Generalized Outcome Scoring Rules: Axiomatic Characterization and Statistical Properties

Lirong Xia
CS, Rensselaer Polytechnic Institute
Troy, NY 12180, USA
xial@cs.rpi.edu

Abstract

In this paper we pursue a "design by social choice, evaluation by statistics" paradigm towards a principled framework for new social choice mechanisms with superior social choice normative properties as well as statistical properties.

We propose a novel and general class of social choice mechanisms called *generalized outcome scoring rules (GOSRs)* that naturally extends *generalized scoring rules* [19] to arbitrary outcome spaces, including sets of alternatives with fixed or unfixed size, rankings, and sets of rankings. We show that GOSRs include a wide range of existing methods including MLEs, Chamberlin and Courant rule, Monroe rule, and resolute, irresolute, and ranking versions of all positional scoring rules, STV, ranked pairs, etc. We characterize GOSRs by two social choice normative properties: *anonymity* and *finite local consistency*. We also provide general characterizations of statistical consistency of any GOSR w.r.t. any statistical model and matching bounds on the convergence rate, which are useful tools for extending the Condorcet Jury Theorem [7] to GOSRs.

1 Introduction

Social choice theory studies how to aggregate agents' opinions or preferences to make a joint decision. Traditional social choice theory concerns how to reach a consensus among subjective preferences of the agents, and evaluates these mechanisms based on agents' *subjective* satisfaction of the joint decision. Ideally, we would like to respect agents' opinions and preferences and make a joint decision that all agents are happy with, which is often impossible due to agents' conflicting preferences. A typical example is political elections.

In many multi-agent scenarios, the goal is to aggregate agents' opinions and preferences to reveal the *ground truth*. For example, online retailers aggregate reviewers' ratings of an item to provide an estimate to the ground truth quality of the item. In such settings, instead of making a joint decision (e.g. the aggregated score) to make agents (e.g. reviewers of an item) happy, we want to make a joint decision with high *objective* quality evaluated w.r.t. the ground truth. Still we often need to respect agents' preferences and use a good mechanism in the social choice sense, for scenarios with strong societal context, e.g. a group of friends vote to decide a restaurant for dinner. Even for scenarios with less societal context, e.g. meta-search engines [10], recommender systems [12], crowdsourcing [14], semantic webs [17], some social choice normative properties are still desired.

These scenarios naturally lead to the following challenge:" *How can we design new voting rules with good statistical properties as well as social choice normative properties [3]?*". Azari Soufiani et al. [3] follow a "design by statistic, evaluation by social choice" paradigm and proposed a statistical decision-theoretic framework to discover new social choice mechanisms as decision rules, then evaluate them by social choice normative properties.

In this paper, we answer this question by pursuing a different paradigm, which is "design by social choice, evaluation by statistics". We propose a novel framework grounded in social choice, and then evaluate statistical properties of new mechanisms in this framework. More precisely, we propose *generalized outcome scoring rules (GOSRs)* that naturally extends *generalized scoring rules* [19] to arbitrary outcome spaces, including sets of alternatives with fixed or unfixed size, rankings, and sets of rankings. We show that GOSRs include a wide range of methods including MLEs, Chamberlin and Courant rule, Monroe rule, and resolute, irresolute, and ranking versions of all positional scoring rules, STV, ranked pairs, etc. We characterize GOSRs by two normative properties: *anonymity* and *finite local consistency*. We also provide general conditions to study the statistical *consistency* of any GOSR w.r.t. any statistical model with a bound on the convergence rate, which provides a convenient tool to extend the Condorcet Jury Theorem [7] to all GOSRs and statistical models.

Related work. The pursuit of truth-revealing social choice can be dated back to the *Condorcet Jury Theorem* in the 18th century [7], which states that fix $p > 1/2$, for two alternatives, suppose agents' votes are generated i.i.d. such that each agent has probability p to be correct, then the majority aggregation of agents' votes converges to the ground truth as the number of agents goes to infinity. In modern statistical terms, Condorcet proposed a parametric model to capture the random generation of votes, and proved that the majority rule, as a statistical estimator, satisfies an important statistical property called *(statistical) consistency*, which should be distinguished from the consistency axiom in social choice.

Most previous work on statistical approaches towards social choice focused on the computation and characterization of the *maximum likelihood estimators (MLEs)* of various ranking models [8, 4, 9, 11, 21, 18, 13, 1, 2, 5]. Most of these works provide case-by-case evaluation of existing social choice mechanisms w.r.t. computational properties and statistical consistency. GOSRs provides a natural framework to obtain *new* social choice mechanisms, and their statistical consistency is established by Theorem 2 in this paper.

Due to the space constraint many discussions and proofs are omitted. Some of them can be found in the COMSOC version of this paper on the author's homepage.

2 Preliminaries

Let $\mathcal{C} = \{c_1, \ldots, c_m\}$ denote a set of m *alternatives* and let $\mathcal{L}(\mathcal{C})$ denote the set of all linear orders over $\mathcal{C}$. Each agent uses a linear order in $\mathcal{L}(\mathcal{C})$ to represent her preferences, called her *vote*. The collection of all agents' votes P is called a *profile*. Let $\mathcal{L}(\mathcal{C})^* = \mathcal{L}(\mathcal{C}) \cup \mathcal{L}(\mathcal{C})^2 \cup \cdots$ denote the set of all profiles.

Let $\mathcal{O}$ denote the set of *outcomes*. A (deterministic) *voting rule* r is a mapping that assigns to each profile a single outcome in $\mathcal{O}$. Common choices of $\mathcal{O}$ are: (1) $\mathcal{C}$, where voting rules are often called *resolute voting rules*; (2) $(2^{\mathcal{C}} \setminus \emptyset)$, where voting rules often are called *irresolute voting rules*; and (3) $\mathcal{L}(\mathcal{C})$, where voting rules are often called *preference functions* (a.k.a. *social welfare function*). A randomized voting rule assigns to each profile a probability distribution over $\mathcal{O}$.

Many commonly-studied voting rules have resolute, irresolute, and preference function versions. For example, an irresolute *positional scoring rule* is characterized by a scoring vector $\vec{s} = (s_1, \ldots, s_m)$ with $s_1 \geq s_2 \geq \cdots \geq s_m$. For any alternative c and any linear order V, we let $\vec{s}(V, c) = s_j$, where j is the rank of c in V. Given a profile P, the irresolute version of positional scoring rule chooses all alternatives c that maximize $\sum_{V \in P} \vec{s}(V, c)$, where P is viewed as a multi-set of votes. The resolute version chooses a single alternative by further applying a tie-breaking mechanism, and the preference function version ranks the alternatives w.r.t. their scores, and sometimes uses a mechanism to break ties.

Given the set of alternatives $\mathcal{C}$ and the number of agents n, a *parametric ranking model* $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$ has two parts: a *parameter space* $\mathcal{O}$, and a set of distributions over $\mathcal{L}(\mathcal{C})$, denoted by $\vec{\pi} = \{\pi_o : o \in \mathcal{O}\}$. Throughout the paper, we let P_n denote an i.i.d.-generated profile of n votes from a distribution π_o that will be clear from the context.

In this paper, we focus on the cases where the parameter space is exactly the same as the outcome space, and require that $\pi_o(V) > 0$ for all $V \in \mathcal{L}(\mathcal{C})$ and all $o \in \mathcal{O}$. Given $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$, a deterministic *estimator* T is a function that maps each profile to a parameter (outcome) in $\mathcal{O}$; a randomized estimator T maps each profile to a distribution over $\mathcal{O}$. An estimator T is *(statistically) consistent* w.r.t. a parametric ranking model $\mathcal{M}_\mathcal{C}$ if for all $o \in \mathcal{O}$, $\lim_{n\to\infty} \Pr(T(P_n) = o) \to 1$, where the n votes in P_n are generated i.i.d. from π_o. That is, a consistent estimator correctly reveals the ground truth with probability 1 as the number of i.i.d. generated votes goes to infinity.

3 Generalized Outcome Scoring Rules

For any $K \in \mathbb{N}$, let $\mathcal{B}_K = \{b_1, \dots, b_K\}$. A *total preorder* (*preorder* for short) is a reflexive, transitive, and total relation. Let $\text{Pre}(\mathcal{B}_K)$ denote the set of all preorders over $\mathcal{B}_K$. For any $\vec{p} \in \mathbb{R}^K$, we let $\text{Order}(\vec{p})$ denote the preorder $\trianglerighteq$ over $\mathcal{B}_K$ where $b_{k_1} \trianglerighteq b_{k_2}$ if and only if $[\vec{p}]_{k_1} \geq [\vec{p}]_{k_2}$. That is, the k_1-th component of $\vec{p}$ is at least as large as the k_2-th component of $\vec{p}$. For any preorder $\trianglerighteq$, if $b \trianglerighteq b'$ and $b' \trianglerighteq b$, then we write $b =_\trianglerighteq b'$. Each preorder $\trianglerighteq$ naturally induces a (partial) strict order $\triangleright$, where $b \triangleright b'$ if and only if $b \trianglerighteq b'$ and $b' \ntrianglerighteq b$.

Definition 1 (Generalized outcome scoring rules) *Given an outcome space $\mathcal{O}$, $K \in \mathbb{N}$, $f : \mathcal{L}(\mathcal{C}) \to \mathbb{R}^K$ and $g : Pre(\mathcal{B}_K) \to \mathcal{O}$, we define a* generalized outcome scoring rule (GOSR)*, denoted by $GOSR_{(f,g)}$, to be a mapping such that for any profile P, $GOSR_{(f,g)}(P) = g(Order(f(P)))$, where $f(P) = \sum_{V\in P}^{n} f(V)$.*

In words, a GOSR first uses the f function to transform the input profile P to a vector $f(P) = \sum_{V\in P} f(V)$ in $\mathbb{R}^K$, then use g to select the winner based on the *preorder* of the components in $f(P)$. For any $V \in \mathcal{L}(\mathcal{C})$, $f(V)$ is called a *generalized scoring vector*, $f(P)$ is called a *total generalized score vector*. To simplify notation, we let $\text{Order}_f(P) = \text{Order}(f(P))$. We note that $\text{Order}_f(P)$ is a preorder over $\mathcal{B}_K$, which means that it may not be a linear order. For any distribution π over $\mathcal{L}(\mathcal{C})$, we define $f(\pi) = \sum_{V\in\mathcal{L}(\mathcal{C})} \pi(V) f(V)$ and $\text{Order}_f(\pi) = \text{Order}(f(\pi))$.

The next proposition shows that $\text{GOSR}_{(f,g)}$ is a general class of voting rules. The proof is by construction and resembles the proof for GSRs [19].

Proposition 1 *Generalized scoring rules [19] are GOSRs with $\mathcal{O} = \mathcal{C}$. The irresolute versions and the ranking versions of positional scoring rules, Bucklin, Copeland, maximin, ranked pairs, STV are GOSRs, with $\mathcal{O} = (2^\mathcal{C} \setminus \emptyset)$ and $\mathcal{O} = \mathcal{L}(\mathcal{C})$, respectively.*[1]

Example 1 (MLE) MLE with a fixed-order tie-breaking[2] of any parametric ranking model $\mathcal{M}_\mathcal{C}$ is a GOSR. Let $K = |\mathcal{O}|$, $\mathcal{O} = \{o_1, o_2, \dots, o_K\}$, and for any $V \in \mathcal{L}(\mathcal{C})$ and any $i \leq K$, $[f(V)]_i = \log \pi_{o_i}(V)$. g outputs the outcome that corresponds to the largest component and uses a the same tie-breaking mechanism as in the MLE. Not all MLEs are GOSRs. For example, the MLE that uses the first agent's vote to break ties is not a GOSR. □

Example 2 (The Chamberlin and Courant rule and the Monroe rul) The Chamberlin and Courant rule [6] is a GOSR, where each dimension in the generalized scoring vector corresponds to a set of k alternatives, and its value for each ranking is the misrepresentation of the set. The Monroe rule [15] is a GOSR, where in addition to the GOSR representation for the Chamberlin and Courant rule, we have dimensions corresponding to (k-set,alternative) pairs recording whether the alternative present the agent given that the k-set is chosen to be the winner; moreover, we also have an additional dimension whose value is always $1/k$ to control the final selection of the k-set. □

4 Axiomatic Characterization

Definition 2 *A set S of profiles is* consistent*, if for any $P_1, P_2 \in S$ with $r(P_1) = r(P_2)$, we have $P_1 \cup P_2 \in S$. A voting rule r is* locally consistent *on a consistent set S if for any $P_1, P_2 \in S$ with $r(P_1) = r(P_2)$, we have $r(P_1 \cup P_2) = r(P_1) = r(P_2)$.*

Definition 3 *For any natural number t, a voting rule r is t-*consistent *if there exists a partition $\{S_1, \dots, S_t\}$ of all profiles such that for all $i \leq t$, r is locally consistent within S_i. A voting rule r is* finitely locally consistent *if it is t-consistent for some finite number t.*

Theorem 1 *Given an outcome space, a voting rule is a GOSR if and only if it satisfies anonymity and finite local consistency.*

[1] Definitions of these rules can be found in [16].

[2] A fixed-order tie-breaking break ties among alternatives w.r.t. a fixed linear order over all alternatives.

The proof is similar to the proof of the axiomatic characterization in [20].

5 Consistency of GOSRs

We first introduce more notation to present the results.

Definition 4 (Extension of a preorder) *We say that* $\trianglerighteq' \in Pre(\mathcal{B}_K)$ *is an* extension *of* $\trianglerighteq \in Pre(\mathcal{B}_K)$, *if for all* $b, b' \in \mathcal{B}_K$, *we have* $(b \rhd b') \Rightarrow (b \rhd' b')$. *For any* $\trianglerighteq, \trianglerighteq' \in Pre(\mathcal{B}_K)$, *we let* $\trianglerighteq \oplus \trianglerighteq'$ *denote the preorder in* $Pre(\mathcal{B}_K)$ *obtained from* $\trianglerighteq$ *by using* $\trianglerighteq'$ *to break ties. That is,* b_i *is strictly preferred to* b_j *in* $(\trianglerighteq \oplus \trianglerighteq')$ *if and only if (1)* $b_i \rhd b_j$, *or (2)* $b_i =_{\trianglerighteq} b_j$ *and* $b_i \rhd' b_j$.

For example, $c_1 \rhd' c_2 \rhd' c_3$ is an extension of $c_1 =_{\trianglerighteq} c_2 \rhd c_3$, but is not an extension of $c_1 =_{\trianglerighteq} c_3 \rhd c_2$.

Definition 5 (Possible linear orders) *Give a generalized scoring function* f, *we define the set of* possible linear orders over $\mathcal{B}_K$, *denoted by* $PL(f)$, *to be the linear orders over* $\mathcal{B}_K$ *that are the orders of the total score vector of* some *profile. Formally,* $PL(f) = \{Order_f(P) : P \in \mathcal{L}(\mathcal{C})^*\} \cap \mathcal{L}(\mathcal{B}_K)$.

Definition 6 (Neighborhood) *For any* $\trianglerighteq \in Pre(\mathcal{B}_K)$, *we define the* neighborhood *of* $\trianglerighteq$ *w.r.t.* f, *denoted by* $Nbr_f(\trianglerighteq)$, *to be all linear orders over* $\mathcal{B}_K$ *that can be obtained from* $\trianglerighteq$ *by using a linear order in* $PL(f)$ *to break ties. That is,* $Nbr_f(\trianglerighteq) = \{\trianglerighteq \oplus \rhd^* : \rhd^* \in PL(f)\}$. *Given* f, *the neighborhood of a distribution* π, *denoted by* $Nbr_f(\pi)$, *is the neighborhood of* $f(\pi)$ *w.r.t.* f, *that is,* $Nbr_f(\pi) = Nbr_f(Order_f(\pi))$.

We note that the definition of neighborhood does not involve the g function.

Theorem 2 *Given* $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$, f, *and* g, $GOSR_{(f,g)}$ *is consistent w.r.t.* $\mathcal{M}_{\mathcal{C}}$ *if and only if for all* $o \in \mathcal{O}$ *and all* $\rhd \in Nbr_f(\pi_o)$, *we have* $g(\rhd) = o$.

We can apply Theorem 2 to prove the statistical consistency of STV w.r.t. the Mallows model and the model for positional scoring rules by [8].

Proposition 2 *STV (preference function) is a consistent estimator w.r.t. the Mallows model. STV (resolute rule) is a consistent estimator for all models for positional scoring rules proposed by [8].*

Given a parametric ranking model $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$ and a consistent $\text{GOSR}_{(f,g)}$, we next give an upper bound on the convergence rate of the outcome of $\text{GOSR}_{(f,g)}$ to reveal the ground truth. Let s_{max} denote the maximum absolute value of the components in all generalized scoring vectors. That is, $s_{max} = \max_{V,j} |[f(V)]_j|$. Let s_{min} denote the minimum non-zero absolute value of the components in all generalized scoring vectors. Let d_{min} denote the smallest non-zero difference between the components in all $f(\pi_o)$. That is, $d_{min} = \min_{i,j \leq K, o}\{|[f(\pi_o)]_i - [f(\pi_o)]_j| : [f(\pi_o)]_i \neq [f(\pi_o)]_j\}$. Let p_{min} denote the minimum probability of any linear order under any parameter, that is, $p_{min} = \min_{V,o} \pi_o(V)$.

Theorem 3 *Suppose* $GOSR_{(f,g)}$ *is a consistent estimator for* $\mathcal{M} = (\mathcal{O}, \vec{\pi})$. *For any* $o \in \mathcal{O}$ *and* $n \in \mathbb{N}$, *we have:*

$$\Pr\left(GOSR_{(f,g)}(P_n) \neq o\right) < K \cdot \exp\left(-n \cdot \tfrac{d_{min}}{8s_{max}^2}\right) + \frac{(K(K-1)s_{max})^3}{(2p_{min})^{1.5}(s_{min})^3\sqrt{n}} = O(n^{-0.5})$$

The next theorem statess that the $O(n^{-0.5})$ bound in Theorem 3 is asymptotically tight.

Theorem 4 *There exists a parametric ranking model* $\mathcal{M}_{\mathcal{C}}$ *where* $\mathcal{O} = \mathcal{C}$ *and a GOSR* r *such that (1)* r *is consistent w.r.t.* $\mathcal{M}_{\mathcal{C}}$, *and (2) there exists* $o \in \mathcal{O}$ *such that for all even numbers* n, $\Pr(r(P_n) \neq o) = \Omega(n^{-0.5})$, *where votes in* P_n *are generated i.i.d. from* π_o.

The next theorem fully characterize all GOSRs that are consistent w.r.t. *some* models.

Theorem 5 *A GOSR is consistent w.r.t.* some *parametric ranking model if and only if for all* $o \in \mathcal{O}$, $g^{-1}(o) \cap PL(f) \neq \emptyset$.

References

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

[2] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Preference Elicitation For General Random Utility Models. In *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, Bellevue, Washington, USA, 2013.

[3] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Statistical decision theory approaches to social choice. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, Montreal, Quebec, Canada, 2014.

[4] Mark Braverman and Elchanan Mossel. Noisy Sorting Without Resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 268–276, Philadelphia, PA, USA, 2008.

[5] 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.

[6] John R. Chamberlin and Paul N. Courant. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *The American Political Science Review*, 77(3):718–733, 1983.

[7] 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.

[8] 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.

[9] Vincent Conitzer, Matthew Rognlie, and Lirong Xia. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 109–115, Pasadena, CA, USA, 2009.

[10] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th World Wide Web Conference*, pages 613–622, 2001.

[11] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Good rationalizations of voting rules. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 774–779, Atlanta, GA, USA, 2010.

[12] Sumit Ghosh, Manisha Mundhe, Karina Hernandez, and Sandip Sen. Voting for movies: the anatomy of a recommender system. In *Proceedings of the third annual conference on Autonomous Agents*, pages 434–435, 1999.

[13] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the Twenty-Eighth International Conference on Machine Learning (ICML 2011)*, pages 145–152, Bellevue, WA, USA, 2011.

[14] 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.

[15] Burt L. Monroe. Fully Proportional Representation. *The American Political Science Review*, 89(4):925–940, 1995.

[16] Hannu Nurmi. *Comparing voting systems*. Springer, 1987.

[17] Daniele Porello and Ulle Endriss. Ontology Merging as Social Choice: Judgment Aggregation under the Open World Assumption. *Journal of Logic and Computation*, 2013.

[18] 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.

[19] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 109–118, Chicago, IL, USA, 2008.

[20] Lirong Xia and Vincent Conitzer. Finite local consistency characterizes generalized scoring rules. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 336–341, Pasadena, CA, USA, 2009.

[21] 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.