Supplementary Material

A. Missing Proofs

A.1. Proof of Lemma 11

Proof. (i) To show this, we do a case analysis. Suppose there exists some $j < j'' \leq j'$ such that $\frac{w_{i_{j''}}}{w_{i_{j''-1}}} \geq \frac{2n}{\epsilon}$. Then, $\rho_{j''} \geq \log_{1+\epsilon/(2n)} \frac{2n}{\epsilon}$ and

$$\frac{\widetilde{w}_{i_j}}{\widetilde{w}_{i_{j'}}} = \left(1 + \frac{\epsilon}{2n}\right)^{-\sum_{\ell=j+1}^{j'} \rho_\ell} \leq \left(1 + \frac{\epsilon}{2n}\right)^{-\rho_{j''}} \leq \frac{\epsilon}{2n}.$$

Otherwise, for each $\ell \in \{j+1, \ldots, j'\}$, it holds $\rho_\ell \geq \log_{1+\epsilon/(2n)} \frac{w_{i_\ell}}{w_{i_{\ell-1}}}$. Thus,

$$\frac{\widetilde{w}_{i_j}}{\widetilde{w}_{i_{j'}}} = \left(1 + \frac{\epsilon}{2n}\right)^{-\sum_{\ell=j+1}^{j'} \rho_\ell} \leq \prod_{\ell=j+1}^{j'} \frac{w_{i_{\ell-1}}}{w_{i_\ell}} = \frac{w_{i_j}}{w_{i_{j'}}} \leq \frac{\epsilon}{2n}.$$

(ii) Suppose for some $j < j'$, it holds $\frac{w_{i_j}}{w_{i_{j'}}} > \frac{\epsilon}{2n}$. Then, for each $\ell \in \{j+1, \ldots, j'\}$, it holds

$$\log_{1+\epsilon/(2n)} \frac{w_{i_\ell}}{w_{i_{\ell-1}}} \leq \rho_\ell \leq 1 + \log_{1+\epsilon/(2n)} \frac{w_{i_\ell}}{w_{i_{\ell-1}}}.$$

Thus,

$$\frac{\widetilde{w}_{i_j}}{\widetilde{w}_{i_{j'}}} = \left(1 + \frac{\epsilon}{2n}\right)^{-\sum_{\ell=j+1}^{j'} \rho_\ell} \leq \prod_{\ell=j+1}^{j'} \frac{w_{i_{\ell-1}}}{w_{i_\ell}} = \frac{w_{i_j}}{w_{i_{j'}}},$$

and

$$\begin{aligned}
\frac{\widetilde{w}_{i_j}}{\widetilde{w}_{i_{j'}}} &= \left(1 + \frac{\epsilon}{2n}\right)^{-\sum_{\ell=j+1}^{j'} \rho_\ell} \geq \left(1 + \frac{\epsilon}{2n}\right)^{j-j'} \cdot \prod_{\ell=j+1}^{j'} \frac{w_{i_{\ell-1}}}{w_{i_\ell}} \\
&= \left(1 + \frac{\epsilon}{2n}\right)^{j-j'} \cdot \frac{w_{i_j}}{w_{i_{j'}}} \geq \left(1 + \frac{\epsilon}{2n}\right)^{-n} \cdot \frac{w_{i_j}}{w_{i_{j'}}} \\
&= \left(\left(1 + \frac{\epsilon}{2n}\right)^{-\frac{2n}{\epsilon}}\right)^{\epsilon/2} \cdot \frac{w_{i_j}}{w_{i_{j'}}} \geq e^{-\epsilon/2} \cdot \frac{w_{i_j}}{w_{i_{j'}}} \\
&\geq \left(1 - \frac{\epsilon}{2}\right) \cdot \frac{w_{i_j}}{w_{i_{j'}}}.
\end{aligned}$$

□

A.2. Proof of Theorem 12

Proof. Let $L^{(1)}, \ldots, L^{(t)}$ be the t MNLs in the support of L and for $i \in [t]$, let $p_i > 0$ be the mixture probability that L assigns to $L^{(i)}$. We will first obtain a succinct representation of the mixture probabilities.

For each $i \in [t]$, let $\tau_i = \frac{\lceil \epsilon^{-1} n t^2 p_i \rceil}{\epsilon^{-1} n t^2}$. Note that $\epsilon^{-1} n t^2 \tau_i$ is a positive integer satisfying $1 \leq \epsilon^{-1} n t^2 \tau_i \leq \epsilon^{-1} n t^2 + 1$. Thus, each τ_i can be represented using $O\left(\log \frac{nt}{\epsilon}\right) = O\left(\log \frac{n}{\epsilon}\right)$ bits. Define $\widetilde{p}_i = \frac{\tau_i}{\sum_{j \in [t]} \tau_j}$. It is easy to see that the total number of bits needed to represent all the $\widetilde{p}_i$'s is $O\left(t \log \frac{n}{\epsilon}\right)$. For each $i \in [t]$, we have

$$p_i \leq \tau_i < \frac{\epsilon^{-1} n t^2 p_i + 1}{\epsilon^{-1} n t^2} = p_i + \frac{\epsilon}{nt^2}.$$

We will show that $|\widetilde{p}_i - p_i| \leq \frac{\epsilon}{nt}$, for each $i \in [t]$. Indeed,

$$\widetilde{p}_i - p_i = \frac{\tau_i}{\sum_{j \in [t]} \tau_j} - p_i \leq \frac{p_i + \frac{\epsilon}{nt^2}}{\sum_{j \in [t]} p_j} - p_i = p_i + \frac{\epsilon}{nt^2} - p_i = \frac{\epsilon}{nt^2}, \text{ and}$$

$$\widetilde{p}_i - p_i \geq \frac{p_i}{\sum_{j \in [t]} (p_j + \frac{\epsilon}{nt^2})} - p_i = \frac{p_i}{1 + \frac{\epsilon}{nt}} - p_i \geq \frac{p_i - p_i - \frac{p_i \epsilon}{nt}}{1 + \frac{\epsilon}{nt}} \geq \frac{-\frac{\epsilon}{nt}}{1 + \frac{\epsilon}{nt}} \geq -\frac{\epsilon}{nt}.$$

From Theorem 10, we have for each $i \in [t]$, an MNL $\widetilde{L}^{(i)}$ that can be represented by $O\left(n \log \frac{n^2}{\epsilon}\right) = O\left(n \log \frac{n}{\epsilon}\right)$ bits and

$\mathrm{dist}(L^{(i)}, \widetilde{L}^{(i)}) \le \frac{\epsilon}{n}$.

Now, let $\widetilde{L}$ be the mixture over the MNLs $\widetilde{L}^{(1)}, \ldots, \widetilde{L}^{(t)}$ assigning probability $\widetilde{p}_i$ to $\widetilde{L}^{(i)}$ for $i \in [t]$. By the above bounds, this mixture can be represented using $t \cdot \left(O(\log \frac{n}{\epsilon}) + O(n \log \frac{n}{\epsilon})\right) = O(tn \log \frac{n}{\epsilon})$ bits. To show $\mathrm{dist}(L, \widetilde{L}) \le \epsilon$, consider a slate $S \subseteq [n]$:

$$
\begin{aligned}
|\widetilde{L}_S - L_S|_{\mathrm{tv}} &= \frac{1}{2}\sum_{s\in S} |\widetilde{L}_S(s) - L_S(s)| = \frac{1}{2}\sum_{s \in S}\left|\sum_{i=1}^{t}\left(\widetilde{p}_i \widetilde{L}_S^{(i)}(s) - p_i L_S^{(i)}(s)\right)\right| \\
&= \frac{1}{2}\sum_{s\in S}\left|\sum_{i=1}^{t}\left(p_i\left(\widetilde{L}_S^{(i)}(s) - L_S^{(i)}(s)\right)\right) + \sum_{i=1}^{t}\left((\widetilde{p}_i - p_i)\widetilde{L}_S^{(i)}(s)\right)\right| \\
&\le \frac{1}{2}\sum_{s\in S}\left(\left|\sum_{i=1}^{t}\left(p_i\left(\widetilde{L}_S^{(i)}(s) - L_S^{(i)}(s)\right)\right)\right| + \left|\sum_{i=1}^{t}\left((\widetilde{p}_i - p_i)\widetilde{L}_S^{(i)}(s)\right)\right|\right) \\
&\le \frac{1}{2}\sum_{s\in S}\left(\frac{\epsilon}{n} + \left|\sum_{i=1}^{t}\left((\widetilde{p}_i - p_i)\widetilde{L}_S^{(i)}(s)\right)\right|\right),
\end{aligned}
$$

where the last step follows from $\sum_{i=1}^{t} p_i = 1$ and $|\widetilde{L}_S^{(i)}(s) - L_S^{(i)}(s)| \le \frac{\epsilon}{n}$. Then, using $|S| \le n$,

$$
\begin{aligned}
|\widetilde{L}_S - L_S|_{\mathrm{tv}} &\le \frac{1}{2}\sum_{s\in S}\left(\frac{\epsilon}{n} + \left|\sum_{i=1}^{t}\left((\widetilde{p}_i - p_i)\widetilde{L}_S^{(i)}(s)\right)\right|\right) \\
&\le \frac{\epsilon}{2} + \frac{1}{2}\sum_{s\in S}\left|\sum_{i=1}^{t}\left((\widetilde{p}_i - p_i)\widetilde{L}_S^{(i)}(s)\right)\right| \le \frac{\epsilon}{2} + \frac{1}{2}\cdot\epsilon = \epsilon,
\end{aligned}
$$

where the penultimate step follows from $|\widetilde{p}_i - p_i| \le \frac{\epsilon}{nt}$, $0 \le \widetilde{L}_S^{(i)}(s) \le 1$, and $|S| \le n$. □

A.3. Proof of Theorem 14

Proof. If $\widetilde{U}$ is supported by t permutations, then $\widetilde{U}$ can produce at most t different winners for any slate S, i.e., $|\mathrm{supp}(\widetilde{U}_S)| \le t$. Let $S = [n]$ and $S' = \mathrm{supp}(\widetilde{U}_S)$; then, $|S'| \le t < o(n)$. Moreover,

$$
|\widetilde{U}_S - U_S|_{\mathrm{tv}} = \max_{T \subseteq S} |\widetilde{U}_S(T) - U_S(T)| \ge |\widetilde{U}_S(S') - U_S(S')| = \left|1 - \frac{|S'|}{n}\right| \ge 1 - \frac{t}{n} > 1 - o(1). \quad \square
$$

A.4. Proof of Theorem 15

Proof. Having access to D, we can take $t = \left\lceil \left(3k\ln(3n) + 3\ln\frac{2}{\delta}\right)\cdot\epsilon^{-2}\right\rceil$ independent samples (i.e., permutations) $\pi_1, \ldots, \pi_t$ from D. After this first step, we let $\widetilde{D}$ be the uniform distribution on the multiset of these samples, i.e., $\widetilde{D}$ will choose $i \in [t]$ uniformly at random, and will return π_i.

Consider any slate $S \subseteq [n]$, and any of its non-empty subsets $S' \subseteq S$. Let $\widetilde{D}_S(S') = \sum_{s\in S'} \widetilde{D}_S(s)$ be the probability that, using the distribution $\widetilde{D}$, the maximum element in the subset S belongs to S'. Then, $\widetilde{D}_S(S') \in [0, 1]$ is a random variable. Clearly, $\mathrm{E}[\widetilde{D}_S(S')] = \sum_{s\in S'} \mathrm{E}[\widetilde{D}_S(s)] = \sum_{s\in S} D_S(s) = D_S(S')$, which is the probability that the maximum element in the slate S with distribution D belongs to S'. Observe, further, that

$$
|D_S - \widetilde{D}_S|_{\mathrm{tv}} = \max_{\varnothing \ne S' \subseteq S} |D_S(S') - \widetilde{D}_S(S')|,
$$

thus, if we show that, with probability at least $1-\delta$, for each $\varnothing \ne S' \subseteq S$ it holds that $|D_S(S') - \widetilde{D}_S(S')| \le \epsilon$, we have shown the main claim.

Pick any slate $S' \subseteq S \subseteq [n]$. Then, by a Chernoff bound,

$$
\Pr\left[|D_S(S') - \widetilde{D}_S(S')| \ge \epsilon\right] \le 2e^{-\frac{\epsilon^2}{3}\cdot t} \le 2e^{-\frac{\epsilon^2}{3}\cdot\left(3k\ln(3n) + 3\ln\frac{2}{\delta}\right)\cdot\epsilon^{-2}} = 2e^{-k\ln(3n) - \ln\frac{2}{\delta}} = \delta\cdot\frac{1}{3^k\cdot n^k}.
$$

Observe that there are at most $\binom{n}{k} \le n^k$ subsets T of size k of $[n]$; for each such T there are at most $3^{|T|} = 3^k$ pairs of slates $S' \subseteq S \subseteq T$ (the generic element can either be part of S', or of $S \setminus S'$, or of $T \setminus S$). Thus,

$$
\Pr\left[\exists \varnothing \ne S' \subseteq S \subseteq [n] \wedge |S| \le k : |D_S(S') - \widetilde{D}_S(S')| \ge \epsilon\right] \le \delta\cdot n^k\cdot 3^k\cdot\frac{1}{3^k\cdot n^k} = \delta. \quad \square
$$

A.5. Proof of Theorem 17

Proof. Let $n = k \cdot t$. We start by partitioning $[n]$ into $P_1 = [k], \ldots, P_i = [ik] \setminus [(i-1)k], \ldots, P_t = [tk] \setminus [(t-1)k]$. Then for $i \in [t]$, let $D^{(i)}$ be any RUM on P_i. We create a RUM D on $[n]$ by sampling a permutation π as follows: for each $i \in [t]$, sample independently $\pi_i \sim D^{(i)}$ and then return the concatenated permutation $\pi = \pi_1 \cdots \pi_t$.

Since for each $i \in [t]$, the RUM $D^{(i)}$ is defined over a set of k items, by Corollary 7, the number of bits required to represent $D^{(i)}$ to within a total variation distance of $\frac{1-\alpha}{4}$ is at least $\ell = \frac{\alpha^3}{6} \cdot k^2 - O(1)$. Note that for each $i \in [t]$, for each slate $S \subseteq P_i$, $D_S^{(i)} = D_S$, i.e., they behave identically.

Moreover, the $D^{(i)}$'s are over disjoint sets of items. Since each $D^{(i)}$ requires at least ℓ bits to be represented with maximum total variation distance of $\frac{1-\alpha}{4}$, and since each of its slates has size at most k, D requires at least $t \cdot (\ell - 1) = \frac{\alpha^3}{6} \cdot n \cdot k - O(n/k)$ bits to be represented with a maximum total variation distance of $\frac{1-\alpha}{4}$ on slates of size up to k. $\square$

B. A Lower Bound on the Size of an Exact Representation

In this section, we show that if one aims to represent a RUM exactly—in fact, even just a RUM that chooses u.a.r. from its support—one needs at least $2^{n-o(n)}$ bits, i.e., exponentially many bits.[10] Thus, exactly representing a RUM has a cost that is exponentially larger than that of approximating it to within ϵ total variation distance.

Let $n \geq 2$ be even. Consider the class F_n of functions $f : \binom{[n]}{n/2} \to [n]$ such that $f(S) \in S$ for each $S \in \binom{[n]}{n/2}$. Then, $|F_n| = (n/2)^{\binom{n}{n/2}}$.

We will show that, for each $f \in F_n$, there exists a RUM D (that chooses a permutation u.a.r. from its support) such that, for each $S \in \binom{[n]}{n/2}$, it holds that $D_S(f(S))$ is slightly larger than $D_S(i)$ for each $i \in S \setminus \{f(S)\}$. Thus, a perfect representation of D has to record $f(S)$ for each $S \in \binom{[n]}{n/2}$; this implies that one needs $\lg_2 |F_n| = 2^{\Omega(n)}$ bits to perfectly represent a RUM (even if the RUM is assumed to choose u.a.r. in its support).

Lemma 18. *Let $n \geq 2$ be even. For each $f \in F_n$, there exists a set $T^{(f)} \subseteq \mathbf{S}_n$ such that the RUM $D^{(f)}$ that chooses a permutation uniformly at random in $T^{(f)}$ satisfies, for each $S \in \binom{[n]}{n/2}$, (i) $D_S^{(f)}(f(S)) = \frac{2}{n} + \frac{1-\frac{2}{n}}{\binom{n}{n/2}}$, and (ii) for each $j \in S \setminus \{f(S)\}$, $D_S^{(f)}(j) = \frac{2}{n} - \frac{2}{n\binom{n}{n/2}}$.*

Proof. Let $T^{(f)} \subseteq \mathbf{S}_n$ be defined as:

$$T^{(f)} = \left\{ \pi \mid \text{if } S \text{ is the set of elements having ranks } \left\{ \frac{n}{2} + 1, \ldots, n \right\} \text{ in } \pi, \text{ then } f(S) \text{ has rank } n/2 + 1 \text{ in } \pi \right\}.$$

Now consider the generic set $S \in \binom{[n]}{n/2}$. Let ξ_S be the event that, if π is sampled from $D^{(f)}$ (that is, if π is chosen u.a.r. in $T^{(f)}$), then the elements of S are those in the $n/2$ lowest ranks of π. Then,

$$\Pr[\xi_S] = \frac{(n/2)! \cdot (n/2-1)!}{\binom{n}{n/2} \cdot (n/2)! \cdot (n/2-1)!} = \binom{n}{n/2}^{-1}.$$

Now, if $\pi \sim D^{(f)}$, then

$$\begin{aligned} D_S^{(f)}(i) &= \Pr[\pi(S) = i] = \Pr[\pi(S) = i \mid \xi_S] \cdot \Pr[\xi_S] + \Pr\left[\pi(S) = i \mid \overline{\xi}_S\right] \cdot \Pr\left[\overline{\xi}_S\right] \\ &= [i = f(S)] \cdot \Pr[\xi_S] + \frac{1}{|S|} \cdot \Pr\left[\overline{\xi}_S\right] = \frac{[i = f(S)]}{\binom{n}{n/2}} + \frac{1 - \binom{n}{n/2}^{-1}}{|S|}, \end{aligned}$$

thus, $D_S^{(f)}(f(S)) = \frac{2}{n} + \frac{1-2/n}{\binom{n}{n/2}}$, and $D_S^{(f)}(j) = \frac{2}{n} - \frac{2/n}{\binom{n}{n/2}}$ for each $j \in S \setminus \{f(S)\}$. $\square$

This yields the following.

[10] We notice that this claim is trivial for RUMs that do not choose u.a.r. over their support. I.e., if $n = 2$, so that there are only 2 permutations $\pi_1 = (1 \succ 2)$ and $\pi_2 = (2 \succ 1)$, the RUM D that assigns probability p to π_1 will have $D_{\{1,2\}}(1) = p$. Thus, to exactly represent the winning distributions of D, one needs to exactly represent p; since p is an arbitrary real number in $[0, 1]$, no finite number of bits is sufficient to exactly represent p, and the RUM.

Corollary 19. *Consider the class of RUMs that choose u.a.r. from their supports. A data structure that can exactly represent the winning distributions of such a RUM on each slate, requires at least* $\log_2 |F_n| = \Omega\left(2^n \cdot n^{-1/2} \cdot \log n\right)$ *bits.*

Proof. For each $\{f, f'\} \in \binom{F_n}{2}$, there exists at least one set $S \in \binom{[n]}{n/2}$, such that $D_S^{(f)} \neq D_S^{(f')}$. Thus, a data structure needs at least $\log_2 |F_n|$ bits to represent the winning distribution of each slate. □

C. Relationships between RUMs and other Choice Models

C.1. RUMs vs PCMC Models

Ragain & Ugander (2016) show that there exist PCMC models that are not RUMs. We observe here that the converse is also true. Recall that a PCMC model is defined by an $n \times n$ matrix Q satisfying $Q_{i,j} + Q_{j,i} > 0$, for each $\{i, j\} \in \binom{[n]}{2}$. Given a slate S, the distribution of the winner of S is the stationary distribution of the continuous-time Markov chain on state space S, and transition rates $q_{i,j} = Q_{i,j}$ for each $i \in S$ and $j \in S \setminus \{i\}$.

The RUM we will be using in our example is somewhat natural. It considers two types of users, and three types of items, e.g., $\{1, 2, 3\}$. The first type of users strongly prefers 1 over the others, and has no strong preference between 2 and 3; the second type of users strongly prefers 2 over the others, and has no strong preference between 1 and 3.[11] We will use a uniform mixture of the two user types.

Observation 20. *There exists a RUM on* $n = 3$ *elements that cannot be represented with PCMC models.*

Proof. Let $n = 3$, and consider the RUM R that chooses a permutation u.a.r. from $\{1 \succ 2 \succ 3, 1 \succ 3 \succ 2, 2 \succ 1 \succ 3, 2 \succ 3 \succ 1\}$; that is, the RUM R chooses a permutation uniformly at random, conditioned on its highest-ranked element to not be 3. Then, $R_{[3]}(3) = 0$ and $R_{[3]}(1) = R_{[3]}(2) = 1/2$; $R_{[2]}(1) = R_{[2]}(2) = 1/2$; $R_{\{1,3\}}(3) = R_{\{2,3\}}(3) = 1/4$ and $R_{\{1,3\}}(1) = R_{\{2,3\}}(2) = 3/4$. We will show that these choice distributions cannot be represented with PCMC models.

Consider the generic Q matrix of a PCMC model for $n = 3$ items:

$$Q = \begin{pmatrix} \cdot & p_{1,2} & p_{1,3} \\ p_{2,1} & \cdot & p_{2,3} \\ p_{3,1} & p_{3,2} & \cdot \end{pmatrix}.$$

The winning distributions for the slate $\{1, 2\}$ according to the PCMC model on Q is given by the solution to the following linear system:

$$(x_1, x_2) \cdot \begin{pmatrix} -p_{1,2} & p_{1,2} & 1 \\ p_{2,1} & -p_{2,1} & 1 \end{pmatrix} = (0, 0, 1).$$

If (x_1, x_2) is the solution, then x_i is the probability that i wins in $[2]$. The first two constraints of the system are equivalent, and the third constraint forces $x_2 = 1 - x_1$. Thus, the system simplifies to $p_{1,2}x_1 = p_{2,1}x_2$; i.e., $p_{1,2}x_1 = p_{2,1}(1 - x_1)$, which implies $x_1 = p_{2,1}/(p_{1,2} + p_{2,1})$, and $x_2 = p_{1,2}/(p_{1,2} + p_{2,1})$. Thus, for a PCMC model to represent the RUM R (and thus to guarantee that the choice in [2] is u.a.r.), one must have have $p_{1,2} = p_{2,1}$.

Now, consider the winning distribution for the slate $\{i, 3\}$, for $i \in [2]$. The winning distribution of PCMC model then is the solution to

$$(x_i, x_3) \cdot \begin{pmatrix} -p_{i,3} & p_{i,3} & 1 \\ p_{3,i} & -p_{3,i} & 1 \end{pmatrix} = (0, 0, 1).$$

As before, we obtain that the unique solution satisfies $x_i = p_{3,i}/(p_{3,i} + p_{i,3})$ and $x_3 = p_{i,3}/(p_{3,i} + p_{i,3})$. For the PCMC model to represent R, one must then have $p_{i,3} = p_{3,i}/3$, since $R_{\{i,3\}}(3) = 1/4$ and $R_{\{i,3\}}(i) = 3/4$.

A PCMC model Q' that satisfies all the constraints given by the slates of size 2 can then be represented as

$$Q' = \begin{pmatrix} \cdot & q & q_1 \\ q & \cdot & q_2 \\ 3q_1 & 3q_2 & \cdot \end{pmatrix},$$

[11]For instance, 1 could be a sports event, 2 could be a concert, and 3 could be a pub—sports fans would strongly prefer 1, music fans would strongly prefer 2, but both types of users would be happy to choose uniformly at random over the remaining available choices.

where $q = p_{1,2} = p_{2,1}$, $q_1 = p_{1,3}$ and $q_2 = p_{2,3}$. By the positivity constraints of PCMC models, we have that $q + q > 0$, $q_1 + 3q_1 > 0$ and $q_2 + 3q_2 > 0$, that is, $q, q_1, q_2 > 0$. That is, each entry of Q' outside the main diagonal has to be strictly positive.

Now, let us consider the full slate $[3] = \{1, 2, 3\}$. Its winning distribution, if the constraints of the 2-slates are satisfied, is the solution to the following linear system

$$(x_1, x_2, x_3) \cdot \begin{pmatrix} -q - q_1 & q & q_1 & 1 \\ q & -q - q_2 & q_2 & 1 \\ 3q_1 & 3q_2 & -3(q_1 + q_2) & 1 \end{pmatrix} = (0, 0, 0, 1).$$

The first and second constraints are $3q_1x_3 - q_1x_1 + q(x_2 - x_1) = 0$ and $3q_2x_3 - q_2x_2 + q(x_1 - x_2) = 0$. Recall that, for Q' to represent R, we must have that $x_1 = x_2$, since $R_{[3]}(1) = R_{[3]}(2)$. We then set $x_1 = x_2 = y$; we will see that this forces x_3 to be different from $R_{[3]}(3)$, so that no PCMC model can represent the RUM R.

Observe that, under $y = x_1 = x_2$, the first and second constraints become $3q_1x_3 - q_1y = 0$ and $3q_2x_3 - q_2y = 0$. Summing them up, we get $3(q_1 + q_2)x_3 = (q_1 + q_2) \cdot y$. Since $q_1 + q_2 > 0$, we can divide the two sides of the equation by $q_1 + q_2$, and get $y = 3x_3$, that is, $y = x_1 = x_2 = 3x_3$.

Finally, the last constraint of the system is equivalent to $x_3 = 1 - x_1 - x_2$, that is, $x_3 = 1 - 2y$; thus, we get $x_3 = 1 - 2y = 1 - 6x_3$, which entails $x_3 = 1/7$, and $x_1 = x_2 = 3/7$. I.e., in a PCMC model that satisfies the constraints induced by slates of size 2, item 3 *must* win in the slate [3] with probability $1/7$. However, in the RUM R, the probability that 3 wins in [3] is $R_{[3]}(3) = 0$. □

C.2. RUMs vs CDMs

Context-Dependent utility Models (CDMs) have recently been introduced by Seshadri et al. (2019) (see also Seshadri et al., 2020) to model irrational choice behavior. We show in this section that CDMs are unable to represent general RUMs.

A CDM is defined by an order-2 tensor—a sequence $w_{1,2}, \ldots, w_{1,n}, w_{2,1}, w_{2,3}, \ldots, w_{n,n-1}$ of $n \cdot (n - 1)$ positive weights—i.e., a weight for each ordered pair of items $i \in [n], j \in [n] \setminus \{i\}$.[12]

In a CDM, the probability that i is chosen in a slate $S \ni i$ is equal to

$$\frac{\prod_{j \in S \setminus \{i\}} w_{j,i}}{\sum_{k \in S} \prod_{j \in S \setminus \{k\}} w_{j,k}}.$$

In the following, we show that there exist RUMs that cannot be represented as CDMs.[13]

The example we will be using is similar to that for PCMC; we will use three types of users and four types of items, $\{1, 2, 3, 4\}$. The first user type strongly prefers 1 to each other item, and would choose uniformly over any subset of $\{2, 3, 4\}$. The second (resp., third) user type strongly prefers 2 (resp., 3) to each other item, and would choose uniformly over any subset of $\{1, 3, 4\}$ (resp., $\{1, 2, 4\}$). Again, we assume to have a uniform mixture of the three user types.

Observation 21. *There exists a RUM on $n = 4$ elements that cannot be represented with CDM models.*

Proof. Let $n = 4$, and consider the RUM R that chooses a permutation u.a.r., conditioned on its top-most element to be different from 4. Then, $R_{[4]}(4) = 0$ and $R_{[4]}(1) = R_{[4]}(2) = R_{[4]}(3) = 1/3$. Moreover, for $\{i, j\} \subset \{1, 2, 3\}$, $R_{\{i,4\}}(i) = \frac{2}{3}$, $R_{\{i,j,4\}}(i) = R_{\{i,j,4\}}(j) = \frac{4}{9}$. And, R_S chooses uniformly at random in S if $4 \notin S$.

A CDM model that guarantees the marginals of the slate $\{i, 4\}$, for each $i \in [3]$, must then have $w_{4,i} = 2w_{i,4}$; we define

[12] We mention that CDMs can be generalized to tensors of order k, for an arbitrary $k = 1, \ldots, n - 1$. For any constant k, an order k CDM requires $\Omega(n^k)$ bits to be represented. CDMs of order $n - 1$ can approximate any choice model on $[n]$, but require exponentially many bits. We show here that the class of CDMs studied in Seshadri et al. (2019; 2020), i.e., those with $k = 2$, is unable to represent a general RUM. Order-2 CDMs can be represented by $\widetilde{\Theta}(n^2)$ bits, that is, they have the same representation complexity of PCMC models and, as we prove in our paper, of RUMs; higher-order CDMs require many more bits than RUMs and PCMC models.

[13] It is easy to see the converse, i.e., that there exist CDMs that cannot be represented as RUMs. For instance, if $w_{1,2} = w_{2,1} = 1$, then the CDM chooses uniformly at random from the slate $\{1, 2\}$, thus 1 wins in $\{1, 2\}$ with probability $1/2$. Now, consider the slate $\{1, 2, 3\}$ and suppose that $w_{3,1} = t$, for some $t > 2$, while $w_{1,2} = w_{2,1} = w_{2,3} = w_{3,2} = w_{1,3} = 1$. Then, the probability that 1 wins in the slate $\{1, 2, 3\}$ is equal to $\frac{t}{t+1+1} = 1 - \frac{2}{t+2} > \frac{1}{2}$. Conversely, each RUM R is such that $R_{\{1,2\}}(1) \geq R_{\{1,2,3\}}(1)$. Thus, this CDM cannot be represented by a RUM.

$w_i = w_{i,4}$ so that $w_{4,i} = 2w_i$. Analogously, if the CDM model guarantees the marginals of the slate $\{i,j\} \in \binom{[3]}{2}$, it must be that $w_{i,j} = w_{j,i}$. We then define $w_{\{i,j\}} = w_{i,j} = w_{j,i}$.

Moreover, for the slate $\{i,j,k\} = \{1,2,3\}$, the probability that i wins is

$$\frac{w_{j,i} \cdot w_{k,i}}{w_{j,i} \cdot w_{k,i} + w_{i,j} \cdot w_{k,j} + w_{i,k} \cdot w_{j,k}}.$$

Since each $i \in [3]$ should win with probability $1/3$ in [3], we get

$$w_{2,1}w_{3,1} = w_{1,2}w_{3,2} = w_{1,3}w_{2,3},$$

which, using the constraints given by the 2-subslates of [3], entails

$$w_{\{1,2\}}w_{\{1,3\}} = w_{\{1,2\}}w_{\{2,3\}} = w_{\{1,3\}}w_{\{2,3\}},$$

which finally entails $w_{\{1,2\}} = w_{\{1,3\}} = w_{\{2,3\}}$. We then define $w = w_{1,2} = w_{2,1} = w_{1,3} = w_{3,1} = w_{2,3} = w_{3,2}$.

Now, consider the slate $\{i,j,4\}$ for $\{i,j\} \in \binom{[3]}{2}$. The probability that 4 wins in this slate is

$$\frac{w_i w_j}{w_i w_j + 2w_i w + 2w_j w},$$

and it should be equal to $R_{\{i,j,4\}}(4) = 1/9$, that is

$$\begin{aligned} 9w_i w_j &= w_i w_j + 2w_i w + 2w_j w \\ 8w_i w_j &= 2w(w_i + w_j) \\ \frac{8}{2w} &= \frac{w_i + w_j}{w_i w_j} \\ \frac{4}{w} &= \frac{1}{w_i} + \frac{1}{w_j}. \end{aligned}$$

Since this holds for any $\{i,j\} \in \binom{[3]}{2}$, we obtain

$$\frac{1}{w_1} + \frac{1}{w_2} = \frac{1}{w_1} + \frac{1}{w_3} = \frac{1}{w_2} + \frac{1}{w_3},$$

thus, $w_1 = w_2 = w_3$. By $w_i^{-1} + w_j^{-1} = 4w^{-1}$, we get $w_1 = w_2 = w_3 = w/2$. Thus, $w_{1,4} = w_{2,4} = w_{3,4} = w/2$ and $w_{4,1} = w_{4,2} = w_{4,3} = w$. I.e., for $i \in [4]$ and $j \in [4] \setminus \{i\}$, $w_{i,j} = w/2$ if $j = 4$, and $w_{i,j} = w$ if $j \neq 4$.

The probability that 4 wins in the slate [4], is then

$$\frac{(w/2)^3}{(w/2)^3 + w^3 + w^3 + w^3} = \frac{1/8}{25/8} = \frac{1}{25},$$

which is different from $R_{[4]}(4) = 0$. $\square$

C.3. RUMs vs Deterministic Choice Models

Deterministic choice models (such as that in Rosenfeld et al., 2020) are unable to represent the random choices made by RUMs. For instance, the uniform RUM U we considered in Section 6 (i.e., the one that chooses u.a.r. from the set of all permutations of $[n]$) chooses the winner uniformly at random in the slate $\{1,2\}$. A deterministic model, instead, will either always choose 1, or always 2, as the winner in that slate. Therefore, the distribution of RUM U on $\{1,2\}$ is at total variation distance $1/2$ from the one given by any deterministic model. In fact, if one considers a larger slate S, then the total variation distance increases to $1 - 1/|S|$.