_i = \sum_{a\in\{a_1,...,a_n\}} \sum_{\{i\in I: a_i=a\}} k_i < \sum_{a\in\{a_1,...,a_n\}} \sum_{\{i\in I: b_i=a\}} k_i \leq \sum_{i\in I} k_i.$$

Finally, we note that $\{a_1,...,a_n\} = \{b_1,...,b_n\}$, since if $b \in \{a_1,...,a_n\} \backslash \{b_1,...,b_n\}$ then $\sum_{\{i\in I: a_i=b\}} k_i -$

$\sum_{\{i\in I:b_i=b\}} k_i > 0$ and if $b \in \{b_1,...,b_n\}\setminus\{a_1,...,a_n\}$ then $\sum_{\{i\in I:a_i=b\}} k_i - \sum_{\{i\in I:b_i=b\}} k_i < 0$. A contradiction again.

□

Lemma 0.4. *Let $A \subseteq X$. If it holds that*

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A)}(P) \le \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A)}(P)$$

for all $P \in \mathscr{P}$ then

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A)}(P).$$

for all $P \in \mathscr{P}$.

Proof. This follows immediately by lemma 0.3. □

Lemma 0.5. *Let $A_1,...,A_n \subseteq X$. If it holds that*

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$ then

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$.

Proof. The claim is proven by induction on the number of distinct sets in the sequence $A_1,...,A_n$ (i.e. on the cardinality of $\{A_1,...,A_n\}$). The base case follows by lemma 0.5 above. Assume that the claim is true for all sequences with $k-1$ distinct sets in $A_1,...,A_n$. Assume that $A_1,...,A_n$ is a sequence with k distinct sets such that

$$\sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$. By lemma 0.3 it follows that

$$\sum_{\{i\in I:a_i=b\}} k_i - \sum_{\{i\in I:b_i=b\}} k_i = 0$$

for all $b \in X$ and moreover $\{a_1,...,a_n\} = \{b_1,...,b_n\}$.

There are $i, j \in \{1,...,n\}$ with $A_i \setminus A_j \neq \emptyset$. Let $b \in A_i \setminus A_j$ and consider the set of orders $P \in \mathscr{P}$ with b as their top element in X, i.e. $\mathscr{P}(b,X)$.

Let $I(2) = \{i \in I(1) : b \notin A_i\}$. If $b \notin \{a_1,...,a_n\}$ then $\sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = 0 = \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$. If $b \in \{a_1,...,a_n\}$ then $\sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{\{i\in I:a_i=b\}} k_i = \sum_{\{i\in I:b_i=b\}} k_i = \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$. It hence follows that

$$\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \le \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}(b, X)$.

But since $b \notin A_i$ for any $i \in I(2)$ this implies that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$.

Since $(A_i)_{i \in I(2)}$ is a sequence with $k-1$ distinct sets it follows by the induction hypotheis that $\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$ for all $P \in \mathscr{P}$.

This further implies that

$$\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. Since there are less than $k-1$ sets in the sequence $(A_i)_{i \in I \setminus I(2)}$ it follows by the induction hypothesis that $\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) = \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$ for all $P \in \mathscr{P}$. This proves the claim. □

Finally, we are ready for the proof of proposition 8.4 in Petri (2021).

Proof. Define a relation $\succsim$ on $X \times \mathscr{A}$ by $(a, A) \succsim (b, B)$ if and only if $\rho(a, A) \geq \rho(b, B)$ and $A = B$. We check that $\succsim$ satisfies $\mathscr{P}$-cancellation. Let $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ be such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ and assume that :

$$\sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i=1}^{n} k_i \mathbf{1}_{\mathscr{P}(b_i, B_i)}(P)$$

for all $P \in \mathscr{P}$.

Since $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ it follows that $A_i = B_i$ for all $i \in \{1, \ldots, n\}$ by definition of $\succsim$. We prove the claim by induction on the number of distinct sets in the sequence $A_1, \ldots, A_n$. The base case is clear. Assume that the claim holds when there are $k-1$ distinct sets in the sequence. Let $A_1, \ldots, A_n$ be a sequence with k distinct sets. Hence there are $i, j \in \{1, \ldots, n\}$ with $A_i \setminus A_j \neq \emptyset$. Let $b \in A_i \setminus A_j$ and consider the set of orders $P \in \mathscr{P}$ with b as their top element in X, i.e. $\mathscr{P}(b, X)$. Let $I(2) = \{i \in I : b \notin A_i\}$ and note that

$$\sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) - \sum_{i \in I \setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P) = \sum_{\{i \in I : a_i = b\}} k_i - \sum_{\{i \in I : b_i = b\}} k_i = 0$$

for all $P \in \mathscr{P}(b, X)$, where the first equality follows since $P \in \mathscr{P}(b, X)$ and the second equality follows by lemma 0.3. Using this, it then follows that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}(b, X)$. But since $b \notin A_i$ for any $i \in I(2)$ we futher have that

$$\sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i \in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i, A_i)}(P)$$

for all $P \in \mathscr{P}$. If $n \in I(2)$ the claim follows by induction hypothesis. Otherwise, note that lemma 0.5 implies that $\sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) = \sum_{i\in I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$ for all $P \in \mathscr{P}$. This further implies that

$$\sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(a_i,A_i)}(P) \leq \sum_{i\in I\setminus I(2)} k_i \mathbf{1}_{\mathscr{P}(b_i,A_i)}(P)$$

for all $P \in \mathscr{P}$. Since there are less than $k-1$ sets in the sequence $(A_i)_{i\in I\setminus I(2)}$ it follows by the induction hypothesis that $(b_n, A_n) \succsim (a_n, A_n)$. As we wanted to show. □

References

H. Petri. Stochastic choice analysis with mismeasured data. *Working Paper*, 2021.