Machine Learning
https://doi.org/10.1007/s10994-021-06026-2

On testing transitivity in online preference learning

Björn Haddenhorst[1] · **Viktor Bengs**[1] · **Eyke Hüllermeier**[2]

Received: 22 November 2020 / Revised: 10 May 2021 / Accepted: 17 June 2021

Abstract
The efficiency of state-of-the-art algorithms for the dueling bandits problem is essentially due to a clever exploitation of (stochastic) transitivity properties of pairwise comparisons: If one arm is likely to beat a second one, which in turn is likely to beat a third one, then the first is also likely to beat the third one. By now, however, there is no way to test the validity of corresponding assumptions, although this would be a key prerequisite to guarantee the meaningfulness of the results produced by an algorithm. In this paper, we investigate the problem of testing different forms of stochastic transitivity in an online manner. We derive lower bounds on the expected sample complexity of any sequential hypothesis testing algorithm for various forms of stochastic transitivity, thereby providing additional motivation to focus on weak stochastic transitivity. To this end, we introduce an algorithmic framework for the dueling bandits problem, in which the statistical validity of weak stochastic transitivity can be tested, either actively or passively, based on a multiple binomial hypothesis test. Moreover, by exploiting a connection between weak stochastic transitivity and graph theory, we suggest an enhancement to further improve the efficiency of the testing algorithm. In the active setting, both variants achieve an expected sample complexity that is optimal up to a logarithmic factor.

Keywords Dueling bandits · Online learning · Pairwise preferences · Stochastic transitivity · Sequential testing

Editors: Annalisa Appice, Sergio Escalera, Jose A. Gamez, Heike Trautmann.

✉ Björn Haddenhorst
bjoernha@mail.uni-paderborn.de

Viktor Bengs
bengs@mail.uni-paderborn.de

Eyke Hüllermeier
eyke@ifi.lmu.de

[1] Heinz Nixdorf Institute and Department of Computer Science, Paderborn University, Paderborn, Germany

[2] Department of Computer Science, University of Munich (LMU), Munich, Germany

Published online: 12 July 2021

1 Introduction

The setting of dueling bandits (Yue and Joachims 2009; Sui et al. 2018; Bengs et al. 2021) is a variant of the standard multi-armed bandit (MAB) problem, in which the learner is allowed to compare pairs of choice alternatives (arms) in a sequential manner. Thus, instead of repeatedly pulling an arm and observing a numerical reward, the learner pulls two arms and observes the winner of the corresponding duel. Like in the standard MAB problem, this feedback is assumed to be stochastic. A typical task of the learner is to find the "best" arm as quickly as possible, or, more generally, to identify a complete ranking of all arms. There is a variety of practically relevant applications for this learning scenario, such as ranking XBox gamers according to duel outcomes (Guo et al. 2012) or rating different objects based on pairwise preferences of users, which can nowadays be gathered quite conveniently by means of crowdsourcing services such as Amazon Mechanical Turk (Chen et al. 2013).

Relaxed assumptions of transitivity, especially different types of stochastic transitivity (Fishburn 1973), play an important role in this regard: If arm a is likely to be preferred over arm b, and b is likely to be preferred over arm c, then a is also likely to be preferred over c. Assumptions of that kind are important for several reasons. First, they assure that the learning task itself is actually well defined, for example that a naturally "best" arm actually exists. Second, they are on the basis of the design of efficient learning algorithms, which exploit generalized transitivity to reduce sample complexity (Yue and Joachims 2011; Mohajer et al. 2017; Falahatgar et al. 2018). This is comparable to how standard sorting algorithms avoid the comparison of all pairs of items and achieve an $\mathcal{O}(n \log n)$ (instead of an $\mathcal{O}(n^2)$) complexity.

Somewhat surprisingly, the problem of testing the validity of transitivity assumptions underlying various algorithms has not been considered so far. Needless to say, this would be important to guarantee the meaningfulness of the results produced by such algorithms. In fact, if the assumptions made by an algorithm are violated by the data-generating process in a concrete application, then neither its prediction nor any of its guarantees can be trusted anymore. In this paper, we therefore propose a method for testing an important form of transitivity, namely weak stochastic transitivity (WST), in an online manner. Being the weakest type of stochastic transitivity, WST is quite natural to start with. Moreover, weak stochastic transitivity of pairwise preferences (winning probabilities) is a necessary and sufficient condition for the existence of a complete ranking (strict total ordering) of all arms that is consistent with all pairwise preferences.

More specifically, we introduce an algorithmic framework consisting of two main components, namely an active sampling strategy π and a sequential test. In this way, the algorithmic framework covers two conceivable scenarios to online hypothesis testing:

- The *passive online testing* scenario, where the sampling strategy π is any dueling bandits algorithm based on a transitivity assumption, and the test component is (passively) monitoring the statistical validity of the transitivity assumption made by the dueling bandits algorithm — in other words, the learning and testing component are working in parallel, independently of each other.
- The *active online testing* scenario, in which the sampling strategy π is specifically constructed to support the test component, i.e., to make a test decision as quickly as possible.

In this paper, we introduce the problem of testing different types of stochastic transitivity within a dueling bandits problem (Sect. 4), both with and without the so-called low noise assumption (Korba et al. 2017). We prove that the expected sample complexity for testing different types of stochastic transitivity stronger than WST in an online manner is infinite in the worst case. These results provide an additional theoretical motivation for focusing on WST, as it is the only type of stochastic transitivity that admits finite expected sample complexity for online testing, which can be inferred by an appropriate reduction to the setting of *pure exploration bandits with multiple correct answers* introduced by Degenne and Koolen (2019) (Sect. 5).

We improve upon the corresponding asymptotic lower bounds on the expected sample complexity for testing WST from the latter reduction by providing instance-wise lower bounds for fixed confidence levels (Sect. 6). For the passive online testing scenario, we construct a test component based on multiple binomial hypothesis tests, for which we show consistency in terms of almost sure termination time and reliability in terms of maintained error bounds under mild assumptions on π. For the active online testing variant, we provide a sampling strategy π, such that the expected sample complexity of the latter test is optimal up to a logarithmic term (Sect. 7). Moreover, by exploiting a connection between WST and graph theory, we suggest an enhancement to further improve the efficiency of the testing algorithm (Sect. 8). The superiority of this variant in the passive setting is illustrated by an empirical evaluation (Sect. 9). The paper starts with a brief account of related work (Sect. 2), followed by a refresher of the dueling bandits problem as well as different types of stochastic transitivity (Sect. 3). Detailed proofs of theoretical results are provided in the supplementary material.

2 Related work

The dueling bandits problem was studied under strong stochastic transitivity in (Yue et al. 2012) and relaxed stochastic transitivity in (Yue and Joachims 2011), in both cases with the goal of regret minimization. In these works, the transitivity assumption is explicitly required for the theoretical guarantees. In other approaches, transitivity properties are assumed in a more indirect way, for example through probabilistic models of the feedback process. This includes the Plackett-Luce model (Luce 1959; Plackett 1975) resp. Bradley-Terry model (Bradley and Terry 1952) considered in (Szörényi et al. 2015) resp. (Maystre and Grossglauser 2017), as well as the Mallows model (Mallows 1957) studied in (Busa-Fekete et al. 2014). Mohajer et al. (2017) consider the goals of finding the best arm as well as the (top-k-)ranking of arms under WST, while Falahatgar et al. (2017a, 2017b, 2018) investigate the impact of various transitivity assumptions on these goals in an online PAC-framework. Finally, transitivity assumptions were also analyzed in batch learning scenarios, for example to estimate the underlying pairwise preference relation (Shah et al. 2016), or for the purpose of rank aggregation (Korba et al. 2017).

The literature on testing transitivity conditions is primarily rooted in the social sciences, psychology, and economics, with a special focus on experimental studies for real data. The only mathematical treatment we found is (Iverson and Falmagne 1985), where the authors provide an asymptotic likelihood-ratio test for WST. The use of Bayes factors for testing stochastic transitivity is proposed in (Cavagnaro and Davis-Stober 2014) . In (McNamara and Diwadkar 1997) and (Waite 2001), multiple binomial tests are conducted to test WST of preferences in different field studies. From a methodological point of view, this is closest

to the sequential testing approach put forward in this paper. Yet, all these works are settled in classical hypothesis testing, assuming all the data to be available beforehand. In contrast to this, the focus of this paper is on hypothesis testing in an online setting, where data arrives sequentially, and test decisions should be taken as quickly as possible while maintaining a predefined level of confidence.

As already mentioned in the introduction the problem of testing stochastic transitivity in an online manner can be tackled by a suitable reduction to the pure exploration bandits with multiple correct answers introduced by Degenne and Koolen (2019), which will be discussed more thoroughly in Sect. 5.

3 Theoretical background

In this section, we concisely recall the main theoretical foundations needed throughout the paper. In the supplementary material, we provide a list of symbols used in the paper for the sake of convenience.

3.1 Dueling bandits

Consider a finite set of m arms identified by the index set $[m] := \{1, \dots, m\}$. In the setting of the dueling bandits problem, two distinct arms $i, j \in [m]$ can be compared with each other at each time step $t \in \mathbb{N}$. Querying a pairwise preference, the learner is provided with binary feedback about the winner of the duel, which is assumed to be generated by a time-stationary iid probabilistic process. The probability $\mathbb{P}(i \succ j)$ that arm i wins against arm j is given by some underlying (unknown) ground truth parameter $q_{i,j} \in [0, 1]$. We suppose that ties are not possible. Thus (assuming w.l.o.g. $q_{i,i} = \frac{1}{2}$ for every $i \in [m]$), we can infer that $\mathbf{Q} = (q_{i,j})_{1 \le i,j \le m}$ is a reciprocal relation on $[m]$, i.e., $\mathbf{Q}$ is an element of

$$\mathcal{Q}_m := \left\{ \boldsymbol{Q} = (q_{i,j})_{1 \le i,j \le m} \in [0,1]^{m \times m} \mid q_{j,i} = 1 - q_{i,j} \text{ for every } i, j \in [m] \right\}.$$

To assimilate the information available at time $t \in \mathbb{N}$, let us write $(\boldsymbol{n}_t)_{i,j}$ for the number of comparisons between i and j until time t, and $(\boldsymbol{w}_t)_{i,j}$ for the number of times i has won against j until time t. This obviously implies $(\boldsymbol{w}_t)_{i,j} + (\boldsymbol{w}_t)_{j,i} = (\boldsymbol{n}_t)_{i,j}$ and $(\boldsymbol{n}_t)_{i,j} = (\boldsymbol{n}_t)_{j,i}$. Then, $\boldsymbol{n}_t = ((\boldsymbol{n}_t)_{i,j})_{1 \le i,j \le m}$ is a symmetric integer-valued matrix with zeros on its diagonal. If $\boldsymbol{w} \in \mathbb{N}_0^{m \times m}$ and $\boldsymbol{n} \in \mathbb{N}_0^{m \times m}$, we denote the matrix $(\frac{w_{i,j}}{n_{i,j}})_{1 \le i,j \le m} \in [0,1]^{m \times m}$ by $\frac{\boldsymbol{w}}{\boldsymbol{n}}$, where we define for convenience $\frac{x}{0} := \frac{1}{2}$ for any $x \in \mathbb{N}_0$. Moreover, we write $[m]_2$ for the set containing all subsets of size 2 of $[m]$ and $(m)_2$ for the set of all $(i, j) \in [m] \times [m]$ with $i < j$. A specific learning algorithm in the realm of dueling bandits can be identified by a sampling strategy as defined in the following.

Definition 3.1 A **sampling strategy** π is a family of random mappings, which, depending on the time t and the observations $\boldsymbol{n}_0, \boldsymbol{w}_0, \dots, \boldsymbol{n}_{t-1}, \boldsymbol{w}_{t-1}$ available before time t, determines the two distinct arms $i(t), j(t) \in [m]$ that are to be compared at time $t \in \mathbb{N}$. Let $\boldsymbol{\Pi}$ denote the set of all sampling strategies, while $\boldsymbol{\Pi}_\infty$ denotes the family of sampling strategies π that sample every pair $\{i, j\}$ almost surely (a.s.) infinitely often, which means that $(\boldsymbol{n}_t)_{i,j} \to \infty$ a.s. as $t \to \infty$.

Note that if $\pi \in \Pi \setminus \Pi_\infty$, then a sampling strategy $\hat{\pi} \in \Pi$ that chooses the same pair as π in each time step with probability $1 - 1/t$, and otherwise (i.e., with probability $1/t$) picks a pair $\{i, j\}$ uniformly at random from $[m]_2$, fulfills $\hat{\pi} \in \Pi_\infty$ and

$$\mathbb{P}\big(\pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \le t' \le t-1}) \neq \hat{\pi}(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \le t' \le t-1}\big) \le \frac{1}{t} \to 0 \text{ as } t \to \infty\,.$$

Thus, $\hat{\pi}$ and π behave similarly in the limit. This shows that the assumption $\pi \in \Pi_\infty$, which is required for theoretical results in our framework, is rather mild.

3.2 Stochastic transitivity

Different types of stochastic transitivity have been used in the realm of dueling bandits problems (Bengs et al. 2021), mainly because they provide a certain degree of regularity of the reciprocal relations in $\mathcal{Q}_m$, and thereby facilitate learning. In particular, the following transitivities are commonly considered in the literature.

Definition 3.2 A reciprocal relation $\boldsymbol{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m$ is said to satisfy

(i) **weak stochastic transitivity** (WST) iff

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge 1/2\,,$$

(ii) **moderate stochastic transitivity** (MST) iff

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \min(q_{i,j}, q_{j,k})\,,$$

(iii) ν-**relaxed stochastic transitivity** ($\nu - RST$) for some $\nu \in (0, 1)$ iff

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \nu \max(q_{i,j}, q_{j,k}) + (1 - \nu)/2\,,$$

(iv) **strong stochastic transitivity** (SST) iff

$$\big(q_{i,j} \ge 1/2 \wedge q_{j,k} \ge 1/2\big) \Rightarrow q_{i,k} \ge \max(q_{i,j}, q_{j,k})\,,$$

where all previous conditions must hold for all distinct $i, j, k \in [m]$.

The set consisting of all stochastic transitive reciprocal relations of a certain type is

$$\mathcal{Q}_m(\mathrm{XST}) := \{\mathbf{Q} \in \mathcal{Q}_m \mid \mathbf{Q} \text{ is XST}\}, \quad \mathrm{XST} \in \{\mathrm{WST}, \mathrm{MST}, \nu - RST, \mathrm{SST}\},$$

and we write $\mathcal{Q}_m(\neg\mathrm{XST}) := \mathcal{Q}_m \setminus \mathcal{Q}_m(\mathrm{XST})$. The following relationships hold between the different types of stochastic transitivities:

$$\mathcal{Q}_m(\mathrm{SST}) \subsetneq \mathcal{Q}_m(\mathrm{MST}) \subsetneq \mathcal{Q}_m(\mathrm{WST}), \quad \mathcal{Q}_m(\mathrm{SST}) \subsetneq \mathcal{Q}_m(\nu - RST) \subsetneq \mathcal{Q}_m(\mathrm{WST})\,,$$

but neither $\mathcal{Q}_m(\nu - RST) \subseteq \mathcal{Q}_m(\mathrm{MST})$ nor $\mathcal{Q}_m(\mathrm{MST}) \subseteq \mathcal{Q}_m(\nu - RST)$.

3.3 Violations of WST

To illustrate the issues that may arise in case of a violation of the WST assumption, and highlight the importance of testing such assumptions, consider algorithms that are based on the idea of (noisy) sorting (Szörényi et al. 2015; Mohajer et al. 2017). Roughly speaking, the

active sampling strategies underlying such algorithms mimic the behavior of sorting algorithms, such as merge sort or quicksort — with the main difference that, due to the assumed stochasticity, deciding the order between two arms may require repeated comparisons.

Obviously, weak stochastic transitivity is the least assumption required by such algorithms. On the other side, it is easy to see that a sorting-based algorithm will always return a complete ranking (with high confidence), regardless of whether the underlying relation contains preferential cycles or not. Yet, this ranking will strongly depend on the order in which the arms are compared, and hence be more or less random and therefore meaningless.

4 Online transitivity testing

We focus on the following testing problem in the context of an underlying dueling bandits problem:

$$\mathbf{H}_0 : \mathbf{Q} \text{ satisfies XST} \quad \mathbf{H}_1 : \mathbf{Q} \text{ does not satisfy XST}, \tag{1}$$

where $\text{XST} \in \{\text{WST}, \text{MST}, \nu - RST, \text{SST}\}$. This test shall be conducted for different types of transitivity in an online manner.

Thus, it is natural to consider sequential hypothesis tests, in which a test decision can be provided at any time during the data generating process. The particular choice of the null hypothesis is motivated by the passive scenario, in which a learning algorithm assumes XST to be fulfilled and the test shall detect a possible violation thereof. As we focus on tests with guarantees on both, its type I and the type II error, it is possible to swap $\mathbf{H}_0$ and $\mathbf{H}_1$, and still obtain qualitatively the same theoretical results as below.

In the course of the paper, we focus on algorithms $\mathcal{A}$ for the testing problem, which might be probabilistic and interact with the underlying dueling bandits environment, as stipulated by the definition of a sampling strategy π (Definition 3.1). In case an algorithm $\mathcal{A}$ terminates, it returns a decision denoted by $\mathbf{D}(\mathcal{A}) \in \{\text{XST}, \neg\text{XST}\}$ with the semantic $\mathbf{D}(\mathcal{A}) = \text{XST}$ resp. $\mathbf{D}(\mathcal{A}) = \neg\text{XST}$ indicate that $\mathcal{A}$ predicts that XST holds resp. is violated. Moreover, we denote by $T^{\mathcal{A}}$ the sample complexity of an algorithm $\mathcal{A}$, i.e., the number of pairwise comparisons $\mathcal{A}$ has made before termination.

For our theoretical analysis of the testing problem, we will consider the following set of relations:

$$\mathcal{Q}_m^h := \left\{ \mathbf{Q} = (q_{i,j})_{1 \le i,j \le m} \in \mathcal{Q}_m \mid |q_{i,j} - 1/2| > h \text{ for all distinct } i,j \in [m] \right\},$$

where $h \in [0, 1/2)$. In case $h > 0$, the relations in $\mathcal{Q}_m^h$ are said to satisfy the *low noise assumption* (Korba et al. 2017). Here, the parameter h determines to some extent the complexity of the testing problem: For instance, the larger h, the easier it becomes to determine the sign of $q_{i,j} - 1/2$, which in turn facilitates checking WST. For $\text{XST} \in \{\text{WST}, \text{MST}, \text{SST}, \nu - RST\}$ and any $h \in [0, 1/2)$, we define

$$\mathcal{Q}_m^h(\text{XST}) := \mathcal{Q}_m^h \cap \mathcal{Q}_m(\text{XST}) \quad \text{and} \quad \mathcal{Q}_m^h(\neg\text{XST}) := \mathcal{Q}_m^h \cap \mathcal{Q}_m(\neg\text{XST}).$$

Moreover, we may regard $\mathcal{Q}_m$ as a subset of $\mathbb{R}^{m(m-1)/2}$ and, in this way, equip it with the standard Euclidean topology of $\mathbb{R}^{m(m-1)/2}$. Therefore, for a subset $\mathcal{Q}'_m \subseteq \mathcal{Q}_m$, we use the standard notation $\partial\mathcal{Q}'_m$ for the boundary of $\mathcal{Q}'_m$ as a subset of this topological space $\mathcal{Q}_m$. The notion of a solution to the XST-testing problem is stated in the following.

Definition 4.1 For given $h \in [0, 1/2)$ and error probabilities $\alpha, \beta \in (0, 1)$, we say that an algorithm[1] $\mathcal{A}$ **solves the** XST**-testing problem on** $\mathcal{Q}_m^h$ **for** α **and** β (in short: $\mathcal{A}$ **solves** $\mathcal{P}_{\mathrm{XST}}^{m,h,\alpha,\beta}$) if $T^{\mathcal{A}}$ is almost surely finite on any instance $\mathbf{Q} \in \mathcal{Q}_m^0$ and the following holds:

$$\begin{aligned} \inf_{\boldsymbol{Q} \in \mathcal{Q}_m^h(\mathrm{XST})} \mathbb{P}_{\boldsymbol{Q}}(\mathbf{D}(\mathcal{A}) = \mathrm{XST}) \geq 1 - \alpha \\ \text{and } \inf_{\boldsymbol{Q} \in \mathcal{Q}_m^h(\neg \mathrm{XST})} \mathbb{P}_{\boldsymbol{Q}}(\mathbf{D}(\mathcal{A}) = \neg\, \mathrm{XST}) \geq 1 - \beta. \end{aligned} \tag{2}$$

Interestingly, as the following theorem reveals, the testing problem (1) for a stochastic type of transitivity stronger than WST turns out be too difficult. Hence, we will focus on the case XST = WST in the rest of the paper.

Theorem 4.2 *Let* $h, \alpha, \beta \in (0, 1/2)$, $m \in \mathbb{N}_{\geq 3}$ *and* $\mathrm{XST} \in \{\mathrm{MST}, \mathrm{SST}, \nu - RST\}$ *be fixed. If an algorithm* $\mathcal{A}$ *solves* $\mathcal{P}_{\mathrm{XST}}^{m,h,\alpha,\beta}$, *then* $\mathbb{E}_{\boldsymbol{Q}}[T^{\mathcal{A}}] = \infty$ *for any* $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{XST}) \cap \partial\mathcal{Q}_m^h(\neg \mathrm{XST}) \neq \emptyset$. *In particular, we have* $\sup_{\boldsymbol{Q} \in \mathcal{Q}_m^h} \mathbb{E}_{\boldsymbol{Q}}[T^{\mathcal{A}}] = \infty$.

To prove this theorem, we show that any solution $\mathcal{A}$ to $\mathcal{P}_{\mathrm{XST}}^{m,h,\alpha,\beta}$ may be used to test, for some $p_0 \in [0, 1]$, *any* $p_1 > p_0$, and with an error probability of at most $\max\{\alpha, \beta\}$ whether a coin $C \sim \mathrm{Ber}(p)$ has bias $p = p_0$ or $p = p_1$. But if p_1 converges to p_0, the number of coin flips necessary to maintain the error probability tends to infinity in expectation. A detailed proof of the theorem is provided in Section B in the supplement.

5 Reduction to Pure Exploration Bandits with Multiple Correct Answers

The testing problem at hand may be reduced to the *Pure Exploration Bandits* scenario with multiple correct answers as presented by Degenne and Koolen (2019), the details of which can be found in Section F of the supplement. This approach leads to the following results: If $\mathcal{A}(\gamma)$ solves $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma,\gamma}$, then for some (known) constant $D_m^h(\mathbf{Q}) > 0$,

$$\liminf_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln(\gamma^{-1})} \geq \frac{1}{D_m^h(\mathbf{Q})}, \tag{3}$$

and there exists a solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma,\gamma}$ with

$$\lim_{\gamma \to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln(\gamma^{-1})} \leq \frac{1}{D_m^h(\mathbf{Q})}. \tag{4}$$

If $\mathbf{Q} \in \mathcal{Q}_m(\mathrm{X})$ (for $\mathrm{X} \in \{\mathrm{WST}, \neg\mathrm{WST}\}$), the complexity term $D_m^h(\mathbf{Q})$ is given as

$$\sup_{\mathbf{v} \in \Delta_{(m)_2}} \inf_{\mathbf{Q}' \in \mathcal{Q}_m^h(\neg \mathrm{X})} \sum\nolimits_{(i,j) \in (m)_2} v_{i,j} d_{\mathrm{KL}}(q_{i,j}, q'_{i,j}),$$

where $\Delta_{(m)_2}$ is the set of all $\mathbf{v} = (v_{i,j})_{1 \leq i < j \leq m}$ with $\min_{i<j} v_{i,j} \geq 0$ and $\sum_{i<j} v_{i,j} = 1$, and $d_{\mathrm{KL}}(p, q) = p \ln(p/q) + (1 - p) \ln((1 - p)/(1 - q))$ is the KL-divergence between two

[1] Possibly probabilistic

Bernoulli distributions with success probability p resp. q. We prove in the supplement[2] (cf. Lemma F.7) that

$$\frac{1/4 - h^2}{192}\binom{m}{2}h^{-2} \leq \sup_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST})} \frac{1}{D_m^h(\mathbf{Q})} \leq \sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \frac{1}{D_m^h(\mathbf{Q})} \leq \frac{1}{8}\binom{m}{2}h^{-2}$$

and

$$\frac{1}{192}\binom{m}{2}h^{-2} \leq \sup_{\mathbf{Q}\in\mathcal{Q}_m^h(\mathrm{WST})} \frac{1}{D_m^0(\mathbf{Q})} \leq \sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \frac{1}{D_m^0(\mathbf{Q})} \leq \frac{1}{2}\binom{m}{2}h^{-2}$$

hold for all $h \in (0, 1/2)$. This indicates that the case $h = 0$ is more complex than the case $h > 0$ and shows that any *optimal* solution $\mathcal{A}(\gamma)$ to $\mathcal{P}_{\mathrm{WST}}^{m,h,\gamma,\gamma}$ or $\mathcal{P}_{\mathrm{WST}}^{m,0,\gamma,\gamma}$ fulfills

$$\sup_{\mathbf{Q}\in\mathcal{Q}_m^h} \lim_{\gamma\to 0} \frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln(\gamma^{-1})} \in \Theta(m^2h^{-2}), \tag{5}$$

respectively, as $\max\{m, h^{-1}\} \to \infty$. Unfortunately, these results do not yield any information on the case where γ is fixed. Moreover, the algorithmic solution $\mathcal{A}(\gamma)$ presented by Degenne and Koolen (2019) is very inefficient for the problem of testing WST, if not infeasible in practice, which is due to a hard min-max problem that has to be solved at each time step (cf. Remark F.1). In the following, we will discuss further lower and upper bounds on the worst-case sample complexity of solutions to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$. Our results are to some extent stronger than (3) and (4), as they are covering the cases of a fixed confidence level γ, which in turn corresponds to the typical setting of (online) testing.

6 Lower bounds for online testing of weak stochastic transitivity

In this section, we provide lower bounds on the expected termination time of any algorithm solving $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$. Similarly to Theorem 4.2, these results are obtained by reducing a testing problem for the biases of independent coins to $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$. A sample complexity analysis of the latter testing problem results in the bounds stated below, the proof of which can again be found in Section B.

In order to state an instance-wise lower bound for the case $h > 0$, let us introduce some more notation: Given $\mathbf{Q} \in \mathcal{Q}_m^0$, we write $\sigma_{\mathbf{Q}}$ for a permutation on $[m]$, which fulfills $q_{\sigma_{\mathbf{Q}}(i),\sigma_{\mathbf{Q}}(i+1)} > 1/2$ for every $i \in [m]$. We show in the appendix (Lemma B.1) that $\sigma_{\mathbf{Q}}$ exists for every $\mathbf{Q} \in \mathcal{Q}_m^0$, even though we only need this for every $\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST})$. In case $\mathbf{Q} \in \mathcal{Q}_m^0(\mathrm{WST})$, $\sigma_{\mathbf{Q}}$ is the underlying ground-truth ranking of $\mathbf{Q}$, and permuting rows and columns according to $\sigma_{\mathbf{Q}}$ results in a reciprocal relation with entries $> 1/2$ above the diagonal.

Theorem 6.1 *Let $h_0, \gamma_0 \in (0, 1/2)$ be fixed, $h \in (0, h_0)$, $\alpha, \beta \in (0, \gamma_0)$ and $m \in \mathbb{N}_{\geq 3}$. Suppose $\mathcal{A}$ is an algorithm that solves $\mathcal{P}_{\mathrm{WST}}^{m,h,\alpha,\beta}$, and let $\mathbf{Q} \in \mathcal{Q}_m^h(\mathrm{WST})$ be arbitrary. Define $h_{i,j} := |q_{i,j} - 1/2|$ for every distinct $i, j \in [m]$, $\gamma := \min\{\alpha, \beta\}$, and $\sigma = \sigma_{\mathbf{Q}}$. Then, there exists a constant $c = c(h_0, \gamma_0) > 0$ such that*

[2] The bounds presented in Lemma F.6 are stronger but omitted here for the sake of brevity.

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \geq c \ln\left(\gamma^{-1}\right) \sum\nolimits_{1 \leq i < j-1 < m} h^{-2}_{\sigma(i),\sigma(j)} \geq c \binom{m-1}{2} \ln\left(\gamma^{-1}\right) h^{-2}. \tag{6}$$

Thus, $\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]$ *is in* $\Omega(m^2 h^{-2} \ln \gamma^{-1})$ *as* $\max\{m, \gamma^{-1}, h^{-1}\} \to \infty$.

Note that the right-hand side of (6) is of the order $m^2 h^{-2} \ln(\gamma^{-1})$, which is coherent with (5). The fact that the instance-wise bound only depends on $\binom{m-1}{2}$ instead of all $\binom{m}{2}$ entries of $\mathbf{Q}$ is due to our proof technique, which is nonetheless of the same order with respect to m.

Let us now consider the more complex case $h = 0$. As any solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{WST}}$ is also a solution to $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$ for any $h \in (0, 1/2)$, Theorem 6.1 is applicable in this case. However, we can slightly improve upon this. In the following, for functions $f, g : X \to (0, \infty)$, we say that $f \in \Omega_{\sup}(g)$ as $x \to x_0$ if $\limsup_{x \to x_0} \frac{g(x)}{f(x)} < \infty$.

Theorem 6.2 *Let* $\alpha, \beta \in (0, 1/2)$ *be fixed and suppose* $\mathcal{A}$ *to be an algorithm that solves* $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{WST}}$. *Then, the following holds*:

(a) $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] = \infty$ *for any* $\mathbf{Q}$ *in a set* $\emptyset \neq \mathcal{Q}^{\dagger}_m \subsetneq \partial\mathcal{Q}_m(\mathrm{WST}) \cap \partial\mathcal{Q}_m(\neg\mathrm{WST})$,
(b) $\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega(m^2 h^{-2}) \cap \Omega_{\sup}(h^{-2} \ln \ln h^{-1})$ *as* $\max\{m, h^{-1}\} \to \infty$.

As we point out in the proof of this theorem, the set $\mathcal{Q}^{\dagger}_m$ in (a) can be chosen as the set of all $\mathbf{Q} \in \mathcal{Q}_m$, for which some permutation σ on $[m]$ exists such that the following conditions are fulfilled:

$$\begin{aligned} \forall 1 \leq i < j \leq m : \; & q_{\sigma(i),\sigma(j)} \geq 1/2, \\ \forall i \in [m-1] : \; & q_{\sigma(i),\sigma(i+1)} > 1/2, \\ \exists 1 \leq i' < j' - 1 \leq m - 1 : \; & q_{\sigma(i'),\sigma(j')} = 1/2. \end{aligned}$$

In the proof of the theorem, to make (b) more explicit, we provide several examples for a family $\{\mathbf{Q}(h)\}_{h \in (0,1/2)} \subseteq \mathcal{Q}^h_m(\mathrm{WST})$, for which

$$\limsup_{h \searrow 0} \mathbb{E}_{\mathbf{Q}(h)}[T^{\mathcal{A}}] / (h^{-2} \ln \ln(h^{-1})) \geq (1 - 2\gamma)/2.$$

Regarding the occurrence of the limes superior in Lemma A.2, this is the best we may infer from Lemma A.2.

At first sight, part (b) of Theorem 6.2 may appear to contradict (5), which does not involve a $\ln \ln h^{-1}$-factor. However, note that (5) only yields a bound on the worst-case of the *asymptotic* of $\frac{\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}(\gamma)}]}{\ln(\gamma^{-1})}$ as $\gamma \searrow 0$, whereas our bound holds for any fixed γ.[3] Thus, there is actually no contradiction.

[3] To illustrate this difference, note that $f : (0,1)^2 \to \mathbb{R}$ defined via $f(\gamma, h) := h^{-2} \ln \ln h^{-1}$ if $h \leq \gamma$ and $f(\gamma, h) := h^{-2}$ if $h > \gamma$ fulfills $\lim_{\gamma \to 0} f(\gamma, h) = h^{-2}$ for all fixed $h \in (0,1)$, but at the same time we have $\lim_{h \to 0} \frac{f(\gamma,h)}{h^{-2} \ln \ln h^{-1}} = 1$.

7 Online testing of WST

Guided by our findings in Sect. 6, we now focus on the testing problem (1) for WST in the framework developed in Sect. 4. Note that weak stochastic transitivity is in any case of particular interest for the ranking problem in dueling bandits, as it is both a sufficient and a necessary condition for the existence of a ranking over the arms consistent with the preference relation $\mathbf{Q}$, in the sense that an arm i is preferred over an arm j if and only if $q_{i,j} \geq 1/2$.

Algorithm 1 $\mathcal{A}_{\text{naive}}$

Parameters: m, a sampling strategy π, a function $C : \mathbb{N} \to [0, \infty]$
Input: $\boldsymbol{n}_0$, $\boldsymbol{w}_0$

$\mathbf{Q}' \leftarrow (1/2)_{1 \leq i,j \leq m} \in \mathcal{Q}_m$
$S \leftarrow \emptyset$
for $t \in \mathbb{N}$ **do**
  $\{i(t), j(t)\} \sim \pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \leq t' \leq t-1})$, w.l.o.g. $i(t) < j(t)$
  Observe $X^{[t]}_{i(t),j(t)} \sim \text{Ber}(q_{i(t),j(t)})$
  Update $\boldsymbol{w}_t$ via $(\boldsymbol{w}_t)_{k,l} \leftarrow (\boldsymbol{w}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i(t),j(t)\} \text{ and } X^{[t]}_{k,l}=1\}}$ $\quad \forall 1 \leq k, l \leq m$
  Update $\boldsymbol{n}_t$ via $(\boldsymbol{n}_t)_{k,l} \leftarrow (\boldsymbol{n}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i(t),j(t)\}\}}$ $\quad \forall 1 \leq k, l \leq m$
  if $\frac{(\boldsymbol{w}_t)_{i(t),j(t)}}{(\boldsymbol{n}_t)_{i(t),j(t)}} > 1/2 + C((\boldsymbol{n}_t)_{i,j})$ **then**
    Set $q'_{i(t),j(t)} \leftarrow 1, q'_{j(t),i(t)} \leftarrow 0$ and $S \leftarrow S \cup \{i(t), j(t)\}$
  else if $\frac{(\boldsymbol{w}_t)_{i(t),j(t)}}{(\boldsymbol{n}_t)_{i(t),j(t)}} < 1/2 + C((\boldsymbol{n}_t)_{i,j})$ **then**
    Set $q'_{i(t),j(t)} \leftarrow 1, q'_{j(t),i(t)} \leftarrow 0$ and $S \leftarrow S \cup \{i(t), j(t)\}$
  if $S = (m)_2$ and $\mathbf{Q}' \in \mathcal{Q}_m(\text{WST})$ **then return** WST
  else if $S = (m)_2$ and $\mathbf{Q}' \in \mathcal{Q}_m(\neg\text{WST})$ **then return** $\neg$WST

A first naïve approach for a testing component for the passive scenario (cf. Section 1) is Algorithm 1, which does the following: Terminate as soon as we can decide, *for every* $(i,j) \in (m)_2$, each with error probability at most $\gamma' = \min\{\alpha, \beta\} \binom{m}{2}^{-1}$, whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds, and output WST if an auxiliary relation $\mathbf{Q}'$ generated during runtime is WST, and $\neg$WST otherwise. To construct $\mathbf{Q}'$, the value $q'_{i,j}$ is set to 1 resp. 0 whenever we are sure enough (for the first time) that $q_{i,j} > 1/2$ resp. $q_{i,j} < 1/2$ holds. Here, testing the sign of $q_{i,j} - 1/2$ with confidence level γ may be done by stopping as soon as $(\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j}$ leaves the interval $[1/2 - C((\boldsymbol{n}_t)_{i,j}), 1/2 + C((\boldsymbol{n}_t)_{i,j})]$, where $C(\cdot)$ is an appropriate any-time confidence bound for $(\boldsymbol{w}_t)_{i,j}/(\boldsymbol{n}_t)_{i,j}$. The term *appropriate* is specified in Definition 7.1 below.

In the initialization step of $\mathcal{A}_{\text{naive}}$, we inform the algorithm about how often every item i has already been compared to every other item j before the start, denoted by $(\boldsymbol{n}_0)_{i,j}$, and how often i has won against j, denoted by $(\boldsymbol{w}_0)_{i,j}$. Our setting allows us to assume that $(\boldsymbol{w}_0)_{i,j} \sim \text{Bin}((\boldsymbol{n}_0)_{i,j}, q_{i,j})$ for all $1 \leq i < j \leq m$. As the theoretical results do not depend on the explicit choice of $\boldsymbol{n}_0$ and $\boldsymbol{w}_0$, we assume w.l.o.g. that $(\boldsymbol{n}_0)_{i,j} = 1$ for all distinct $i, j \in [m]$ throughout the paper.

Definition 7.1 For any $p \in [0,1]$, suppose $\{X_n^{(p)}\}_{n\in\mathbb{N}}$ to be a family of iid random variables with distribution Ber(p). We say that a function $C : \mathbb{N} \to [0,\infty]$ is (h,γ)-**correct** for given $h \in [0,1/2)$ and $\gamma \in (0,1/2)$, if the following holds:

(a) For any $p \neq 1/2$, the following stopping time is almost surely finite:

$$\mathcal{N}^{(p)} := \mathcal{N}^{(p)}(C) := \min\left\{ n \in \mathbb{N} \,:\, \frac{1}{n}\sum\nolimits_{k=1}^{n} X_k^{(p)} \notin [1/2 - C(n), 1/2 + C(n)] \right\}.$$

(b) For all $p > 1/2 + h$, we have

$$\mathbb{P}\Big(\frac{1}{\mathcal{N}^{(p)}} \sum\nolimits_{k=1}^{\mathcal{N}^{(p)}} X_k^{(p)} < 1/2 - C\big(\mathcal{N}^{(p)}\big) \Big) \leq \gamma \,,$$

and similarly for all $p < 1/2 - h$,

$$\mathbb{P}\Big(\frac{1}{\mathcal{N}^{(p)}} \sum\nolimits_{k=1}^{\mathcal{N}^{(p)}} X_k^{(p)} > 1/2 + C\big(\mathcal{N}^{(p)}\big) \Big) \leq \gamma.$$

In case $h > 0$, a first example for an (h,γ)-correct function $C_{h,\gamma}$ can be inferred from Hoeffding's inequality, by means of

$$C_{h,\gamma}^{\text{Hoeffding}}(n) := \begin{cases} 1/2, & \text{if } n \leq \lceil h^{-2}\ln(\gamma^{-1})/2 \rceil \\ 0, & \text{otherwise} \end{cases} . \tag{7}$$

With this, the decision whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ is not made in a sequential manner, but instead after exactly $\lceil h^{-2}\ln(\gamma^{-1})/2 \rceil$ duels of i and j have been conducted. At the end of this section, we will introduce more sophisticated any-time confidence bounds admitting decisions in a sequential manner, and also treat the case $h = 0$.

Theorem 7.2 *Let* $m \in \mathbb{N}_{\geq 3}$, $\alpha, \beta \in (0,1)$, *and* $h \in (0,1/2)$ *be fixed, and define* $\gamma' := \min\{\alpha,\beta\}\binom{m}{2}^{-1}$. *For any* $\pi \in \Pi_\infty$ *and* (h,γ')*-correct function* C, *Algorithm* 1 *instantiated with parameters* m, π, *and* C *is a solution to* $\mathcal{P}_{\text{WST}}^{m,h,\alpha,\beta}$.

By construction, the sample complexity of Algorithm 1 is exactly the number of iterations that are required for testing the signs of all $q_{i,j} - 1/2$, $(i,j) \in (m)_2$. By choosing C according to (7), testing the sign of $q_{i,j} - 1/2$ requires in any case exactly $N := \lceil h^{-2}\ln(\gamma^{-1})/2 \rceil$ iid samples governed by Ber($q_{i,j}$). However, the explicit time at which a pair has been sampled at least N times highly depends on the underlying sampling strategy π, so that an analysis of the sample complexity of $\mathcal{A}_{\text{naive}}$ can only be done w.r.t. the corresponding sampling strategy π. As the testing component is working in parallel to π in the passive setting, i.e., it has no influence on the behavior of π, the minimum requirement for a test component in the passive online test seems to be consistency in terms of an a.s. finite termination time and the adherence to predefined error bounds for a general class of sampling strategies. Both requirements are met by the test underlying $\mathcal{A}_{\text{naive}}$ by Theorem 7.2 for the class Π_∞ if $\mathcal{A}_{\text{naive}}$ is instantiated with an (h,γ')-correct C.

Remark 7.3 In the passive online testing scenario, i.e., the sampling strategy π is instantiated in a black-box fashion by some dueling bandits algorithm based on a transitivity

assumption (such as those by Falahatgar et al. (2017a, 2018)), it might happen that π terminates before the testing algorithm came to a decision, and in particular that π is not defined any more. In this case, if one is still interested in whether transitivity was fulfilled in hindsight, one may continue sampling according to the strategy $\hat{\pi}$, which picks each query $\{i,j\} \in [m]_2$ with probability $1/\binom{m}{2}$.

The other way around, if the testing algorithm came to a positive decision ($\mathbf{D}(\mathcal{A}) = \mathrm{XST}$), although the online ranking algorithm has not yet terminated, one can simply continue the sampling strategy without the testing component.

In case of a negative decision ($\mathbf{D}(\mathcal{A}) = \neg\mathrm{XST}$), the online ranking algorithm should be interrupted due to violating the assumptions.

In the active online testing scenario (cf. Section 1), on the other side, we have the possibility to choose π in a favorable way and consequently analyze the sample complexity of Algorithm 1. For this purpose, we consider a sampling strategy $\pi = \pi(m, C)$ depending on the other parameters of $\mathcal{A}_{\mathrm{naive}}$, which focuses on the time-dependent set consisting of all pairs $\{i,j\}$, for which it is not yet sure with confidence level γ' whether $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ holds. Formally, the following set is considered:

$$U_C(t) := \left\{ \{i,j\} \in [m]_2 \,\middle|\, \forall t' < t \,:\, (\boldsymbol{w}_{t'})_{i,j}/(\boldsymbol{n}_{t'})_{i,j} \in \left[1/2 \pm C((\boldsymbol{n}_{t'})_{i,j})\right] \right\}, \quad t \in \mathbb{N}.$$

In each time t, the sampling strategy $\pi(m, C)$ queries $\{i,j\} \in [m]_2$ uniformly at random from $U_C(t)$, if $U_C(t)$ is non-empty, and otherwise queries $\{i,j\} \in [m]_2$ uniformly at random from $[m]_2$. Note that the second case (i.e., $U_C(t)$ is empty) is only defined in order to ensure that $\pi \in \Pi_\infty$, which in turn allows for applying Theorem 7.2. In light of this, we obtain the following corollary.

Corollary 7.4 *Let $m \in \mathbb{N}_{\geq 3}$, $h \in (0, 1/2)$, $\alpha, \beta \in (0, \gamma_0)$ for some $\gamma_0 \in (0,1)$, and choose $\gamma' := \min\{\alpha, \beta\}/\binom{m}{2}$. Let $\pi = \pi(m, C^{\mathrm{Hoeffding}}_{h,\gamma'})$ be the sampling strategy of the above type and suppose $\mathcal{A}$ to be Algorithm 1 called with parameters m, π, and $C = C^{\mathrm{Hoeffding}}_{h,\gamma'}$ from (7). Then, $\mathcal{A}$ solves $\mathcal{P}^{n,h,\alpha,\beta}_{\mathrm{WST}}$ and fulfills*

$$T^{\mathcal{A}} = \binom{m}{2} \left\lceil \frac{h^{-2}}{2} \ln\left(\frac{m(m-1)}{2\min\{\alpha,\beta\}} \right) \right\rceil \quad \mathbb{P}_{\mathbf{Q}}\text{-almost surely for all } \mathbf{Q} \in \mathcal{Q}^h_m.$$

In particular, if $\gamma := \min\{\alpha, \beta\}$, we have that

$$\sup_{\mathbf{Q} \in \mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}\left[T^{\mathcal{A}}\right] \in \mathcal{O}\big((m^2 \ln m) h^{-2} \ln \gamma^{-1}\big)$$

as $\max\left\{m, h^{-1}, \gamma^{-1}\right\} \to \infty$.

With regard to Theorem 6.1, the testing algorithm from Corollary 7.4 is already asymptotically optimal up to logarithmic factors for the WST testing problem in (1) for instances $\mathbf{Q} \in \mathcal{Q}^h_m$. Nevertheless, one may ask, firstly, whether termination is only possible as soon as being sure about the signs of $q_{i,j} - 1/2$ of *all* the $\binom{m}{2}$ many $\{i,j\} \in [m]_2$, and secondly, if the rough correction term in the error probability (i.e., $\binom{m}{2}$) for the sign test of any

$q_{i,j} - 1/2$, is optimal. In the following section, we answer both questions negatively, giving rise to more sophisticated testing procedures. Moreover, we also present a solution to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{WST}}$ and develop instance-wise upper bounds for $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$.

We conclude this section with a discussion of further suitable anytime confidence bounds, the proofs of which are deferred to the supplement for the sake of convenience. In the following, if $p \in [0,1]$ and $C : \mathbb{N} \to \mathbb{R}$ are fixed, let us define $\mathcal{N}^{(p)}(C)$ as in Definition 7.1. Inspired by the sequential probability ratio test (Wald and Wolfowitz 1948) for testing whether a coin has bias $1/2 + h$ or $1/2 - h$, we may define

$$C^{\mathrm{SPRT}}_{h,\gamma}(n) := \frac{1}{2n}\left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil$$

for any $h \in (0, 1/2)$ and $\gamma \in (0, 1/2)$. Then, $C^{\mathrm{SPRT}}_{h,\gamma}$ is (h,γ)-correct and fulfills

$$\sup_{p:|p-1/2|\geq h} \mathbb{E}[\mathcal{N}^{(p)}(C^{\mathrm{SPRT}}_{h,\gamma})] = (2h)^{-1}\left\lceil \frac{\ln((1-\gamma)/\gamma)}{\ln((1/2+h)/(1/2-h))} \right\rceil (1-2\gamma).$$

This is shown in Lemma A.1 in the supplement. In contrast to $C^{\mathrm{Hoeffding}}_{h,\gamma}$, choosing $C^{\mathrm{SPRT}}_{h,\gamma}$ leads to a sequential test, where the runtime depends on the (unknown) ground-truth p, which makes the question of instance-dependent bounds actually interesting. But on the other side, for any $p \in (0,1)$, the random variable $\mathcal{N}^{(p)}(C^{\mathrm{SPRT}}_{h,\gamma})$ is not bounded in the sense that $\mathcal{N}^{(p)}(C^{\mathrm{SPRT}}_{h,\gamma}) \leq N$ a.s. for some $N \in \mathbb{N}$. However, as we also point out in Lemma A.1, the optimality of the sequential probability ratio test assures us that choosing $C = C^{\mathrm{SPRT}}_{h,\gamma}$ is optimal w.r.t. $\mathbb{E}[\mathcal{N}^{(1/2\pm h)}(C)]$.

We now turn to the more complex case of preference relations in $\mathcal{Q}^0_m$. In the following, we write $\ln_2(\cdot) := \ln\ln(\cdot)$ and $\ln_3(\cdot) := \ln\ln\ln(\cdot)$ for the sake of convenience. From a result by Farrell (1964) we can infer that, for some appropriate value[4] $n_0 \in \mathbb{N}$, the function

$$C^{\mathrm{Farrell}}_{0,\gamma}(n) := \begin{cases} \sqrt{\ln_2(n+e) + c\ln_3(n+e^e)}/\sqrt{8n}, & \text{if } n \geq n_0 + 1 \\ 1/2, & \text{otherwise} \end{cases}$$

is $(0,\gamma)$-correct and fulfills

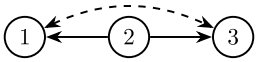
$$\lim_{h\to 0} \frac{\mathbb{E}\left[\mathcal{N}^{(1/2\pm h)}(C^{\mathrm{Farrell}}_{0,\gamma})\right]}{h^{-2}\ln\ln h^{-1}} = \frac{1}{2}\mathbb{P}_{1/2}(\mathcal{N}^{(0)}(C^{\mathrm{Farrell}}_{0,\gamma}) = \infty) > 0,$$

which is shown in Lemma A.3 in the supplement. With the help of $C^{\mathrm{Farrell}}_{0,\gamma}$, we will be able to present a solution $\mathcal{A}$ to $\mathcal{P}^{m,0,\alpha,\beta}_{\mathrm{WST}}$, in which the term $h^{-2}\ln\ln h^{-1}$ will naturally appear in the sample-complexity bound (cf. in Theorem 8.6). As we have seen in Theorem 6.2, the $\ln\ln h^{-1}$-factor may not be avoided here.

[4] To define n_0, suppose S'_n to be a symmetric random walk on $\mathbb{Z}$, i.e., $S'_n = \sum_{i=1}^n X'_i$ where $\{X'_i\}_{i\in\mathbb{N}}$ is a family of iid random variables X'_i with $\mathbb{P}(X'_i = 1) = \mathbb{P}(X'_i = -1) = 1/2$. Then, $n_0 := \min\left\{n \in \mathbb{N} \,\middle|\, \mathbb{P}\left(\exists \tilde{n} \geq n+1 \,:\, |S'_{\tilde{n}}| \geq 2^{-\frac{1}{2}}\tilde{n}\ln_2(\tilde{n}+e) + c\ln_3(\tilde{n}+e^e)\right) \leq \gamma\right\}$.

8 Enhanced online WST testing

In this section, we will exploit the connection between graph theory and WST in order to improve the algorithm from Corollary 7.4. The main idea for improvement is the following: Suppose we wanted to test whether $\mathbf{Q} \in \mathcal{Q}_3$ is WST.



If we are sure enough that $q_{2,1}, q_{2,3} > 1/2$ holds (depicted by the edges $2 \to 1, 2 \to 3$ in the picture to the right), then we can infer that $\mathbf{Q}$ is WST, since the definition of weak stochastic transitivity is fulfilled in both cases ($q_{1,3} < 1/2$ and $q_{1,3} > 1/2$). Thus, testing $q_{1,3}$ is in some sense superfluous. To generalize this kind of reasoning to the case $m > 3$, we first introduce a graph-theoretical interpretation of the problem.

8.1 Graph-theoretical considerations

Throughout this section, we let $G = ([m], E_G)$ be some directed graph (digraph) on $[m]$, i.e., $E_G \subseteq [m] \times [m]$ and whenever $(i,j) \in E_G$ holds then $(j,i) \notin E_G$. We call G a tournament (or complete digraph), if for all distinct $i,j \in [m]$ either $(i,j) \in E_G$ or $(j,i) \in E_G$ holds. A graph $G \in \mathcal{G}_m$ is called acyclic if it does not contain any cycle.

Note that, for every $\mathbf{Q} \in \mathcal{Q}_m^0$ and every distinct $i,j \in [m]$, either $q_{i,j} > 1/2$ or $q_{j,i} > 1/2$ holds. Hence, each $\mathbf{Q} \in \mathcal{Q}_m^0$ can be identified by a tournament $G_{\mathbf{Q}} := G = ([m], E_G)$ with $E_G := \left\{(i,j) \in [m] \times [m] \mid i \neq j \text{ and } q_{i,j} > 1/2\right\}$. It can be shown that $\mathbf{Q} \in \mathcal{Q}_m^0$ is WST iff the corresponding identifying tournament $G_{\mathbf{Q}}$ is acyclic (Proposition D.2).

In the toy example above, note that the identifying tournament of $\mathbf{Q}$ is acyclic in any case, i.e., regardless whether $q_{1,3} < \frac{1}{2}$ or $q_{1,3} > \frac{1}{2}$ holds, making one edge of the identifying tournament superfluous for inferring WST of $\mathbf{Q}$ and allowing a correct decision merely on the digraph given by $2 \to 1, 2 \to 3$. The following two definitions generalize the idea of superfluous edges for general digraphs.

Definition 8.1 A digraph G is called **transitive in expansion** if each of its extensions to a tournament is acyclic. In other words, no tournament $\tilde{G}$ on $[m]$ with $E_G \subseteq E_{\tilde{G}}$ contains any cycle.

Definition 8.2 Let $G \in \mathcal{G}_m$. We call a pair $\{i,j\} \in [m]_2$ **negligible for** G if for every $k \in [m] \setminus \{i,j\}$ either $(i,k), (j,k) \in E_G$ or $(k,i), (k,j) \in E_G$ holds.

Regarding Proposition D.2, we may write $\mathcal{G}_m(\text{WST})$ for the set of all digraphs G on $[m]$, which are transitive in expansion. The following result provides a link between transitivity in expansion and the notion of negligibility.

Proposition 8.3 *Let $G \in \mathcal{G}_m$. If G does not contain a cycle and every $\{i,j\} \in [m]_2$ with $(i,j), (j,i) \notin E_G$ is negligible for G, then $G \in \mathcal{G}_m(\text{WST})$ holds.*

This result together with the connection of preference relations and tournaments, brings us closer to answering the questions raised at the end of Sect. 7, as we show the following: If G is transitive in expansion, then there exists some graph $\tilde{G}$, which is transitive in expansion, satisfying $E_{\tilde{G}} \subseteq E_G$ and $|E_{\tilde{G}}| = \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ (Proposition D.5), i.e., in particular we have $|E_G| \geq |E_{\tilde{G}}| = \binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$. Thus, it is possible to infer WST of $\mathbf{Q}$ by merely considering $\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ edges of the identifying tournament, while a violation of WST by $\mathbf{Q}$ can be confirmed if the identifying tournament contains a cycle.

8.2 Exploiting transitivity in expansion

Equipped with these insights, we suggest Algorithm 2 as a testing procedure for $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$. In the next theorem, we verify that this algorithm has in fact the desired theoretical guarantees; the proof is given in Section D in the supplement.

Algorithm 2 : $\mathcal{A}_{\text{improved}}$

Parameters: m, a sampling strategy π, a function $C : \mathbb{N} \to [0, \infty]$
Input: $\boldsymbol{n}_0$, $\boldsymbol{w}_0$

$\hat{E}_0 \leftarrow \emptyset$
for $t \in \mathbb{N}$ **do**
  $\{i(t), j(t)\} \sim \pi(t, (\boldsymbol{n}_{t'}, \boldsymbol{w}_{t'})_{0 \leq t' \leq t-1})$, w.l.o.g. $i(t) < j(t)$
  Observe $X^{[t]}_{i(t),j(t)} \sim \text{Ber}(q_{i(t),j(t)})$
  Update $\boldsymbol{w}_t$ via $(\boldsymbol{w}_t)_{k,l} \leftarrow (\boldsymbol{w}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i(t),j(t)\} \text{ and } X^{[t]}_{k,l}=1\}}$ $\quad \forall 1 \leq k, l \leq m$
  Update $\boldsymbol{n}_t$ via $(\boldsymbol{n}_t)_{k,l} \leftarrow (\boldsymbol{n}_{t-1})_{k,l} + \mathbf{1}_{\{\{k,l\}=\{i(t),j(t)\}\}}$ $\quad \forall 1 \leq k, l \leq m$
  $\hat{E}_t \leftarrow \hat{E}_{t-1}$
  if $\frac{(\boldsymbol{w}_t)_{i(t),j(t)}}{(\boldsymbol{n}_t)_{i(t),j(t)}} > 1/2 + C((\boldsymbol{n}_t)_{i(t),j(t)})$ and $(i(t), j(t)) \notin \hat{E}_t$ and $(j(t), i(t)) \notin \hat{E}_t$ **then**
    $\hat{E}_t \leftarrow \hat{E}_t \cup \{(i(t), j(t))\}$
  else if $\frac{(\boldsymbol{w}_t)_{i(t),j(t)}}{(\boldsymbol{n}_t)_{i(t),j(t)}} < 1/2 + C((\boldsymbol{n}_t)_{i,j})$ and $(i(t), j(t)) \notin \hat{E}_t$ and $(j(t), i(t)) \notin \hat{E}_t$ **then**
    $\hat{E}_t \leftarrow \hat{E}_t \cup \{(j(t), i(t))\}$
  if $([m], \hat{E}_t)$ is transitive in expansion **then return** WST
  else if $([m], \hat{E}_t)$ contains a cycle **then return** $\neg$WST

Theorem 8.4 *Let* $\pi \in \Pi_\infty$, $\alpha, \beta \in (0, 1)$ *and* $h \in [0, 1/2)$ *be fixed and define* $\gamma' := \min\{\alpha/m, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$. *Suppose* $C : \mathbb{N} \to [0, \infty]$ *is* (h, γ')*-correct, and let* $\mathcal{A}$ *denote Algorithm* 2 *called with parameters* m, π *and* C. *Then,* $\mathcal{A}$ *solves* $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$. *In case* $C = C^{\text{X}}_{h,\gamma'}$ *for* $\text{X} \in \{\text{Hoeffding}, \text{SPRT}, \text{Farrell}\}$ *and* $\tilde{\mathcal{A}}$ *is Algorithm* 1 *called with parameters* m, π *and* $C_{h,\tilde{\gamma}}$ *with* $\tilde{\gamma} := \min\{\alpha, \beta\} / \binom{m}{2}$ *(as suggested by Theorem 7.2),* $T^{\mathcal{A}} \leq T^{\tilde{\mathcal{A}}}$ *holds almost surely w.r.t.* $\mathbb{P}_{\mathbf{Q}}$ *for any* $\mathbf{Q} \in \mathcal{Q}^0_m$.

Lemma D.9 indicates that one can not expect to choose a correction term smaller than $\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor$ for the desired type II error within the choice of γ in Algorithm 2. Furthermore, the fact that the graph $G \in \mathcal{G}_m$ with edges $1 \to 2 \to \ldots \to m \to 1$ contains a cycle, unlike any of its proper subgraphs, demonstrates optimality of the correction term m for the desired type I error within the choice of γ. As a direct consequence of Theorem 8.4, we obtain a result analogous to the one stated in Corollary 7.4 for Algorithm 2 called with m, the sampling strategy π from Corollary 7.4, and $C^{\text{Hoeffding}}_{h,\gamma'}$ with $\gamma' = \min\{\alpha/m, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$, so that it achieves an optimal worst-case runtime (up to a logarithmic term of m) in the active online testing scenario as well.

8.3 Instance-wise upper bounds and exploiting negligibility of edges

We conclude this section with more sophisticated solutions to $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$ in the active setting, which take into account that those queries $\{i,j\}$, which are negligible with high probability, are superfluous and should be avoided. To this end, we define the sampling strategy $\pi^*(m, C)$ as the sampling strategy which, similarly to the sampling strategies $\pi(m, C)$ considered in Corollary 7.4, keeps track of a specific subset of $[m]_2$ consisting of all $\{i,j\}$ for which $q_{i,j} > 1/2$ or $q_{i,j} < 1/2$ can be decided with enough confidence (with regard to C) at time t. In contrast to the latter, the used subset by $\pi^*(m, C)$ takes also the negligibility of edges into account. Formally, $\pi^*(m, C)$ considers the following set at time t:

$$U^*_C(t) := \big\{ \{i,j\} \in [m]_2 \,\big|\, (i,j), (j,i) \notin \hat{E}_t \text{ and } \{i,j\} \text{ is not negligible for } ([m], \hat{E}_t) \big\}.$$

The sampling procedure of $\pi^*(m, C)$ is just like $\pi(m, C)$, but only replacing $U_C(t)$ by $U^*_C(t)$. Note that $\hat{E}_t$ may be defined in terms of $\boldsymbol{n}_0, \boldsymbol{w}_0, \ldots, \boldsymbol{n}_{t-1}, \boldsymbol{w}_{t-1}$ as the set of all $(i,j) \in [m] \times [m]$ for which some $t' < t$ exists, such that

$$(\boldsymbol{w}_{t'}/\boldsymbol{n}_{t'})_{i,j} > 1/2 + C((\boldsymbol{n}_{t'})_{i,j}) \quad \text{and} \quad \forall t'' < t' : (\boldsymbol{w}_{t''}/\boldsymbol{n}_{t''})_{i,j} \in \big[1/2 \pm C((\boldsymbol{n}_{t''})_{i,j})\big],$$

whence $\pi^*(m, C)$ is in fact a sampling strategy as stipulated in Definition 3.1.

From Theorem 8.4, we immediately obtain that Algorithm 2 called with parameters m, $\pi^*(m, C)$ and C is a solution to $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$. But even if this guarantee holds for any (h, γ')-correct function C, it is desirable to choose C in such a way that the sample complexity of the corresponding algorithm is low. According to Lemma A.1, Lemma A.3, and Lemma A.2, the choices $C = C^{\text{SPRT}}_{h,\gamma'}$ resp. $C = C^{\text{Farrell}}_{h,\gamma'}$ are to some extent optimal in this regard for the cases $h > 0$ resp. $h = 0$. With these, we obtain the following instance-wise upper bounds on the expected termination time for solutions to $\mathcal{P}^{m,h,\alpha,\beta}_{\text{WST}}$. They show that the values $|q_{i,j} - 1/2|$ determine the complexity of testing whether $\mathbf{Q}$ is weakly stochastic transitive or not. In comparison to the lower bound stated in Theorem 6.1, our instance-wise upper bounds depend on all $\binom{m}{2}$ instead of only $\binom{m-1}{2}$ entries of $\mathbf{Q}$. Needless to say, in terms of the asymptotic behavior as $m \to \infty$, this difference is negligible.

Theorem 8.5 *Suppose* $m \in \mathbb{N}_{\geq 3}$, $\alpha, \beta \in (0, 1/2)$, $h \in (0, 1/2)$, *and define* $\gamma' := \min\{\alpha/m, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$. *Let* $\mathcal{A}$ *be Algorithm* 2 *called with parameters* m, *the sampling strategy* $\pi^*(m, C^{\mathrm{SPRT}}_{h,\gamma'})$ *and* $C = C^{\mathrm{SPRT}}_{h,\gamma'}$ *as the function* C. *Then,* $\mathcal{A}$ *solves* $\mathcal{P}^{m,h,\alpha,\beta}_{\mathrm{WST}}$. *Suppose* $\mathbf{Q} \in \mathcal{Q}^h_m$ *is fixed and write* $h_{i,j} := |q_{i,j} - 1/2|$ *for all distinct* $i, j \in [m]$. *Then, with* $e(h, \gamma') := \left\lceil \frac{\ln((1-\gamma')/\gamma')}{\ln((1/2+h)/(1/2-h))} \right\rceil$, *we have that* $\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}]$ *is bounded from above by*

$$\sum_{(i,j)\in(m)_2} \frac{e(h,\gamma')}{2h_{i,j}} \left| 1 - 2\left(1 + (1/2 + h_{i,j})^{e(h,\gamma')}(1/2 - h_{i,j})^{-e(h,\gamma')}\right)^{-1} \right|. \tag{8}$$

By means of Lemma A.1, it immediately follows that algorithm $\mathcal{A}$ from Theorem 8.5 fulfills $\sup_{\mathbf{Q}\in\mathcal{Q}^h_m} \mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \in \Omega(m^2 \ln(m) h^{-2} \ln(\gamma^{-1}))$ as $\max\{m, h^{-1}, \gamma^{-1}\} \to \infty$, i.e., it is asymptotically optimal up to a $\ln(m)$-factor. In order to compare the result of Theorem 8.5 with the instance-wise lower bound from Theorem 6.1 more thoroughly, suppose $\mathbf{Q} \in \mathcal{Q}^h_m(\mathrm{WST})$ and $(i, j) \in (m)_2$ with $|\sigma_{\mathbf{Q}}(i) - \sigma_{\mathbf{Q}}(j)| > 1$ to be fixed for the moment and let $\alpha = \beta = \gamma$ for simplicity. Due to $e(h, \gamma') \in \Theta(h^{-1})$ as $h \searrow 0$, the dependency of (8) on the (i, j)-entry of $\mathbf{Q}$ is approximately $h^{-1}_{i,j} h^{-1}$, whereas this dependency in (6) is of the form $h^{-2}_{i,j}$. This suggests, that the two bounds are closest in case $h \approx h_{i,j}$. Considering that the choice $C = C^{\mathrm{SPRT}}_{h,\gamma'}$ assures optimal early detection of $\mathrm{sign}(q_{i,j} - 1/2)$ only in case $|q_{i,j} - 1/2| = h$, the appearance of h^{-1} in (8) may not come as a surprise. Moreover, the scaling $\gamma' \approx \gamma/m^2$ leads to an additional factor of $2\ln(m)$ in (8) compared to (6).

Theorem 8.6 *Let* $m \in \mathbb{N}_{\geq 3}$, $\alpha, \beta \in (0, 1/2)$ *be fixed and define* $\gamma' := \min\{\alpha/m, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$. *Suppose* $\mathcal{A}$ *is Algorithm* 2 *called with parameters* m, $\pi^*(m, C^{\mathrm{Farrell}}_{0,\gamma'})$ *and* $C^{\mathrm{Farrell}}_{0,\gamma'}$. *Then* $\mathcal{A}$ *solves* $\mathcal{P}^{m,0,\gamma,\gamma}_{\mathrm{WST}}$, *and there exists some* $h_0 \in (0, 1/2)$ *with the following property*: *If* $\mathbf{Q} \in \mathcal{Q}^0_m$ *is such that* $h_{i,j} := |q_{i,j} - 1/2| \leq h_0$ *for all distinct* $i, j \in [m]$, *then*

$$\mathbb{E}_{\mathbf{Q}}[T^{\mathcal{A}}] \leq \frac{1}{2} \sum\nolimits_{(i,j)\in(m)_2} h^{-2}_{i,j} \ln\ln(h^{-1}_{i,j}).$$

9 Experiments

In this section, we compare the WST testing procedures from Theorems 7.2 and 8.4. Since the solution obtained by Degenne and Koolen (2019) appears infeasible in practice (Remark F.1), we do not consider it in our experiments. For the sake of simplicity, we focus on the passive testing scenario, with $\pi \in \Pi_\infty$ being such that it chooses its queries at each time step uniformly at random from $[m]_2$. We also fix $\alpha = \beta = 0.05$ as well as $h = 0.01$ in the following. Further, we will write $\mathcal{A}_{\mathrm{naive}}$ for Algorithm 1 instantiated with the parameters m, π and $C^{\mathrm{SPRT}}_{h,\gamma'}$ with $\gamma' := \min\{\alpha, \beta\}/\binom{m}{2}$, and $\mathcal{A}_{\mathrm{improved}}$ for Algorithm 2 called with parameters m, π and $C^{\mathrm{SPRT}}_{h,\gamma''}$ with $\gamma'' := \min\{\alpha/m, \beta(\binom{m}{2} - \lfloor \frac{m+1}{3} \rfloor)^{-1}\}$. Here, we have chosen the boundary function C due to its optimal behavior w.r.t. the expected runtime on some instances as stated in Lemma A.1.

In the first experiment, we investigate the termination time of $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}_{\text{improved}}$ for preference relations in $\mathcal{Q}_m^{0.05}(\text{WST})$ or $\mathcal{Q}_m^{0.05}(\neg\text{WST})$. To this end, we sample $\mathbf{Q}$ uniformly at random from $\mathcal{Q}_m^{0.05}(\text{WST})$ (resp. $\mathcal{Q}_m^{0.05}(\neg\text{WST})$), run the test algorithms until termination, respectively, and repeat this process for 100 times. Here, both $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}_{\text{improved}}$ — started with some $\mathbf{Q}$ — observe the same duel chosen by π in each time step, as well as the same outcome of the duel. As stated in Theorem 8.4, $\mathcal{A}_{\text{improved}}$ may thus terminate earlier than $\mathcal{A}_{\text{naive}}$ in any case. In the following table we report the obtained average termination times (and the corresponding standard error in brackets) for varying values of m.

	WST		¬WST	
	$\mathcal{A}_{\text{naive}}$	$\mathcal{A}_{\text{improved}}$	$\mathcal{A}_{\text{naive}}$	$\mathcal{A}_{\text{improved}}$
$m = 4$	5540 (329.3)	**2936** (245.5)	5273 (325.4)	**3468** (315.6)
$m = 5$	11,670 (601.7)	**9862** (596.3)	12,041 (581.5)	**4380** (367.1)
$m = 6$	20,420 (789.1)	**17,951** (810.7)	20,374 (921.3)	**4903** (235.6)
$m = 7$	36,149 (1403.7)	**32,429** (1408.6)	35,261 (1535.9)	**6203** (342.1)
$m = 8$	52,214 (2050.0)	**48,216** (2009.0)	55,727 (1910.8)	**7066** (191.6)

The results reveal that $\mathcal{A}_{\text{improved}}$ needs significantly less samples for checking WST than $\mathcal{A}_{\text{naive}}$ throughout, and the effect is strongest if $\mathbf{Q}$ is not WST and m is large. In particular, if the underlying preference relation is not WST, the termination time of $\mathcal{A}_{\text{improved}}$ is mostly decreasing with the number of available arms, while the termination time of $\mathcal{A}_{\text{naive}}$, on the other side, increases rapidly with the number of arms. Moreover, both test algorithms did not make any error in deciding whether WST holds or not for the underlying preference relation $\mathbf{Q}$, i.e., the observed accuracy of both test algorithms was 100% throughout. Last but not least, it is worth mentioning that $\mathcal{A}_{\text{improved}}$ (as well as $\mathcal{A}_{\text{naive}}$) terminates for each problem scenario much earlier than the derived worst-case upper bound $(2h)^{-1}\left\lceil \frac{\ln((1-\gamma'')/\gamma'')}{\ln((1/2+h)/(1/2-h))} \right\rceil (1-2\gamma'')\binom{m}{2}$, which is $\geq 4370\binom{m}{2}$ for any $m \geq 3$ (cf. Theorem 8.5).

Next, we analyze the impact of the degree of violation of WST within a preference relation $\mathbf{Q}$ — measured by the number of cycles[5] in the identifying tournament $G_{\mathbf{Q}}$ — on the sample complexities of $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}_{\text{improved}}$, respectively. For this purpose, we choose $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$ and $\mathbf{Q}_4$ as

$$\begin{pmatrix} - & x & x & x & x & x \\ & - & x & x & x & x \\ & & - & x & x & x \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix}, \begin{pmatrix} - & x & y & x & x & x \\ & - & x & x & x & x \\ & & - & x & x & x \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix}, \begin{pmatrix} - & x & y & x & y & x \\ & - & x & y & x & x \\ & & - & x & x & x \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix} \text{ and } \begin{pmatrix} - & x & y & x & y & x \\ & - & x & y & x & x \\ & & - & x & x & y \\ & & & - & x & x \\ & & & & - & x \\ & & & & & - \end{pmatrix},$$

respectively, where $x := 0.6$ and $y := 0.4$. The following table shows the number of cycles in $G_{\mathbf{Q}_i}$ together with the average runtimes (as well as the empirical standard errors in

[5] Recall that, according to our definition above, any cycle is of the form $i_1 \to \ldots \to i_k \to i_1$, where $i_1, \ldots, i_k$ are *distinct*.

brackets) of $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}_{\text{improved}}$, if started with $\mathbf{Q}_i$, over 100 runs. We also added the average elapsed time T_{elapsed} (in seconds) per run as an indicator of the computational costs of $\mathcal{A}_{\text{naive}}$ and $\mathcal{A}_{\text{improved}}$. All experiments were run on a single CPU.[6]

		$\mathcal{A}_{\text{naive}}$		$\mathcal{A}_{\text{improved}}$	
i	# cycles in $G_{\mathbf{Q}_i}$	$T^{\mathcal{A}}$	T_{elapsed}	$T^{\mathcal{A}}$	T_{elapsed}
1	0	25,919 (332.3)	0.60	**25**,**639** (340.8)	2.16
2	1	25,170 (296.4)	0.58	**10**,**609** (187.4)	0.44
3	9	25,599 (366.1)	0.60	**8988** (110.3)	0.31
4	28	26,014 (355.7)	0.60	**9063** (110.7)	0.31

These results support the following conclusions. Firstly, the larger the number of cycles in the identifying tournament $G_{\mathbf{Q}_i}$ of the underlying preference relation $\mathbf{Q}_i$ (i.e., the more severe the WST property is violated), the lower the sample complexity of $\mathcal{A}_{\text{improved}}$ is on average. Secondly, the latter effect reveals an "elbow" dependency in the sense that the decrease of the termination time is rapidly declining with the number of cycles, with the strongest decline if at least one cycle is present. Thirdly, $\mathcal{A}_{\text{naive}}$ does not seem to benefit from stronger violations of WST and in fact does not exploit structural properties of the current estimated preference relation for an early termination such as $\mathcal{A}_{\text{improved}}$ does. Finally, the results for $\mathbf{Q}_1$ with regard to the averaged elapsed time demonstrate that checking the transitive in expansion property of the internal graph maintained by $\mathcal{A}_{\text{naive}}$ (i.e., line 7 in Algorithm 2) increases the computational cost per iteration step by a factor of $\approx \frac{2.16}{25639}\frac{25919}{0.6} \approx 3.64$. However, the superiority of $\mathcal{A}_{\text{improved}}$ over $\mathcal{A}_{\text{naive}}$ in terms of sample complexity is so strong, that it outperforms $\mathcal{A}_{\text{naive}}$ even with regard to computational costs on $\mathbf{Q}_2$, $\mathbf{Q}_3$ and $\mathbf{Q}_4$.

In summary, the experiments empirically confirm our theoretical results on the superiority of the enhanced testing algorithm $\mathcal{A}_{\text{improved}}$ compared to $\mathcal{A}_{\text{naive}}$.

10 Conclusion

In this paper, we have analyzed the problem of testing stochastic transitivity assumptions within the dueling bandits framework. For various types of stochastic transitivity, we provided instance-dependent lower bounds on the expected number of samples needed by any sequential test to come to a test decision obeying predefined error bounds. These results indicate that testing a stochastic transitivity assumption stronger than weak stochastic transitivity is hopeless in worst case scenarios.

In light of these results, we have introduced a flexible algorithmic framework, which allows one to either monitor the validity of the weak stochastic transitivity assumption made by a dueling bandit algorithm during its sampling process in a passive way, or to actively query pairs of arms in order to confirm or refute this assumption as quickly as possible. To this end, we designed a sequential testing method within the algorithmic framework and provided theoretical guarantees for its type I and type II error as well as an almost surely finite termination time within the passive testing scenario, if it is instantiated

[6] For our experiments, we used a machine with an Intel® Core™ i7-4700MQ Processor.

with an appropriate function to measure the confidence of pairwise probability estimates. In addition, we have provided some examples for appropriate confidence functions and have shown optimality of the resulting algorithm up to a logarithmic factor in terms of the expected runtime for a suitable sampling strategy, which is actively supporting the test component. Finally, we enhanced the testing method by incorporating graph-theoretical considerations, resulting in faster decisions on the validity or violation of WST, and provided instance-dependent upper bounds on the expected runtime of this testing procedure.

Based on our findings, it would be of interest to transfer the ideas for WST testing as developed in this paper to weaker yet still practically relevant assumptions in the realm of dueling bandits, such as the existence of a Condorcet Winner. Furthermore, a more thorough experimental study for the suggested algorithmic framework would be important to gain more insights into the actual degree of support provided by the testing component to already established sampling strategies for ranking problems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10994-021-06026-2.

Acknowledgements The authors gratefully acknowledge financial support by the German Research Foundation (DFG).

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Bengs, V., Busa-Fekete, R., El Mesaoudi-Paul, A., & Hüllermeier, E. (2021). Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research, 22*(7), 1–108.

Bermond, J. C. (1972). Ordres à distance minimum d'un tournoi et graphes partiels sans circuits maximaux. *Mathématiques et Sciences humaines, 37,* 5–25.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika, 39*(3/4), 324–345.

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning, 5*(1), 1–122.

Busa-Fekete, R., Hüllermeier, E., Szörényi, B. (2014). Preference-based rank elicitation using statistical models: The case of Mallows. In: Proceedings of the International Conference on Machine Learning (ICML), pp 1071–1079.

Cantelli, F. P. (1933). Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del sig. *Paul Lévy. Giornale dell'Istituto Italiano degli Attuari, 4*(3), 327–350.

Cavagnaro, D. R., & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision, 1*(2), 102.

Chen, L., & Li, J. (2015). On the optimal sample complexity for best arm identification. *CoRR*. (**abs/1511.03774**).

Chen, X., Bennett, P.N., Collins-Thompson, K., Horvitz, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In: Proceedings of ACM International Conference on Web Search and Data Mining (WSDM), pp 193–202.

Degenne, R., Koolen, W.M. (2019). Pure exploration with multiple correct answers. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), pp 14591–14600.

Falahatgar, M., Hao, Y., Orlitsky, A., Pichapati, V., Ravindrakumar, V. (2017a). Maxing and ranking with few assumptions. In: Proceedings of Advances in Neural Information Processing Systems (NIPS), pp 7060–7070.

Falahatgar, M., Orlitsky, A., Pichapati, V., Suresh, A.T. (2017b). Maximum selection and ranking under noisy comparisons. In: Proceedings of International Conference on Machine Learning (ICML), pp 1088–1096.

Falahatgar, M., Jain, A., Orlitsky, A., Pichapati, V., Ravindrakumar, V. (2018). The limits of maxing, ranking, and preference learning. In: Proceedings of International Conference on Machine Learning (ICML), pp 1426–1435.

Farrell, R. H. (1964). Asymptotic behavior of expected sample size in certain one sided tests. *The Annals of Mathematical Statistics, 35*(1), 36–72.

Ferguson, T. S. (1967). *Mathematical statistics: A decision theoretic approach Probability and mathematical statistics*. Cambridge: Academic Press.

Fishburn, P. C. (1973). Binary choice probabilities: On the varieties of stochastic transitivity. *Journal of Mathematical Psychology, 10*(4), 327–352.

Garivier, A., Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In: Proceedings of Annual Conference on Learning Theory (COLT), pp 998–1027.

Guo, S., Sanner, S., Graepel, T., Buntine, W. (2012). Score-based Bayesian skill learning. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD), pp 106–121.

Iverson, G., & Falmagne, J. C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10*(2), 131–153.

Korba, A., Clémençon, S., Sibony, E. (2017). A learning theory of ranking aggregation. In: Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS), pp 1001–1010.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Amsterdam: Wiley.

Mallows, C. L. (1957). Non-null ranking models. *Biometrika, 44*(1), 114–130.

Maystre, L., Grossglauser, M. (2017). Just sort it! A simple and effective approach to active preference learning. In: Proceedings of International Conference on Machine Learning (ICML), pp 2344–2353.

McNamara, T. P., & Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology, 34*(2), 160–190.

Mohajer, S., Suh, C., Elmahdy, A. (2017). Active learning for top-k rank aggregation from noisy comparisons. In: Proceedings of International Conference on Machine Learning (ICML), pp 2488–2497.

Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society Series C (Applied Statistics), 24*(1), 193–202.

Ross, S. (1996). *Stochastic processes. Wiley series in probability and statistics probability and statistics*. Amsterdam: Wiley.

Sachs, H. (1971). *Einführung in die Theorie der endlichen Graphen*. Hanser.

Shah, N., Balakrishnan, S., Guntuboyina, A., Wainwright, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In: Proceedings of the International Conference on Machine Learning (ICML), pp 11–20.

Siegmund, D. (1985). *Sequential analysis: Tests and confidence intervals. Springer series in statistics*. Berlin: Springer.

Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika, 48,* 303–312.

Sui, Y., Zoghi, M., Hofmann, K., Yue, Y. (2018). Advancements in dueling bandits. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp 5502–5510.

Szörényi, B., Busa-Fekete, R., El Mesaoudi-Paul, A., Hüllermeier, E. (2015). Online rank elicitation for Plackett-Luce: A dueling bandits approach. In: Proceedings of Advances in Neural Information Processing Systems (NIPS), pp 604–612.

Waite, T. A. (2001). Intransitive preferences in hoarding gray jays (Perisoreus canadensis). *Behavioral Ecology and Sociobiology, 50*(2), 116–121.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics, 16*(2), 117–186.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*(3), 326–339.

Yue, Y., Joachims, T. (2009). Interactively optimizing information retrieval systems as a dueling bandits problem. In: Proceedings of International Conference on Machine Learning (ICML), pp 1201–1208.
Yue, Y., Joachims, T. (2011). Beat the mean bandit. In: Proceedings of International Conference on Machine Learning (ICML), pp 241–248.
Yue, Y., Broder, J., Kleinberg, R., & Joachims, T. (2012). The *k*-armed dueling bandits problem. *Journal of Computer and System Sciences, 78*(5), 1538–1556.