Preference Functions That Score Rankings and Maximum Likelihood Estimation

Vincent Conitzer Matthew Rognlie Lirong Xia

Abstract

A preference function (PF) takes a set of votes (linear orders over a set of alternatives) as input, and produces one or more rankings (also linear orders over the alternatives) as output. Such functions have many applications, for example, aggregating the preferences of multiple agents, or merging rankings (of, say, webpages) into a single ranking. The key issue is choosing a PF to use. One natural and previously studied approach is to assume that there is an unobserved "correct" ranking, and the votes are noisy estimates of this. Then, we can use the PF that always chooses the maximum likelihood estimate (MLE) of the correct ranking. In this paper, we define simple ranking scoring functions (SRSFs) and show that the class of neutral SRSFs is exactly the class of neutral PFs that are MLEs for some noise model. We also define extended ranking scoring functions (ERSFs) and show a condition under which these coincide with SRSFs. We study key properties such as consistency and continuity, and consider some example PFs. In particular, we study Single Transferable Vote (STV), a commonly used PF, showing that it is an ERSF but not an SRSF, thereby clarifying the extent to which it is an MLE function. This also gives a new perspective on how ties should be broken under STV. We leave some open questions.

1 Introduction

In a typical social choice setting, there is some set of alternatives, and multiple rankings of these alternatives are provided. These input rankings are called the *votes*. Based on these votes, the goal is either to choose one alternative, or to create an aggregate ranking of all the alternatives. In this paper, we will be interested in the latter goal; if it is desired to choose one alternative, then we can simply choose the top-ranked alternative in the aggregate ranking. Formally, a *preference function (PF)*[1] takes a set of votes (linear orders over the alternatives) as input, and produces one or more aggregate rankings (also linear orders over the alternatives) as output. (The only reason for allowing multiple aggregate rankings is to account for the possibility of ties.)

The key issue is to choose a rule for determining the aggregate ranking, that is, a preference function. So, we may ask the following (vague) question: *What is the optimal preference function?* This has been (and will likely continue to be) a topic of debate for centuries among social choice theorists. Many different PFs have been proposed, each with its own desirable properties; some of them have elegant axiomatizations. Presumably, which PF is optimal depends on the setting at hand. For example, in some settings, the voters are agents that each have their own idiosyncratic preferences over the alternatives, and the only purpose of voting is to reach a compromise. In such a setting, no alternative can be said to be better than another alternative in any *absolute* sense: an alternative's quality is defined relative to the votes. In such a setting, it makes sense to pay close attention to issues such as the manipulability of the PF.

In other settings, however, there is more of an absolute sense in which some alternatives are better than others. For example, when we wish to aggregate rankings of webpages, provided by multiple search engines in response to the same query, it is reasonable to believe that some of these pages are in fact more relevant than others. The reason that not all of the search engines agree on

[1] We use "preference function" rather than "social welfare function" because the resulting set of strict rankings need not correspond to a weak ranking (where a set of strict rankings "corresponds" to a weak ranking if it consists of all the strict rankings that can be obtained by breaking the ties in the weak ranking). The term "preference function" has previously been used in this context [14].

the ranking is that the search engines are unable to directly perceive this absolute relevance of the pages. Here, it makes sense to think of each vote as a *noisy estimate* of the correct, absolute ranking. Our goal is to find an aggregate ranking that is as close as possible to the correct ranking, based on these noisy estimates. This is the type of setting that we will study in this paper.

In a 2005 paper, Conitzer and Sandholm considered the following way of making this precise [3]. There is a correct ranking r of the alternatives; given r, for every ranking v, there is a conditional probability $P(v|r)$ that a given voter will cast vote v. (In this paper, we do not consider the possibility that different voters' votes are drawn according to different conditional distributions.) Votes are conditionally independent given r. Put another way, the noise that each voter experiences is i.i.d. The Bayesian network in Figure 1 illustrates this setup.

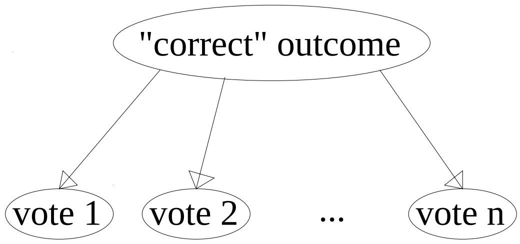

Figure 1: A Bayesian network representation.

The votes are the observed variables, and the noise that a voter experiences is represented by the conditional probability table of that vote. Under this setup, a natural goal is to find the maximum likelihood estimate (MLE) of the correct ranking. (If r is drawn uniformly at random, this maximum likelihood estimate also maximizes the posterior probability.) The function that takes the votes as input and produces the MLE ranking(s) as output is a preference function; in a sense, it is the optimal one for the particular noise model at hand.

As pointed out by Conitzer and Sandholm, they were not the first to consider this type of setup. In fact, the basic idea dates back over two centuries to Condorcet [4], who studied one particular noise model. He solved for the MLE PF for two and three alternatives under this model; the general solution was given two centuries later by Young [13], who showed that the MLE PF for Condorcet's model coincides with a function proposed by Kemeny [7]. This has frequently been used as an argument in favor of using Kemeny's PF; however, different noise models will in general result in different MLE PFs. Several generalizations of this basic noise model have been studied [6, 5, 8, 9]. Conitzer and Sandholm considered the opposite direction: they studied a number of specific well-known PFs and they showed that for some of them, there exists a noise model such that this PF becomes the MLE, whereas for others, no such noise model can be constructed. This shows that the former PFs are in a sense more natural than the latter. Also, when a noise model can be constructed, it gives insight into the PF; moreover, if the noise model is unreasonable in a certain way, it can be modified, resulting in an improved PF.

In this paper, we continue this line of work. We provide an exact characterization of the class of (neutral) PFs for which a noise model can be constructed: we show that this class is equal to the class of (neutral) *simple ranking scoring functions (SRSFs)*, which, for every vote, assign a score to every potential aggregate ranking, and the ranking(s) with the highest total score win(s). We show that several common PFs are SRSFs (these proofs resemble the corresponding proofs by Conitzer and Sandholm that these PFs are MLEs, but the proofs are significantly simpler in the language of SRSFs). We also consider *extended ranking scoring functions (ERSFs)*, which coincide with SRSFs except they can break ties according to another SRSF, and remaining ties according to another SRSF, *etc*. We show that if there is a bound on the number of votes, then the two classes (SRSFs and ERSFs) coincide. We study some basic properties of SRSFs and ERSFs, some of them closely related to Conitzer and Sandholm's proof techniques. Finally, we study one PF, Single Transferable

Vote (STV), also known as Instant Runoff Voting, in detail. STV is used in many elections around the world; additionally, it illustrates a number of key points about our results. A noise model for STV was given by Conitzer and Sandholm. However, this noise model involves probabilities that are infinitesimally smaller than other probabilities. We show that such infinitesimally small probabilities are in a sense necessary, by showing that STV is in fact not an SRSF (when there is no bound on the number of votes). Still, we do show that STV is an ERSF (in a way that resembles the noise model with infinitesimally small probabilities). Hence, STV is in fact an MLE PF if there is an upper bound on the number of votes. Along the way, some interesting questions arise about how ties should be broken under STV. We propose two ways of breaking ties that we believe are perhaps more sensible than the common way, although at least one of the ways leads to computational difficulties. We also leave some open questions.

2 Definitions

In the below, we let A be the set of alternatives, $|A| = m$, and $L(A)$ the set of linear orders over (that is, strict rankings of) these alternatives. A *preference function (PF)* is a function $f : \bigcup_{i=0,1,2,\ldots} L(A)^i \rightarrow 2^{L(A)} - \emptyset$. That is, f takes as input a vector (of any length) V of linear orders (votes) over the alternatives, and as output produces one or more linear orders over (aggregate rankings of) the alternatives. (On many inputs, only a single ranking is produced, but it is possible that there are ties.) Input vectors are also called *profiles*. We restrict our attention to PFs that are *anonymous*, that is, they treat all votes equally; hence, a profile can be thought of as a multiset of votes. Below are the PFs that we will study in this paper.

- *Positional scoring functions.* A positional scoring function is defined by a vector $(s_1, \ldots, s_m) \in \mathbb{R}^m$, with $s_1 \geq s_2 \geq \ldots \geq s_m$. An alternative receives s_i points every time it is ranked ith. Alternatives are ranked by how many points they receive; if some alternatives end up tied, then they can be ranked in any order (and all the complete rankings that can result from this will be produced by the PF). Examples include *plurality* ($s_1 = 1, s_2 = s_3 = \ldots = s_m = 0$), *veto* or *anti-plurality* ($s_1 = s_2 = \ldots = s_{m-1} = 1, s_m = 0$), and *Borda* ($s_1 = m-1, s_2 = m-2, \ldots, s_m = 0$).

- *Kemeny.* Given a vote v, a possible ranking r, and two alternatives a, b, let $\delta(v, r, a, b) = 1$ if $a \succ_v b$ and $a \succ_r b$, and $\delta(v, r, a, b) = 0$ otherwise. Then, $f(V) = \arg\max_{r \in L(A)} \sum_{a,b \in A} \sum_{v \in V} \delta(v, r, a, b)$. That is, we choose the ranking(s) that maximize(s) the total number of times that the ranking agrees with a vote on a pair of alternatives.

- *Single Transferable Vote (STV).* The alternative with the lowest plurality score (that is, the one that is ranked first by the fewest votes) is ranked last, and is removed from all the votes (so that the plurality scores change). The remainder of the ranking is determined recursively. (We will have more to say about how ties are broken later.)

A PF is *neutral* if treats all alternatives equally. To be precise, a PF is neutral if for any votes V and any permutation π on the alternatives, $f(\pi(V)) = \pi(f(V))$. Here, a permutation is applied to a vector or set of rankings of the alternatives by applying it to each individual alternative in those rankings. Naturally, neutrality is a common requirement. Another common requirement for an anonymous PF is *homogeneity*: if we multiply the profile by some natural number $n > 0$ (that is, replace each vote by n duplicates of it), then the outcome should not change. All of the above PFs are anonymous, neutral, and homogenous.

We now define noise models and MLE PFs formally.

Definition 1 *A* noise model ν *specifies a probability* $P_\nu(v|r)$ *for every* $v, r \in L(A)$.

Definition 2 *A noise model ν is* neutral *if for any v, r, and permutation π on A, we have $P_\nu(v|r) = P_\nu(\pi(v)|\pi(r))$.*

Definition 3 *A PF f is a* maximum likelihood estimator (MLE) *if there exists a noise model ν so that $f(V) = \arg\max_{r \in L(A)} \prod_{v \in V} P_\nu(v|r)$.*

We now define simple ranking scoring functions. Effectively, every vote gives a number of points to every possible aggregate ranking, and the ranking(s) with the most points win(s).

Definition 4 *A PF f is a* simple ranking scoring function (SRSF) *if there exists a function $s : L(A) \times L(A) \to \mathbb{R}$ such that for all V, $f(V) = \arg\max_{r \in L(A)} \sum_{v \in V} s(v, r)$.*

Definition 5 *A function $s : L(A) \times L(A) \to \mathbb{R}$ is* neutral *if for any v, r, and permutation π on A, $s(v, r) = s(\pi(v), \pi(r))$.*

An SRSF can be run by explicitly computing each ranking's score, but because there are $m!$ rankings this is impractical for all but the smallest numbers of alternatives. However, such explicit computation is generally not necessary. For example, we will see that positional scoring functions as well as the Kemeny function are SRSFs. Positional scoring functions are of course easy to run; running the Kemeny function is in fact NP-hard [1], but can in practice be done quite fast [2, 9].

3 Equivalence of neutral MLEs and SRSFs

We now show the equivalence of MLEs and SRSFs. We only show this for neutral PFs; in fact, it is not true for PFs that are not neutral. For example, a PF that always chooses the same ranking r^* regardless of the votes is an SRSF, simply by setting $s(v, r^*) = 1$ for all v and setting $s(v, r) = 0$ everywhere else. However, this PF is not an MLE: given a noise model ν, if we take another ranking $r \neq r^*$, we must have $\sum_{v \in L(A)} P_\nu(v|r) = 1 = \sum_{v \in L(A)} P_\nu(v|r^*)$, hence there exists some v such that $P_\nu(v|r) \geq P_\nu(v|r^*)$; it follows that r^* is not the (sole) winner if v is the only vote.

Lemma 1 *A neutral PF f is an MLE if and only if it is an MLE for a neutral noise model.*

Proof: The "if" direction is immediate. For the "only if" direction, given a noise model ν for f, construct a new noise model ν' as follows: $P_{\nu'}(v|r) = (1/m!) \sum_\pi P_\nu(\pi(v)|\pi(r))$. (Here, π ranges over permutations of A.) This is still a valid noise model because $\sum_{v \in L(A)} P_{\nu'}(v|r) = \sum_{v \in L(A)} (1/m!) \sum_\pi P_\nu(\pi(v)|\pi(r)) = (1/m!) \sum_\pi \sum_{v \in L(A)} P_\nu(\pi(v)|\pi(r)) = 1$. ν' is also neutral because $P_{\nu'}(\pi(v)|\pi(r)) = (1/m!) \sum_{\pi'} P_\nu(\pi'(\pi(v))|\pi'(\pi(r))) = (1/m!) \sum_{\pi''} P_\nu(\pi''(v)|\pi''(r)) = P_{\nu'}(v|r)$. Also, if $r^* \in \arg\max_{r \in L(A)} \prod_{v \in V} P_\nu(v|r)$, then by the neutrality of f, for any π, $\pi(r^*) \in \arg\max_{r \in L(A)} \prod_{v \in V} P_\nu(\pi(v)|r)$. Hence, $r^* \in \arg\max_{r \in L(A)} (1/m!) \sum_\pi \prod_{v \in V} P_\nu(\pi(v)|\pi(r)) = \arg\max_{r \in L(A)} \prod_{v \in V} (1/m!) \sum_\pi P_\nu(\pi(v)|\pi(r)) = \arg\max_{r \in L(A)} \prod_{v \in V} P_{\nu'}(v|r)$. Conversely, it can similarly be shown that if $r^* \notin \arg\max_{r \in L(A)} \prod_{v \in V} P_\nu(v|r)$, then $r^* \notin \arg\max_{r \in L(A)} \prod_{v \in V} P_{\nu'}(v|r)$. Hence, ν' is a valid noise model for f. ∎

Lemma 2 *A neutral PF f is an SRSF if and only if it is an SRSF for a neutral function s'.*

Proof: The "if" direction is immediate. For the "only if" direction, given a function s, construct a new function s' as follows: $s'(v, r) = \sum_\pi s(\pi(v), \pi(r))$. s' is neutral because $s'(\pi(v), \pi(r)) = \sum_{\pi'} s(\pi'(\pi(v)), \pi'(\pi(r))) = \sum_{\pi''} s(\pi''(v), \pi''(r)) = s'(v, r)$. Also, if $r^* \in \arg\max_{r \in L(A)} \sum_{v \in V} s(v, r)$, then by the neutrality of f, for any π, $\pi(r^*) \in \arg\max_{r \in L(A)} \sum_{v \in V} s(\pi(v), r)$. Hence, $r^* \in \arg\max_{r \in L(A)} \sum_\pi \sum_{v \in V} s(\pi(v), \pi(r)) =$

$\arg\max_{r \in L(A)} \sum_{v \in V} \sum_{\pi} s(\pi(v), \pi(r)) = \arg\max_{r \in L(A)} \sum_{v \in V} s'(v, r)$. Conversely, it can similarly be shown that if $r^* \notin \arg\max_{r \in L(A)} \sum_{v \in V} s(v, r)$, then $r^* \notin \arg\max_{r \in L(A)} \sum_{v \in V} s'(v, r)$. Hence, s' is a valid function for f. ■

We can now prove the characterization result:

Theorem 1 *A neutral PF is an MLE if and only if it is an SRSF.*

Proof: If f is an MLE, then for some neutral ν, $f(V) = \arg\max_{r \in L(A)} \prod_{v \in V} P_\nu(v|r) = \arg\max_{r \in L(A)} \log(\prod_{v \in V} P_\nu(v|r)) = \arg\max_{r \in L(A)} \sum_{v \in V} \log(P_\nu(v|r))$. Hence it is the SRSF where $s(v, r) = \log(P_\nu(v|r))$ (here, s is neutral).

Conversely, if f is an SRSF, then for some neutral s, $f(V) = \arg\max_{r \in L(A)} \sum_{v \in V} s(v, r) = \arg\max_{r \in L(A)} 2^{\sum_{v \in V} s(v,r)} = \arg\max_{r \in L(A)} \prod_{v \in V} 2^{s(v,r)}$. Because s is neutral, we have that $\sum_{v \in L(A)} 2^{s(v,r)}$ is the same for all r. (This is because for any r_1, r_2, there exists a permutation π on A such that $\pi(r_1) = r_2$, so that we have $\sum_{v \in L(A)} 2^{s(v,r_1)} = \sum_{v \in L(A)} 2^{s(\pi(v),r_2)}$ by neutrality, which by changing the order of the summands is equal to $\sum_{v \in L(A)} 2^{s(v,r_2)}$.) It follows that $f(V) = \arg\max_{r \in L(A)} \prod_{v \in V} (2^{s(v,r)}) / (\sum_{v' \in L(A)} 2^{s(v',r)})$. Hence f is the maximum likelihood estimator for the noise model ν defined by $P_\nu(v|r) = (2^{s(v,r)}) / (\sum_{v' \in L(A)} 2^{s(v',r)})$. ■

4 Examples of SRSFs

We now show that some common PFs are SRSFs. These proofs resemble the corresponding proofs by Conitzer and Sandholm that these functions are MLEs, but they are simpler. These propositions also follow from the work of Zwicker [15].

Proposition 1 *Every positional scoring function is an SRSF.*

Proof: Given a positional scoring function, let $t : L(A) \times A \to \mathbb{R}$ be defined as follows: $t(v, a)$ is the number of points that a gets for vote v. Then, let $s(v, r) = \sum_{i=1}^{m} (m - i) t(v, r(i))$, where $r(i)$ is the alternative ranked ith in r. Let us consider the SRSF defined by this function s; it selects $\arg\max_{r \in L(A)} \sum_{v \in V} s(v, r) = \arg\max_{r \in L(A)} \sum_{v \in V} \sum_{i=1}^{m} (m - i) t(v, r(i)) = \arg\max_{r \in L(A)} \sum_{i=1}^{m} (m - i) \sum_{v \in V} t(v, r(i))$. Here, $\sum_{v \in V} t(v, r(i))$ is the total score that alternative $r(i)$ receives under the positional scoring function. Because $m - i$ is decreasing in i, to maximize $\sum_{i=1}^{m} (m - i) \sum_{v \in V} t(v, r(i))$, we should rank the alternative with the highest total score first, the one with the next-highest total score second, *etc.* If some of the alternatives are tied, they can be ranked in any order. ■

Not only positional scoring functions are SRSFs, however.

Proposition 2 *The Kemeny PF is an SRSF.*

Proof: This is almost immediate: we defined the Kemeny PF by $f(V) = \arg\max_{r \in L(A)} \sum_{a,b \in A} \sum_{v \in V} \delta(v, r, a, b)$, so we simply let $s(v, r) = \sum_{a,b \in A} \delta(v, r, a, b)$. ■

5 Extended ranking scoring functions

An *extended ranking scoring function (ERSF)* starts by running an SRSF, then (potentially) breaks ties according to another SRSF, and (potentially) any remaining ties according to yet another SRSF, *etc.* Formally:

Definition 6 *An* ERSF f of depth k *consists of an ERSF* f' *of depth* $k-1$ *and a function* s_d : $L(A) \times L(A) \rightarrow \mathbb{R}$. *It chooses* $f(V) = \arg\max_{r \in f'(V)} \sum_{v \in V} s_d(v, r)$. *An ERSF of depth* 0 *returns the set of all rankings* $L(A)$.

So, an ERSF of (finite) depth d is defined by a sequence $f_1, \ldots, f_d$ of SRSFs. We can think of the scores at each depth as being infinitesimally smaller than the ones at the previous depths. We can multiply the scores at depth l by ϵ^l for some small ϵ and then add all the scores together to obtain an SRSF; however, this SRSF will in general be different from the ERSF. Nevertheless, if ϵ is small relative to the number of votes, then the two will coincide. This is the intuition behind the following result:

Proposition 3 *For any ERSF, for any natural number* N*, there exists an SRSF that agrees with the ERSF as long as there are at most* N *votes.*

Proof: Let the sequence of SRSFs $f_1, \ldots, f_d$, defined by scoring functions $s_1, \ldots, s_d$, define the ERSF f; we prove the claim by induction. The claim is trivial for $d = 1$. Let us assume that we have proven the result for $d = k - 1$; we will show it for $d = k$. Let f' be the ERSF corresponding to the first $d - 1$ SRSFs, and, by the induction assumption, let s define the SRSF that agrees with f' when there are at most N votes. There are only finitely many profiles V of size at most N; hence, there must be some ϵ such that $\sum_{v \in V} s(v, r) < \sum_{v \in V} s(v, r')$ and $|V| \leq N$ implies that $\sum_{v \in V} s(v, r) + \epsilon < \sum_{v \in V} s(v, r')$. Now let us consider s_d; there must exist some $H \in \mathbb{R}$ such that $|V| \leq N$ implies $\sum_{v \in V} s_d(v, r) < H$. Then, let s' be defined by $s'(v, r) = s(v, r) + (\epsilon/H) s_d(v, r)$. On profiles of size at most N, the second term will contribute at most ϵ to the total score of any r, so if r receives a strictly lower total score than r' under s, it will also receive a strictly lower score under s'. Hence, the only effect of the second term is to break ties according to s_d; so the SRSF defined by s' coincides with the original ERSF f when there are at most N votes. ■

Thus, for all practical purposes, we can simulate an ERSF with an SRSF. (Of course, every SRSF is also an ERSF.)

6 Properties of SRSFs and ERSFs

In this section, we study some important properties of SRSFs and ERSFs. Specifically, we study *consistency* and *continuity*. There are several related works that study similar properties and derive related results, but there are significant differences in the setup. Smith [11] and Young [12] study these properties in *social choice rules*, which select one or more alternatives as the winner(s); we will discuss their results in more detail in Section 8. However, consistency in the context of preference functions (studied previously by Young and Levenglick [14]) is significantly different from consistency in the context of social choice rules. Other related work includes Myerson [10], who extends the Smith and Young result to settings where voters do not necessarily submit a ranking of the alternatives, and Zwicker [15], who studies a general notion of scoring rules and shows these rules are equivalent to *mean proximity rules*, which compute the mean location of the votes according to some embedding in space, and then choose the closest outcome(s).

An anonymous PF f is *consistent* if for any pair of profiles V_1 and V_2, if $f(V_1) \cap f(V_2) \neq \emptyset$, then $f(V_1 + V_2) = f(V_1) \cap f(V_2)$ (where addition is defined in the natural way). That is, if the rankings that f produces given V_1 overlap with those that f produces given V_2, then when V_1 and V_2 are taken together, f must produce the rankings that were produced in both cases, and no others.

Proposition 4 *Any ERSF is consistent.*

Proof: Let f be an ERSF of depth k, defined by a sequence of SRSFs $f_1, \ldots, f_k$ with score functions $s_1, \ldots, s_k$. For any $i \leq k$, let F_i be the ERSF of depth i defined by the sequence $f_1, \ldots, f_i$. Let V_1,

V_2 be profiles such that $f(V_1) \cap f(V_2) \neq \emptyset$; this also implies that $F_i(V_1) \cap F_i(V_2) \neq \emptyset$ for all $i \leq k$. We use induction on i to prove that for any $i \leq k$, $F_i(V_1 + V_2) = F_i(V_1) \cap F_i(V_2)$. When $i = 1$, $F_1(V_1) = f_1(V_1)$ is the set of rankings r that maximize $s_1(V_1, l)$; $F_1(V_2) = f_1(V_2)$ is the set of rankings r that maximize $s_1(V_2, l)$. Therefore, $F_1(V_1) \cap F_1(V_2)$ (which we know is nonempty) is the set of rankings r that maximize $s_1(V_1 + V_2, r)$. Now, suppose that for some $i \leq k$, $F_i(V_1 + V_2) = F_i(V_1) \cap F_i(V_2)$. $F_{i+1}(V_1)$ ($F_{i+1}(V_2)$) is the set of rankings $r \in F_i(V_1)$ ($r \in F_i(V_2)$) that maximize $s_{i+1}(V_1, r)$ ($s_{i+1}(V_2, r)$). Hence, $F_{i+1}(V_1) \cap F_{i+1}(V_2)$ (which we know is nonempty) is the set of rankings $r \in F_i(V_1) \cap F_i(V_2)$ that maximize $s_{i+1}(V_1, r) + s_{i+1}(V_2, r) = s_{i+1}(V_1 + V_2, r)$. By the induction assumption, we have that $F_i(V_1) \cap F_i(V_2) = F_i(V_1 + V_2)$, and we know that the set of rankings $r \in F_i(V_1 + V_2)$ that maximize $s_{i+1}(V_1 + V_2, r)$ is equal to $F_{i+1}(V_1 + V_2)$. It follows that $F_{i+1}(V_1) \cap F_{i+1}(V_2) = F_{i+1}(V_1 + V_2)$, completing the induction step. When $i = k$, $F_k = f$, which completes the proof. ■

The proofs by Conitzer and Sandholm [3] that several PFs are not MLEs effectively come down to showing examples where these PFs are not consistent. By the above result, this implies that they are not ERSFs (and hence not SRSFs, and hence not MLEs). Formally (we will not define these PFs in this paper):

Proposition 5 *The Bucklin, Copeland, maximin, and ranked pairs PFs are not ERSFs.*

Proof: None of these PFs are consistent: counterexamples can be found in the proofs of Conitzer and Sandholm [3]. ■

Let $L(A) = \{l_1, \ldots, l_{m!}\}$. For any anonymous PF f, any profile V can be rewritten as a linear combination of the linear orders in $L(A)$. Let $V = \sum_{i=1}^{m!} t_i l_i$, where for any $i \leq m!$, t_i is a non-negative integer. If f is also homogenous, then the domain of f can be extended to the set of all *fractional profiles* $V = \sum_{i=1}^{m!} t_i l_i$ where each t_i is a nonnegative rational number, as follows. We choose $N_V > 0, N_V \in \mathbb{N}$ such that for every $i \leq m!$, $t_i N_V$ is a integer. Then, we let $f(V) = f(N_V V)$. (This is well-defined because of the homogeneity.)

A fractional profile V can be viewed as a point in the $m!$-dimensional space $(\mathbb{Q}^{\geq 0})^{m!}$ where the coefficient t_i is the component of the ith dimension. Thus, in a slight abuse of notation, we can apply f to vectors of $m!$ nonnegative rational numbers, under the interpretation that $f(t_1, \ldots, t_{m!}) = f(\sum_{i=1}^{m!} t_i l_i)$. The extension of f to $(\mathbb{Q}^{\geq 0})^{m!}$ allows us to define continuity. An anonymous PF f is *continuous* if for any sequence of points $p_1, p_2, \ldots \in (\mathbb{Q}^{\geq 0})^{m!}$ with 1. $\lim_{i\to\infty} p_i = p$, and 2. for all $i \in \mathbb{N}$, $r \in f(p_i)$, we have $r \in f(p)$. That is, if f produces some ranking r on every point along a sequence that converges to a limit point, then f should also produce r at the limit point.[2]

Proposition 6 *Any SRSF is continuous.*

Proof: For any sequence of points $p_1, p_2, \ldots \in (\mathbb{Q}^{\geq 0})^{m!}$ with $\lim_{i\to\infty} p_i = p$, we have that for all $r \in L(A)$, $\lim_{i\to\infty} s(p_i, r) = s(p, r)$. If $r \in f(p_i)$ for all i, then for any $r' \in L(A)$, $s(p_i, r) \geq s(p_i, r')$, hence we have $s(p, r) = \lim_{i\to\infty} s(p_i, r) \geq \lim_{i\to\infty} s(p_i, r') = s(p, r')$. It follows that $r \in f(p)$. ■

In contrast, ERSFs are not necessarily continuous, as shown by the following example. Let f_1 be the SRSF defined by the score function s_1, which is defined by $s_1(v, r) = 1$ if $v = r$ and $s_1(v, r) = 0$ if $v \neq r$. Let f_2 be the Borda function. Let f be the ERSF defined by the sequence f_1, f_2. Let $m = 3$ with alternatives A, B, and C, and let $p = \{A \succ B \succ C, B \succ C \succ A, C \succ B \succ A\}$. We have $f(p) = \{B \succ C \succ A\}$, but for any $\epsilon > 0$, $f(p + \epsilon(A \succ B \succ C)) = f_1(p + \epsilon(A \succ B \succ$

[2]Our definition of continuity is equivalent to the correspondence being *upper hemicontinuous*, or *closed* (the two are equivalent in this context).

$C)) = \{A \succ B \succ C\}$. Therefore, if we let $p_i = p + \frac{1}{i}(A \succ B \succ C)$, it follows that $\lim_{i\to\infty} p_i = p$ and for any i, $A \succ B \succ C \in f(p_i)$, but $A \succ B \succ C \notin f(p)$.

As we have noted before, there is generally a possibility of ties for PFs, and sometimes a PF is not defined for these cases (for example, we have not defined how they should be broken for STV). We can use the continuity property to gain some insight into how ties should be broken. For any $S \subseteq (\mathbb{Q}^{\geq 0})^{m!}$, let $C(S)$ be the *closure* of S, that is, $C(S)$ is the smallest set such that for any infinite sequence $p_1, p_2, \ldots$ in S, if $\lim_{i\to\infty} p_i = p$, then $p \in C(S)$. Let f_S be a PF that satisfies anonymity and homogeneity, defined over S. That is, $f_S : S \to 2^{L(A)} - \emptyset$. The *minimal continuous extension* of f_S is the PF $f_{C(S)} : C(S) \to 2^{L(A)} - \emptyset$ such that for any $p \in C(S)$ and any $r \in L(A)$, $r \in f_{C(S)}(p)$ if and only if there exists a sequence $p_1, p_2, \ldots$ in S such that $\lim_{i\to\infty} p_i = p$ and for any i, $r \in f_S(p_i)$. The following lemma will be useful in our study of STV.

Lemma 3 *Suppose we have two SRSFs f, f_S that have the same score function s, but f is defined over $(\mathbb{Q}^{\geq 0})^{m!}$, and f_S over a set $S \subseteq (\mathbb{Q}^{\geq 0})^{m!}$ such that $C(S) = (\mathbb{Q}^{\geq 0})^{m!}$. If for any $r \in L(A)$, there exists a profile p_r such that $f(p_r) = \{r\}$, then f is the minimal continuous extension of f_S.*

Proof: By Proposition 6, f is continuous. On the other hand, for any $p \in (\mathbb{Q}^{\geq 0})^{m!}$ with $r \in f(p)$, for any $i \in \mathbb{N}$, $f(p + \frac{1}{i}p_r) = \{r\}$. Because $C(S) = (\mathbb{Q}^{\geq 0})^{m!}$, for every $i \in \mathbb{N}$, there exists a point $p_i \in S$ sufficiently close to $p + \frac{1}{i}p_r$ such that $f(p_i) = \{r\}$, because s is continuous and at $p + \frac{1}{i}p_r$, for any $r' \in L(A)$ with $r \neq r'$, $s(p + \frac{1}{i}p_r, r) - s(p + \frac{1}{i}p_r, r') > 0$. So, $p_1, p_2, \ldots$ is a sequence of points in S with for any i, $r \in f_S(p_i)$; therefore any continuous extension must have $r \in f(p)$. ∎

7 Single Transferable Vote (STV)

In this section, we study the Single Transferable Vote (STV) PF in detail, for two reasons. First, it is a commonly used PF, so it is of interest in its own right. Second, it gives a good illustration of a number of subtle technical phenomena, and a precise understanding of these phenomena is likely to be helpful in the analysis of other PFs. We recall that under STV, in each round, the alternative that is ranked first (among the remaining alternatives) the fewest times is removed from all the votes and ranked the lowest among the remaining alternatives, that is, just above the previously removed alternative. We note that when an alternative is removed, all the votes that ranked it first *transfer* to the next remaining alternative in that vote. The number of votes ranking an alternative first is that alternative's *plurality score* in that round. One key issue is determining how ties in a round should be broken, that is, what to do if multiple alternatives have the lowest plurality score in a round. We will at first ignore this and show that STV is an ERSF. (This proof resembles the earlier Conitzer-Sandholm noise model but is much clearer in the language of scoring functions.)

Theorem 2 *When restricting attention to profiles without ties, STV is an ERSF.*

Proof: For $l \in L(A)$, let $l(i)$ be the ith-ranked alternative in l. Let $s_1(v, r) = 0$ if $r(m) = v(1)$, and $s_1(v, r) = 1$ otherwise. That is, a ranking receives a point for a vote if and only if the ranking does not rank the alternative ranked first in the vote last. Consider the alternative a with the lowest plurality score; the rankings that win under s_1 are exactly the rankings that rank a last. Now, let $s_2(v, r) = 0$ if *either* $r(m-1) = v(1)$, *or* $r(m) = v(1)$ and $r(m-1) = v(2)$; and $s_2(r, v) = 1$ otherwise. That is, a ranking receives a point for a vote *unless* the ranking ranks the first alternative in the vote second-to-last, or the ranking ranks the first alternative in the vote last and the second alternative in the vote second-to-last. If we look at rankings that survived s_2—the rankings that ranked the alternative a with the lowest plurality score last—a ranking that ranks b ($\neq a$) second-to-last will fail to receive a point for every vote that ranks b first, and for every vote that ranks a first and b second. That is, it fails to receive a point for every vote that ranks b first in the second iteration of STV. Hence, the rankings that survive s_2 are the ones that rank the alternative that receives the fewest votes in the second iteration of STV second-to-last. More generally, let $s_k(v, r) = 0$ if, letting $b = r(m-k+1)$, for every a such that $v^{-1}(a) < v^{-1}(b)$, $r^{-1}(a) > r^{-1}(b) = m-k+1$; and $s_k(v, r) = 0$ otherwise. That is, a ranking receives a point for a vote *unless* the alternative b ranked kth-to last by r is preceded in v only by alternatives ranked after b in r. Given that r has not yet been eliminated and is hence consistent with STV so far, the latter condition holds if and only if b receives v's vote in the kth iteration of STV. ∎

In fact, we can break ties in STV simply according to the scoring functions used in the proof of Theorem 2. We will call the resulting PF *ERSF-STV*. ERSF-STV is an ERSF and hence consistent. By Theorem 1 and Proposition 3, this means that ERSF-STV is an MLE when there is an upper bound on the number of votes. Does there exist a tiebreaking rule for STV such that it is an SRSF, that is, so that it is an MLE without a bound on the number of votes? We will show that the answer is negative. To do so, we consider one particular tiebreaking rule. Under this rule, when multiple alternatives are tied to be eliminated, we have a choice of which one is eliminated. A ranking is among the winning rankings if and only if there is some sequence of such choices that results in this ranking. We call the resulting PF *parallel-universes tiebreaking STV (PUT-STV)*. (Every choice can be thought of as leading to a separate parallel universe in which STV is executed.)

Lemma 4 *PUT-STV is the minimal continuous extension of STV defined on non-tied profiles.*

Proof: Let f_{STV} be the STV PF restricted to the set S of non-tied profiles, and let $f_{PUT-STV}$ be PUT-STV. We first prove that for any tied profile $p = (t_1, \ldots, t_{m!})$ and any $r \in f_{PUT-STV}(p)$, there exists a sequence of points $p_1, p_2, \ldots \in S$ such that $\lim_{i\to\infty} p_i = p$ and for any i, $r \in f_{STV}(p_i)$. From this, it will follow that any continuous extension of f_{STV} must include all of the rankings that win under $f_{PUT-STV}$ among the winners. Let N be a positive integer such that for any $i \leq m!$, $Nt_i \in \mathbb{Z}$. Let $n = |Np|$, that is, $n = \sum_{i=1}^{m!} Nt_i$. For any $a \in A$ and any $r \in f_{PUT-STV}(p)$ such that $r = a_{i_1} \succ \ldots \succ a_{i_m}$, where $(i_1, \ldots, i_m)$ is a permutation of $(1, \ldots, m)$, let $v_{a,r} = a \succ a_{i_1} \succ others$ if $a \neq a_{i_1}$, and $v_{a,r} = a_{i_1} \succ others$ if $a = a_{i_1}$. (These are complete linear orders in which the order of the others does not matter.) We let $p_r = \sum_{j=0}^{m-1} 2^j \sum_{k<m-j} v_{a_{i_k},r}$. We now show that for any $\epsilon > 0$, $p + \epsilon p_r \in S$ and $f_{STV}(p + \epsilon p_r) = \{r\}$.

For any $A' \subseteq A$ and any profile p over A, let $p|_{A'}$ be the profile over A' obtained by removing all alternatives in $A - A'$ from p. For any $j \leq m$, let $A_j = \{a_{i_1}, \ldots, a_{i_{m-j}}\}$. For any profile p^*, subset of alternatives $A' \subseteq A$, and any alternative a, let $Pl(p^*|_{A'}, a)$ be the number of times that a is ranked first in the votes in $p^*|_{A'}$. We note that because $r \in f_{PUT-STV}(p)$, for any $j \leq m-1$, any $k < m-j$, $Pl(p|_{A_j}, a_{i_k}) \geq Pl(p|_{A_j}, a_{i_{m-j}})$. We have:

$$
\begin{aligned}
&Pl((p+\epsilon p_r)|_{A_j}, a_{i_k}) = Pl(p|_{A_j}, a_{i_k}) + \epsilon Pl(p_r|_{A_j}, a_{i_k})\\
\geq &Pl(p|_{A_j}, a_{i_k}) + \epsilon 2^j > Pl(p|_{A_j}, a_{i_k}) + \epsilon \sum_{q=0}^{j-1} 2^q\\
= &Pl(p|_{A_j}, a_{i_k}) + \epsilon Pl(p_r|_{A_j}, a_{i_{m-j}})\\
\geq &Pl(p|_{A_j}, a_{i_{m-j}}) + \epsilon Pl(p_r|_{A_j}, a_{i_{m-j}})\\
= &Pl((p+\epsilon p_r)|_{A_j}, a_{i_{m-j}})
\end{aligned}
$$

Hence, for any $j \leq m-1$, in round j, $a_{i_{m-j}}$ is the alternative in A_j that is ranked first in the votes in $(p+\epsilon p_r)|_{A_j}$ (strictly) the fewest times. It follows that $f_{STV}(p + \epsilon p_r) = \{r\}$.

All that remains to show is that $f_{PUT-STV}$ is continuous, that is, for any sequence $p_1, p_2, \ldots \in S$ for which $\lim_{i\to\infty} p_i = p$ and there exists an $r \in L(A)$ such that for any i, $r \in f_{STV}(p_i)$, we have that $r \in f_{PUT-STV}(p)$. Again, let $r = a_{i_1} \succ \ldots \succ a_{i_m}$ and $A_j = \{a_{i_1}, \ldots, a_{i_{m-j}}\}$. Because for all i and $k < m-j$, $Pl(p_i|_{A_j}, a_{i_{m-j}}) < Pl(p_i|_{A_j}, a_{i_k})$, by the continuity of Pl, we have $Pl(p|_{A_j}, a_{i_{m-j}}) \leq Pl(p|_{A_j}, a_{i_k})$. Hence, under PUT-STV, it is possible to eliminate $a_{i_{m-j}}$ in the $j+1$th round, completing the proof. ■

Lemma 5 *PUT-STV is not consistent.*

Proof: Consider the following profile of votes, where A, B, and C are alternatives: $2(A \succ B \succ C) + 0(A \succ C \succ B) + 1(B \succ A \succ C) + 1(B \succ C \succ A) + 1(C \succ A \succ B) + 1(C \succ B \succ A)$. All alternatives are tied in the first round, and we split into three parallel universes. In the universe where A is eliminated, the $A \succ B \succ C$ votes transfer to B, and B is left as the only possible winner, producing the ranking $B \succ C \succ A$. In the universe where B is eliminated, the $B \succ A \succ C$ and $B \succ C \succ A$ votes transfer evenly to A and C, leaving us with another tie between A and C, and hence the rankings $A \succ C \succ B$ and $C \succ A \succ B$ are produced. Similarly, in the universe where C is eliminated first, the rankings $A \succ B \succ C$ and $B \succ A \succ C$ are produced. Ultimately, every ranking *except* $C \succ B \succ A$ is in the set of winning rankings.

By symmetry, under the profile $0(A \succ B \succ C) + 2(A \succ C \succ B) + 1(B \succ A \succ C) + 1(B \succ C \succ A) + 1(C \succ A \succ B) + 1(C \succ B \succ A)$, every ranking except $B \succ C \succ A$ wins. If we add the two profiles together, we obtain $2(A \succ B \succ C) + 2(A \succ C \succ B) + 2(B \succ A \succ C) + 2(B \succ C \succ A) + 2(C \succ A \succ B) + 2(C \succ B \succ A)$, which has all rankings in its output. But this violates consistency (which would require all rankings but $C \succ B \succ A$ and $B \succ C \succ A$ to win). ■

Corollary 1 *PUT-STV is not an ERSF (and hence not an SRSF).*

Proof: This follows immediately from Proposition 4 and Lemma 5. ■

This allows us to prove a property of STV in general:

Theorem 3 *STV is not an SRSF, even when restricting attention to non-tied profiles.*

Proof: Suppose that f_{STV} (restricted to the set S of non-tied profiles) is an SRSF defined by the score function s. By Lemma 4, $f_{PUT-STV}$ is the minimal continuous extension of f_{STV}. Also, for every $r \in L(A)$, it is easy to construct a (non-tied) profile p_r such that $f_{STV}(p_r) = \{r\}$. So, we can use Lemma 3 to conclude that $f_{PUT-STV}$ is the SRSF that results from using s on all profiles. However, by Corollary 1, we know that PUT-STV is not an SRSF, and we have the desired contradiction. ■

Incidentally, PUT-STV is also computationally intractable (in a sense); we omit the proof due to space constraint. (We do not know if an analogous result holds for ERSF-STV.)

Theorem 4 *It is NP-complete to determine whether, given a profile p and an alternative a, one of the winning rankings under PUT-STV ranks a first.*

As it turns out, neither PUT-STV nor ERSF-STV corresponds to how ties are commonly broken under STV: rather, usually, if there is a tie, all of these alternatives are simultaneously eliminated. Mathematically, this leads to bizarre discontinuities; we omit further discussion due to space constraint.

8 Axiomatic characterization of SRSFs and ERSFs

Examining *social choice rules (SCRs)*, that is, functions that output one or more alternatives as the winner(s) (rather than one or more rankings), Young found the following axiomatic characterization of positional scoring functions [12]. (A similar characterization was given by Smith [11].)

He showed that all SCRs satisfying consistency, continuity, and neutrality—SCR analogues of the properties we considered—must be positional scoring functions, and all positional scoring functions satisfy these properties. Further, dropping continuity, he found that any consistent and neutral SCR must be equivalent to what in the language of this paper would be called an "extended" positional scoring function. These results lead to two natural analogous conjectures about PFs.

Conjecture 1 *Any PF that is consistent, continuous, and neutral must be an SRSF (and therefore an MLE).*

Conjecture 2 *Any PF that is consistent and neutral must be an ERSF (and therefore an MLE when the number of votes is bounded).*

It does not appear that these conjectures can be easily proven using Smith and Young's techniques.

9 Conclusions

The maximum likelihood approach provides a natural way for choosing a PF in settings where it makes sense to think there is a "correct" ranking. In this paper, we gave a characterization of the neutral MLE PFs, showing they coincide with the neutral SRSFs. We also considered ERSFs as a slight generalization and showed that for bounded numbers of votes they coincide with SRSFs. We considered key properties such as continuity and consistency, and gave several examples of SRSFs and ERSFs. We studied STV in detail, showing that it is an ERSF but not an SRSF, and discussed the implications for breaking ties under STV. Finally, we left some open questions concerning the complexity of ERSF tiebreaking for STV and whether consistency can be used to characterize the class of SRSFs/ERSFs.

We believe that these results will greatly facilitate the use of the maximum likelihood approach in (computational) social choice. Similar results can be obtained for social choice settings other than PFs—for example, for social choice rules that only choose the winning alternative(s), or for settings in which the inputs are not linear orders (but rather, for example, labelings of the alternatives as "approved" or "not approved", or partial orders, *etc.*).

Acknowledgements

We thank Felix Brandt, Zheng Li, Ariel Procaccia, and Bill Zwicker for very helpful discussions, and the reviewers for very helpful comments. The authors are supported by an Alfred P. Sloan Research Fellowship, a Johnson Research Assistantship, and a James B. Duke Fellowship, respectively; they are also supported by the NSF under award number IIS-0812113.

References

[1] John Bartholdi, III, Craig Tovey, and Michael Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.

[2] Vincent Conitzer, Andrew Davenport, and Jayant Kalagnanam. Improved bounds for computing Kemeny rankings. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 620–626, Boston, MA, 2006.

[3] 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.

[4] Marie Jean Antoine Nicolas de Caritat (Marquis de Condorcet). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. 1785. Paris: L'Imprimerie Royale.

[5] Mohamed Drissi and Michel Truchon. Maximum likelihood approach to vote aggregation with variable probabilities. Technical Report 0211, Departement d'economique, Universite Laval, 2002.

[6] Michael Fligner and Joseph Verducci. Distance based ranking models. *Journal of the Royal Statistical Society B*, 48:359–369, 1986.

[7] John Kemeny. Mathematics without numbers. In *Daedalus*, volume 88, pages 571–591. 1959.

[8] Guy Lebanon and John Lafferty. Cranking: Combining rankings using conditional models on permutations. In *International Conference on Machine Learning (ICML)*, pages 363–370, Sydney, Australia, 2002.

[9] Marina Meila, Kapil Phadnis, Arthur Patterson, and Jeff Bilmes. Consensus ranking under the exponential model. In *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, Vancouver, BC, Canada, 2007.

[10] Roger B. Myerson. Axiomatic derivation of scoring rules without the ordering assumption. *Social Choice and Welfare*, 12(1):59–74, 1995.

[11] John H. Smith. Aggregation of preferences with variable electorate. *Econometrica*, 41(6):1027–1041, November 1973.

[12] H. Peyton Young. Social choice scoring functions. *SIAM Journal of Applied Mathematics*, 28(4):824–838, 1975.

[13] H. Peyton Young. Optimal voting rules. *Journal of Economic Perspectives*, 9(1):51–64, 1995.

[14] H. Peyton Young and Arthur Levenglick. A consistent extension of Condorcet's election principle. *SIAM Journal of Applied Mathematics*, 35(2):285–300, 1978.

[15] William S. Zwicker. Consistency without neutrality in voting rules: When is a vote an average? *Mathematical and Computer Modelling*, 2008.

Vincent Conitzer
Departments of Computer Science and Economics
Duke University
Durham, NC 27708, USA
Email: `conitzer@cs.duke.edu`

Matthew Rognlie
Duke University
Durham, NC 27708, USA
Email: `matthew.rognlie@duke.edu`

Lirong Xia
Department of Computer Science
Duke University
Durham, NC 27708, USA
Email: `lxia@cs.duke.edu`