Deciphering Young's interpretation of Condorcet's model

Lirong Xia
Computer Science Department
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, USA
xial@cs.rpi.edu

Abstract

The Condorcet Jury theorem [5] has been widely regarded as the first attempt to aggregating agents' noisy preferences to reveal the ground truth in social choice settings. The original theorem was stated for two alternatives, and Condorcet himself was not clear how it can be extended to the case of three or more alternatives. Two centuries later, Young's seminal paper [15] pointed out that Condorcet's model can indeed be extended to more than 2 alternatives. Young gave the credit to Condorcet, and among other observations, mentioned that the Kemeny rule can be viewed as a consequence of statistical reasoning under Condorcet's model. Young's interpretation may seem vague because it was not formally defined as a standard statistical model. Instead, Young gave descriptions of independence in Condorcet's model for more than two alternatives, and showed calculations of some probability as examples.

Recently, there are two popular viewpoints of Young's interpretation of Condorcet's model: the maximum likelihood estimator (MLE) of Condorcet's model is the Kemeny rule (a.k.a. *Kemeny-Young method*), and Condorcet's model is equivalent to Mallows model. In this note, we falsify both by showing that these viewpoints are *not* compatible with Young's calculations. Moreover, we showed that there exists a formally defined natural statistical model that is compatible with all Young's descriptions and calculations. This may help us distinguish Young's interpretation of Condorcet's model from other statistical models, especially the Mallows model.

1 Young's interpretation of Condorcet's model

1.1 Condorcet's Jury theorem

Condorcet described a statistical model to explain the Condorcet Jury Theorem: suppose there are two alternatives a and b, the ground truth is composed of two parts: a ground truth pairwise comparison between a and b (called an *opinion*), and a number $0.5 < p < 1$. Then, each agent (jury) generates an pairwise comparison between a and b (opinion) i.i.d. such that with probably p it is the same as the ground truth, and with probability $1 - p$ it is different (that means, the comparison is flipped). The Condorcet Jury theorem states that when the number of agents goes to infinity, the majority aggregation of the agents' opinions converges to the ground truth with probability 1.

1.2 Statistical models and maximum likelihood estimator (MLE)

To apply statistic analysis, the standard way to define a statistical model is composed of the following three parts:[1]

[1] In principle the parameter space is not necessary and the model is simply defined by the sample space and a group of distributions over the sample space. Parameters are used to index these distributions. We focus on parametric models where the distributions are indexed by parameters in finite-dimensional Euclidean space.

1. a sample space S composed of all possible observations (preference profiles in the social choice setting); an element in S is called an observation or data and is denoted by D;

2. a parameter space Θ, which is a subset of finite or infinite dimensional Euclidean space;

3. a probability distribution over S for each $\theta \in \Theta$, denoted by $\Pr(D|\theta)$, such that $\sum_{D \in S} \Pr(D|\theta) = 1$. Taking a Bayesian viewpoint that considers the joint probability distribution between the sample space and the parameter space, $\Pr(D|\theta)$ is the conditional probability given any parameter θ. However, taking a frequentist's viewpoint $\Pr(D|\theta)$ is just a distribution over S indexed by θ, and it is *not* the conditional probability since there is no distribution over the parameter space Θ. We will abuse the notation and use $\Pr(D|\theta)$, but we want to emphasize that this should be read as "the probability distribution indexed by θ" rather than "the conditional probability distribution given θ", unless in the Bayesian setting.

In many cases an observation is composed of n i.i.d. generated data, each of which is drawn from a space X with probability $\Pr(\cdot|\theta)$, so that $S = X \times \cdots \times X$ and for each $D = (V_1, \ldots, V_n) \in S$, $\Pr(D|\theta) = \prod_{j=1}^{n} \Pr(V_j|\theta)$.

Let $\mathcal{C}$ denote the set of alternatives (or candidates). Let $\mathcal{L}(\mathcal{C})$ denote the set of all linear orders (a.k.a. total orders or rankings) over $\mathcal{C}$, that is, $\mathcal{L}(\mathcal{C})$ is the set of all transitive, antisymmetric, and total binary relations over $\mathcal{C}$. Let $\mathcal{O}(\mathcal{C})$ denote the set of all *opinions* over $\mathcal{C}$, where an opinion is composed of pairwise comparisons between all $m(m-1)/2$ pairs of alternatives in $\mathcal{C}$. Mathematically $\mathcal{O}(\mathcal{C})$ is the set of all asymmetric and total binary relations over $\mathcal{C}$. $\mathcal{O}(\mathcal{C}) = \prod_{\{a,b\}} \{a \succ b, b \succ a\}$. In Young's paper, an "opinion" by default refers to a linear order, and in some rare cases an "impossible opinion" refers to a cyclic order. In this note, we distinguish an opinion from a linear order. Throughout the paper, m is the number of alternatives and n is the number of agents (voters). We give two examples of formally-defined statistical models in preference aggregation.

Example 1 (Condorcet's model for two alternatives)**.** *Let $\{a, b\}$ denote the two alternatives and suppose there are n agents. Fix $0.5 < p < 1$.*

- *The parameter space $\Theta = \{a \succ b, b \succ a\}$ that represents the ground truth opinion between a and b. This can be seen as a discrete subspace of $\mathbb{R}$.*

- *The sample space is $S = X^n = \{a \succ b, b \succ a\}^n$, where $X = \{a \succ b, b \succ a\}$. For each $D = (V_1, \ldots, V_n) \in S$ and any $j \leq n$, V_j represents agent j's vote.*

- *For any $D \in S$ and $W \in \Theta$, $\Pr(D|W) = \prod_{j=1}^{n} I_p(V_j, W)$, where $I_p(V_j, W) = \begin{cases} p & \text{if } V_j = W \\ 1-p & \text{if } V_j \neq W \end{cases}$.*

It is not clear where p is a component of the parameters in Condorcet's original definition. Adding p to the parameter space will give us another model where statistical inference can be different.

Another example is the *Mallows* model.

Example 2 (Mallows model [9])**.**

- *The parameter space Θ is a subspace of $\mathbb{R}^2$: the first component is $\mathcal{L}(\mathcal{C})$, which is discrete and it represents the ground truth* ranking *among the alternatives in $\mathcal{C}$; the second component is denoted by φ with $0 < \varphi < 1$.*

- *The sample space is $S = \mathcal{L}(\mathcal{C})^n$.*

- *For any $D \in S$ and $\theta = (W, p)$, $\Pr(D|\theta) = \prod_{j=1}^{n} I(V_j, \theta)$, where $I(V_j, \theta) = \frac{1}{Z}\varphi^{\text{Kendall}(V_j, W)}$, and Kendall$(V_j, W)$ is the Kendall tau distance between V_j and W, and Z is a normalization factor so that $\sum_{V \in \mathcal{L}(\mathcal{C})} \varphi^{\text{Kendall}(V,W)} = Z$.*

In both models, each agents' vote is generated i.i.d. from X. Given an observation D in the sample space, the likelihood function is a mapping whose domain is the parameter space. For any $\theta \in \Theta$, $\mathrm{L}(\theta, D) = \Pr(D|\theta)$. A *maximum likelihood estimator (MLE)* chooses a parameter that maximizes the likelihood. That is,

$$\mathrm{MLE}(D) \in \arg\max_{\theta \in \Theta} L(\theta, D)$$

We note that while the term "maximum likelihood" has been extensively used, the standard textbook definition of MLE refers to the function that must output a parameter in the unrestricted parameter space with the maximum likelihood. We are not aware of terms like "MLE for a subset of the parameter space" or "MLE restricted on a subset of the parameter space".[2]

1.3 Young's description of Condorcet's model for $m \geq 3$

Young did not give a formal definition of Condorcet's model by specifying the sample space and parameter space, but instead, he gave some descriptions and calculations. Below are some of Young's quotes in [15] and our interpretation w.r.t. the three components of the formal definition of a statistical model. We will discuss Young's calculations in Section 4.

1. Sample space.
   - **Quote:** Young wrote *"Consider a list of candidates (or decisions) that are voted upon pairwise* Then, later he wrote *"Normally, each individual voter is able to rank all of the candidates in a consistent order."*
   - **Interpretation:** The first sentence seems to suggest that the sample space is composed of all combinations of opinions, one for each agent. The second sentence makes it a little vague: it is not clear whether each agent's opinion is required to be a linear order.
   - **Possibilities:** The sample space can be either (1) $\mathcal{O}(\mathcal{C})^n$, that is, each agent votes for a (possibly cyclic) opinion; or (2) $\mathcal{L}(\mathcal{C})^n$, that is, each agent votes for a linear order.
2. Parameter space:
   - **Quote:** Young wrote *"a state of nature corresponds to a true or correct ordering of the candidates"*.
   - **Interpretation:** From the context, it seems that "a state of nature" refers to a parameter. Therefore, it seems reasonable to assume that the parameter space should be all orderings over candidates (alternatives). However, it is not clear whether here an "ordering" is a linear order or an opinion (possibly cyclic order).
   - **Possibilities:** The parameter space can be either (1) $\mathcal{O}(\mathcal{C})$, that is, the ground truth is a (possibly cyclic) opinion; or (2) $\mathcal{L}(\mathcal{C})$, that is, the ground truth must be a linear order.
3. Probability distributions:

[2]There is a term called "restricted maximum likelihood (REML)", which maximizes a different function rather than putting a restriction on the parameter space. There are some studies on estimating parameters in a restricted parameter space, for example see [11] for a recent survey. However, we do not see an occurrence of "restricted MLE" or "MLE on a restricted domain" in the survey, though "unrestricted MLE" was sometimes used to refer to the usual MLE.

- **Quote:** Young wrote the following:
  - *"Condorcet assumed, first, that in any pairwise comparison each voter will choose the better candidate with some fixed probability p, where $1/2 < p < 1$ and p is the same for all voters."*
  - *"Second, every voter's judgment on every pair of candidates is independent of his or her judgment on every other pair.*
  - *"Finally, each voter's judgment is independent of the other voters' judgments (a significant assumption).*
- **Interpretation:** The first part identifies the probability of each pairwise comparison for any given parameter. The way that the first part was phrased has already suggested that pairwise comparisons are mutually independent (for both within-voter comparisons and across-voter comparisons) in any distribution indexed by a parameter. Given this, independence relationships stated in the second and third parts become less clear: our best guess is that they further emphasize that the second part says that in the distribution associated with any parameter, the pairwise comparisons are independent for the same voter, and the third part is the usual i.i.d. assumption.

2 (Mis)understandings of Young's interpretation

It is quite easy to come up with the following two understandings w.r.t. MLE and Condorcet's model, though they are not written down exactly word-by-word in any formal publications. Specifically, Young did not mention them in his paper [15].

Viewpoint 1. *The Kemeny rule is the maximum likelihood estimator (MLE) of Condorcet's model.*

When mentioning the Kemeny rule, Young never said that it is equivalent to the MLE of Condorcet's model. Instead, he pointed out that:

This procedure [finding the most likely ranking of the alternatives]*, which is sometimes known as "Kemeny's rule,"*

A potential source for Viewpoint 1 is that having the definition of MLE in mind, it is quite easy to come up with this when reading various restatements of Young's contribution. For example, Conitzer and Sandholm [6] commented:

"The basic idea of using an MLE approach to voting was introduced as early as the 18th century, by Condorcet...The solution for arbitrary numbers of candidates was given two centuries later by Young: he showed that it [MLE approach to voting] *coincided with a voting rule proposed by Kemeny."*

Pivato [12] states:

"Young showed that the Kemeny rule can be seen as the maximum-likelihood estimator (MLE) of the 'true' preference ordering over a set of candidates, while the Borda rule is the MLE of the best candidate."

Azari Soufiani, Parkes, and Xia [2] mentioned:

"Later, Kemenys rule was shown to provide the maximum likelihood estimator (MLE) for this [Condorcet's] *model."*

The wikipedia entry for "Kemeny-Young method"[3] states:

"Young adopted an epistemic approach to preference-aggregation: he supposed that there was an objectively 'correct', but unknown preference order over the alternatives, and voters receive noisy signals of this true preference order (cf. Condorcet's jury theorem.) Using a simple probabilistic model for these noisy signals, Young showed that the Kemeny-Young method was the maximum likelihood estimator of the true preference order."

[3] `http://en.wikipedia.org/wiki/Kemeny-Young_method`

These descriptions might be misleading because as we commented above, MLEs may output any element in the parameter space. Therefore, "maximum likelihood estimator of the true preference order" sounds like: (1) the parameter space of Condorcet's model consists of all preference orders (from the context, a preference order should be a linear order), or (2) even though the parameter space may be larger, the output of an MLE is always a linear order. We will see later that none of these interpretations are correct.

Since it is quite straightforward to see that Kemeny-Young method is equivalent to the MLE of the Mallows model, a closely relevant viewpoint is the following.

Viewpoint 2. *Condorcet's model is equivalent to Mallows.*

In [14], Xia and Conitzer described Condorcet's model as Mallows by *"In Condorcets model, given the correct ranking V, for each pair of alternatives c and c' where $c \succ_V c'$, with probability $p > 1/2$ the voter also prefers c to c' (and with probability $1 - p$, she prefers c' to c)."* The *noisy order model* defined by Procaccia, Reddi, and Shah [13] is essentially Mallows model with uniform prior, and the authors commented "In this noisy orders model (also known as the Condorcet noise model)". Mao, Procaccia, and Chen [10] noted *"the maximum likelihood estimator in the Mallows model–the Kemeny rule–had to wait another two centuries to receive due recognition [Young 1988]"*. Also in [4], Caragiannis, Procaccia, and Shah wrote *"We note that this model* [Mallows model] *is equivalent to the Condorcet noise model."*

3 Why should we care?

From a historical point of view it is important to identify Condorcet's model. Beyond that, should we care whether Condorcet's model is different from Mallows model? Practically, does this make any difference in decision-making?

We think the answer is affirmative, which is the main purpose of writing this paper. We will see that Kemeny's rule is not the MLE of Condorcet's model when $m \geq 3$, but rather, it is the estimator that chooses parameters in a subspace (the set of linear orders) with maximum likelihood. This type of estimators may not be well-justified and one criticism is that if we think the ground truth must be a linear order, then why not build a model whose parameter space is exactly the set of all linear orders as in Mallows model? Is there a principled way we can justify this approach beyond the hypothesis testing idea Young proposed in the paper?

Our answer is yes, and the key idea is a clear separation between statistical inference and decision. Usually statistical inference are used to make some argument about the true parameter, for example which parameter is more probable. However, knowing this might not be the best interest for decision making, and in many cases the criterion should be quality of decision, whose relationship with the most probable parameter is not very clear and can be defined in many natural ways.

This separation can be better seen when the parameter space is different from the decision space. For example, let us reexamine the situation in Calculation 2 and Calculation 3 in Section 4. Suppose the data are generated from $\mathcal{M}_p$ and we want to select the most probable winner. In this case the parameter space is $\mathcal{O}(\mathcal{L})$ and the decision space is $\mathcal{C}$. There are many seemingly natural ways we can select the winner, including the following.

1. Choose the top-ranked alternative in the Kemeny winning ranking, which was proposed and analyzed by Fishburn [7]. Young commented that this approach is "not necessarily the most convincing answer".

2. Choose the alternative c with the highest $\Pr(c|M)$, where $\Pr(c|M) = \sum_{W \in \mathcal{O}(\mathcal{C}): c \text{ is ranked in the top of } W} \Pr(W|M)$. According to Proposition 3, this seems to be what Young meant in his paper.

3. Choose the alternative c with the highest $\Pr^*(c|M)$, where $\Pr^*(c|M) = \sum_{R\in\mathcal{L}(\mathcal{C}):c \text{ is ranked in the top of } R} \Pr(R|M)$. This is also natural but it can be different from 2.

Though it may be easy to say that 2 seems more natural than 1, it is probably much harder say which of 2 and 3 is better. To understand what is a good way to made decisions in this case, we think statistical decision theory can help. Statistical decision theory [3] provides principled guidelines and methods towards making decisions with a statistical model, especially for cases where the parameter space is different from the decision space, and the quality of a decision is modeled by a loss function whose inputs are the true parameter and the decision. In fact, 2 and 3 corresponds to the Bayesian approach in statistical decision theory with different loss functions. See [1] for more discussions.

4 Young's three calculations

In Young's calculations, we do not directly observe each agent's vote but instead, we are given *voting matrices*. A voting matrix is a summary of pairwise comparisons in a preference profile. Thus a voting matrix loses some information.

While the pairwise-comparison nature of Young's interpretation of Condorcet's model has been well-perceived (see e.g. [8]), the independence (or more precisely, conditional independence in the Bayesian setting) between pairwise comparisons seems to be less clear. In this section we recall three of Young's key calculations. These calculations will help us in the next section to decipher Young's interpretation of Condorcet's model.

	a	b	c
a	—	8	6
b	5	—	11
c	7	2	—

(i)

	a	b	c
a	—	23	29
b	37	—	29
c	31	31	—

(ii)

Table 1: Two voting matrices.

Calculation 1 (p.1235 [15])**.** *Suppose there are* 13 *voters and three candidates* $\{a, b, c\}$*, and the voting matrix* M_1 *is shown in Table 1 (i). Then, given that the true ranking is* $a \succ b \succ c$*, the probability of observing the voting matrix is*

$$L(a \succ b \succ c) = \Pr(M_1|a \succ b \succ c) = \frac{13!}{8!5!}[p^8(1-p)^5]\frac{13!}{6!7!}[p^6(1-p)^7]\frac{13!}{11!2!}[p^{11}(1-p)^2] = \frac{(13!)^3}{8!5!6!7!11!2!}[p^{25}(1-p)^{14}]$$

This calculation itself is not hard to understand by treating each pairwise comparison in the voting matrix separately. Then the combinatorial numbers $\frac{13!}{8!5!} = \binom{13}{5}$ is the number of times $a \succ b$ appears 8 times and $b \succ a$ appears 5 times in 13 i.i.d. generated pairwise comparisons. We will provide a formal calculation using the formally defined model in Section 5.

Calculation 2 (p.1238 [15])**.** *Suppose there are* 60 *voters and three candidates* $\{a, b, c\}$*, and the voting matrix* M_2 *is shown in Table 1 (ii). Taking a Bayesian approach, the probability for c to be ranked in the top of the ground truth is calculated as follows.*

$$\Pr(c \textit{ ranked in the top of the ground truth}|M_2) = \Pr(c \succ a \textit{ and } c \succ b|M_2) = \Pr(c \succ a|M_2)\times\Pr(c \succ b|M_2)$$

Young explained the meaning for c to be ranked in the top of the ground truth by saying that it is *"top-ranked in the true ranking"* (p. 1237). This seems to suggest that we are computing

$$\sum_{\text{ranking } R \text{ where } c \text{ is ranked in the top}} \Pr(R|M_2)$$

We will see that in the hypothetical Condorcet model defined in Section 5, this holds for 3 alternatives. For more than three alternatives we think "c is ranked in the top of the ground truth" means that c is ranked in the ground truth *opinion*, that is, $\sum_{\text{opinion } W \text{ where } \forall a \in \mathcal{C}-\{c\}, c \succ a} \Pr(W|M_2)$.

Another feature of the calculation is that it assumes that event $c \succ a$ and event $c \succ b$ are independent in the parameter space given the voting matrix. There are two potential confusions about this calculation.

1. Even though it is typical to assume independence in data given parameter, usually we do automatically obtain independence in the parameter space given data. In fact, once the prior distribution over the parameter space and the conditional distribution are given (which is the case in Young's paper), the joint distribution is fully determined. Thus such independence should be a consequence of statistical inference rather than assumption.

2. Another natural question is whether a similar calculation can be done for more than three alternatives, namely, for any $c \in \mathcal{C}$ and voting matrix M, do we have the following under the uniform prior?

$$\Pr(c \text{ ranked in the top}|M) = \prod_{d \neq c} \Pr(c \succ d|M)$$

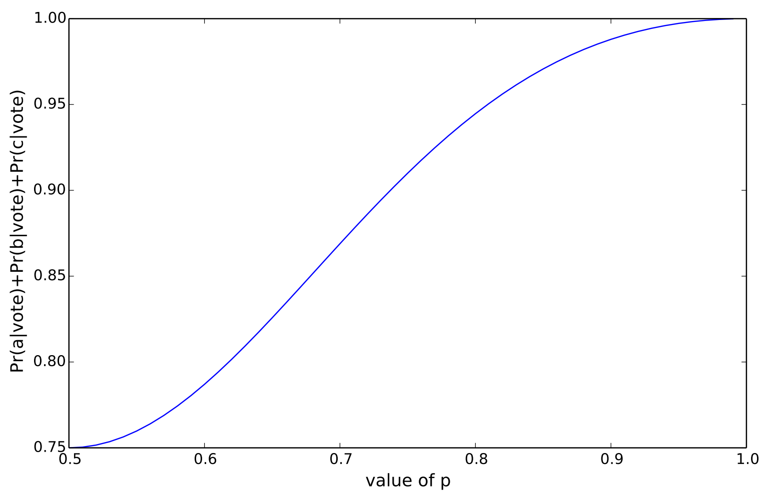

Figure 1: $\Pr(a|M_2) + \Pr(b|M_2) + \Pr(c|M_2)$ for different p.

Calculation 3 (p.1238 [15]). *For the voting matrix M_2 defined in Table 1 (ii),*

$$\begin{aligned}
&\Pr(a|M_2) + \Pr(b|M_2) + \Pr(c|M_2)\\
=&\frac{[p^{31}(1-p)^{29}]^2}{[p^{31}(1-p)^{29} + p^{29}(1-p)^{31}]^2} + \frac{p^{37}(1-p)^{23}p^{29}(1-p)^{31}}{[p^{37}(1-p)^{23} + p^{23}(1-p)^{37}] \times [p^{29}(1-p)^{31} + p^{31}(1-p)^{29}]}\\
&+ \frac{p^{23}(1-p)^{37}p^{29}(1-p)^{31}}{[p^{23}(1-p)^{37} + p^{37}(1-p)^{23}] \times [p^{29}(1-p)^{31} + p^{31}(1-p)^{29}]}
\end{aligned}$$

For different p, the probability of $\Pr(a|M_2) + \Pr(b|M_2) + \Pr(c|M_2)$ in Calculation 3 is plotted in Figure 1. We can see that this probability is smaller than 1 for all $p < 1$. This might be confusing since if the ground truth is always a ranking, then such posterior distributions should always sum up to one.

5 Deciphering Young's interpretation

We now present a model that fits all Young's description and calculations mentioned above. For any opinion $W \in \mathcal{O}(\mathcal{C})$ and any pair of alternatives $\{a, b\}$, we let $W_{\{a,b\}}$ denote the pairwise comparison between a and b in W.

Definition 1 (Hypothetical Condorcet's model)**.** *For any $0.5 < p < 1$, we define the following model $\mathcal{M}_p$.*

- *Sample space: $\mathcal{O}(\mathcal{C})^n$. That is, each agent votes for an opinion.*
- *Parameter space: $\mathcal{O}(\mathcal{C})$. That is, a ground truth is an opinion.*
- *Probability distributions: for any parameter $W \in \mathcal{O}(\mathcal{C})$ and $D \in \mathcal{O}(\mathcal{C})^n$, let $\Pr(D|W) = \prod_{O\in D} \Pr(O|W)$, where for some fixed $0.5 < p < 1$,*

$$\Pr(O|W) = \prod_{\{a,b\}\subseteq\mathcal{C}} I_p(O_{\{a,b\}}, W_{\{a,b\}})$$

We recall that $I_p(O_{\{a,b\}}, W_{\{a,b\}}) = \begin{cases} p & O_{\{a,b\}} = W_{\{a,b\}} \\ 1-p & O_{\{a,b\}} \neq W_{\{a,b\}} \end{cases}$

That is, we assume that agents' votes are generated i.i.d. Moreover, for any agent and any pair of candidates $\{a, b\}$, the pairwise comparison between a and b only depends on the pairwise comparison between a and b in the ground truth. So the model can be thought of as $m(m-1)/2$ independent Condorcet's models, one for each pair of alternatives.

5.1 How does this model explain Young's calculation?

We first formally specify the events in Young's calculations. Let M be a voting matrix, e.g. the voting matrix in Table 1(i). For any pair of alternatives (a, b), let $M(a \succ b)$ denote the number of times $a \succ b$ in M. Observing M is the event (formally, a subset of the sample space) that is composed of all $D \in \mathcal{O}(\mathcal{C})^n$ where for all pairs of alternatives (a, b), the number of times that $a \succ b$ in D is exactly $M(a, b)$. Therefore it is reasonable to formally define the event of observing M as follows.

Definition 2. *For any voting matrix M, define Event$(M) \subseteq \mathcal{O}(\mathcal{C})^n$ so that for all $D \in$ Event(M) and all pairs of alternatives (x, y), the number of times that $x \succ y$ in D is exactly $M(x \succ y)$.*

In the sequel, when computing the posterior distribution in the parameter space, we sometimes write $\Pr(\theta|M)$ instead of $\Pr(\theta|\text{Event}(M))$.

Calculation 1. The following proposition verifies Calculation 1 under $\mathcal{M}_p$.

Proposition 1. *Calculation 1 holds under $\mathcal{M}_p$.*

Proof. Let M_1 denote the voting matrix in Table 1. Note that for any $D \in \text{Event}(M_1)$, we have $\Pr(D|a \succ b \succ c) = p^{25}(1-p)^{14}$. The total number of such D's can be calculated as follows: for any pair of alternatives $\{x, y\}$, let $Q_{\{x,y\}}$ denote the subset of $\{x \succ y, y \succ x\}^{13}$ where in each element the number of occurrence of $x \succ y$ is $M_1(x \succ y)$. There is a one-one correspondence between $\text{Event}(M_1)$ and $\prod_{\{x,y\}} Q_{\{x,y\}}$: for any $\{x, y\}$, $D \in \text{Event}(M_1)$ corresponds to its sub-opinion profile of D for pairwise comparison between $\{x, y\}$, denoted by $D|_{\{x,y\}}$, where the jth component of $D|_{\{x,y\}}$ is $x \succ y$ if and only if the jth agent prefers x to y. Therefore, $|\text{Event}(M_1)| = |\prod_{\{x,y\}} Q_{\{x,y\}}| = \frac{(13!)^3}{8!5!6!7!11!2!}$. This proves the proposition. $\square$

Calculation 2. For any number of alternatives and any preference profile $D \in \mathcal{O}(\mathcal{C})^n$, the next proposition shows that under $\mathcal{M}_p$, the events $c \succ y$ for all $y \in \mathcal{C} - \{c\}$ are mutually independent conditioned on D.

Proposition 2. *Let $0.5 < p < 1$. Taking a Bayesian viewpoint of $\mathcal{M}_p$ and assuming uniform prior, for any $c \in \mathcal{C}$, the $m-1$ events in $\{c \succ y : y \in \mathcal{C} - \{c\}\}$ are conditionally mutually independent given any $D \in \mathcal{O}(\mathcal{C})^n$. That is, for any set $C \subseteq \mathcal{C} - \{c\}$, we have:*

$$\Pr(\bigcap_{y \in C} c \succ y | D) = \prod_{y \in C} \Pr(c \succ y | D)$$

Proof. Given a preference profile $D \in \mathcal{O}(\mathcal{C})^n$, for any pair of alternatives x, y we let $\#(x \succ y)$ denote the number of times $x \succ y$ in D. It follows that $\#(x \succ y) + \#(y \succ x) = n$. We let $T_c(C)$ denote the set of opinions D' where for all $y \in C$, we have $c \succ y$ in D'. That is, $T_c(C) = \{W \in \mathcal{O}(\mathcal{C}) : \forall y \in C, c \succ y \text{ in } W\}$. We also let $R_c(C)$ denote the set of unordered pairs $\{x, y\}$ such that $\{x, y\} \not\subseteq C \cup \{c\}$. For example, when $\mathcal{C} = \{c, c_1, c_2, c_3\}$ and $C = \{c_1, c_2\}$, then $R_c(C) = \{\{c, c_3\}, \{c_1, c_2\}, \{c_1, c_3\}, \{c_2, c_3\}\}$.

$$\begin{aligned}
&\Pr(\bigcap_{y \in C} c \succ y | D) \\
=&\sum_{W \in T_c(C)} \Pr(W|D) \\
=&\sum_{W \in T_c(C)} \frac{\Pr(D|W) \cdot \Pr(W)}{\Pr(D)} \qquad \text{(Bayes Formula)} \\
=&\frac{1}{2^{m(m-1)/2}\Pr(D)} \sum_{W \in T_c(C)} \Pr(D|W) \qquad \text{(uniform prior)} \\
=&\frac{1}{2^{m(m-1)/2}\Pr(D)} \sum_{W \in T_c(C)} \left\{ \prod_{(x,y): x \succ y \text{ in } W} \left\{ p^{\#(x \succ y)} \cdot (1-p)^{\#(y \succ x)} \right\} \right\} \\
=&\frac{1}{2^{m(m-1)/2}\Pr(D)} \prod_{y \in C} p^{\#(c \succ y)}(1-p)^{\#(y \succ c)} \sum_{W \in T_c(C)} \left\{ \prod_{(x,y): x \succ y \text{ in } W \text{ s.t. } [x \neq c \text{ or } y \notin C]} \left\{ p^{\#(x \succ y)} \cdot (1-p)^{\#(y \succ x)} \right\} \right\} && (1) \\
=&\frac{1}{2^{m(m-1)/2}\Pr(D)} \prod_{y \in C} p^{\#(c \succ y)}(1-p)^{\#(y \succ c)} \prod_{\{x,y\} \in R_c(C)} \left(p^{\#(x \succ y)} \cdot (1-p)^{\#(y \succ x)} + p^{(\#y \succ x)} \cdot (1-p)^{\#(x \succ y)} \right) && (2)
\end{aligned}$$

The intuition behind the last equation is that when computing $\Pr(D|W)$ for any $W \in T_c(C)$, a common factor is $\prod_{y \in C} p^{\#(c \succ y)}(1-p)^{\#(y \succ c)}$ since all pairwise comparisons between c and C are fixed as $c \succ y$ for all $y \in C$. For each remaining pairwise comparisons $\{x, y\}$ in $R_c(C)$, if $x \succ y$ in W, then it contributes a $p^{\#(x \succ y)} \cdot (1-p)^{\#(y \succ x)}$ multiplicative factor to $\Pr(D|W)$; otherwise it contributes a $p^{\#(y \succ x)} \cdot (1-p)^{\#(x \succ y)}$ factor to $\Pr(D|W)$. Since opinions in $T_c(C)$ cover all combinations of pairwise comparisons in $R_c(C)$, it follows that expanding (2) gives us exactly (1).

Following a similar calculation, we have the following:

$$\begin{aligned}
\Pr(D) =& \sum_{W \in \mathcal{O}(\mathcal{C})} \Pr(W) \times \Pr(D|W) \\
=&\frac{1}{2^{m(m-1)/2}} \prod_{\{x,y\}} \left(p^{\#(x \succ y)} \cdot (1-p)^{\#(y \succ x)} + p^{\#(y \succ x)} \cdot (1-p)^{\#(x \succ y)} \right) && (3)
\end{aligned}$$

Similarly, we can calculate $\Pr(c \succ y|D)$ for any $y \in C$ as follows:

$$\Pr(c \succ y|D) = \sum_{W \in T_c(\{y\})} \Pr(W|D)$$

$$= \frac{1}{2^{m(m-1)/2}\Pr(D)} p^{\#(c \succ y)}(1-p)^{\#(y \succ c)} \prod_{\{x,z\} \neq \{a,c\}} \left(p^{\#(x \succ z)} \cdot (1-p)^{\#(z \succ x)} + p^{\#(z \succ x)} \cdot (1-p)^{\#(x \succ z)}\right)$$

$$= \frac{p^{\#(c \succ y)}(1-p)^{\#(y \succ c)}}{p^{\#(c \succ y)}(1-p)^{\#(y \succ c)} + p^{\#(y \succ c)}(1-p)^{\#(c \succ y)}} \tag{4}$$

Substituting $\Pr(D)$ by Equation (3) in (2), we have:

$$\Pr(\bigcap_{y \in C} c \succ y|D) = \text{Equation } (2)$$

$$= \prod_{y \in C} \frac{p^{\#(c \succ y)}(1-p)^{\#(y \succ c)}}{p^{\#(c \succ y)}(1-p)^{\#(y \succ c)} + p^{\#(y \succ c)}(1-p)^{\#(c \succ y)}} \tag{5}$$

$$= \prod_{y \in C} \Pr(c \succ y|D)$$

This proves the proposition. □

Let $C = \mathcal{C}$. We have:

$$\Pr(c \text{ is ranked in the top of an opinion}|D) = \prod_{y \neq c} \Pr(c \succ y|D)$$

A similar conditional mutual independence can be proved given any voting matrix M.

Proposition 3. *Let $0.5 < p < 1$. Taking a Bayesian viewpoint of $\mathcal{M}_p$ and assuming uniform prior, for any $c \in \mathcal{C}$, for any set $C \subseteq \mathcal{C} - \{c\}$, and any voting matrix M, we have:*

$$\Pr(\bigcap_{y \in C} c \succ y|M) = \prod_{y \in C} \Pr(c \succ y|M)$$

Proof. We cannot prove the independence directly after Proposition 2, since if we know events A and B are conditionally independent given each of the two events E_1 and E_2, in general A and B are not conditionally independent given $E_1 \cup E_2$. We will use the calculations in Proposition 2 to prove the independence.

Notice that in (4) and (5) the calculation only depends on the number of occurrences of $c \succ y$ in D. Therefore, for all $D_1, D_2 \in \text{Event}(M)$ and for all y, $\Pr(c \succ y|D_1) = \Pr(c \succ y|D_2)$ and $\Pr(\bigcap_{y \in C} c \succ y|D_1) = \Pr(\bigcap_{y \in C} c \succ y|D_2)$. Since D_1 and D_2 are disjoint in the sample space, they are also disjoint in the joint space of sample and parameter. Therefore, we have:

$$\Pr(c \succ y|M) = \Pr(M|c \succ y)\frac{\Pr(c \succ y)}{\Pr(M)}$$

$$= \sum_{D \in \text{Event}(M)} \Pr(D|c \succ y)\frac{\Pr(c \succ y)}{\Pr(M)}$$

$$= \sum_{D \in \text{Event}(M)} \Pr(c \succ y|D)\frac{\Pr(D)}{\Pr(M)} \qquad \text{(Bayes' rule)}$$

$$= \Pr(c \succ y|D^*) \qquad \text{(for any } D^* \in \text{Event}(M))$$

Similarly we can prove that $\Pr(\bigcap_{y\in C} c \succ y|M) = \Pr(\bigcap_{y\in C} c \succ y|D^*)$ for any $D^* \in \text{Event}(M)$. Therefore, choosing an arbitrary $D^* \in \text{Event}(M)$ we have:

$$\begin{aligned}
&\Pr(\bigcap_{y\in C} c \succ y|M) = \Pr(\bigcap_{y\in C} c \succ y|D^*) \\
=& \prod_{y\in C} \Pr(c \succ y|D^*) \qquad \text{(Proposition 2)} \\
=& \prod_{y\in C} \Pr(c \succ y|M)
\end{aligned}$$

This proves the proposition. □

Remark: Young described the calculation as " Candidate c will be best [i.e., top-ranked in the true ranking] if two propositions hold: namely, if c is better than a ($c \succ a$) and c is better than b ($c \succ b$)". This is true for the case of three alternatives, since if $c \succ a$ and $c \succ b$, then no matter the comparison between a and b is, we will always have a linear order. As we see in Proposition 2 and Proposition 3, when then number of alternatives is more than three, such independence holds if "c is the best" means that c is top-ranked in the truth *opinion*. The next section will answer two closely relevant questions:

1. For $m > 3$, does the conditional independence still holds if Pr(c is the best$|D$) is defined as

$$\sum_{\text{ranking} R \text{ where } c \text{ is ranked in the top}} \Pr(R|D)$$

2. Does the conditional independence still holds when the parameter space is composed of all rankings, as in the Mallows model?

Calculation 3. Let M_2 denote the voting matrix as in Table 1(ii). According to Proposition 3, when $m = 3$, we have:

$$\Pr(c|M_2) = \Pr(c \text{ ranked in the top } |M_2) = \Pr(c \succ a \succ b|M_2) + \Pr(c \succ b \succ a|M_2)$$

Therefore, $Pr(c|M_2) + Pr(a|M_2) + Pr(b|M_2) = 1 - \Pr(c \succ a, a \succ b, b \succ c|M_2) - \Pr(a \succ c, b \succ a, c \succ b|M_2)$. This explains why the total probability of either a,b b, or c is ranked in the top of the ground truth is not 1. One explanation of the remaining probability can be "there is no clear winner".

5.2 An unsolved mystery

On page 1237, Young wrote "Then the probability of observing the vote 31 for c, 29 for a is"

$$\Pr(vote) = (60!/31!29!)(p^{31}(1-p)^{29} + p^{29}(1-p)^{31})/2 \tag{6}$$

Then in the following paragraphs Young used "vote" for multiple times in his calculations. It might not be very clear what event does "vote" refers to. Clearly in the above calculation "vote" is the event where 31 agents voted for c, and the rest 29 voted for a, regardless of their other pairwise comparisons. However, on the bottom of the page Young wrote $\Pr(c \succ b|vote)$, which seems to suggest that "vote" in this context means observing the corresponding pairwise comparison between c and b (as opposed to c and a in (6)) in the voting matrix. Continuing, on the top of the next page Young did calculations for $\Pr(c|vote)$, $\Pr(b|vote)$, $\Pr(a|vote)$, and interpret them as "posterior distributions" to determine which alternative is more likely to be the true winner. For the sake of fair comparison, it is natural to assume that these three occurrences of "vote" refer to the same event, which should be the voting matrix.

Is it possible that "vote" refers the same event in all these calculations?

Our answer is negative. If "vote" denotes a single event, then it must be the voting matrix as argued above for $\Pr(c|vote)$, $\Pr(a|vote)$ and $\Pr(b|vote)$. However, Young explicitly mentioned that for (6), "vote" means "observing the vote 31 for c, 29 for a" rather than the full voting matrix, which is the event as we defined in Definition 2. Another evidence is, using our calculation in (3), we have $\Pr(M) = \sum_{D\in \text{Event}(M)} \Pr(D)$. Then (6) would become:

$$\frac{(60!)^3}{8\times 31!29!31!29!37!23!}\left[p^{31}(1-p)^{29}+p^{29}(1-p)^{31}\right]^2\left[p^{37}(1-p)^{23}+p^{23}(1-p)^{37}\right]$$

Hence in different calculations "vote" should refer to different events. We now present a conjecture of the meanings of "vote" under $\mathcal{M}_p$. First, let us formally define the event Young described in (6).

Definition 3. *For any voting matrix M and any pair of alternatives $\{x,y\}$, define $M_{\{x,y\}} \subseteq \mathcal{O}(\mathcal{C})^n$ so that for all $D \in M_{\{x,y\}}$, the number of times that $x \succ y$ in D is exactly $M(x \succ y)$ and the number of times that $M(y \succ x)$ in D is exactly $M(y \succ x)$.*

Define $M_{\{x\}} \subseteq \mathcal{O}(\mathcal{C})^n$ so that for all $D \in M_{\{x\}}$ and all $y \neq x$, the number of times that $x \succ y$ in D is exactly $M(x \succ y)$ and the number of times that $M(y \succ x)$ in D is exactly $M(y \succ x)$.

By definition, $M_{\{x\}} = \bigcap_{y\neq x} M_{\{x,y\}}$. Note that both $M_{\{x\,y\}}$ are $M_{\{x\}}$ are proper supersets of $\text{Event}(M)$. We first show that (6) holds for this interpretation under $\mathcal{M}_p$. The proof is by summing (3) over all $D \in M_{\{c,a\}}$.

Proposition 4. *For any voting matrix M and any pair of alternatives $\{c,a\}$, we have the following calculation under $\mathcal{M}_p$.*

$$\Pr(M_{\{c,a\}}) = \binom{n}{M(c\succ a)}(p^{M(c\succ a)}(1-p)^{M(a\succ c)} + p^{M(a\succ c)}(1-p)^{M(c\succ a)})/2$$

Proof. For any q, we let $F(p,q) = (p^q(1-p)^{n-q} + p^{n-q}(1-p)^q$. Note that $F(p,q) = F(p,n-q)$.

$$\begin{aligned}
&\Pr(M_{\{c,a\}})\Pr(D) = \sum_{W\in M_{\{c,a\}}} \Pr(W)\times\Pr(D|W)\\
=&\frac{1}{2^{m(m-1)/2}}\binom{n}{M(c\succ a)}F(p,M(c\succ a))\prod_{\{x,y\}\neq\{c,a\}}\left(\sum_{q=0}^{n}\binom{n}{q}F(p,q)\right)\\
=&\frac{1}{2^{m(m-1)/2}}\binom{n}{M(c\succ a)}F(p,M(c\succ a))2^{m(m-1)/2-1}\\
=&\binom{n}{M(c\succ a)}F(p,M(c\succ a))/2
\end{aligned}$$

□

Proposition 5. *For any voting matrix M and any pair of alternatives $\{c,a\}$, under $\mathcal{M}_p$ we have the following equation.*

$$\Pr(c\succ a|M_{\{c,a\}}) = \Pr(c\succ a|M)$$

Proof. The proof is similar to the proof of Proposition 3. From (4) we observe that $\Pr(c \succ a|D)$ only depends on the number of times $c \succ a$ and the number of times $a \succ c$ in D. Therefore, for all $D \in M_{\{c,a\}}$, $\Pr(c \succ a|D)$ is the same. Recall that $\text{Event}(M)$ is a proper subset of $M_{\{c,a\}}$. Hence we have $\Pr(c \succ a|M_{\{c,a\}}) = \Pr(c\succ a|M) = \Pr(c\succ a|D)$. □

For $M_{\{c\}}$ a similar proposition can be proved.

Proposition 6. *For any voting matrix M any alternative $\{c\}$, under $\mathcal{M}_p$ we have the following equation.*

$$\Pr(c|M_{\{c\}}) = \Pr(c|M)$$

Proof. The proof is similar to the proof of Proposition 3. From (5) we observe that $\Pr(c \succ a|D)$ only depends on the number of times $c \succ y$ and the number of times $a \succ y$ in D for all $y \neq c$. Therefore, for all $D \in M_{\{c\}}$, $\Pr(c|D)$ is the same. Recall that Event(M) is a proper subset of $M_{\{c\}}$. Hence we have $\Pr(c|M_{\{c\}}) = \Pr(c|M) = \Pr(c|D)$. □

With the above two propositions we can propose the following conjecture on the meanings of "vote" in Young's calculations.

Conjecture 1. *The following meanings of "vote" holds under $\mathcal{M}_p$ can explain Young's calculations.*

- *In $\Pr(vote)$, it means $M_{\{x,y\}}$, that is, the event of observing the results of comparisons between $\{x, y\}$.*
- *In $\Pr(x \succ y|vote)$, it still means $M_{\{x,y\}}$, though changing the meaning to M does not affect the calculation (Proposition 5).*
- *In $\Pr(x|vote)$, it means $M_{\{x\}}$, though changing the meaning to M does not affect the calculation (Proposition 6).*

The conjecture states that only in $\Pr(vote)$ it is required that "vote" means $M_{\{x,y\}}$. Due to Proposition 5 and Proposition 6, in other places it does not matter whether "vote" means $M_{\{x,y\}}$ (sometimes $M_{\{x\}}$) or M, since the conditional probabilities under these two meanings are the same. Young's calculations are so concise that it is hard to figure out what was the answer. Even though Conjecture 1 with Proposition 5 and Proposition 6 are compatible with Young's calculations, the exact meaning of "vote" in Young's calculations is still an unsolved mystery.

5.3 Is there another natural model that can explain Calculations 1–3?

Suppose Young has just one statistical model in mind throughout the paper. As we briefly discussed in Section 1.3, for the parameter space, there are two possibilities.

1. $\mathcal{O}(\mathcal{C})$ as in $\mathcal{M}_p$.
2. $\mathcal{L}(\mathcal{C})$ as in Mallows. That is, a parameter (ground truth) must be a linear orders over $\mathcal{C}$.

For the sample space, there are also two possibilities.

1. $(\mathcal{O}(\mathcal{C}))^n$ as in $\mathcal{M}_p$.
2. $(\mathcal{L}(\mathcal{C}))^n$ as in Mallows. That is, each agent must report a linear orders over $\mathcal{C}$.

Therefore we have four models corresponding to the four combinations of parameter space and sample space. All of them seem to be compatible with Young's descriptions in Section 1.3. As we have already proved in the last subsection, $\mathcal{M}_p$ (which corresponds to case 1 for both sample space and parameter space) can explain all Young's calculations. We will show that $\mathcal{M}_p$ is the only one among the four that can explain all Young's calculations.

For models with case 2 of the parameter space, that is, the parameter space is composed of all rankings, for each parameter there is always an alternative ranked in the top. Therefore, as we mentioned after Calculation 3, $\Pr(a|M_1) + \Pr(b|M_1) + \Pr(c|M_1)$ should always be 1 for all p. However, this is not true as shown in Figure 1. Therefore we know that in Young's mind the parameter space cannot be $\mathcal{L}(\mathcal{C})$.

For models with case 2 of the sample space, we will show that this is not compatible with Calculation 1. Let $S' \subseteq \mathcal{L}(\mathcal{C})^n$ denote the set of preference profiles whose voting matrix is the same as M_1. It is easy to check that if the sample space is $\mathcal{L}(\mathcal{C})^n$, then $L(a \succ b \succ c) = \frac{|S'|}{Z^{13}}[p^{25}(1-p)^{14}]$, where Z is the normalization factor so that $\sum_{V \in \mathcal{L}(\mathcal{C})} p^{m(m-1)/2-\text{Kendall}(V, a \succ b \succ c)}(1-p)^{\text{Kendall}(V, a \succ b \succ c)} = Z$. Notice that Z is a function of p but in Calculation 1 the likelihood of $a \succ b \succ c$ is a constant multiplied by $[p^{25}(1-p)^{14}]$. Hence the sample space cannot be $\mathcal{L}(\mathcal{C})^n$.

Also in the model where the parameter space is $\mathcal{O}(\mathcal{C})$ and the sample space is $\mathcal{L}(\mathcal{C})^n$, the independence between $c \succ a$ and $c \succ b$ given M illustrated in Calculation 2 does not hold in general. This is shown in the following proposition.

Proposition 7. *Let $\mathcal{C} = \{a, b, c\}$, $n = 1$ and $D = \{[c \succ b \succ a]\}$. That is, the only vote is $[c \succ b \succ a]$. In the model where the parameter space is $\mathcal{O}(\mathcal{C})$ and the sample space is $\mathcal{L}(\mathcal{C})$, we have* $\Pr(c \succ a$ *and* $c \succ b|D) \neq \Pr(c \succ a|D) \times \Pr(c \succ b|D)$ *for some* $0.5 < p < 1$.

Proof. Let $\varphi = \frac{1-p}{p}$. For any $R \in \mathcal{L}(\mathcal{C})$ as the ground truth, we have $\Pr(D|R) = \dfrac{\varphi^{\text{Kendall}(D,R)}}{1+2\varphi+2\varphi^2+\varphi^3} = \dfrac{\varphi^{\text{Kendall}(D,R)}}{(1+\varphi)(1+\varphi+\varphi^2)}$.

For any For any $W \in (\mathcal{O}(\mathcal{C}) - \mathcal{L}(\mathcal{C}))$ as the ground truth, we have $\Pr(D|W) = \dfrac{\varphi^{\text{Kendall}(D,W)}}{3\varphi+3\varphi^2} = \dfrac{\varphi^{\text{Kendall}(D,W)}}{3\varphi(1+\varphi)}$.

We have

$$\begin{aligned}
&\Pr(c \succ a|D)\\
=&\Pr(b \succ c \succ a|D) + \Pr(c \succ b \succ a|D) + \Pr(c \succ a \succ b|D) + \Pr(c \to a \to b \to c|D)\\
=&\frac{1}{8\Pr(D)}(\Pr(D|b \succ c \succ a) + \Pr(D|c \succ b \succ a) + \Pr(D|c \succ a \succ b) + \Pr(D|c \to a \to b \to c))
\end{aligned}$$

Here $c \to a \to b \to c$ represents the binary relation where $c \succ a$, $a \succ b$, and $b \succ c$. We recall that $D = \{[c \succ b \succ a]\}$. The Kendall-tau distance between D and the ground truth in the above calculation is shown in the following table.

$b \succ c \succ a$	$c \succ b \succ a$	$c \succ a \succ b$	$c \to a \to b \to c$
1	0	1	2

Table 2: Calculation of Kendall-tau.

Therefore $\Pr(c \succ a|D) = \frac{1}{8\Pr(D)}(\frac{1+2\varphi}{(1+\varphi)(1+\varphi+\varphi^2)} + \frac{\varphi^2}{3\varphi(1+\varphi)})$.
Similarly,

$$\begin{aligned}
&\Pr(c \succ b|D)\\
=&\Pr(a \succ c \succ b|D) + \Pr(c \succ a \succ b|D) + \Pr(c \succ b \succ a|D) + \Pr(c \to b \to a \to c|D)\\
=&\frac{1}{8\Pr(D)}(\Pr(D|a \succ c \succ b) + \Pr(D|c \succ a \succ b) + \Pr(D|c \succ b \succ a) + \Pr(D|c \to b \to a \to c))
\end{aligned}$$

Give that $D = \{[c \succ b \succ a]\}$, we the Kendall-tau distance between D and the ground truth in the above calculation is shown in the following table.

$a \succ c \succ b$	$c \succ a \succ b$	$c \succ b \succ a$	$c \to b \to a \to c$
2	1	0	1

Table 3: Calculation of Kendall-tau.

Therefore $\Pr(c \succ a|D) = \frac{1}{8\Pr(D)}(\frac{1+\varphi+\varphi^2}{(1+\varphi)(1+\varphi+\varphi^2)} + \frac{\varphi}{3\varphi(1+\varphi)}) = \frac{1}{8\Pr(D)}\frac{4}{3(1+\varphi)}$.
Given uniform prior over $\mathcal{O}(\mathcal{C})$, $\Pr(D) = \frac{1}{6}$. Therefore

$$\Pr(c \succ a|D) \times \Pr(c \succ b|D) = \frac{3}{4}(\frac{1+2\varphi}{(1+\varphi)(1+\varphi+\varphi^2)} + \frac{\varphi}{3(1+\varphi)}) \cdot \frac{1}{1+\varphi}$$

On the other hand,

$$\begin{aligned}&\Pr(c \succ a \text{ and } c \succ b|D)\\ =&\Pr(c \succ a \succ b|D) + \Pr(c \succ b \succ a|D)\\ =&\frac{1}{8\Pr(D)}(\Pr(D|c \succ a \succ b) + \Pr(D|c \succ b \succ a))\\ =&\frac{3}{4}(\frac{1+\varphi}{(1+\varphi)(1+\varphi+\varphi^2)})\\ =&\frac{3}{4} \times \frac{1}{1+\varphi+\varphi^2}\end{aligned}$$

Clearly this does not equal to $\Pr(c \succ a|D) \times \Pr(c \succ b|D)$ because the denominator in $\Pr(c \succ a \text{ and } c \succ b|D)$ does not contain a $(1+\varphi)$ factor. $\square$

Now it seems reasonable to assume that the model in Young's mind was $\mathcal{M}_p$. Even though the parameter space is $\mathcal{O}(\mathcal{C})$, the meaning of the following two events in the parameter space may still not be very clear: "c is ranked in the top" and the event "$c \succ x$" as in Calculation 2. For each event there are two possibilities.

For the event "c is ranked in the top":

1. It is composed of $W \in \mathcal{O}(\mathcal{C})$ where for all $x \neq c$, $c \succ x$ in W, as in Proposition 2.
2. It is composed of $R \in \mathcal{L}(\mathcal{C})$ where c is ranked in the top in R.

The second may result from Young's comment "i.e. top-ranked in the true ranking" on p. 1237.

Similarly for the event $c \succ x$ there are two possibilities:

1. It is composed of $W \in \mathcal{O}(\mathcal{C})$ where $c \succ x$, as in Proposition 2.
2. It is composed of $W \in \mathcal{L}(\mathcal{C})$ where $c \succ x$ in R.

As we briefly discussed in the remark after Proposition 3, when there are three alternatives, the two understandings for "c is ranked in the top" are equivalent. It is not hard to see that the probability for understanding 1 of "$c \succ x$" is strictly larger than the probability for understanding 2 of "$c \succ x$". It follows from Proposition 3 that understanding 2 of "$c \succ x$" is not true, otherwise $\Pr(c|M_2)$ would be strictly larger than $\Pr(c \succ a|M_2) \times \Pr(c \succ b|M_2)$.

Given that understanding 1 is true, namely, the event "$c \succ x$" is composed of all $W \in \mathcal{O}(\mathcal{C})$ where $c \succ x$, we can now infer which of 1 and 2 for the event "c is ranked in the top" is true for $m > 3$. It is not hard to see that when there are more than three alternatives, 2 is a proper subset of 1. For example, when there are four alternatives $\{c, a, b, d\}$, then 1 contains the following (none-linear) opinion $\{c \succ a, c \succ b, c \succ d, a \succ b, b \succ d, d \succ a\}$ but 2 does not. Hence by Proposition 3, 2 should not hold, otherwise we would have $\Pr(c|M) < \prod_{x \neq c} \Pr(c \succ x|M)$.

6 Main Messages

In the last subsection we showed that in Young's interpretation, the parameter space of Condorcet's model is a proper superset of $\mathcal{L}(\mathcal{C})$. This means that the MLE of Condorcet's model does not always choose a linear order. Therefore, we can falsify Viewpoint 1. This is our first main message.

Message 1. *The MLE of Condorcet's model is not the Kemeny rule. We can probably say that the MLE* restricted to linear orders *in Condorcet's model is the Kemeny rule.*

Since Kemeny is the MLE of Mallows model. Therefore, we can also falsify Viewpoint 2. This is our second main message.

Message 2. *Condorcet's model is different from Mallows model.*

Our third message concerns the following events in the parameter space: "c ranked in the top".

Message 3. *In the hypothetical Condorcet's model $\mathcal{M}_p$, "c ranked in the top" is the event in the parameter space that is composed of $W \in \mathcal{O}(\mathcal{L})$ where for any other alternative y, $c \succ y$ in W.*

References

[1] Hossein Azari Soufiani, David Parkes, and Lirong Xia. *Statistical Decision Theory Approaches to Social Choice*. Working paper, `cs.rpi.edu//~xial/Files/SDT.pdf`, 2014.

[2] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Random utility theory for social choice. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 126–134, Lake Tahoe, NV, USA, 2012.

[3] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 2nd edition, 1985.

[4] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, Philadelphia, PA, 2013.

[5] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale, 1785.

[6] Vincent Conitzer and Tuomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, Edinburgh, UK, 2005.

[7] Peter C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

[8] Christin List and Robert E. Goodin. Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy*, 9(3):277—306, 2001.

[9] Colin L. Mallows. Non-null ranking model. *Biometrika*, 44(1/2):114–130, 1957.

[10] Andrew Mao, Ariel D. Procaccia, and Yiling Chen. Better human computation through principled voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, Bellevue, WA, USA, 2013.

[11] Eric Marchand and William E. Strawderman. Estimation in restricted parameter spaces: A review. In *A Festschrift for Herman Rubin*, volume 45 of *Lecture Notes–Monograph Series*, pages 21–44. 2004.

[12] Marcus Pivato. Voting rules as statistical estimators. *Social Choice and Welfare*, 40(2):581–630, 2013.

[13] Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.

[14] Lirong Xia and Vincent Conitzer. A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pages 446–451, Barcelona, Catalonia, Spain, 2011.

[15] H. Peyton Young. Condorcet's theory of voting. *American Political Science Review*, 82:1231–1244, 1988.