Statistical Models of Top-k Partial Orders

Amel Awadelkarim
ameloa@stanford.edu
Stanford University
Stanford, CA, USA

Johan Ugander
jugander@stanford.edu
Stanford University
Stanford, CA, USA

Abstract

In many contexts involving ranked preferences, agents submit *partial* orders over available alternatives. Statistical models often treat these as marginal in the space of *total* orders, but this approach overlooks information contained in the list length itself. In this work, we introduce and taxonomize approaches for jointly modeling distributions over top-k partial orders and list lengths k, considering two classes of approaches: *composite models* that view a partial order as a truncation of a total order, and *augmented ranking models* that model the construction of the list as a sequence of choice decisions, including the decision to stop. For composite models, we consider three dependency structures for joint modeling of order and truncation length. For augmented ranking models, we consider different assumptions on how the stop-token choice is modeled. Using data consisting of partial rankings from San Francisco school choice and San Francisco ranked choice elections, we evaluate how well the models predict observed data and generate realistic synthetic datasets. We find that composite models, explicitly modeling length as a categorical variable, produce synthetic datasets with accurate length distributions, and an augmented model with position-dependent item utilities jointly models length and preferences in the training data best, as measured by negative log loss. Methods from this work have significant implications on the simulation and evaluation of real-world social systems that solicit ranked preferences.

CCS Concepts

• **Computing methodologies** → **Model development and analysis**; • **Information systems** → **Learning to rank**.

Keywords

rankings; partial orders; statistical modeling; discrete choice

ACM Reference Format:
Amel Awadelkarim and Johan Ugander. 2024. Statistical Models of Top-k Partial Orders. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3637528.3672014

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08
https://doi.org/10.1145/3637528.3672014

1 Introduction

Ranked preferences serve as input to many consequential social systems. In election contexts, ranked choice voting (RCV) asks voters to express their preferences for candidates with a ranked list of available candidates, and these rankings are aggregated to select a winner. In school choice contexts, families express preferences for their child's education by submitting a ranked list of available program offerings to their district, and these preferences are considered when matching students to programs.

In the vast majority of such contexts, participants often submit partial rankings, specifically top-k rankings. For example, a voter given a slate of m alternatives may only express strict preferences for their first $k < m$ alternatives, and not bother to provide a strict ordering of the rest. In many preference elicitation contexts, truncating rankings in this manner can be highly consequential. For example, in San Francisco school choice, families often submit truncated rankings over only k programs, but around 11% of households do not get assigned to any of their chosen alternatives [1]. Submitting longer rankings would only increase their placement chances, but the decision to truncate a preference list can be easily understood in terms of high search costs, a misunderstanding of the mechanism, or inflated subjective placement chances [4, 12, 13, 21, 22, 32]. In school choice, the decision to truncate can thus mean the difference between gaining admittance to a preferred school and being relegated to participating in a later round of assignment. As a lesser but still meaningful consequence, in ranked choice voting, truncation can mean the difference between having a say in later rounds of instant run-off voting vs. having one's ballot "exhausted."

Statistical models of ranking data are often used by researchers to model demand, understand trends in preferences, or simulate counterfactual outcomes [2, 3, 10, 20, 28, 34]. In counterfactual simulations, inquiries center around what would happen (e.g., to overall welfare) if people's alternatives or preferences changed in some structured way: some new candidate or school is added or removed from the slate of alternatives, the school-age population changes, or school prioritization changes. But investigating these counterfactuals invariably involves strong assumptions about preference lengths: either that lengths are directly taken from observed data or otherwise chosen independent of the ordered preferences themselves [2, 20].

In this work, we argue that the problem of suitably modeling length in top-k partial orders is an overlooked and highly consequential component of real-world deployments of ranking data models. We consider two distinct approaches to modelling top-k orders: *composite models* and *augmented ranking models*. First, consider the probability of a partial order Q of length k. The probability of the order can be written as the probability of its length times the marginal probability of total orders whose first k elements match

Q. Informally (see Section 3.2 for a formal treatment), we have

$$\Pr(Q) = \Pr(k) \cdot \sum_{R \in \bar{Q}} \Pr(R|k),$$

where $\bar{Q}$ denotes the set of total orders of m items whose first k elements are Q, and R is an element of that set. Decoupling these two component probabilities, a length model and a conditional ranking model can indeed be combined to form a model of partial rankings, a model class we call *composite models*. Under an independence assumption, such a composite model is simply:

$$\Pr(Q) = \Pr(k) \cdot \sum_{R \in \bar{Q}} \Pr(R),$$

dropping the conditional dependence in the ranking model. Moving beyond independence, we consider situations where the component models (of length, of order) can be have structured dependence, and the models themselves be simple or complex.

Beyond composite models, we also take an alternative approach that we call *augmented ranking*. In this approach, we augment the choice universe with an END token representing end-of-list, and consider a partial order as arising from a sequential choice process that models termination itself as one of the available alternatives. The agent selects k alternatives and then "chooses" the END token at position $k+1$. This approach extends earlier ideas on modeling non-choice/non-purchase from choice settings to the ranking setting. Using a choice-based approach to ranking, as developed in the theory of L-decomposable ranking distributions [9, 25], the probability of a partial ranking Q becomes

$$\begin{aligned}\Pr(Q) &= \Pr(q_1 \succ q_2 \succ \ldots \succ q_k) \\ &= \Pr(q_1) \cdot \Pr(q_2|q_1) \cdot \ldots \cdot \Pr(q_k|q_1, \ldots, q_{k-1}) \cdot \Pr(\texttt{END}|q_1, \ldots, q_k).\end{aligned}$$

This lens *implicitly* models list length by augmenting the choice space to include a token alternative representing end-of-list. This token can be modeled as having fixed or position-dependent selection probabilities, and we evaluate such modelling decisions in our work.

To model general dependence between ranking and length, we utilize two model stratification techniques [5]. In the composite case, the ranking model can be made dependent on the length by learning K separate ranking models that cover K disjoint subsets of the space of partial orders, partitioned by length. The probability of a particular partial order Q would then be the probability of its length, k, times the probability of the ordering Q under the corresponding ranking model. In the case of augmented ranking models, the END alternative (and/or the other choice alternatives) are given K distinct choice probabilities depending on its choice position within the ranking. This way, the probability of choosing an alternative, including the END token, is allowed to vary down the ranking.

The primary contribution of our work is the development and evaluation of the composite and augmented ranking models, two classes of models tailored to this consequential task. We consider several instances and sub-classes of such models, and evaluate their performances across a range of datasets with considerable variation in the size of their choice universes, m. We find that the composite approaches, which model list length explicitly, can produce more accurate sampled length distributions when generating synthetic datasets from the models, and that model stratification—a strategy we employ in both composite and augmented models—improves simulated demand over alternatives. While our results do not promote a universally dominant model nor model class, this work begins to formalize and taxonomize approaches for modeling distributions over the space of partial orders, and the results showcase how these components can lead to improved demand modeling and more realistic synthetic datasets.

In Section 2, we survey the related literature. In Section 3, we present notation and definitions relevant to statistical models of ranking. In Sections 4 and 5, we introduce and define the two classes of partial order models discussed in this work, composite models and augmented ranking models, respectively. Model selection and estimation is discussed in Section 6. Datasets and experiments are presented in Section 7. Section 8 concludes.

2 Related work

Traditional statistical models for rankings focus on modeling total orders. Notable among these are variations of the Plackett–Luce (PL) model, which has been extensively used in economics, marketing, and revenue management, to learn consumer preferences from choice and rank data. The Plackett–Luce model is a multi-stage model whereby rankings are broken into sequential choices, which are then modeled by a random utility choice model, the multinomial logit (MNL) model [33].

Several works apply these models to learn preferences from partial order datasets for the purposes of rank aggregation or explanation of user preferences. For example, Zhao and Xia [37] develop theory around the task of learning mixtures of Plackett–Luce models from partial order data. In a school choice setting, Abdulkadiroğlu et al. [2] learn an MNL choice model from partial order data to evaluate how much parents value certain school attributes. Li et al. [24] develop Bayesian techniques to learn from total and partial ranking data with heterogeneous ranker preferences and item covariates. While these works enlist the Plackett–Luce model which produces a distribution over the space of total rankings, the present work seeks to develop models that produce distributions over a larger sample space, namely the set of all partial orderings (which subsumes the former).

At other times, ranking models are specifically used to simulate real agent behavior, *including* the submission of partial orders in some preference elicitation contexts. In these instances, researchers tend to learn models of total orders like the Plackett–Luce, simulate sequences of choices, and then make some assumption about the truncation of the data to achieve a dataset of partial orders. For example, Pathak and Shi [2] learn an MNL choice model from partial order data and then simulate partial orders by manually truncating total orders to a predetermined length of ten. Laverde [20] models partial orders to simulate counterfactual outcomes for Black and Hispanic students in Boston, but chooses to impose the original list lengths from the data. As a significant improvement, the models developed in this work jointly model preferences and order lengths from data, and we show this model enrichment is consequential for counterfactual simulations.

Our augmented ranking model modifies choice-based ranking to incorporate an alternative in the choice universe representing

end-of-list. The idea has roots in demand modeling and revenue management for modeling the value of outside options or non-choice/non-purchase behavior [8, 11, 17, 29]. To our knowledge, the modelling of non-choice has not been previously employed in the ranking context for models of partial orders.

The problem of modeling partial orders is related to the problem of selecting a subset of a choice universe, as is studied in marketing, operations research, information retrieval and data mining. Such works focus on predicting or generating subsets of size $k > 1$ of a universe set of m alternatives, but not on ordering these subsets to yield rankings [6, 7, 23, 31]. A notable subset selection model is that of Regenwetter et al. [31]. In that work, their "subset model" is identical to our independent composite model, but the work is (1) interested in modeling the total ordering, not modeling the subsets themselves, and (2) does not consider any dependence between size and preferences as we do in this work.

Finally, we apply model stratification in this work, a common technique in machine learning for understanding heterogeneity between different demographic and other characteristic groups in the population [14, 18, 19]. Several preference modeling works use model stratification to develop tailored models for various demographic groups of interest [12, 20]. Here, we apply *regularized* stratification [35] as a modeling tool for encoding dependence (in the composite case) or enriching the expressivity of a ranking model (in the augmented case), even within the same populations. Most recently, Awadelkarim et al. utilized a similar rank-based stratification technique to model heterogeneity in school choice preferences within the rankings of single individuals [5]. We adopt and expand upon their idea in the present work.

3 Preliminaries

Rankings, or ordered preferences, are seemingly intuitive objects that we engage with everyday. We routinely express ordered preferences over where we want to eat, who we want to lead an organization, or which movie we would like to watch with friends. In Section 3.1, we define the notions of total and partial orders, and their related outcome spaces. We define statistical models, or distributions over these spaces, in Section 3.2.

3.1 Rankings

Given a collection of m alternatives, $\mathcal{A} = \{a_1, ..., a_m\}$, let $\mathcal{L}(\mathcal{A})$ denote[1] the set of complete rankings, or total orders, of the elements of $\mathcal{A}$, which contains $|\mathcal{L}(\mathcal{A})| = m!$ elements. As an illustration, for a collection $\mathcal{A} = \{a, b, c\}$ of $m = 3$ items, we have

$$\mathcal{L}(\{a,b,c\}) = \begin{Bmatrix} a \succ b \succ c & b \succ a \succ c & c \succ a \succ b \\ a \succ c \succ b & b \succ c \succ a & c \succ b \succ a \end{Bmatrix}.$$

Here the element $R = a \succ b \succ c$ is the event that a is preferred to b and b is preferred to c.

Borrowing nomenclature of Xia [36] and Zhao and Xia [37], we focus in this work on *top-k* partial orders, $Q = q_1 \succ q_2 \succ \ldots \succ q_k[\succ \text{others}]$. In this case, an agent orders their top k most-preferred elements of $\mathcal{A}$ and leaves the rest unordered. We denote by $\Omega_k(\mathcal{A})$ the set of all top-k partial orders of $\mathcal{A}$, and $\Omega(\mathcal{A})$ to be the union of these sets, for any k. The size of $\Omega(\mathcal{A})$ is

$$|\Omega(\mathcal{A})| = \sum_{i=1}^{m} |\Omega_i(\mathcal{A})| = \sum_{i=1}^{m} \frac{m!}{(m-i)!}.$$

A total ordering is a special case of a top-k partial ordering where $k = m$, so we have $\mathcal{L}(\mathcal{A}) = \Omega_m(\mathcal{A}) \subseteq \Omega(\mathcal{A})$. For the collection $\mathcal{A} = \{a, b, c\}$ as above, we then have the space of top-2 partial orders and the (full) space of partial orders,

$$\Omega_2(\{a,b,c\}) = \begin{Bmatrix} a \succ b & b \succ a & c \succ a \\ a \succ c & b \succ c & c \succ b \end{Bmatrix},$$

$$\Omega(\{a,b,c\}) = \begin{Bmatrix} a & b & c \\ a \succ b & b \succ a & c \succ a \\ a \succ c & b \succ c & c \succ b \\ a \succ b \succ c & b \succ a \succ c & c \succ a \succ b \\ a \succ c \succ b & b \succ c \succ a & c \succ b \succ a \end{Bmatrix}.$$

Note that the top-2 partial order $a \succ b$ and the total order $a \succ b \succ c$ represent the same overall preference profile over the items $\mathcal{A} = \{a, b, c\}$, and as such would be equal-probability events under a ranking model. These two preferences would also figure identically in many uses of ranked preferences, e.g., ranked choice voting. But the models in this work assign these two elements different probabilities on purpose, and we emphasize that this is not a bug, but a feature. The event of reporting a list of length two is different than that of listing all three and the distinction is not informationless.

Given a k-length partial order $Q \in \Omega_k(\mathcal{A})$, we define by $\text{Ext}(Q)$ the set of *completions* of Q in $\mathcal{L}(\mathcal{A})$. Specifically, let R_k be shorthand for the first k elements of $R \in \mathcal{L}(\mathcal{A})$: $R_k = r_1 \succ r_2 \succ \ldots \succ r_k$. We have

$$\text{Ext}(Q) = \{R \in \mathcal{L}(\mathcal{A}) \mid R_k = Q\}. \tag{1}$$

For $\mathcal{A} = \{a, b, c\}$ and $Q = a$, the completions of this partial order are given by

$$\text{Ext}(a) = \{a \succ b \succ c \quad a \succ c \succ b\}.$$

Finally, for any $Q \in \Omega(\mathcal{A})$, let k_Q be shorthand for the length of the partial order, $k_Q = |Q|$.

3.2 Statistical models

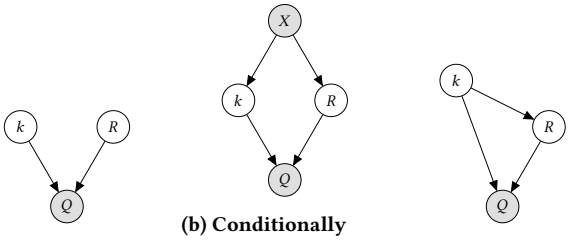
A statistical model defines a probability distribution over a sample space $\mathcal{S}$. Specifically, let π_θ be that distribution, parameterized by θ, such that $\pi_\theta : \mathcal{S} \mapsto [0, 1]$. The distribution π_θ maps events in $\mathcal{S}$ to probabilities and therefore must sum to 1, $\sum_{s \in \mathcal{S}} \pi_\theta(s) = 1$. Given a distribution over a sample space π_θ and a subset of its sample space $C \subseteq \mathcal{S}$, we slightly overload notation and denote the probability of C under π_θ as the sum of its constituent probabilities:

$$\pi_\theta(C) \coloneqq \sum_{c \in C} \pi_\theta(c). \tag{2}$$

This expression holds mathematically as the events in C are disjoint, so the probability of their union is the sum of their individual probabilities.

In this work, we are concerned with the task of modeling distributions over the space of top-k partial orders, $\mathcal{S} = \Omega(\mathcal{A})$. One modeling approach would be to assign a distinct probability to every element in $\Omega(\mathcal{A})$, learning $|\Omega(\mathcal{A})| = \sum_{i=1}^{m} m!/(m-i)!$ model

[1] There is a bijection between the elements of $\mathcal{L}(\mathcal{A})$ and the symmetric group, S_m. However, we need not invoke the group properties and operations of S_m, and as such, refer to the space in this work as $\mathcal{L}(\mathcal{A})$.



(a) Independent (b) Conditionally independent (c) Length-dependent

Figure 1: Three dependence structures for composite models. $k \in [m]$ is a random variable representing length, $R \in \mathcal{L}(\mathcal{A})$ is a random variable representing a total order, $X \in \mathbb{R}^d$ are covariates, and $Q \in \Omega(\mathcal{A})$ is a top-k partial order. Observed quantities are shaded in grey.

parameters, but this strategy quickly becomes intractable as m grows.

Another strategy commonly employed for learning models from partial order data is to treat these events as marginal in the space of total orders. That is, given a distribution over linear orders, $\pi_\theta : \mathcal{L}(\mathcal{A}) \mapsto [0,1]$, consider a distribution over partial orders $\pi'_\theta : \Omega(\mathcal{A}) \mapsto [0,1]$ that is defined as

$$\pi'_\theta(Q) \coloneqq \pi_\theta(\text{Ext}(Q)) = \sum_{R \in \text{Ext}(Q)} \pi_\theta(R). \tag{3}$$

This perspective alone does not allow a researcher to sample partial orders from the space, since the list length itself is taken as exogenous and left unmodeled. In the following two sections, we present our two classes of models for efficiently parameterizing distributions over $\Omega(\mathcal{A})$.

4 Composite models

Recall our illustrative expression of the composite model from the introduction. Now we formally define the two component models, one of length and one of preferences. To signify these different model types, we use a superscript of k to indicate a length distribution and a superscript of R to indicate a ranking distribution,

$$\pi^k_\theta : [m] \mapsto [0,1], \qquad \pi^R_\theta : \mathcal{L}(\mathcal{A}) \mapsto [0,1].$$

Through the composite lens, a partial order arises from the realization of these two component variables—a length and an order—where the order is then truncated to the sampled length to generate a partial order $Q \in \Omega(\mathcal{A})$.

We now consider both *independent* or *conditionally* independent variations of the composite model, given agent covariates, graphically modeled in Fig. 1(a)-(b). Namely, given a realization of covariates X, independence assumes that the distribution is decomposable in one of the following ways, with or without agent and item covariates:

$$\textbf{C-I:} \quad \pi_\theta(Q) = \pi^k_\theta(|Q|) \cdot \pi^R_\theta(\text{Ext}(Q)), \tag{4}$$

$$\textbf{C-CI:} \quad \pi_\theta(Q|x) = \pi^k_\theta(k_Q|x) \cdot \pi^R_\theta(\text{Ext}(Q)|x). \tag{5}$$

where by Eq. (2), we have $\pi^R_\theta(\text{Ext}(Q)) = \sum_{R \in \text{Ext}(Q)} \pi^R_\theta(R)$. We refer to Eq. (4) and Eq. (5) as the **independent** (C-I) and **conditionally-independent** (C-CI) **composite models**. In a recent preference modeling work where sampling partial orders was required, Laverde [20] effectively used an independent composite model, which serves as one of two baselines in this work for the task of modeling top-k partial orders.

Beyond independence, we also consider length-dependence, as depicted in Fig. 1(c), where the ranking is dependent on the length. That composite distribution is modelled as:

$$\textbf{C-LD:} \quad \pi_\theta(Q) = \pi^k_\theta(k_Q) \cdot \pi^R_\theta(\text{Ext}(Q)|k_Q). \tag{6}$$

This assumption posits that *how many* elements an agent chooses to rank provides additional signal as to *what* they may choose to rank. In the school choice setting, this modeling decision allows families who rank one school program to have a different utility structure over the available alternatives than a household that ranks, say, five alternatives. Eq. (6) is the **length-dependent** (C-LD) **composite model**.

It is logical to ask at this point: what about a composite model where the length k is dependent on the ranking R? While mathematically coherent, conditioning a simple length distribution on a total order is rather unwieldy. When there are many items, it requires spelling out a total order R before finding that the length distribution, conditional on R, may actually be concentrated on very short lengths. Rather than condition the length on a total order, our augmented ranking approach in this next section can be thought of as a nuanced approach to this type of dependence, through a choice-based perspective on ranking. As an intuition for the augmented ranking model to come in Section 5, it effectively conditions the probability of different length outcomes upon the sequential "choices" made in the assembly of a ranking, up to a given truncation length.

4.1 Model specification

4.1.1 Ranking model. We enlist the classic Plackett–Luce (PL) model [26, 30] as the ranking model, π^R_θ, of the composite class. Under PL, we model the probability of a ranking as the product of sequential choice probabilities from shrinking choice sets. Partial orders are modeled as marginal events in $\mathcal{L}(\mathcal{A})$ as given by Eq. (3). Specifically, the probability of a partial order $Q = q_1 \succ \ldots \succ q_k$ is

$$\pi^R_\theta(Q) = \pi^R_\theta(\text{Ext}(Q)) \tag{7}$$

$$= \prod_{j=1}^{k} \frac{\exp(\theta_{q_j})}{\sum_{a \in \mathcal{A} \setminus \{q_1,\ldots,q_{j-1}\}} \exp(\theta_a)}. \tag{8}$$

The parameters $\theta \in \mathbb{R}^m$ are directly interpretable as latent item utilities. However, given features $x_{ij} \in \mathbb{R}^d$ of agent i and item j, we can further model the utilities as linear in these attributes, parameterizing item j's utility to agent i as $\delta_j + \beta^T x_{ij}$, where $\theta = (\delta, \beta) \in \mathbb{R}^m \times \mathbb{R}^d$ are the model parameters. The probability of agent i producing partial order Q is then

$$\pi^R_\theta(Q; X_i) = \prod_{j=1}^{k} \frac{\exp(\delta_j + \beta^T x_{ij})}{\sum_{a \in \mathcal{A} \setminus \{q_1,\ldots,q_{j-1}\}} \exp(\delta_a + \beta^T x_{ia})}.$$

When covariates are available, we enlist the linear model version of Plackett–Luce. If not covariates are available, we use the simpler model with only fixed effects δ_j.

We choose to model the length-dependence of C-LD by stratifying the preference model by list length. In this case, we use stratification to train separate preference models for people who ranked exactly one item, people who ranked exactly two items, and so on, all the way up to those who ranked exactly $K-1$ items. The final strata we reserve for those who ranked K or more items.

Taking the number of strata to be K, a stratified ranking model is then the composition of K sub-models, each with its own parameters: $\theta = \{\theta_1, ..., \theta_K\} \in \mathbb{R}^{m\times K}$. Specifically, the likelihood of a partial order Q is modeled as

$$\pi_\theta^R(Q|k_Q) = \pi_{\theta_{k'}}^R(Q).$$

where $k' = \min(k_Q, K)$ and k_Q is the length of Q; orders of length $k \in [1, K]$ are modeled by Eq. (8) with parameters θ_k, and rankings of length $k > K$ are modeled with parameters θ_K.

A possible concern with this approach is that we end up with considerably less data for each model, compared to estimating a single common model. The thinning of the training data can lead to over-parameterization which leads to overfitting. To address these concerns, we encourage models of neighboring lengths to be close to one another via Laplacian regularization [35], borrowing the predictive power of neighboring models. The regularization is "Laplacian" because the K sub-models are regularized towards each other as dictated by an accompanying *regularization graph* with weights governed by the graph's Laplacian. Length-based stratification lends itself well to a common path graph—the model of top-1 partial orders should be similar to a model of top-2, and so on—so this regularization takes on a simple form. See Section 6 for presentation of the stratified objective function.

4.1.2 Length model. The model of length within a composite model is a distribution over $\mathcal{S} = [m]$. For full generality, we first choose to model the length as a categorical variable over these m discrete categories for two of the three composite models. Namely, given parameters $\theta \in \mathbb{R}^m$, we have

$$\pi_\theta^k(k) = p_k = \frac{\exp(\theta_k)}{\sum_{i=1}^m \exp(\theta_i)}.$$

For the conditionally-independent composite model (C-CI), we model length as a Poisson random variable. That is, given agent covariates x, we define the rate parameter as $\lambda(x) = \exp(\theta^T x)$, and the resulting length distribution is given by

$$\pi_\theta^k(k; x) = \frac{\lambda(x)^k e^{-\lambda(x)}}{k!}.$$

The domain of the Poisson model is not bounded between $[1, m]$, so we clip extreme values into this range. Parameters of this model are then $\theta \in \mathbb{R}^d$ where d is the dimension of covariates x.

5 The augmented model

Consider a set of alternatives, $\mathcal{A}$, augmented with an end-of-list alternative END as an additional alternative in the choice universe. A similar idea has been utilized in the revenue management literature for modeling no-purchase options [8, 11, 17, 29]. In this model, a partial order $Q \in \Omega(\mathcal{A})$ is taken to arise from the following process: let $\mathcal{A}^+$ be the universe of alternatives plus an END alternative representing end-of-list, $\mathcal{A}^+ = \mathcal{A} \cup \{\text{END}\}$. A partial ranking Q of length $k = |Q|$ is viewed as the sequential selection of each item q_i from a sequence of shrinking choice sets $C_i \subseteq \mathcal{A}^+$, followed by the selection of END at position $k+1$.

Let $\text{Ext}^+(Q)$ represent all linear extensions of Q in $\mathcal{L}(\mathcal{A}^+)$ that begin with $Q \succ \text{END}$. Namely,

$$\text{Ext}^+(Q) = \{R : R \in \mathcal{L}(\mathcal{A}^+), R_{k+1} = Q \succ \text{END}\}.$$

For example, given $\mathcal{A} = \{a, b, c\}$ and $Q = a$, we have

$$\text{Ext}(a) = \{a \succ b \succ c \quad a \succ c \succ b\}$$
$$\text{Ext}^+(a) = \{a \succ \text{END} \succ b \succ c \quad a \succ \text{END} \succ c \succ b\}.$$

Under an augmented ranking model, the probability of observing partial order Q is given by

$$\textbf{A: } \pi_\theta(Q) = \pi_\theta^R(\text{Ext}^+(Q)). \tag{9}$$

Here π_θ^R represents a probability distribution over $\mathcal{L}(\mathcal{A}^+)$, since elements R here order the augmented universe that includes the END alternative.

Taking the Plackett–Luce model as our baseline ranking model, we evaluate three approaches to parameterizing this new choice system. Specifically, we consider various ways to parameterize the END alternative relative the rest in $\mathcal{A}$, yielding the (1) naïve, (2) position-dependent, and (3) K-stratified augmented models. The (naïve) **augmented model** (A) assigns the END alternative a fixed utility, resulting in a standard Plackett–Luce over $(m+1)$ alternatives with parameters $\theta \in \mathbb{R}^{m+1}$:

$$\pi_\theta^R(Q) = \pi_\theta^R(\text{Ext}^+(Q)) = \prod_{j=1}^k \frac{\exp(\theta_{q_j})}{\sum_{a\in\mathcal{A}^+\setminus\{q_1,\ldots,q_{j-1}\}}\exp(\theta_a)} \cdot \frac{\exp(\theta_{\text{END}})}{\sum_{a\in\mathcal{A}^+\setminus Q}\exp(\theta_a)}.$$

Similar to the independent composite model, our naive augmented model has been implemented for the purpose of modeling end-of-list in at least one preference modeling application [29] and we consider it as the second of two baselines for our modeling task.

The **position-dependent augmented model** (A-PD) endows the END alternative with position-dependent utilities. Specifically, the END token has m utilities, $\gamma \in \mathbb{R}^m$, depending on how far the choice process has gotten without terminating. The m ordinary alternatives have a single position-invariant utility each, $\theta \in \mathbb{R}^m$.

$$\pi_\theta^R(Q) = \prod_{j=1}^k \frac{\exp(\theta_{q_j})}{\exp(\gamma_j) + \sum_{a\in\mathcal{A}\setminus\{q_1,\ldots,q_{j-1}\}}\exp(\theta_a)} \cdot \frac{\exp(\gamma_{k+1})}{\exp(\gamma_{k+1}) + \sum_{a\in\mathcal{A}\setminus Q}\exp(\theta_a)}.$$

Finally, the **K-stratified augmented model** (A-S) allows all alternatives position-dependent utilities, up to position K,

$$\pi_\theta^R(Q) = \prod_{j=1}^k \frac{\exp(\theta_{q_j}^{(j')})}{\sum_{a\in\mathcal{A}^+\setminus\{q_1,\ldots,q_{j-1}\}}\exp(\theta_a^{(j')})} \cdot \frac{\exp(\theta_{\text{END}}^{(k'+1)})}{\sum_{a\in\mathcal{A}^+\setminus Q}\exp(\theta_a^{(k'+1)})},$$

where $i' = \min(i, K)$. This stratification is adopted from [5] and is in contrast with the length-dependent stratification of C-LD, detailed in Section 4.1. There, the model assigns each *user* to one of K PL models, depending on how long their list was. Here, each

Table 1: Summary of models; m denotes the number of choice alternatives and k_Q denotes the length of partial order $Q \in \Omega(\mathcal{A})$.

Model	Distribution, $\pi_\theta(Q)$	Description
Composite, Independent (C-I)	$\pi^k_\theta(k_Q) \cdot \pi^R_\theta(\text{Ext}(Q))$	Categorical length and PL ranking.
Composite, Conditionally-independent (C-CI)	$\pi^k_\theta(k_Q\|x) \cdot \pi^R_\theta(\text{Ext}(Q)\|x)$	Poisson length and PL ranking, both conditional on covariates.
Composite, Length-dependent (C-LD)	$\pi^k_\theta(k_Q) \cdot \pi^R_\theta(\text{Ext}(Q)\|k_Q)$	Categorical length and length-stratified PL ranking.
Augmented (A)	$\pi^R_\theta(\text{Ext}^+(Q))$	PL over $m+1$ alternatives: $\theta \in \mathbb{R}^{m+1}$.
Augmented, Position-dependent (A-PD)	$\pi^R_\theta(\text{Ext}^+(Q))$	PL over $m+1$ alternatives, but `END` gets m utilities: $\theta \in \mathbb{R}^{2m}$.
Augmented, Stratified (A-S)	$\pi^R_\theta(\text{Ext}^+(Q))$	Rank-stratified PL over $m+1$ alternatives: $\theta \in \mathbb{R}^{K(m+1)}$.

sequential *choice* up to position K is governed by a unique PL model over the elements of $\mathcal{A}^+$. The resulting model has parameters $\theta = (\theta^{(1)}, ..., \theta^{(K)}) \in \mathbb{R}^{(m+1)\times K}$. Similarly to the length-based stratification, we also apply Laplacian regularization between the K adjacent models. Rank-based stratification also lends itself well to a common path graph as regularization graph—in this case saying the model of top-choices should be similar to a model second choices, and so on. Of course, these two stratifications (by length and by rank) are not mutually exclusive and could be applied within the same ranking model using a two-dimensional grid as the regularization graph. We leave the investigation of how such multiple stratifications interact as future work. See Table 1 for a summary of our six proposed methods for modeling top-k partial orders.

6 Model selection

Here we present our procedure for selecting model parameters and hyperparameters. We estimate all model parameters using ℓ_2-regularized maximum likelihood estimation. Given a dataset of partial orders from n participants, $D = \{Q_1, ..., Q_n\}$ where $Q_i \in \Omega(\mathcal{A})$, and an optional matrix of covariates on the agents and items, $X \in \mathbb{R}^{n\times m\times d}$, model parameters θ are chosen to maximize the likelihood of the observed dataset by minimizing the regularized negative log-likelihood (NLL),

$$F(D;\theta) = \ell(D;\theta) + r(\theta), \tag{10}$$

where $\ell(D;\theta)$ is the NLL loss and $r(\theta)$ is the ℓ_2 penalty on parameters:

$$\ell(D;\theta) = -\frac{1}{|D|}\sum_{Q\in D}\log\left(\pi_\theta(Q)\right), \quad r(\theta) = \lambda||\theta||_2^2.$$

The regularization strength λ is a hyperparameter of the model. Our proceedure for selecting λ and other optimization details are provided in Section 7.1.

The ℓ_2 regularization also makes our model *identifiable*; a statistical model is identifiable if no two distinct sets of parameters, θ and θ', produce the same probability distributions over the sample space. The traditional PL family of ranking distributions are non-identifiable due to their shift-invariance: $\pi^{\text{PL}}_\theta(s) = \pi^{\text{PL}}_{\theta+c\vec{1}}(s)$ for all $s \in \mathcal{S}$ and $c \in \mathbb{R}$. In this case, strategies for achieving identifiability are to fix one of the parameters, constrain their sum, or to apply regularization and obtain the minimum-norm parameter estimates [36], as we have chosen to do in this work.

For stratified models, notably the length-dependent composite model (C-LD) and the stratified augmented model (A-S), we learn K PL models, over $\mathcal{A}$ and $\mathcal{A}^+$ respectively, with parameters $\theta = (\theta_1, ..., \theta_K)$, regularized toward each other via Laplacian regularization. These models are trained on K stratified bands of the dataset, $D = \{D_1, ..., D_K\}$, but the two models stratify the data differently. For C-LD, D_i contains all partial orders of length i for all $i < K$, and D_K contains lists of length at least K. Specifically, $D_i = \{Q \,|\, Q \in D, |Q| = i\}, \forall i < K$, and $D_K = \{Q \,|\, Q \in D, |Q| \geq K\}$. For the A-S model class, for all $i < K$, D_i contains all *choices* made at position i across all partial orders $Q \in D$, and D_K contains choices made in positions K onward across all partial orders.

Laplacian regularization in both cases is defined as:

$$r_{\mathcal{L}}(\theta) = \lambda_{\mathcal{L}}\sum_{i=2}^{K}||\theta_i - \theta_{i-1}||_2^2,$$

where $r_{\mathcal{L}}$ is convex in θ and $\lambda_{\mathcal{L}}$ is a chosen Laplacian regularization strength. Compared to the non-stratified objective in Eq. (10), the regularized, stratified objective function is the sum of K decoupled model losses (each with a local ℓ_2 regularization) and the Laplacian regularization term:

$$F(D;\theta) = \sum_{k=1}^{K}\left[\ell(D_k;\theta_k) + r(\theta_k)\right] + r_{\mathcal{L}}(\theta). \tag{11}$$

Both the C-LD and A-S model introduce two additional hyperparameters, the number of strata and the Laplacian regularization gain, $(K, \lambda_{\mathcal{L}})$. Selection of these hyperparameters is discussed in Section 7.1.

7 Experiments

In this section we present an evaluation of six models—three composite models (C-I, C-CI, C-LD) and three augmented models (A, A-PD, A-S)—using seven partial order datasets, summarized in Table 2. Our main motivation for modelling partial ranking data is to provide a principled way to simulate partial order data for the purposes of counterfactual policy simulation or demand modeling. In these instances, researchers tend to learn models of total orders like the Plackett-–Luce, simulate sequences of choices, and then make some (unexplored) assumption about the truncation of the data to achieve a dataset of partial orders [2, 20]. The method implemented in [20] is actually equivalent to our independent-composite model, and another work recently implemented what we consider as a (naive) augmented model [29] for the same task. Of our combined six variants of the composite and augmented classes of models, these two basic models act as our working "baselines" for modeling tractable distributions over $\Omega(\mathcal{A})$.

We structure our experiments to answer two key questions:

Table 2: Dataset summary statistics for ranked-choice voting (RCV) and school choice (SC) data. 2018 and 2019 Board of Supervisor elections correspond to San Francisco Districts 8 and 5, respectively. Here n is the number of orders in the dataset, m is the number of available alternatives, and $\bar{k}$ is the average order length.

Name	Label	n	m	$\bar{k}$
2018 Board of Supervisors	RCV1	33,394	3	2.04
2019 District Attorney	RCV2	193,492	4	2.32
2019 Board of Supervisors	RCV3	23,698	4	2.06
2019 Mayor	RCV4	178,924	7	2.58
2018 Mayor	RCV5	253,866	8	2.52
2017-18 Kindergarten	SC1	3,503	152	9.28
2018-19 Kindergarten	SC2	3,544	152	9.97

(1) How likely is observed data, in its heterogeneously truncated form, under each of the models?
(2) How well do synthetic datasets, sampled from models with parameters estimated via maximum likelihood, simulate real demand patterns?

In Section 7.2, we report on overall goodness-of-fit on held out data as measured by negative log loss, addressing (1). To address (2), in Section 7.3 we sample datasets from the models and present demand statistics as compared with ground truth preferences in the observed data, namely in terms of length and alternative-specific demand distributions. Finally, in Section 7.4, we demonstrate a downstream application of our models in the school choice realm, simulating SFUSD policy assignments using preference data generated by our models and comparing assignment outcomes under the synthetic data with assignment outcomes under the true data.

7.1 Setup

Our data comes from two sources: we use four publicly-available San Francisco ranked-choice voting (RCV) datasets from the online library of preference data, `preflib` [27], and two (non-public) school choice (SC) datasets from the San Francisco Unified School District (SFUSD). Both types of datasets contain strict (ie. no ties) partial orders that we treat as top-k orders. The school choice data features covariates on households and programs, which we incorporate through linear utility models. These datasets contain PII and as such are not publicly available. See Table 2 for dataset summary statistics.

Models that enlist regularized stratification, namely the length-dependent composite (C-LD) and stratified augmented (A-S) models, introduce two additional hyperparameters, the number of stratification buckets K and the regularization strength $\lambda_{\mathcal{L}}$. We conducted a grid search across suitable values of each of these parameters for each model class (composite vs. augmented), on a representative dataset of each type (RCV vs. SC). Each $(K, \lambda_{\mathcal{L}})$ pair was trained and evaluated using 5-fold cross validation on these datasets, and the values yielding the lowest average validation loss were selected. For the school choice datasets, we choose $(K, \lambda_{\mathcal{L}}) = (15, 1e-3)$ for the length-dependent composite model, and $(K, \lambda_{\mathcal{L}}) = (15, 0)$ for the stratified augmented model. For the ranked-choice voting datasets, we choose $(K, \lambda_{\mathcal{L}}) = (10, 0)$ for both models.

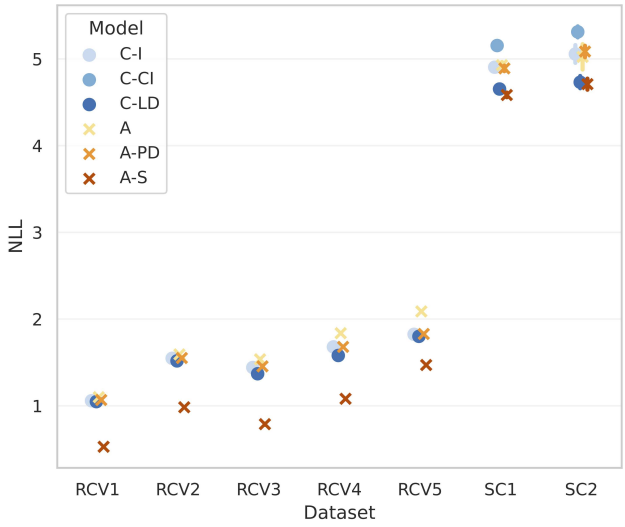


Figure 2: NLL loss (lower is better) of our test datasets under the six models in Table 1. C-I (lightest blue) and A (yellow) are the baselines.

For model optimization, we run Adam [15], implemented in PyTorch, with default parameters, lr $= 0.001$, $\beta = (0.9, 0.999)$, $\epsilon = 1e-8$, adding ℓ_2 regularization with weight $\lambda = 1e-5$. Model parameters are updated over batches of training data until reaching `max_epoch` $= 2000$ or convergence, i.e., when the absolute difference in losses is less than $\epsilon = 1e-4$. Within each dataset, we conducted 5-fold cross-validation, training on four-fifths and evaluating on the held-out fifth. Evaluation metrics are averaged across each of these held-out portions and presented in this section. A repository containing the ranked choice voting datasets, code for our models, and a notebook for recreating the RCV results in this section can be found at https://github.com/ameloa/partial-orders.

7.2 Goodness of fit

We present the negative log likelihood (NLL) loss on the test set of our seven datasets under the six models in Figure 2. Recall that the naive augmented (A) and independent composite (C-I) models serve as our baseline models for modeling distributions over $\Omega(\mathcal{A})$. We see some main themes emerge from these plots. Of the augmented models, assigning only a single fixed-effect to the `END` token, as does the naïve augmented model (A), results in the worst NLL losses on observed data. The conditionally-independent composite model (C-CI), only evaluated on the school choice datasets, yields the second worst losses; the Poisson distributional assumption on length in this model is not representative of our datasets, and thus results in poor NLL losses on observed data. Obviously, researchers should avoid making distributional assumptions that do not hold. Of the remaining four models, the stratified augmented model (A-S) demonstrates the lowest NLL loss on the RCV datasets, and all four show similar NLL results on the school choice datasets.

7.3 Synthetic datasets

A main objectives of this work is to advance the available statistical models for sampling partial orders, key to generating synthetic datasets for forecasting demand or simulating counterfactual policy outcomes. Algorithms 1 and 2 provide pseudocode of the sampling procedure used for the composite and augmented model classes, respectively.

Algorithm 1 Sampling partial orders from the composite model

Input: Learned parameters θ of a composite model.
Output: $Q \in \Omega(\mathcal{A})$ from composite model, π_θ.

Sample length, $l \sim \pi_\theta^k$.
Initialize $Q = \emptyset$ and $A = \mathcal{A}$.
for $i = 1$ to l **do**
    Sample an alternative a from A according to π_θ^R
    $Q \leftarrow Q \succ a$
    $A \leftarrow A \setminus \{a\}$
end for
return Q

Algorithm 2 Sampling partial orders from the augmented model

Input: Learned parameters θ of the augmented model.
Output: $Q \in \Omega(\mathcal{A})$ from augmented model, π_θ.

Initialize $Q = \emptyset$ and $A = \mathcal{A}^+$.
loop
    Sample an alternative a from A according to π_θ^R
    if a is END **then return** Q
    else
        $Q \leftarrow Q \succ a$
        $A \leftarrow A \setminus \{a\}$.
    end if
end loop

We sample $N = 100$ synthetic datasets, $\tilde{D}^{(i)} = \{Q_1^{(i)}, ..., Q_n^{(i)}\}$ for $i \in [N]$, from each of the six models, for each of the seven datasets in Table 2. We use the same covariates $X \in \mathbb{R}^{n \times m \times d}$ for sampling as in training, so the synthetic datasets are seen as simulated preferences of those n households.

Each dataset $\tilde{D}^{(i)}$ produces an empirical length distribution over n examples, and we summarize averaged statistics in Figure 3, taking one RCV and one SC dataset as representative examples. We see that the biggest swings in either direction — longer or shorter lists — come from the augmented class. The naive augmented model (yellow) samples orders that are shorter than the rest, with higher standard deviations on RCV dataset. The single END token of this model learns a high utility to explain shorter-than-full-length lists seen in the observed data. The stratified augmented model, where all alternatives get position-dependent utilities, produces full-length lists in the RCV case. In the school choice case it produces longer, and more dispersed lists. The latter observation suggests that if the sampled list happens to get longer than the modal length of 1, the utility of other alternatives remains higher than END token and the list length continues to grow. Surprisingly, the position-dependent

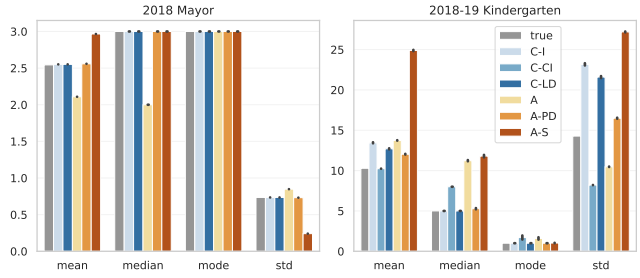


Figure 3: Sampled length distributions statistics on a representative RCV (left) and SC (right) dataset. Statistics of their true distributions in grey. C-CI was not evaluated on the RCV datasets as no voter covariates were available with the data.

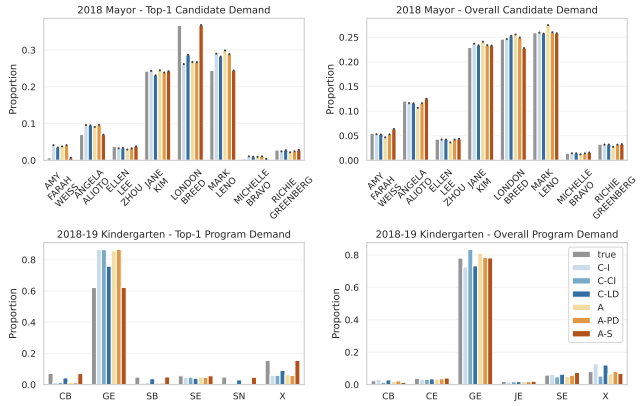


Figure 4: Synthetic demand over 2018 SF Mayoral candidates (top row) and 2018-19 SFUSD program types (bottom row). True demand in grey. Left plots show proportion of choices in first position, right plots show proportion of choices overall.

augmented model, where END token gets position-dependent utilities but the other alternatives utility is fixed down the ranking, matches true lengths distributions as well as or better than when lengths are explicitly modeled as done by the composite class. Overall, the augmented class produces more varied length distributions than the composite class, but still has potential to fit lengths well. The composite class, which models list length explicitly, fits the true length distribution reliably well.

In Figure 4 we showcase the overall and top-choice demand of each candidate in the chosen RCV dataset in the top row, and of the most popular program types in the SFUSD school choice dataset in the bottom row. In the latter plots, rather than presenting every available program, the x-axis buckets school programs by program type — General Education, Chinese Biliteracy, etc. In the left plots of Figure 4, we see that the stratified augmented model (red) fits top choice alternatives best in both the RCV and SC datasets. It is the only model to learn position-1 alternative utilities that are independent of those down rank. The models are showcasing similar

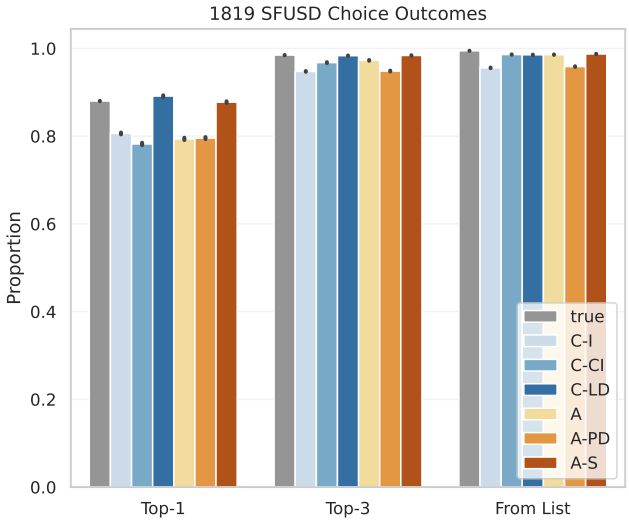


Figure 5: Assignment outcomes using synthetic school choice datasets sampled from our 6 models compared with true outcomes in grey. Proportion of students who were assigned to their top-1, a top-3, or any one of their listed alternatives (as opposed to non-assignment).

performance to one another and matching the overall distribution of demand well in the right plots.

7.4 School assignment outcomes from simulated preferences

Our final analysis aims to demonstrate downstream simulation ability of our proposed methods. We generated $N = 100$ synthetic school choice datasets and simulated real assignment outcomes using SFUSD's 2023 assignment policy. In Figure 5, we summarize assignment outcomes of our six models compared with simulated outcomes using true preference lists. Overall, we find that the stratified models, C-LD and A-S, mimic true outcomes best as they are more expressive preference models than their non-stratified counterparts. The independent composite (C-I) and position-dependent augmented (A-PD) models seem to result in datasets where fewer students gain access to their top choices than in assignments based on the real preferences. These simpler models learn universal weights over certain agent-item attributes, and may lead to "monoculture" effects [16] and greater competition for program offerings than their stratified counterparts. If C-I and A are our baselines, we see there is great potential in using some of the ideas developed in this work, esp. the more expressive stratified models.

8 Conclusions

In this work, we study the statistical modeling of partial orders *not* as marginal events in the space of total orders, but as deliberate individual events. We developed two approaches to modeling top-k partial orders with three implementations each. The sample space of these models are the space of partial orders directly, $\Omega(\mathcal{A})$, and they provide researchers with the ability to sample meaningful synthetic datasets of this type. Under the augmented class of models, whereby end-of-list is modeled as another alternative in the universe, the new alternative should have position dependent fixed-effects, not simply one latent utility. The composite class, which models list length directly, generally produces synthetic datasets that exhibit the most accurate length distributions. Model stratification, applied in both the length-dependent composite model (C-LD) and the K-stratified augmented model (A-S), is an important aspect of preference modeling, improving the accuracy of simulated demand and mimicking true school choice outcomes best in our downstream application experiments.

There are some known limitations with the models and analyses presented in this work. Plackett-Luce ranking models learn ranking distributions over a fixed choice universe, $\mathcal{A}$. As such, the methods presented here are not capable of simulating counterfactuals that add or change the properties of the alternatives in the item set. They are, however, fully applicable to counterfactual evaluations that study when, e.g., the distribution of household covariates changes. Approaches to relaxing these two requirements on the alternative set would increase the range of applications of our models and is an area of future work. Additionally, further analysis is needed to characterize the ranking distributions that are expressible by one or both of our two methods, understand how canonical the representations of each type are, and furnish theoretical guarantees around identifiability and convergence.

Acknowledgments

We thank Irene Lo and Itai Ashlagi for being our liaisons to the San Francisco Unified School District's preference data and Arjun Seshadri for his ongoing modeling support and advice. This work was supported by NSF CAREER Award #2143176.

References

[1] SFUSD annual enrollment highlights. https://www.sfusd.edu/schools/enroll/student-assignment-policy/annual-enrollment-highlights, Feb 2024.
[2] Atila Abdulkadiroğlu, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. Do parents value school effectiveness? *American Economic Review*, 110(5):1502–39, May 2020.
[3] Nikhil Agarwal and Paulo Somaini. Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444, 2018.
[4] Felipe Arteaga, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. Smart matching platforms and heterogeneous beliefs in centralized school choice. Working Paper 28946, National Bureau of Economic Research, June 2021.
[5] Amel Awadelkarim, Arjun Seshadri, Itai Ashlagi, Irene Lo, and Johan Ugander. Rank-heterogeneous preference models for school choice. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 47–56, New York, NY, USA, 2023. Association for Computing Machinery.
[6] Austin R. Benson, Ravi Kumar, and Andrew Tomkins. A discrete choice model for subset selection. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, page 37–45, New York, NY, USA, 2018. Association for Computing Machinery.
[7] Austin R. Benson, Ravi Kumar, and Andrew Tomkins. Sequences of sets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 1148–1157, New York, NY, USA, 2018. Association for Computing Machinery.
[8] Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114(1):175–201, 2006.
[9] Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3):294–318, 1991.
[10] Winship C. Fuller, Charles F. Manski, and David A. Wise. New evidence on the economic determinants of postsecondary schooling choices. *The Journal of Human Resources*, 17(4):477–498, 1982.
[11] Guillermo Gallego, Richard Ratliff, and Sergey Shebalov. A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research*, 63(1):212–232, 2015.

[12] Justine S. Hastings and Jeffrey M. Weinstein. Information, School Choice, and Academic Achievement: Evidence from Two Experiments*. *The Quarterly Journal of Economics*, 123(4):1373–1414, 11 2008.

[13] Matt Kasman and Jon Valant. The opportunities and risks of k-12 student placement algorithms, 2019.

[14] Walter N. Kernan, Catherine M. Viscoli, Robert W. Makuch, Lawrence M. Brass, and Ralph I. Horwitz. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, 52(1):19–26, 1999.

[15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[16] Jon Kleinberg and Manish Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118, 2021.

[17] Meir G Kohn, Charles F Manski, and David S Mundel. An empirical investigation of factors which influence college-going behavior. report no. r-1470-nsf. 1974.

[18] Douglas J. Lanska. Epidemiology and Biostatistics: An Introduction to Clinical Research. *JAMA*, 303(18):1869–1869, 05 2010.

[19] VanderWeele Tyler J. Haneuse Sebastien Lash, Timothy L. and Kenneth J. Rothman. *Modern Epidemiology*. Lippincott Williams and Wilkins, 3 edition, 1986.

[20] Mariana Laverde. Distance to schools and equal access in school choice systems. Working Papers 2022-002, Human Capital and Economic Opportunity Working Group, January 2022.

[21] Min Kyung Lee and Su Baykal. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 1035–1048, New York, NY, USA, 2017. Association for Computing Machinery.

[22] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 3365–3376, New York, NY, USA, 2017. Association for Computing Machinery.

[23] Benjamin Letham, Cynthia Rudin, and Katherine Heller. Growing a list. *Data Mining and Knowledge Discovery*, 27, 11 2013.

[24] Xinran Li, Dingdong Yi, and Jun S. Liu. Bayesian Analysis of Rank Data with Covariates and Heterogeneous Rankers. *Statistical Science*, 37(1):1 – 23, 2022.

[25] R. Duncan Luce. *Individual Choice Behavior: A Theoretical analysis*. Wiley, New York, NY, USA, 1959.

[26] R. Duncan Luce. The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15(3):215–233, 1977.

[27] Nicholas Mattei and Toby Walsh. Preflib: A library for preferences http://www.preflib.org. In Patrice Perny, Marc Pirlot, and Alexis Tsoukiàs, editors, *Algorithmic Decision Theory*, pages 259–270, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[28] Daniel McFadden. The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328, 1974.

[29] Shanjukta Nath. Preference Estimation in Deferred Acceptance with Partial School Rankings. Papers 2010.15960, arXiv.org, October 2020.

[30] R. L. Plackett. Random permutations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(3):517–534, 1968.

[31] Michel Regenwetter and Bernard Grofman. Choosing subsets: a size-independent probabilistic model and the quest for a social welfare ordering. *Social Choice and Welfare*, 15:423–443, 1998.

[32] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[33] Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2 edition, 2009.

[34] Kenneth E. Train, Daniel L. McFadden, and Moshe Ben-Akiva. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, 18(1):109–123, 1987.

[35] Jonathan Tuck, Shane Barratt, and Stephen Boyd. A distributed method for fitting laplacian regularized stratified models. *J. Mach. Learn. Res.*, 22(1), Jan 2021.

[36] Lirong Xia. Learning and decision-making from rank data. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 13:1–159, 02 2019.

[37] Zhibing Zhao and Lirong Xia. Learning mixtures of plackett-luce models from structured partial orders. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.