
Learning a Ranking from Human Feedback in Log-Concave Random Utility Models

Diego Aloviseti
Politecnico di Milano

Marco Mussi
Politecnico di Milano

Alberto Maria Metelli
Politecnico di Milano

Abstract

We study the problem of recovering the ranking of a fixed set of items according to their unknown numerical utilities. At each interaction with the environment, a learner presents the item set to a human and receives comparative feedback of two types. Under *full-ranking* feedback, each interaction reveals a noisy ranking of all items, whereas under *winner-only* feedback, it reveals only the item ranked first. In both settings, we model human feedback using a *random utility model* with log-concave noise and study the number of observations needed to recover an ε -accurate ranking with high probability. This novel criterion tolerates ordering errors only between items whose utilities differ by less than ε . For both feedback types, we establish worst-case sample-complexity lower bounds and develop algorithms that match these bounds up to logarithmic factors. Neither algorithm requires knowledge of the noise distribution, while only requiring an upper bound on its variance. Our results show that the ranking problem under winner-only feedback is intrinsically harder by exposing the sample complexity dependence on the minimum winning probability across the item set.

1 INTRODUCTION

In many machine learning problems, an algorithm deals with a finite set of objects, each associated with a numerical score representing its quality. For example, recommendation systems are concerned with products whose scores reflect users' preferences (Hajek et al., 2014). Similarly, in Reinforcement Learning from Human Feedback (RLHF) (Christiano et al., 2017), an agent can take multiple actions whose scores are reward values. In both cases, although every object is assigned a scalar quantity, the latter typically remains latent.

Indeed, since the algorithm interacts directly with a human evaluator, such as the user in recommendation systems or the problem designer in RLHF, scalar feedback is either unavailable or unreliable. As the psychometric literature suggests, humans struggle to assign numerical scores, while they can reliably provide ordinal feedback. Consequently, the algorithm must address the set of numerical scores while relying on non-scalar feedback signals. In this work, we formalize this to study a ranking-recovery learning problem. We consider a finite set of items associated with numerical scores that we call *utilities* and a learning algorithm that has no prior knowledge about them and aims to recover the ranking induced by the utilities. To do so, the algorithm sequentially interacts with a human evaluator who provides (noisy) feedback in the form of a (partial) ranking of the items. To model human behavior, we use a Random Utility Model (RUM), a standard paradigm for human preferences (Thurstone, 1927), which includes the well-studied Plackett-Luce (PL) model (Plackett, 1975; Luce, 1959). Within this framework, we assume that the latent utilities are perturbed by i.i.d. noise to obtain a random utility vector. Then, a *feedback filter* transforms this vector into the observed feedback signal. We make the additional assumption that the noise distribution is log-concave (e.g., Gaussian, logistic), which is standard in the literature (Shah et al., 2016) and ensures several desirable properties (Critchlow et al., 1991). The feedback filter transforms a vector with scalar entries into a partial ranking. Specifically, we consider two types of filters:

- **Full-Ranking** (FR), where the complete ordering of the random utilities is observed.
- **Winner-Only** (WO), in which the learner observes the items with the highest sampled utility.

We evaluate the ranking recovery problem within the Probably Approximately Correct (PAC) framework (Valiant, 1984), where the approximately correct part is formalized using the novel notion of ε -accuracy. The latter, which measures the distance between the recovered ranking and the target, is fundamentally different from standard discrete distances such as Kendall tau

(Kendall, 1938) or maximum rank (Busa-Fekete et al., 2014b), in that these are insensitive to the underlying utilities. In fact, our metric effectively penalizes misranking two items depending on their utility gap. By considering the proposed ranking problem within the PAC framework, we aim to address the following research question:

What is the minimum number of *ordinal* feedback signals needed for the learner to recover the ranking induced by the items' *scalar* utilities?

By answering this question for two common human-generated feedback types, our goal is to quantify their intrinsic learning difficulties. This will result in a quantitative understanding of the difference between using full-ranking rather than winner-only feedback. We stress that the tension between the learning goal, rooted in scalar utilities, and the ordinal feedback used to reach it is precisely what makes the proposed ranking problem non-trivial. The notion of ε -accuracy is the very element that allows this tension to be resolved. Our proposed solution to the ranking problem is the Variance-aware Utility Ranker (VUR), a unified framework comprising two algorithms for recovering the target ranking under full-ranking and winner-only signals, respectively. In summary, our main contributions are:

- **Novel Distance Metric.** We introduce the notion of ε -accuracy through Definition 1, which generalizes the ε -optimality metric used in best-item identification (Saha and Gopalan, 2020) to learning a full utility ranking.
- **Unified Framework for Diverse Feedback.** We design two PAC algorithms relying on a common design principle, which allows us to view them as part of a single, unified framework for learning the utility ranking across diverse feedback types.
- **Distribution-agnostic Algorithms.** We propose two algorithms requiring no exact knowledge of the noise distribution. Instead, they rely solely on the log-concavity assumption plus a known upper bound on the noise variance.
- **Fundamental Limits of Learnability.** We establish information-theoretic lower bounds for the ranking problem under log-concave RUMs, demonstrating that our algorithms are order-optimal. Crucially, Theorem 4 identifies the minimum winning probability, $P_{\min}$, as the fundamental informational bottleneck of WO feedback.

The related works are discussed in Appendix A.

2 PRELIMINARIES

We use standard big-O notation throughout, with $\tilde{O}$ suppressing logarithmic factors, except for the confidence-dependent term $\log(1/\delta)$. For any $k \in \mathbb{N}$, we denote $[k] = \{1, \dots, k\}$. The set of all permutations on $[k]$, i.e., the bijective functions $\sigma : [k] \rightarrow [k]$, is denoted by $\mathcal{S}_k$. We interpret $\sigma \in \mathcal{S}_k$ as the *ranking* that assigns position $\sigma(i)$ to item i . The notation $(\Omega, \mathcal{F}, \mathbb{P})$ refers to a generic probability space, while $\mathbb{1}_A$ is the indicator function of any event $A \in \mathcal{F}$. The set of all probability measures on $\mathbb{R}$ is denoted by $\mathcal{P}(\mathbb{R})$. If not stated otherwise, the real line is equipped with the Borel sigma-algebra. If $\mu \in \mathcal{P}(\mathbb{R})$ is absolutely continuous with respect to the Lebesgue measure λ , we write $\mu \ll \lambda$. If f is the density of μ , we denote its support by $\text{Supp}(f)$. For a non-decreasing function $g : \mathbb{R} \rightarrow \mathbb{R}$, the *generalized inverse* is defined by:

$$g^{-1}(y) = \sup\{x \in \mathbb{R} : g(x) \leq y\}.$$

When well-defined, the inverse coincides with the generalized inverse; moreover, the generalized inverse is always non-decreasing.

Linear extensions and graphs. A *strict partial order* $<$ on the finite set $\mathcal{S}$ is a homogeneous binary relation that is irreflexive, asymmetric, and transitive. A general result by Szpilrajn (1930) states that, given any strict partial order $<$ on $\mathcal{S}$, there exists a *linear extension* $<^*$, i.e., a total order such that $v <^* w$ whenever $v < w$. A generic directed graph is denoted by $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set and $\mathcal{E}$ is the edge set. A (directed) edge is of the form $v \rightarrow w$, where v and w are distinct vertices. A *tournament* is a directed graph with exactly one directed edge between any pair of distinct vertices. When G is acyclic, its *transitive closure* is the strict partial order “ $<$ ” on the vertex set satisfying: for any two vertices v and w , $v < w$ if and only if there exists a path in G that goes from v to w .

Log-concavity. Given a probability distribution $\nu \in \mathcal{P}(\mathbb{R})$, we say that ν is *log-concave* if:

$$\nu(\lambda A + (1 - \lambda)B) \geq \nu(A)^\lambda \nu(B)^{1-\lambda},$$

whenever the sets A and B , and $\lambda \in [0, 1]$ are chosen so that $\lambda A + (1 - \lambda)B \subseteq \mathbb{R}$ is measurable. A log-concave distribution on $\mathbb{R}$ is either a Dirac measure or absolutely continuous with respect to the Lebesgue measure (Borell, 1975). In this case, its density and cumulative distribution function (CDF) are log-concave (Lovász and Vempala, 2007).

3 PROBLEM FORMULATION

We fix a positive integer $k \geq 2$ and consider a collection of $k \in \mathbb{N}$ *items*, which we identify with the set $[k]$. Each item i is associated with a scalar utility $u_i \in \mathbb{R}$, resulting in the utility vector $\mathbf{u} = (u_1, \dots, u_k) \in \mathbb{R}^k$. The corresponding *utility ranking* is defined as the unique permutation $\sigma_u \in \mathcal{S}_k$ satisfying:¹

$$\sigma_u(i) < \sigma_u(j) \iff u_i > u_j \vee (u_i = u_j \wedge i < j).$$

In this paper, we address the problem of approximating the utility ranking based on noisy *feedback signals* generated by a human evaluator. In our formalization, these arise from a RUM with a centered log-concave noise distribution. To avoid degenerate cases, we assume that the latter is absolutely continuous². Log-concavity is a standard assumption in the literature (Shah et al., 2016), and implies several desirable properties for ranking models (Critchlow et al., 1991). To ensure learnability, we impose a sufficient bounded-variance condition. For a fixed $V > 0$, we let $\mathcal{N}_V$ be the set of all absolutely continuous centered log-concave distributions ν satisfying:

- (V) The variance of distribution ν is bounded above by V .

Intuitively, this can be interpreted as a reliability constraint. While keeping the exact noise distribution unknown, it ensures that the feedback remains a reliable, albeit noisy, reflection of the latent utility relationships. The bound V represents a parameter of the proposed algorithms. For simplicity, we drop the subscript and write $\mathcal{N}$, whose elements are referred to as *noise distributions*.

3.1 Feedback-generation Process

Given $\mathbf{u} \in \mathbb{R}^k$ and $\nu \in \mathcal{N}$, we let $N_1, \dots, N_k$ be i.i.d. random variables drawn from ν , and consider the *random utility vector* $U = (U_1, \dots, U_k)$ given by:

$$U_i = u_i + N_i, \quad \forall i \in [k].$$

For $t \in \mathbb{N}$, we let $U^t = (U_1^t, \dots, U_k^t)$ be an independent copy of U , and denote the stochastic process $(U^t)_{t \in \mathbb{N}}$ by RUM($\mathbf{u}, \nu$), standing for random utility model with utility $\mathbf{u}$ and noise distribution ν . The feedback signal observed at time t is obtained by applying a *feedback filter* to the utility vector U^t . Specifically, we consider two types of filters.

¹If $u_i = u_j$, the tie is broken by ordering the two items according to their indices.

²Because a log-concave distribution is either a Dirac measure or abs. continuous, this amounts to excluding the case in which the noise has a Dirac distribution. This results in no loss of generality as the support of the noise can be taken arbitrarily narrow.

- The **full-ranking** (FR) filter is the function $\varphi_{FR} : \mathbb{R}^k \rightarrow \mathcal{S}_k$ such that $\varphi_{FR}(\mathbf{x})$ is the unique permutation σ defined by: $\sigma(i) < \sigma(j)$ if and only if $x_i > x_j$ or $x_i = x_j$ and $i < j$.
- The **winner-only** (WO) filter is the function $\varphi_{WO} : \mathbb{R}^k \rightarrow [k]$ given by $\varphi_{WO}(\mathbf{x}) := \min\{i \in [k] : x_i \geq x_j \forall j \neq i\}$.³

We assume that the feedback filter is fixed throughout the learning episode. Thus, given a feedback filter $\varphi \in \{\varphi_{FR}, \varphi_{WO}\}$, we let $\varphi_t = \varphi(U^t)$ for every $t \in \mathbb{N}$. The *feedback filtration* is defined as $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{N}}$, where $\mathcal{F}_t$ is the σ -algebra generated by $\varphi_1, \dots, \varphi_t$.

3.2 The PAC Objective

Oftentimes, estimation of the utility ranking comes as a byproduct of utility estimation (Negahban et al., 2012; Azari Soufiani et al., 2014). In contrast, our approach focuses on the ranking itself, thus avoiding the need for explicit utility reconstruction. We achieve this by defining the objective in terms of a novel metric for measuring the distance between two rankings, which effectively relates a misranked pair (i, j) to the latent *utility gap* $\Delta_{ij} := u_i - u_j$.

Definition 1 (ε -accuracy). Let $\mathbf{u} \in \mathbb{R}^k$ with utility ranking σ_u , and let $\varepsilon > 0$. Given $\sigma \in \mathcal{S}_k$, we say that σ is an ε -accurate representation of σ_u if the following holds for any two distinct i and j :

$$(\sigma(i) - \sigma(j))(\sigma_u(i) - \sigma_u(j)) < 0 \implies |\Delta_{i,j}| < \varepsilon. \quad (1)$$

We remark that the left-hand side of the implication is equivalent to saying that the relative position of items i and j in σ is opposite to the one induced by σ_u . In (Saha and Gopalan, 2019), a similar notion is used, where the gap between the PL parameters plays the role of the utility gap in ε -accuracy. In their work, Falahatgar et al. (2018) also propose a metric related to ours; however, their setting is fundamentally different as no utilities are required for the ranking distribution to be defined. We elaborate on these connections in Appendix A.

To define the PAC objective, we first define an *algorithm* as a pair $\mathcal{A} = (T, \hat{\sigma})$, where $T \in \mathbb{N}$ is an almost-surely finite stopping time with respect to the feedback filtration $\mathbb{F}$, and $\hat{\sigma}$ is an $\mathcal{F}_T$ -measurable random permutation of $[k]$.

Definition 2 (PAC objective). Let $\mathcal{A} = (T, \hat{\sigma})$ be an algorithm that learns a ranking based on the sequential observations φ_t generated by a fixed unknown

³The minimum in the definition of φ_{WO} is used for tie-breaking. However, when the noise distribution is absolutely continuous with respect to the Lebesgue measure, ties are a zero-probability event.

RUM($\mathbf{u}, \nu$) with feedback filter φ . Given $\delta \in (0, 1)$ and $\varepsilon > 0$, we say that $\mathcal{A}$ is (ε, δ) -PAC if:

$$\mathbb{P}[\hat{\sigma} \text{ is an } \varepsilon\text{-accurate representation of } \sigma_{\mathbf{u}}] \geq 1 - \delta.$$

If $\mathcal{A}$ is an (ε, δ) -PAC algorithm, we define its *sample complexity* as the expected value $\mathbb{E}[T]$, where the expectation is taken with respect to the probability measure induced by RUM($\mathbf{u}, \nu$).

4 FULL-RANKING FEEDBACK

In this section, we study the ranking problem under full-ranking feedback ($\varphi = \varphi_{\text{FR}}$). Specifically, we introduce an (ε, δ) -PAC algorithm named Full-Ranking Variance-aware Utility Ranker (FR-VUR) and derive an upper bound on its sample complexity. Subsequently, we provide an information-theoretic lower bound, showing that FR-VUR achieves an optimal rate, up to logarithmic dependence. The feedback-generating random utility model RUM($\mathbf{u}, \nu$) is fixed throughout, where the noise distribution ν satisfies Assumption (V).

4.1 The Pairwise-Preference Probabilities

Under FR feedback, every observation can be converted into $\binom{k}{2}$ pairwise-comparison outcomes by *rank-breaking* (Azari Soufiani et al., 2014). Our algorithm will do so, using these values to estimate, for any two distinct items i and j , the *pairwise-preference probability*:

$$P_{ij} := \mathbb{P}[U_i > U_j].$$

The mutual independence of the noise random variables implies that $P_{ij} > 1/2$ if and only if $u_i > u_j$ (see Lemma 5 in Appendix C). For every $t \in \mathbb{N}$, we define the event $E_{ij}^t := \{\varphi_t(i) < \varphi_t(j)\}$, the set of outcomes for which item i is ranked before item j at time t . Then, for every integer $n \geq 2$, we consider the unbiased estimator of P_{ij} given by:

$$P_{ij}^n := \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{E_{ij}^t}.$$

The (unbiased) *sample variance* is the quantity:

$$V_{ij}^n := \frac{1}{n(n-1)} \sum_{1 \leq t_1 < t_2 \leq n} (\mathbb{1}_{E_{ij}^{t_1}} - \mathbb{1}_{E_{ij}^{t_2}})^2.$$

For $\delta \in (0, 1)$ and $n \in \mathbb{N}$, we define the *full-ranking time- n confidence* as $\delta_n := 6\delta/k(k-1)n^2\pi^2$. Following (Maurer and Pontil, 2009), we derive concentration bounds for P_{ij}^n and V_{ij}^n in Appendix C.2 (see Lemma 6). These include a standard concentration inequality for the sample variance and an empirical Bernstein-type inequality for P_{ij}^n .

4.2 The Full-Ranking Resolution Factor

A priori, it is not clear whether the pairwise-preference probabilities are the right estimation target to achieve ε -accuracy. Below, we show that this is indeed the case by formalizing the intuition that, when $u_i - u_j \geq \varepsilon$, the quantity P_{ij} is large enough to reveal the true preference relation between i and j . Consequently, by accurately approximating P_{ij} , the sign of $u_i - u_j$ can be inferred. To make this precise, we derive a mapping between utility gaps and pairwise-preference probabilities. Let $i, j \in [k]$ be two items with $u_i > u_j$, and let $\xi : [0, \infty) \rightarrow [1/2, 1]$ be the function given by:

$$\xi(t) := \int_{\mathbb{R}} F(x+t)f(x) dx, \quad (2)$$

where F and f are the CDF and the density of ν , respectively. Then, we write:

$$P_{ij} = \xi(u_i - u_j) = \frac{\eta(u_i - u_j)}{2} + \frac{1}{2}, \quad (3)$$

where we defined the *full-ranking resolution factor* as $\eta(\varepsilon) := 2\xi(\varepsilon) - 1$, for $\varepsilon > 0$. Intuitively, as Equation (3) suggests, this quantifies how *strongly* an item i with $u_i = u_j + \varepsilon$ is preferable to item j . In Appendix C.3, we show that $\eta(\varepsilon) \geq \eta^*(\varepsilon)$ for every $\varepsilon > 0$, where:

$$\eta^*(\varepsilon) := \min \left\{ \frac{2}{3}, \frac{\varepsilon}{\sqrt{6V}} \right\}, \quad (4)$$

This fact is crucial to guarantee that the FR-VUR algorithm is agnostic to the noise distribution.

4.3 The FR-VUR Algorithm

We fix $\delta \in (0, 1)$ and let δ_n be the time- n confidence defined in the previous section. For a fixed accuracy $\varepsilon > 0$, we write η^* to denote the lower bound $\eta^*(\varepsilon)$ given by Equation (4). To formally describe FR-VUR and show that it is (ε, δ) -PAC, we will refer to it as $\mathcal{A} = (T, \hat{\sigma})$. Pseudocode is given in Appendix B.2.

Stopping rule. The stopping time T is defined as the minimum positive integer $n \in \mathbb{N}$ such that the following holds for any two distinct $i, j \in [k]$:

$$\sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} < \frac{\eta^*}{4}. \quad (5)$$

To compute the right-hand side, $\mathcal{A}$ uses η^* , so it needs to know the variance upper bound V . However, it does not require exact knowledge of the noise distribution.

Using the concentration properties P_{ij}^n and V_{ij}^n , the stopping condition ensures that, with high probability:

$$|P_{ij}^T - P_{ij}| < \eta^*/4.$$

We can think of it as a *multiplicative* requirement⁴ by observing that $\max(P_{ij}^T, P_{ji}^T) \geq 1/2$. This implies:

$$|P_{ij}^T - P_{ji}^T| = |P_{ji}^T - P_{ij}^T| \leq \max(P_{ij}^T, P_{ji}^T) \cdot \frac{\eta^*}{2}.$$

To define the output of $\mathcal{A}$, for every $n \in \mathbb{N}$, we define the *pairwise-preference graph* $G_n = ([k], \mathcal{E}_n)$ as the directed graph where:

$$i \longrightarrow j \in \mathcal{E}_n \iff P_{ij}^n > \frac{1}{2} + \frac{\eta^*}{4}.$$

Additionally, we let the *utility graph* be transitive tournament $G_u = ([k], \mathcal{E}_u)$ whose edge set is characterized as follows: $i \longrightarrow j \in \mathcal{E}_u$ if and only if either $u_i > u_j$, or $u_i = u_j$ and $i < j$.

Output recommendation. The $\mathcal{F}_T$ -measurable random ranking $\hat{\sigma}$ is determined as follows:

- If G_T contains a cycle, choose $\hat{\sigma} \in \mathcal{S}_k$ arbitrarily.
- If G_T is a DAG, let G^* be its transitive closure, corresponding to a strict partial order of its vertices. Then, let $\hat{\sigma} \in \mathcal{S}_k$ be any linear extension of this partial order.

In particular, when G_T contains no cycle, $i \longrightarrow j \in \mathcal{E}_T$ implies $\hat{\sigma}(i) < \hat{\sigma}(j)$.

We show that $\mathcal{A} = (T, \hat{\sigma})$ satisfies the PAC property given by Definition 2. To this end, we first observe that, with probability at least $1 - \delta$, G_T is a subgraph of G_u . This follows from the concentration properties of P_{ij}^n and V_{ij}^n and is proven by Lemma 8 in Appendix C.4. Second, we note that the ε -accuracy target is met if, for any two items i and j with $u_i - u_j \geq \varepsilon$, we have $\hat{\sigma}(i) < \hat{\sigma}(j)$. By definition, this is the case if $i \longrightarrow j$ is an edge of G_T . As demonstrated by Lemma 9 in Appendix C.4, this occurs with probability at least $1 - \delta$, thus proving that $\mathcal{A}$ is (ε, δ) -PAC.

The next result provides an upper bound on the sample complexity of the proposed algorithm. The full derivation is provided in Appendix C.5.

Theorem 1 (Sample complexity upper bound for FR-VUR). *Let $\mathbf{u} \in \mathbb{R}^k$ and $\nu \in \mathcal{N}$ be such that $\text{RUM}(\mathbf{u}, \nu)$ satisfies Assumption (V), and let $\mathcal{A}$ be the FR-VUR algorithm. Define $Q := \max_{i \neq j} P_{ij}(1 - P_{ij})$. Then, for every $\delta \in (0, 1)$ and $\varepsilon > 0$, $\mathcal{A}$ is (ε, δ) -PAC, with sample complexity:*

$$\mathbb{E}[T] = \mathcal{O}\left(\frac{Q}{\eta^{*2}} \log\left(\frac{Qk}{\eta^{*2}\delta}\right) + \frac{1}{\eta^*} \log\left(\frac{k}{\eta^*\delta}\right)\right). \quad (6)$$

When Q is non-zero, the upper bound is $\tilde{\mathcal{O}}(Q/\eta^{*2})$. The term Q follows from the empirical Bernstein-type

⁴We use the term *multiplicative* bound to denote the fact that the bound itself depends on the quantity to estimate.

stopping rule, which adapts to the intrinsic variance of the pairwise comparisons and accelerates convergence when outcomes are highly skewed. To illustrate this advantage, consider the extreme case with distinct utilities and a vanishing noise variance, implying $Q = 0$. In this scenario, the sample complexity reduces to $\tilde{\mathcal{O}}(1/\eta^*)$. As V goes to zero, thus matching the true variance, η^* saturates at $2/3$, and this residual term reduces to the purely logarithmic cost $\tilde{\mathcal{O}}(\log(k/\delta))$. In contrast, a Hoeffding-based approach would rely on the global variance upper bound $Q \leq 1/4$. For $\sqrt{V} \geq 3\varepsilon/(2\sqrt{6})$, so that $\eta^* = \varepsilon/(2\sqrt{6V})$, this would result in the worst-case sample complexity:

$$\mathbb{E}[T] = \mathcal{O}\left(\frac{V}{\varepsilon^2} \log\left(\frac{V}{\varepsilon^2} \cdot \frac{k}{\delta}\right)\right). \quad (7)$$

For a fixed $V > 0$, the right-hand side becomes $\tilde{\mathcal{O}}(V/\varepsilon^2)$, where the term $1/\varepsilon^2$ reflects the ε -accuracy objective. The V scaling captures the intuition that, if the noise variance increases, the observed rankings become less reliable. As we will establish in Theorem 2, the V/ε^2 dependence is information-theoretically optimal.

4.4 Information-theoretic Lower Bound

In this part, we demonstrate an information-theoretic worst-case lower bound matching the upper bound given in Equation (7). Precisely, we address the *small-gap regime* $\varepsilon < \sqrt{6V}/(2\pi)$, which is the most interesting case.⁵ To do so, we identify a class of “hard” problem instances, defined by a Gumbel noise paired with a utility vector where one utility gap is exactly ε . Our proof, provided in Appendix C.6, employs a change-of-measure argument following (Kaufmann et al., 2016). Below, $\mathcal{A}^* = (T^*, \hat{\sigma}^*)$ is any ranking algorithm based on FR feedback signals. We assume $\mathcal{A}^*$ is (ε, δ) -PAC for any $\delta \in (0, 1)$.

Theorem 2 (Lower bound for full-ranking feedback). *Let $\delta \in (0, 1/4]$. Then, there exists a noise-utility pair $(\nu, \mathbf{u})$ satisfying Assumption (V) such that the sample complexity of algorithm $\mathcal{A}^* = (T^*, \hat{\sigma}^*)$ satisfies:*

$$\mathbb{E}[T^*] = \Omega\left(\frac{V}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right),$$

where the expectation is taken with respect to the probability measure induced by $\text{RUM}(\mathbf{u}, \nu)$.

This result establishes the V/ε^2 scaling as a fundamental learning cost of the ranking problem. In particular,

⁵The threshold $\sqrt{6V}/(2\pi) \approx 0.39\sqrt{V}$ is the one under which the Gumbel instances used in the proof satisfy $\varepsilon/\beta < 1/2$, where β is the scale parameter. Its precise value is immaterial: our results are asymptotic, thus justifying the restriction to the small-gap regime.

Theorem 2 shows that the exact dependence on the noise variance is linear.

5 WINNER-ONLY FEEDBACK

We now study the ranking recovery problem under WO feedback, meaning $\varphi = \varphi_{WO}$. For this setting, we design and analyze yet another learning algorithm that differs from FR-VUR in that it estimates winning probabilities rather than pairwise-preference probabilities. Nonetheless, both algorithms rely on a (feedback-specific) resolution factor to relate the estimated probabilities to underlying utility gaps. To emphasize this continuity, we overwrite the symbols η , η^* , and $\mathcal{A}$. As in the full-ranking case, U^t is the random utility vector drawn at time t from a fixed RUM($\mathbf{u}, \nu$), where ν satisfies Assumption (V). Additionally, we let $k \geq 3$, as the winner-only setting reduces to full-ranking when only two items are available.

5.1 Observability Assumption

A full-ranking signal provides information about the relative position of any two items, which was exploited by FR-VUR through rank-breaking. In contrast, under winner-only feedback, information about the relative position of two items is obtained only if one of them wins. Consequently, to correctly rank these, the learner must rely on their winning frequency. Theorem 4 below quantifies this by providing an information-theoretic lower bound, demonstrating that the sample complexity of any PAC algorithm is inversely proportional to the minimum winning probability. Therefore, to avoid ending in a non-learnable regime, we enforce an additional assumption. To this end, we begin by defining the *utility range*:

$$\Delta_u := \max_{i \in [k]} u_i - \min_{j \in [k]} u_j.$$

Then, denoting by f the density of ν , we let $a := \inf \text{Supp}(f)$ and $b := \sup \text{Supp}(f)$, and call $b - a$ the *width* of the noise support. To ensure that every item have a non-zero winning probability, we impose the following condition:

(WO1) The utility range is smaller than the width of the support of ν , i.e., $\Delta_u < b - a$.

Under this assumption, $P_{\min} := \min_{i \in [k]} P_i > 0$, meaning that, as the episode horizon T grows to infinity, every item will eventually win in at least one round with non-zero probability. With this in mind, we can interpret (WO1) as an *observability assumption*. We anticipate that, while taking the variance upper bound V as input, our algorithm does not require knowledge of either a or b .

5.2 The Winner-Only Resolution Factor

The probability that an item is the winner plays a fundamental role in this setting. Formally, for every $i \in [k]$, we define its *winning probability* as:

$$P_i := \mathbb{P}[U_i > U_j, \forall j \in [k] \setminus \{i\}]. \quad (8)$$

For any two items $i, j \in [k]$ with $u_i \geq u_j$, their winning probabilities are *consistent*, in the sense that $P_i > P_j$ if and only if $u_i > u_j$ (see Lemma 15 in Appendix D.3). Since the winning probabilities P_i and P_j depend on the magnitude of *all* utility gaps, it is unclear whether they can be used to infer the magnitude of Δ_{ij} . Remarkably, this turns out to be the case in the (generalized) PL model, a RUM with Gumbel noise with scale parameter $s > 0$, where:⁶

$$u_i - u_j = s \log \left(\frac{P_i}{P_j} \right). \quad (9)$$

Saha and Gopalan (2019) exploited this and used estimates of the winning-probability ratios to approximate the ranking induced by the PL parameters. In this work, we generalize their approach to a broader class of RUMs, while addressing the latent utilities instead of the PL parameters. We do so by establishing a functional relation between winning-probability ratios and utility gaps under the sufficient condition of log-concavity. To illustrate this, we consider i and j such that $u_i > u_j$, and observe that:

$$P_i - P_j \geq P_i \cdot \frac{1 - e^{-(u_i - u_j)/\sqrt{3V}}}{2}. \quad (10)$$

This fact is proven in Lemma 13 in Appendix D.2 using log-concavity. Now we define the function $(1, 2)$ -valued function ξ by setting:

$$\xi(t) := \frac{2}{1 + e^{-t/\sqrt{3V}}}, \quad \forall t \in (0, b - a). \quad (11)$$

Assumption (WO1) ensures $P_j > 0$, so we can rearrange the terms of Equation (10) to obtain $P_i/P_j \geq \xi(u_i - u_j)$. Since ξ is strictly increasing, taking its generalized inverse yields:

$$u_i - u_j \leq \xi^{-1} \left(\frac{P_i}{P_j} \right). \quad (12)$$

Following the approach used in Section 4, we define the *winner-only resolution factor* $\eta : (0, b - a) \rightarrow (0, 1/3)$ as:

$$\eta(t) := \frac{\xi(t) - 1}{\xi(t) + 1} = \frac{1 - e^{-t/\sqrt{3V}}}{3 + e^{-t/\sqrt{3V}}}. \quad (13)$$

⁶This property is typically referred to as Independence of Irrelevant Alternatives (IIA) (Ray, 1973).

The latter is defined using the known variance bound V and is independent of the particular noise distribution in $\mathcal{N}_V$. For a simpler expression with the same scaling in the $\varepsilon \approx 0$ regime, we let:

$$\eta^*(\varepsilon) := \frac{e-1}{3e+1} \min \left\{ 1, \frac{\varepsilon}{\sqrt{3V}} \right\}. \quad (14)$$

Then, $\eta(\varepsilon) \geq \eta^*(\varepsilon)$ for every $\varepsilon \in (0, b-a)$, which is proven using standard analytical tools in Lemma 14 in Appendix D.2.

5.3 The Winning-Probability Estimators

For each item $i \in [k]$ and every time index $t \in \mathbb{N}$, we define the event $E_i^t := \{\varphi_t = i\}$, consisting of all outcomes for which i is the winner at time t . Then, for $n \in \mathbb{N}$, we let:

$$P_i^n := \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{E_i^t},$$

which is an unbiased estimator for the winning probability P_i . For $n \geq 2$, we define the (unbiased) sample variance as:

$$V_i^n := \frac{1}{n(n-1)} \sum_{1 \leq t_1 < t_2 \leq n} (\mathbb{1}_{E_i^{t_1}} - \mathbb{1}_{E_i^{t_2}})^2.$$

Given a fixed $\delta \in (0, 1)$, we tune the *winner-only time- n confidence* to be $\delta_n := 3\delta/\pi^2 n^2 k$, for all $n \in \mathbb{N}$. Concentration inequalities for these estimators are given by Lemma 11 in Appendix D.1 based on (Maurer and Pontil, 2009).

5.4 The WO-VUR Algorithm

Given $\delta \in (0, 1)$ and $\varepsilon \in (0, b-a)$ we let δ_n be the winner-only time- n confidence defined in Section 5.3, while we write η^* for the lower bound $\eta^*(\varepsilon)$ given by Equation (14). Below, we introduce the (ε, δ) -PAC algorithm WO-VUR $\mathcal{A} = (T, \hat{\sigma})$, where the notation overload emphasizes the continuity with the full-ranking setting. This algorithm uses empirical winning frequencies to deliver its final recommendation. The resulting sample-complexity guarantee holds for any noise-utility pair $(\nu, \mathbf{u})$ such that $\nu \in \mathcal{N}_V$ and assumption (WO1) holds. Knowledge of V is required, although the algorithm does not require the noise density or its support. Pseudocode is given in Appendix B.2.

Stopping rule. The algorithm stops at time T , which is defined as the minimum integer $n \geq 2$ such that the following holds for every $i \in [k]$:

$$\sqrt{\frac{2V_i^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} < \frac{\eta^*}{2} \cdot P_i^n. \quad (15)$$

Given the estimates $P_1^n, \dots, P_k^n$ at time n , we let σ_n be the unique permutation of $[k]$ satisfying:

$$\sigma_n(i) < \sigma_n(j) \iff P_i^n > P_j^n \vee (P_i^n = P_j^n \text{ and } i < j).$$

Output recommendation. The recommendation is defined as the $\mathcal{F}_T$ -measurable permutation $\hat{\sigma} := \sigma_T$.

To show that $\mathcal{A}$ has the PAC property given by Definition 2, we suppose that, for every $i \in [k]$:

$$|P_i^T - P_i| \leq \sqrt{\frac{2V_i^T \log(4/\delta_T)}{T}} + \frac{7 \log(4/\delta_T)}{3(T-1)}.$$

As shown by Lemma 11 in Appendix D.1, this is true with probability at least $1 - \delta$. Using the definition of T and the fact that $\eta^* < 1$, the following holds for every item i :

$$|P_i^T - P_i| < P_i^T / 2.$$

This highlights the *multiplicative* nature of the bound. Now, suppose that i and j are misranked at time T , meaning $(P_i - P_j)(P_i^T - P_j^T) \leq 0$. Without loss of generality, we may assume $u_i > u_j$, thus:

$$0 < P_i - P_j < \frac{\eta^*}{2} (P_i^T + P_j^T) \leq \eta^* \cdot (P_i + P_j).$$

Rearranging these terms yields:

$$\frac{P_i}{P_j} < \frac{1 + \eta^*}{1 - \eta^*} \leq \frac{1 + \eta(\varepsilon)}{1 - \eta(\varepsilon)} = \xi(\varepsilon),$$

where $\eta(\varepsilon)$ is the feedback-specific resolution factor defined by Equation (13). Our earlier calculations give $\xi(u_i - u_j) \leq P_i/P_j$, thus implying $\xi(u_i - u_j) < \xi(\varepsilon)$. Since ξ is an increasing function, we conclude that $u_i - u_j < \varepsilon$. In other words, we have shown that: with probability at least $1 - \delta$, if i and j are misranked, then their absolute utility gap is bounded by ε . Hence, algorithm $\mathcal{A}$ is (ε, δ) -PAC.

5.4.1 Sample Complexity Upper Bound

The next result derives an upper bound on the sample complexity of the WO-VUR algorithm. Formal derivations are given in Appendix D.4.

Theorem 3 (Sample complexity upper bound for WO-VUR). *Let $\mathbf{u} \in \mathbb{R}^k$ and $\nu \in \mathcal{N}_V$ be a noise-utility pair satisfying Assumption (WO1), and let $\mathcal{A} = (T, \hat{\sigma})$ be the WO-VUR algorithm. Then, for every $\delta \in (0, 1)$ and $\varepsilon \in (0, b-a)$, $\mathcal{A}$ is (ε, δ) -PAC with sample complexity:*

$$\mathbb{E}[T] = \mathcal{O} \left(\frac{1}{P_{\min} \eta^{*2}} \log \left(\frac{1}{P_{\min} \eta^{*2}} \cdot \frac{k}{\delta} \right) \right). \quad (16)$$

Using the explicit formula (14), in the small-gap regime $\varepsilon \leq \sqrt{3V}$, the sample-complexity upper bound takes the form:

$$\mathbb{E}[T] = \mathcal{O}\left(\frac{V}{P_{\min} \varepsilon^2} \log\left(\frac{V}{P_{\min} \varepsilon^2} \cdot \frac{k}{\delta}\right)\right). \quad (17)$$

As we highlighted through our informal discussion, the term $P_{\min}$ exposes a fundamental learning limit. Since the relative position of i and j is estimated based on their empirical winning frequencies, if these items rarely win, reliable estimation will require a large number of samples. Crucially, the (inverse) linear dependence on $P_{\min}$ is a consequence of the Bernstein-type stopping rule, whereas using a Hoeffding-based approach would have resulted in a suboptimal quadratic dependence. We make this explicit through Equation (41) in Appendix D.4. The $1/\varepsilon^2$ scaling given in the upper bound reflects the ε -accuracy target and mirrors full-ranking case. Similarly, the V term confirms the linear dependence on the reliability of the feedback signals. Finally, the quantity $\log(k)$ results in a cost of lower order. As we shall see, Theorem 4 establishes the $V/(\varepsilon^2 P_{\min})$ scaling as information-theoretically optimal.

5.5 Information-theoretic Lower Bound

The subsequent lower-bound analysis ultimately demonstrates that learning becomes unattainable when $P_{\min}$ goes to zero. We specifically analyze sample complexity over all RUMs with minimum winning probability $P_{\min} \in (0, 1/k)$, excluding the case $P_{\min} = 1/k$, which corresponds to the trivial instance where all utilities are equal. To formalize our result, we let $\mathcal{A}^* = (T^*, \hat{\sigma}^*)$ be a PAC algorithm that outputs a ranking based on winner-only feedback. Full derivation of the lower bound is provided in Appendix D.5.

Theorem 4 (Lower bound for winner-only feedback). *Let $k \geq 3$, $P_{\min} \in (0, 1/k)$, and $\delta \in (0, 1/4]$ be fixed. Then, there exists a noise-utility pair $(\nu, \mathbf{u})$ satisfying Assumptions (V) and (WO1) such that the sample complexity of algorithm $\mathcal{A}^* = (T^*, \hat{\sigma}^*)$ satisfies:*

$$\mathbb{E}[T^*] = \Omega\left(\frac{V}{P_{\min} \varepsilon^2} \log\left(\frac{1}{\delta}\right)\right),$$

where the expectation is taken with respect to the probability measure induced by $\text{RUM}(\nu, \mathbf{u})$.

We begin by observing that our lower bound contrasts with the $\tilde{\mathcal{O}}(k/\varepsilon^2)$ rate claimed by Saha and Gopalan (2019) under the PL model. The discrepancy is due to a fundamental objective misalignment. Indeed, their framework seeks an ε -optimal ranking, where misclassification is penalized based on the difference between the PL parameters e^{u_ℓ} . If i and j have utilities $u_i, u_j \ll 0$, then $e^{u_i}, e^{u_j} \approx 0$. Consequently, under the ε -optimality

metric, misranking these two is tolerable. Therefore, by choosing to use this metric, the true informational bottleneck of $P_{\min}$ established in Theorem 4 is deliberately circumvented. Hence, the claim of full ranking recovery follows without actually resolving the relative order of items with a low winning probability. In contrast, our analysis effectively exposes the difficulties of learning the utility ranking under winner-only feedback, while treating every item pair as equally important. To do so, we chose the ε -accuracy metric to depend on the underlying utilities rather than the PL parameters.

6 OPEN CHALLENGES

We characterized the learning difficulty of the utility-ranking recovery problem under two types of RUM-generated feedback. However, full-ranking and winner-only are the extremes of a wide spectrum. Therefore, understanding how the learning task changes across the whole range of feedback types in between (e.g., top- m (Hajek et al., 2014)) is an open question. Second, we relied on log-concavity to derive the utility-to-probability mappings used by both our VUR algorithms. Without this assumption, the feedback-specific resolution factors can no longer be defined. Understanding whether the same approach can be employed under a milder condition on the noise distribution is not understood yet. Third, a more complex interaction protocol may be considered. In our work, we focused on a setting where the feedback signal regards all items at every learner-environment interaction because our goal was to quantify the learning difficulty due to using winner-only vs. full-ranking feedback. However, allowing for subset selection (Saha and Gopalan, 2019) would accommodate more realistic types of interaction. Finally, our ε -accuracy metric effectively links ranking information to the utilities of the ranked items. We believe that in this lies a potential tool for addressing the regret minimization problem in the context of multi-armed bandits with ranking feedback.

References

- Hossein Azari Soufiani, William Z. Chen, David C. Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. In *Advances in Neural Information Processing Systems (NIPS)*, volume 26, pages 2706–2714, 2013.
- Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Computing parametric ranking models via rank-breaking. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 32 of *Proceedings of Machine Learning Research*, pages 360–368. PMLR, 2014.
- Mark Bagnoli and Ted Bergstrom. Log-concave proba-

- bility and its applications. *Economic theory*, 26(2): 445–469, 2005.
- Linus Baltrunas, Tadas Makcinskas, and Francesco Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the ACM Conference on Recommender Systems*, pages 119–126, 2010.
- Sergey Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics, and kantorovich transport distances. *Memoirs of the American Mathematical Society*, 261(1259), 2019.
- Christer Borell. Convex set functions in d-space. *Periodica Mathematica Hungarica*, 6(2):111–136, 1975.
- Róbert Busa-Fekete, Eyke Hüllermeier, and Balázs Szörényi. Preference-based rank elicitation using statistical models: The case of Mallows. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 32 of *Proceedings of Machine Learning Research*, pages 1071–1079. PMLR, 2014a.
- Róbert Busa-Fekete, Balázs Szörényi, and Eyke Hüllermeier. PAC rank elicitation through adaptive sampling of stochastic pairwise preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pages 1701–1707, 2014b.
- Yuxin Chen and Changho Suh. Spectral MLE: Top-K rank aggregation from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 37 of *Proceedings of Machine Learning Research*, pages 371–380. PMLR, 2015.
- Paul F. Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems (NIPS)*, volume 30, pages 4299–4307, 2017.
- Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3):294–318, 1991.
- Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. The limits of maxing, ranking, and preference learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 80 of *Proceedings of Machine Learning Research*, pages 1427–1436. PMLR, 2018.
- Bruce Hajek, Sewoong Oh, and Jiaming Xu. Minimax-optimal inference from partial rankings. In *Advances in Neural Information Processing Systems (NIPS)*, volume 27, pages 1475–1483, 2014.
- Roxana A. Ion, Chris A. J. Klaassen, and Edwin R. van den Heuvel. Sharp inequalities of Bienaymé-Chebyshev and Gauss type for possibly asymmetric intervals around the mean. *arXiv preprint arXiv:2208.08813*, 2022.
- Anders Jonsson, Emilie Kaufmann, Pierre Ménard, Omar Darwiche Domingues, Edouard Leurent, and Michal Valko. Planning in Markov decision processes with gap-dependent sample complexity. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 1253–1263, 2020.
- Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.
- Maurice G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.
- Ashish Khetan and Sewoong Oh. Data-driven rank breaking for efficient rank aggregation. *Journal of Machine Learning Research*, 17(193):1–54, 2016.
- László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.
- Tyler Lu and Craig Boutilier. Learning Mallows models with pairwise preferences. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 145–152. Omnipress, 2011.
- R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley New York, 1959.
- Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample-variance penalization. In *Annual Conference on Learning Theory (COLT)*, 2009.
- Daniel McFadden. Econometric models for probabilistic choice among products. *Journal of Business*, 53(3): S13–S29, 1980.
- Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems (NIPS)*, volume 25, pages 2474–2482, 2012.
- Robin L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, 1975.
- Paramesh Ray. Independence of irrelevant alternatives. *Econometrica*, 41(5):987–991, 1973.
- Aadirupa Saha and Aditya Gopalan. Active ranking with subset-wise preferences. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *Proceedings of Machine Learning Research*, pages 3312–3321. PMLR, 2019.
- Aadirupa Saha and Aditya Gopalan. Best-item learning in random utility models with subset choices. In

Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), volume 108 of *Proceedings of Machine Learning Research*, pages 4281–4291. PMLR, 2020.

Nihar B. Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin J. Wainwright. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *Journal of Machine Learning Research*, 17(58):1–47, 2016.

Edward Szpilrajn. Sur l’extension de l’ordre partiel. *Fundamenta Mathematicae*, 16(1):386–389, 1930.

Louis L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927.

Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Christian Wirth, Riad Akrou, Gerhard Neumann, and Johannes Fürnkranz. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18(136):1–46, 2017.

Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Masrour Zoghi, Shimon Whiteson, Remi Munos, and Maarten de Rijke. Relative upper confidence bound for the k-armed dueling bandit problem. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 32 of *Proceedings of Machine Learning Research*, pages 10–18. PMLR, 2014.

A RELATED WORKS

Our work relates to the existing literature along three primary dimensions: the way of generating feedback, the definition of the objectives and metrics, and the types of feedback used to achieve the target.

A standard paradigm in preference learning is inferring latent properties from human judgments, such as distinguishing “good” from “bad” trajectories in RLHF (Wirth et al., 2017; Christiano et al., 2017) or predicting user choices in recommender systems (McFadden, 1980; Baltrunas et al., 2010). While the PL model (Plackett, 1975; Luce, 1959) is the ubiquitous choice for these tasks, it is a specific instance of the broader Random Utility Model (RUM) framework (Thurstone, 1927). Our work adopts the RUM framework with log-concave noise, which provides a unifying lens for these models. Several works implicitly utilize log-concavity by considering the PL model (Saha and Gopalan, 2019; Chen and Suh, 2015; Negahban et al., 2012; Khetan and Oh, 2016; Azari Soufiani et al., 2014), while others do this explicitly (Hajek et al., 2014). Remarkably, log-concavity of the noise ensures several desirable properties of the implied ranking distribution (Critchlow et al., 1991). In this paper, we exploit this regularity to derive uniform lower bounds on the resolution factors.

The definition of “preference recovery” varies significantly across the literature. Multi-armed and dueling bandit frameworks often focus on identifying a single best item (Saha and Gopalan, 2020; Yue et al., 2012; Zoghi et al., 2014), a paradigm known as best-item identification. Many approaches focus on statistical parameter estimation, with particular attention devoted to the estimation of PL parameters, typically measured by the L_2 -norm (Azari Soufiani et al., 2013; Hajek et al., 2014; Shah et al., 2016). When the goal is a ranking, metrics like Kendall-tau distance (Kendall, 1938) or maximum rank difference (Busa-Fekete et al., 2014b) are standard. When the utility ranking is the target, it is possible to bypass utilities entirely by assuming distributions over permutations, such as the Mallows model (Lu and Boutilier, 2011; Busa-Fekete et al., 2014a). In contrast to these approaches, we operate within the PAC framework (Valiant, 1984), defining success through the notion of ε -accuracy. This objective bridges the gap between pure parameter estimation and exact ordinal recovery, by incorporating utility information into the target ranking in an intuitive manner.

Finally, we distinguish works by the density of the feedback signal. In line with the present paper, some works consider winner-only feedback signals (Saha and Gopalan, 2019, 2020), which are generalized through the notion of top- m feedback. For full-ranking feedback, a common technique is rank-breaking, which decomposes permutations into pairwise comparisons modeled via a comparison graph (Azari Soufiani et al., 2014; Khetan and Oh, 2016; Negahban et al., 2012).

Distance Metrics for Ranking Accuracy. There exist several ways of measuring the distance between two rankings, including the popular Kendall tau distance and the maximum rank difference (Kendall, 1938; Busa-Fekete et al., 2014b). These do not require the ranked items to be associated with utility scores and penalize misclassification uniformly across all item pairs, thus making the comparison with our notion of ε -accuracy vacuous. In contrast, we can provide a more meaningful comparison for the following measures.

- *ε -optimality.* This was introduced by Saha and Gopalan (2020) and is similar to the one given by Definition 1. Indeed, an item is ε -optimal if its utility value is smaller than the largest utility by less than ε . However, in contrast to our metric, this notion of distance induces a much simpler objective as it is only concerned with determining a nearly optimal item.
- *ε -best ranking.* This notion was introduced by Saha and Gopalan (2019) and is intrinsically gap-dependent. However, despite its similarities with our notion of ε -accuracy, it does not refer to utility gaps. Instead, it is concerned with the gaps between the Plackett-Luce parameters. The latter correspond, up to constants, to the winning probabilities and can be mapped to the utility values through a non-linear function. Consequently, the objective implied by ε -optimality is essentially different from ours. In particular, the former is much more tolerant towards items with a small utility because their winning probabilities are small. Optimizing with respect to this metric obscures the true difficulty of the WO feedback, that is, its sensitivity to $P_{\min}$.
- *ε -ordering.* This notion was introduced by Falahatgar et al. (2018). For a fixed noise distribution, using the full-ranking resolution factor η introduced in Section 4, one can show that:

$$\varepsilon - \text{accurate (ours)} \iff \frac{\eta(\varepsilon)}{2} - \text{ranking (Falahatgar et al. (2018))}.$$

This equivalence can only be derived if we choose a ranking model induced by a RUM (thus diverging from the setting of (Falahatgar et al., 2018)) and use the true resolution factor η instead of the lower bound η^* given by Equation 4. Most importantly, the notion of ε -ordering was given in the context of ranking models where there are no underlying utilities that generate the ranking.

Comparison with Saha and Gopalan (2020). Focusing on winner-only feedback, we elaborate on the differences and similarities between our work and that of Saha and Gopalan (2020). While both frameworks seek to lower-bound the winning-probability ratio P_i/P_j in terms of the utility gap $u_i - u_j$, the methodologies and objectives differ in two key aspects. First, Saha and Gopalan (2020) focus on identifying the ε -best item, whereas we target the recovery of the full utility ranking. Under WO feedback, the latter becomes a significantly more challenging problem. This is due to the informational bottleneck represented by the minimum winning probability $P_{\min}$, as established by our lower bound analysis formalized through Theorem 4. Second, their approach uses the so-called minimum advantage ratio and relies on a distribution-specific constant c , which is derived via a so-called “variational lower bound”. This constant must be computed for each noise model and provided to the algorithm as a parameter. In contrast, our approach leverages the structural regularity of log-concave distributions. Consequently, our algorithms only require an upper bound on the noise variance, regardless of any other distribution-specific parameters.

B VARIANCE-AWARE UTILITY RANKERS

B.1 Distribution Table

The table below lists several examples of log-concave distributions and, for each of them, provides explicit conditions ensuring that the mean is zero and Assumption (V) is satisfied. If these two are met, the given distribution is allowed by the FR-VUR algorithm presented in Section 4. Provided that the additional utility-range-dependent condition (WO1) is also met, the derived sample-complexity guarantees hold for both FR-VUR and WO-VUR.

Noise distribution	Centeredness	Bounded-variance (V)	FR-VUR	WO-VUR
Gumbel(μ, s)	$\mu = -s\gamma$	$s \leq \sqrt{6V}/\pi$	✓	✓
Normal(μ, σ^2)	$\mu = 0$	$\sigma \leq \sqrt{V}$	✓	✓
Logistic(μ, s)	$\mu = 0$	$s \leq \sqrt{3V}/\pi$	✓	✓
Laplace(μ, s)	$\mu = 0$	$s \leq \sqrt{V/2}$	✓	✓
Uniform(c, d)	$c = -d$	$d - c \leq \sqrt{12V}$	✓	✓ [†]

Table 1: The parameter γ in the centeredness condition of the Gumbel distribution is the Euler-Mascheroni constant, whose value is approximately 0.58. All five families satisfy the noise assumptions for both algorithms under the stated parameter conditions. For WO feedback, [†] additionally requires $\Delta_u < d - c$ for uniform noise. The other four families have full support, so Assumption (WO1) holds automatically for any set of utilities.

B.2 Pseudocode for Variance-aware Utility-Rankers

Below is the pseudocode for the VUR algorithms presented in Sections 4 and 5, respectively.

C TECHNICAL RESULTS AND PROOFS FOR SECTION 4

C.1 The Pairwise-preference Probabilities

We begin by showing that the pairwise-preference probabilities P_{ij} behave consistently with the underlying utilities.

Lemma 5. *Let $P_{ij} = \mathbb{P}[U_i > U_j]$ be the pairwise-preference probability defined in Section 4. Then, $P_{ij} > 1/2$ if and only if $u_i > u_j$.*

Algorithm 1 Full-Ranking Variance-aware Utility Ranker

Input: Items $[k]$, confidence $\delta \in (0, 1)$, accuracy $\varepsilon > 0$, variance bound $V > 0$

```

1: Initialize:  $n \leftarrow 0$ ,  $G \leftarrow ([k], \emptyset)$ ,  $\eta^* \leftarrow \min(\frac{2}{3}, \frac{\varepsilon}{\sqrt{6V}})$ 
2:    $P_{ij} \leftarrow 0$ ,  $V_{ij} \leftarrow 0$ , and  $I_{ij} \leftarrow \eta^*/4$  for every  $i \neq j$  in  $[k]$ 
3: repeat
4:    $n \leftarrow n + 1$ 
5:   Observe full ranking  $\sigma_n$  from RUM( $\mathbf{u}, \nu$ ) and set  $\delta_n \leftarrow \frac{6\delta}{k(k-1)n^2\pi^2}$ 
6:   for each pair  $i \neq j$  do
7:      $P_{ij} \leftarrow \frac{n-1}{n}P_{ij} + \frac{1}{n}\mathbb{1}\{i \text{ is ranked before } j \text{ in } \sigma_n\}$ 
8:     if  $n > 1$  then
9:        $V_{ij} \leftarrow \frac{n}{n-1}P_{ij}(1 - P_{ij})$ 
10:       $I_{ij} \leftarrow \sqrt{\frac{2V_{ij} \ln(4/\delta_n)}{n}} + \frac{7 \ln(4/\delta_n)}{3(n-1)}$ 
11:    end if
12:  end for
13: until  $I_{ij} < \eta^*/4$  for every  $i \neq j$ 
14: for  $i, j$  such that  $1 \leq i < j \leq k$  do
15:   if  $P_{ij} > 1/2 + \eta^*/4$  then
16:     Add  $i \rightarrow j$  to  $G$ 
17:   else if  $P_{ji} > 1/2 + \eta^*/4$  then
18:     Add  $j \rightarrow i$  to  $G$ 
19:   end if
20: end for
21: if  $G$  is acyclic then
22:    $\hat{G} \leftarrow$  transitive closure of  $G$  and  $\hat{\sigma} \leftarrow$  topological sort of  $\hat{G}$ 
23: else
24:    $\hat{\sigma} \leftarrow$  arbitrary permutation of  $[k]$ 
25: end if
Output:  $\hat{\sigma}$ 

```

Algorithm 2 Winner-Only Variance-aware Utility Ranker

Input: Items $[k]$, confidence $\delta \in (0, 1)$, accuracy $\varepsilon \in (0, b - a)$, variance bound $V > 0$

```

1: Initialize:  $n \leftarrow 0$ ,  $\eta^* \leftarrow \frac{\varepsilon-1}{3e+1} \min\left\{1, \frac{\varepsilon}{\sqrt{3V}}\right\}$ 
2:    $P_i \leftarrow 1$ ,  $V_i \leftarrow 1/4$ , and  $I_i \leftarrow \eta^*/2$  for every  $i \in [k]$ 
3: repeat
4:    $n \leftarrow n + 1$ 
5:   Observe winner  $w_n$  from RUM( $\mathbf{u}, \nu$ ) and set  $\delta_n \leftarrow \frac{3\delta}{k\pi^2n^2}$ 
6:   for each  $i \in [k]$  do
7:      $P_i \leftarrow \frac{n-1}{n}P_i + \frac{1}{n}\mathbb{1}\{w_n = i\}$ 
8:     if  $n > 1$  then
9:        $V_i \leftarrow \frac{n}{n-1}P_i(1 - P_i)$ 
10:       $I_i \leftarrow \sqrt{\frac{2V_i \ln(4/\delta_n)}{n}} + \frac{7 \ln(4/\delta_n)}{3(n-1)}$ 
11:    end if
12:  end for
13: until  $I_i < P_i \cdot \eta^*/2$  for every  $i \in [k]$ 
14:  $\hat{\sigma} \leftarrow$  unique permutation  $\sigma \in \mathcal{S}_k$  such that  $\sigma(i) < \sigma(j) \iff P_i > P_j \vee (P_i = P_j \text{ and } i < j)$ 
Output:  $\hat{\sigma}$ 

```

Proof. We denote by f and F the density and the CDF of ν , respectively. Additionally, we let (a, b) be an interval fully contained in $\text{Supp}(f)$.

If $u_i = u_j$, then, by symmetry, we have $P_{ij} = 1/2$. If $u_i > u_j$, then

$$\begin{aligned} P_{ij} &= \int_{\mathbb{R}} \mathbb{P}[U_j < U_i \mid U_i = x] f(x - u_i) dx \\ &= \int_{\mathbb{R}} F(x - u_j) f(x - u_i) dx \\ &\stackrel{(i)}{=} \int_{\mathbb{R}} F(y + u_i - u_j) f(y) dy, \end{aligned}$$

where we used the change of variable $y = x - u_i$ for (i). Now consider the function $g : [0, \infty) \rightarrow \mathbb{R}$ defined by

$$g(t) := \int_{\mathbb{R}} F(y + t) f(y) dy.$$

Taking the first derivative, we find

$$g'(t) = \int_{\mathbb{R}} f(y + t) f(y) dy.$$

Since $(a, b) \subset \text{Supp}(f)$, for $t < b - a$, there exists a subinterval of $\text{Supp}(f)$ where both $f(y + t)$ and $f(y)$ are strictly positive. Thus,

$$g'(t) > 0 \quad \text{for every } t \in [0, b - a).$$

Moreover, $g'(t) \geq 0$ for every $t \geq 0$, so g is non-decreasing on the whole half-line. Together with the fact that $g(0) = 1/2$, this implies $g(t) > 1/2$ for every $t \in (0, b - a)$ and, by monotonicity, $g(t) \geq g(t_0) > 1/2$ for every $t \geq b - a$ and any fixed $t_0 \in (0, b - a)$. Since $P_{ij} = g(u_i - u_j)$ and $u_i - u_j > 0$, we conclude that $P_{ij} > 1/2$.

To prove the converse implication, suppose $P_{ij} > 1/2$. As noted above, we cannot have $u_i = u_j$, so either $u_i < u_j$ or $u_i > u_j$. The former would imply $P_{ji} > 1/2$, but since $P_{ij} + P_{ji} = 1$, this yields a contradiction. $\square$

C.2 Concentration Bounds

The following result provides anytime-concentration bounds for the unbiased estimators introduced in Section 4. Our inequalities are based on (Maurer and Pontil, 2009).

Lemma 6. Fix $\delta \in (0, 1)$ and let δ_n be the time- n confidence parameter defined in Section 4.1. Then, with probability at least $1 - \delta$, the following hold simultaneously for every pair (i, j) with $i \neq j$ and every $n \in \mathbb{N} \setminus \{1\}$:

$$|P_{ij} - P_{ij}^n| \leq \sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)}, \quad (18)$$

$$\left| \sqrt{V_{ij}^n} - \sqrt{P_{ij}(1 - P_{ij})} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}}. \quad (19)$$

Proof. For a fixed pair (i, j) with $1 \leq i < j \leq k$ and $n \geq 2$, we define the events

$$A_{ij}^n := \left\{ |P_{ij} - P_{ij}^n| \leq \sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} \right\}$$

and

$$B_{ij}^n := \left\{ \left| \sqrt{V_{ij}^n} - \sqrt{P_{ij}(1 - P_{ij})} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}} \right\}.$$

By Theorem 4 and Theorem 10 in (Maurer and Pontil, 2009), we have $\mathbb{P}[A_{ij}^n] \geq 1 - \delta_n$ and $\mathbb{P}[B_{ij}^n] \geq 1 - \delta_n$. By

taking a union bound, we obtain

$$\begin{aligned}
 \mathbb{P} \left[\bigcap_{n \in \mathbb{N}} \left(\bigcap_{1 \leq i < j \leq k} (A_{ij}^n \cap B_{ij}^n) \right) \right] &= 1 - \mathbb{P} \left[\bigcup_{n \in \mathbb{N}} \left(\bigcup_{1 \leq i < j \leq k} (A_{ij}^n)^c \cup (B_{ij}^n)^c \right) \right] \\
 &\geq 1 - \frac{k(k-1)}{2} \sum_{n=1}^{\infty} 2\delta_n \\
 &= 1 - \frac{6\delta}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \\
 &\stackrel{(i)}{=} 1 - \delta,
 \end{aligned}$$

where (i) follows from the well-known fact that $\sum_{n \in \mathbb{N}} 1/n^2 = \pi^2/6$. To conclude, it suffices to observe that, for any two distinct items i and j , one has

$$|P_{ij} - P_{ij}^n| = |P_{ji} - P_{ji}^n| \quad \text{and} \quad V_{ij}^n = V_{ji}^n.$$

□

C.3 Uniform Lower Bound on the Resolution Function

In this section, we formally derive a uniform lower bound on the FR resolution function that holds across all noise distributions satisfying Assumption (V).

Lemma 7. *Let $\nu \in \mathcal{N}$ be a noise distribution satisfying the bounded-variance condition (V). Then*

$$\eta(t) \geq \eta^*(t) \quad \text{for every } t \in [0, \infty),$$

where η^* is defined by (4).

Proof. Let X and Y be two i.i.d. random variables with distribution ν , and let $Z := Y - X$ denote their difference, which is also log-concave. Additionally, Z is centered and symmetric, and its variance satisfies $\mathbb{V}\text{ar}[Z] = \mathbb{V}\text{ar}[X] + \mathbb{V}\text{ar}[X] \leq 2V$. Let us denote by $\sigma > 0$ the standard deviation of Z , and let H be its CDF. Then, for any $t > 0$, we have:

$$\begin{aligned}
 \int_{\mathbb{R}} F(x+t)f(x) dx &= \int_{\mathbb{R}} \mathbb{P}[Y < x-t] f(x) dx = \\
 &= \int_{\mathbb{R}} \mathbb{P}[Y < X-t | X=x] f(x) dx = \\
 &= \mathbb{P}[Y - X < -t] = \\
 &= \mathbb{P}[Z < -t] = H(-t).
 \end{aligned}$$

The left-hand side is equal to $\xi(t)$, where ξ is the function defined by Equation (2), which satisfies $\eta(t) = 2\xi(t) - 1$. By symmetry, we have:

$$\xi(t) = 1 - H(-t), \quad \forall t > 0.$$

using Lemma 19, we can bound the term $H(-t)$ as follows:

$$H(-t) \leq \begin{cases} \frac{1}{2} - \frac{t}{2\sqrt{3}\sigma} & \text{if } 0 \leq t < 2\sigma/\sqrt{3} \\ \frac{2\sigma^2}{9t^2} & \text{if } t \geq 2\sigma/\sqrt{3} \end{cases}.$$

We now consider two cases.

Case 1. If $t \in (0, 2\sigma\sqrt{3})$, then $H(-t) \leq 1/2 - t/2\sqrt{3}\sigma$. Hence, we obtain:

$$\begin{aligned} 1 - H(-t) &\geq \frac{1}{2} + \frac{t}{2\sqrt{3}\sigma} \stackrel{\sigma \leq \sqrt{2V}}{\geq} \\ &\geq \frac{1}{2} + \frac{t}{2\sqrt{6V}} = \\ &= \frac{1}{2} \left(1 + \frac{t}{\sqrt{6V}} \right). \end{aligned}$$

Using the relation $\eta(t) = 2\xi(t) - 1$, we conclude that $\eta(t) \geq t/\sqrt{6V}$.

Case 2. If $t \geq 2\sigma/\sqrt{3}$, we have

$$H(-t) \leq \frac{2\sigma^2}{9t^2} \leq \frac{2\sigma^2}{9(2\sigma/\sqrt{3})^2} = \frac{1}{6}.$$

This implies

$$\xi(t) \geq \frac{1}{2} \left(1 + \frac{2}{3} \right),$$

from which we obtain that $\eta(t) \geq 2/3$.

Combining the conclusions of both cases yields the desired bound. $\square$

C.4 The PAC Property

In this section, we prove two results showing that the FR-VUR algorithm presented in Section 4 is (ε, δ) -PAC.

Lemma 8. *With probability at least $1 - \delta$, the pairwise-preference graph G_T is a subgraph of the utility graph.*

Proof. With probability at least $1 - \delta$, the inequality given by (18) holds for all $i \neq j$ and $n \geq 2$. In particular, taking $n = T$, this further implies

$$|P_{ij}^T - P_{ij}| < \frac{\eta^*}{4} \quad \forall i \neq j.$$

If $i \rightarrow j \in \mathcal{E}_T$, i.e., there is an edge from i to j in G_T , then

$$P_{ij} = P_{ij}^T + (P_{ij} - P_{ij}^T) > \frac{1}{2} + \frac{\eta^*}{4} - \frac{\eta^*}{4} = \frac{1}{2}.$$

Since $P_{ij} > 1/2$ if and only if $u_i > u_j$ (see Lemma 5), we conclude that $i \rightarrow j$ is an edge of G_u . $\square$

Lemma 9. *With probability at least $1 - \delta$, if i and j are any two items such that $u_i - u_j \geq \varepsilon$, then $\hat{\sigma}(i) < \hat{\sigma}(j)$.*

Proof. With probability at least $1 - \delta$, inequality (18) holds for all $i \neq j$ and $n \geq 2$. In this regime, the condition of Lemma 8 is satisfied, so, taking $n = T$,

$$|P_{ij}^T - P_{ij}| < \frac{\eta^*}{4} \quad \forall i \neq j.$$

Now let i and j be any two items such that $|u_i - u_j| \geq \varepsilon$, say $u_i \geq u_j + \varepsilon$. If $\eta(\varepsilon)$ denotes the value of the resolution function computed for the true noise distribution, then Lemma 7 ensures $\eta(\varepsilon) \geq \eta^*$. By (3) and the monotonicity of ξ , $P_{ij} \geq \xi(\varepsilon)$. Using the definition of $\eta(\varepsilon)$, we have

$$P_{ij} \geq \frac{1}{2} + \frac{\eta(\varepsilon)}{2} \geq \frac{1}{2} + \frac{\eta^*}{2}.$$

Hence,

$$P_{ij}^T = P_{ij} + (P_{ij}^T - P_{ij}) > \frac{1}{2} + \frac{\eta^*}{2} - \frac{\eta^*}{4} = \frac{1}{2} + \frac{\eta^*}{4}.$$

This implies $i \rightarrow j \in \mathcal{E}_T$. Since G_T is a DAG, from the definition of $\hat{\sigma}$, it follows that $\hat{\sigma}(i) < \hat{\sigma}(j)$. $\square$

C.5 Sample Complexity Upper Bound

In this section, we prove Theorem 1. To do so, we first define $\mathcal{E}$ as the set of outcomes for which the following hold simultaneously for every $n \in \mathbb{N} \setminus \{1\}$ and any two distinct i and j :

$$|P_{ij}^n - P_{ij}| \leq \sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)}, \quad (20)$$

$$\left| \sqrt{V_{ij}^n} - \sqrt{P_{ij}(1 - P_{ij})} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}}, \quad (21)$$

where δ_n is the one given in Section 4.1.

If (21) holds, then multiplying both sides of $\sqrt{V_{ij}^n} \leq \sqrt{P_{ij}(1 - P_{ij})} + \sqrt{2 \log(2/\delta_n)/(n-1)}$ by $\sqrt{2 \log(4/\delta_n)/n}$ gives

$$\sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} \leq \sqrt{\frac{2P_{ij}(1 - P_{ij}) \log(4/\delta_n)}{n}} + \frac{2 \log(4/\delta_n)}{n-1},$$

where we used $\log(2/\delta_n) \leq \log(4/\delta_n)$ and $n(n-1) \geq (n-1)^2$. Adding $7 \log(4/\delta_n)/(3(n-1))$ to both sides, we obtain

$$\sqrt{\frac{2V_{ij}^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} \leq \sqrt{\frac{2P_{ij}(1 - P_{ij}) \log(4/\delta_n)}{n}} + \frac{13 \log(4/\delta_n)}{3(n-1)}, \quad (22)$$

whose left-hand side is exactly the quantity appearing in the stopping rule (5). Thus, if $\mathcal{E}$ occurs, the stopping rule is satisfied if the following hold simultaneously for every $i \neq j$:

$$\sqrt{\frac{2P_{ij}(1 - P_{ij}) \log(4/\delta_n)}{n}} < \frac{\eta^*}{8}, \quad (23)$$

$$\frac{13 \log(4/\delta_n)}{3(n-1)} < \frac{\eta^*}{8}. \quad (24)$$

Condition (23) holds for every $i \neq j$ if the stronger requirement below is satisfied:

$$n > \max_{i \neq j} (P_{ij}(1 - P_{ij})) \cdot \frac{128 \log(4/\delta_n)}{\eta^{*2}}.$$

In particular, it holds for n_0 , which we define as the minimum positive integer n such that

$$n \geq \max \left\{ 2, p(1-p) \cdot \frac{128 \log(4/\delta_n)}{\eta^{*2}} \right\},$$

where $p \in [0, 1]$ is chosen so that

$$p(1-p) = \max_{i \neq j} (P_{ij}(1 - P_{ij})).$$

Using the explicit expression of δ_n given in Section 4.1, we have

$$n_0 \leq p(1-p) \cdot \frac{128 \log(4/\delta_{n_0})}{\eta^{*2}} + 1 \leq \max \left(p(1-p) \cdot \frac{256}{\eta^{*2}} \log \left(\frac{2kn_0\pi}{\sqrt{6\delta}} \right), 2 \right).$$

By Lemma 18, there exists a universal constant C_0 such that

$$n_0 \leq \max \left(\frac{C_0 p(1-p)}{\eta^{*2}} \log \left(\frac{C_0 p(1-p)}{\eta^{*2}} \cdot \frac{k}{\delta} \right), 2 \right), \quad (25)$$

where $\log(0)$ is understood as $-\infty$.

Now we consider the condition given by (24), and define n_1 as the minimum positive integer n such that

$$n \geq \left\lceil \frac{104 \log(4/\delta_n)}{3\eta^*} + 1 \right\rceil.$$

Such an integer exists, since the right-hand side grows logarithmically in n , while the left-hand side grows linearly. By minimality, the inequality above fails at $n_1 - 1$, and since δ_n is non-increasing in n , this yields

$$n_1 \leq \frac{104 \log(4/\delta_{n_1})}{3\eta^*} + 2.$$

It follows that

$$n_1 \leq \max \left(\frac{208 \log(4/\delta_{n_1})}{3\eta^*}, 4 \right).$$

Using the definition of δ_n , Lemma 18 ensures that there exists a constant $C_1 > 0$ such that

$$n_1 \leq \max \left(\frac{C_1}{\eta^*} \log \left(\frac{C_1}{\eta^*} \cdot \frac{k}{\delta} \right), 4 \right). \quad (26)$$

If we let $n_2 := n_0 + n_1$, then for every $n \geq n_2$, the conditions given by (23) and (24) are both satisfied.

Now, the sample complexity can be decomposed as

$$\mathbb{E}[T] = \mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}}] + \mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}^c}]. \quad (27)$$

Using the tail-sum formula for the expectation, we can rewrite

$$\mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}^c}] = \sum_{n=1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n].$$

If $n > n_2$, then necessarily $\mathbb{1}_{\mathcal{E}^c} = 1$. Indeed, we defined n_2 in such a way that, if $\mathcal{E}$ occurs, then the stopping condition is met for all $n \geq n_2 + 1$. Therefore, we have

$$\sum_{n=n_2+1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n] = \sum_{n=n_2+1}^{\infty} \mathbb{P}[T \geq n].$$

If $T \geq n$, then the stopping condition was not satisfied at time $n - 1$. Thus, at least one of (20) and (21) is violated for at least one pair (i, j) with $i \neq j$. By a union bound over the $k(k - 1)/2$ unordered pairs and the two conditions, this occurs with probability at most $k(k - 1)\delta_{n-1}$. Hence, using the definition of δ_n given in Section 4.1,

$$\sum_{n=n_2+1}^{\infty} \mathbb{P}[T \geq n] \leq k(k - 1) \sum_{n=n_2}^{\infty} \delta_n = \sum_{n=n_2}^{\infty} \frac{6\delta}{\pi^2 n^2} \leq C_2,$$

for a universal constant $C_2 > 0$. It follows that

$$\sum_{n=1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n] \leq n_2 \cdot \mathbb{P}[\mathcal{E}^c] + C_2.$$

Using the decomposition given by (27), this implies

$$\mathbb{E}[T] \leq n_2 + C_2.$$

To prove the claim, we use the upper bounds given by (25) and (26) together with the definition of η^* . $\square$

C.6 Information-theoretic Lower Bound

We derive an information-theoretic lower bound for the ranking problem in the FR setting, thus proving Theorem 2. To this end, we fix a variance upper bound $V > 0$ and consider the centered Gumbel distribution ν with scale parameter $\beta = \sqrt{6V}/\pi$ and location parameter $\mu = -\beta\gamma$, where γ is the Euler–Mascheroni constant. This ensures that ν is centered and that its variance is equal to V . We fix an accuracy level ε in the small-gap regime $\varepsilon < \sqrt{6V}/(2\pi)$, which is equivalent to requiring $\tilde{\varepsilon} := \varepsilon/\beta \in (0, 1/2)$. Then, consider the utility vectors $\mathbf{u}$ and $\mathbf{v}$ given by

$$u_1 = v_2 = \varepsilon$$

and

$$u_\ell = v_m = 0 \quad \text{for every } \ell \in [k] \setminus \{1\} \text{ and } m \in [k] \setminus \{2\}.$$

Since the location parameter shifts all sampled utilities by the same constant, the induced ranking law does not depend on μ , and we can consider $\text{RUM}(\mathbf{u}, \nu)$ and $\text{RUM}(\mathbf{v}, \nu)$. These are seen as two instances of a multi-armed bandit problem, where the number of arms is $n = 1$. At each round t , the learner plays the only available arm. The observed reward is the feedback signal, consisting of a full ranking of items $1, \dots, k$. The *ranking distributions* P and Q correspond to the reward distributions, and are given by

$$P(\sigma) = \mathbb{P}[U_{(i)} = U_{\sigma^{-1}(i)} \mid \forall i \in [k] \mid \text{RUM}(\mathbf{u}, \nu)]$$

and

$$Q(\sigma) = \mathbb{P}[U_{(i)} = U_{\sigma^{-1}(i)} \mid \forall i \in [k] \mid \text{RUM}(\mathbf{v}, \nu)],$$

for every $\sigma \in \mathcal{S}_k$.

We denote by $\mathbb{P}_u$ and $\mathbb{P}_v$ the probability measures on $(\Omega, \mathcal{F})$ induced by $\text{RUM}(\mathbf{u}, \nu)$ and $\text{RUM}(\mathbf{v}, \nu)$, respectively. The next result provides an upper bound on the KL divergence between the two ranking distributions.

Lemma 10. *Let P and Q be the ranking distributions introduced above. Then, the KL divergence is upper-bounded as*

$$D_{\text{KL}}(P \parallel Q) \leq \frac{\pi^2 \varepsilon^2}{V}.$$

Proof. In what follows, we abuse notation and use $P[\cdot]$ and $\mathbb{P}_u[\cdot]$ to denote the same probability measure.

Consider any sample from $\text{RUM}(\mathbf{u}, \nu)$. Let $\sigma \in \mathcal{S}_k$ and denote by i_j the index of the j -th order statistic. Then, we can factorize the probability of observing σ as

$$P(\sigma) = \mathbb{P}_u[i_1 = \sigma^{-1}(1)] \cdot \prod_{t=2}^{k-1} \mathbb{P}_u[i_t = \sigma^{-1}(t) \mid i_j = \sigma^{-1}(j) \text{ for } j \leq t-1]. \quad (28)$$

A similar factorization holds for Q , and these allow us to think of the sampling process as a sequential one: first, the top item i_1 is chosen among all items $1, \dots, k$; then, item i_2 is picked among all the remaining items; and so on, until only one item is left and is chosen as the last element of the ranking.

According to this interpretation, if $S_t \subseteq [k]$ is the set of items left after the first $t-1$ items have been chosen, we define

$$P[i \mid S_t] := \frac{e^{u_i/\beta}}{\sum_{\ell \in S_t} e^{u_\ell/\beta}} \quad \text{for all } i \in S_t.$$

Now, using (28), we can write

$$D_{\text{KL}}(P \parallel Q) = \sum_{t=1}^{k-1} \mathbb{E}_{\nu, \mathbf{u}}[D_{\text{KL}}(P[\cdot \mid S_t] \parallel Q[\cdot \mid S_t])], \quad (29)$$

where the expectation is taken with respect to $\mathbb{P}_u$ over all pairs (i, S_t) where necessarily $i \in S_t$.

For simplicity, we introduce the parameter $\tilde{\varepsilon} := \varepsilon/\beta$, which allows us to rewrite

$$u_1/\beta = \tilde{\varepsilon} \quad \text{and} \quad u_\ell/\beta = 0 \quad \text{for all } \ell > 1,$$

as well as

$$v_2/\beta = \tilde{\varepsilon} \quad \text{and} \quad v_\ell/\beta = 0 \quad \text{for all } \ell \neq 2.$$

Recall that $\tilde{\varepsilon} \in (0, 1/2)$, which is exactly the small-gap condition $\varepsilon < \sqrt{6V}/(2\pi)$ assumed above. To estimate the KL divergence between P and Q , we consider three different cases. In the following analysis, we fix any $t \in [k-1]$.

Case 1: $\{1, 2\} \cap S_t = \emptyset$.

In this case, 1 and 2 have already been picked as the t -th decision round starts, so the conditional winning probabilities given S_t induced by $\mathbf{u}$ and $\mathbf{v}$ are the same:

$$D_{\text{KL}}(P[\cdot \mid S_t] \parallel Q[\cdot \mid S_t]) = 0.$$

Case 2: $1, 2 \in S_t$.

In this case, we have

$$\begin{aligned}
 D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t]) &= \sum_{i_t \in S_t} P[i_t | S_t] \log \frac{P[i_t | S_t]}{Q[i_t | S_t]} \\
 &= \sum_{i_t \in S_t \setminus \{1, 2\}} P[i_t | S_t] \log \frac{P[i_t | S_t]}{Q[i_t | S_t]} + \sum_{\ell=1, 2} P[\ell | S_t] \log \frac{P[\ell | S_t]}{Q[\ell | S_t]} \\
 &\stackrel{(i)}{=} \frac{e^{\tilde{\varepsilon}}}{|S_t| - 1 + e^{\tilde{\varepsilon}}} \log(e^{\tilde{\varepsilon}}) + \frac{1}{|S_t| - 1 + e^{\tilde{\varepsilon}}} \log(e^{-\tilde{\varepsilon}}) \\
 &= \frac{(e^{\tilde{\varepsilon}} - 1)\tilde{\varepsilon}}{|S_t| - 1 + e^{\tilde{\varepsilon}}} \\
 &\stackrel{(ii)}{\leq} \frac{\tilde{\varepsilon}(\tilde{\varepsilon} + \tilde{\varepsilon}^2)}{|S_t|},
 \end{aligned}$$

where (i) follows from the fact that the first sum is equal to zero and (ii) is due to $e^x - 1 \leq x + x^2$ for $x \in (0, 1/2)$.

Case 3: $|S_t \cap \{1, 2\}| = 1$.

In this case, we consider two further subcases corresponding to: (a) $1 \in S_t$ and $2 \notin S_t$, and (b) $2 \in S_t$ and $1 \notin S_t$.

Consider (a) first. Setting $W := \sum_{j \in S_t} e^{u_j/\beta}$ and $W' := \sum_{j \in S_t} e^{v_j/\beta}$, we obtain

$$\begin{aligned}
 D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t]) &= \sum_{j \in S_t} P[i_t = j | S_t] \log \frac{P[i_t = j | S_t]}{Q[i_t = j | S_t]} \\
 &= \frac{e^{\tilde{\varepsilon}}}{W} \log \frac{W' \cdot e^{\tilde{\varepsilon}}}{W} + \sum_{j \in S_t \setminus \{1\}} \frac{1}{W} \log \frac{W'}{W} \\
 &= \frac{e^{\tilde{\varepsilon}}}{W} \left(\tilde{\varepsilon} + \log \frac{W'}{W} \right) + \sum_{j \in S_t \setminus \{1\}} \frac{1}{W} \log \frac{W'}{W} \\
 &= \frac{e^{\tilde{\varepsilon}} \tilde{\varepsilon}}{W} + \log \frac{W'}{W}.
 \end{aligned}$$

To bound the term with the logarithm, observe that $\log(1/y) \leq 1/y - 1$ for every $y > 0$. Thus, we have

$$\log \frac{W'}{W} = \log \frac{|S_t|}{W} \leq \frac{|S_t|}{W} - 1 = \frac{1 - e^{\tilde{\varepsilon}}}{W},$$

where we used the fact that $W' = |S_t|$ and $W = |S_t| - 1 + e^{\tilde{\varepsilon}}$. It follows that

$$\begin{aligned}
 D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t]) &\leq \frac{e^{\tilde{\varepsilon}} \tilde{\varepsilon}}{W} + \frac{1 - e^{\tilde{\varepsilon}}}{W} \\
 &= \frac{e^{\tilde{\varepsilon}} \tilde{\varepsilon} + 1 - e^{\tilde{\varepsilon}}}{W} \\
 &\leq \frac{\tilde{\varepsilon}^2}{|S_t|},
 \end{aligned}$$

using the inequality $W \geq |S_t|$ and observing that

$$e^x x + 1 - e^x = e^x(x - 1) + 1 \leq x^2 \text{ for } x \in (0, 1/2).$$

The calculations for scenario (b) are similar and yield

$$\begin{aligned}
 D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t]) &= \log \frac{W'}{W} - \frac{\tilde{\varepsilon}}{W} \\
 &= \log \left(1 + \frac{e^{\tilde{\varepsilon}} - 1}{|S_t|} \right) - \frac{\tilde{\varepsilon}}{|S_t|} \\
 &\stackrel{(i)}{\leq} \frac{e^{\tilde{\varepsilon}} - 1}{|S_t|} - \frac{\tilde{\varepsilon}}{|S_t|} \\
 &= \frac{e^{\tilde{\varepsilon}} - 1 - \tilde{\varepsilon}}{|S_t|} \\
 &\stackrel{(ii)}{\leq} \frac{\tilde{\varepsilon}^2 + \tilde{\varepsilon} - \tilde{\varepsilon}}{|S_t|} \\
 &= \frac{\tilde{\varepsilon}^2}{|S_t|},
 \end{aligned} \tag{30}$$

where (i) is due to $\log(1+x) \leq x$ for $x > 0$, whereas for (ii), we relied again on the fact that $e^{\tilde{\varepsilon}} - 1 \leq \tilde{\varepsilon} + \tilde{\varepsilon}^2$.

Now, observe that $\mathbb{E}_{\nu, \mathbf{u}}[D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t])]$ can be computed by averaging the KL divergence inside the expectation over all choices of S_t , where weights are induced by P . Using the bounds derived in the three cases above and the fact that $\tilde{\varepsilon}^3 \leq \tilde{\varepsilon}^2$, we obtain

$$\begin{aligned}
 \mathbb{E}_{\nu, \mathbf{u}}[D_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t])] &\leq P[1, 2 \in S_t] \frac{\tilde{\varepsilon}^2 + \tilde{\varepsilon}^3}{|S_t|} + P[|\{1, 2\} \cap S_t| = 1] \frac{\tilde{\varepsilon}^2}{|S_t|} \\
 &\leq (1 - P[1, 2 \notin S_t]) \frac{2\tilde{\varepsilon}^2}{|S_t|},
 \end{aligned} \tag{31}$$

where, with abuse of notation, $P(E)$ denotes the probability of event E under the PL model associated with $\mathbf{u}$. Next, let us observe that

$$P[1 \in S_t] = \prod_{\ell=1}^{t-1} \frac{k - \ell}{k - \ell + e^{\tilde{\varepsilon}}},$$

where $k - \ell + e^{\tilde{\varepsilon}} \geq k - \ell + 1$ for every $\ell \in [t-1]$. This implies

$$P[1 \in S_t] \leq \prod_{\ell=1}^{t-1} \frac{k - \ell}{k - \ell + 1} = \frac{k - t + 1}{k} = \frac{|S_t|}{k}.$$

On the other hand, one has

$$P[2 \in S_t] \leq \prod_{\ell=1}^{t-1} \frac{k - (\ell + 1) + e^{\tilde{\varepsilon}}}{k - \ell + e^{\tilde{\varepsilon}}} = \frac{|S_t| + e^{\tilde{\varepsilon}} - 1}{k + e^{\tilde{\varepsilon}} - 1}.$$

By using a union bound, we have

$$(1 - P[1, 2 \notin S_t]) = P[1 \in S_t \text{ or } 2 \in S_t] \leq \frac{|S_t|}{k} + \frac{|S_t| + e^{\tilde{\varepsilon}} - 1}{k + e^{\tilde{\varepsilon}} - 1}.$$

Since $\tilde{\varepsilon} > 0$, $k + e^{\tilde{\varepsilon}} - 1 > k$; moreover, $|S_t| \geq 1$. Therefore, we can bound

$$\frac{|S_t| + e^{\tilde{\varepsilon}} - 1}{k + e^{\tilde{\varepsilon}} - 1} \leq \frac{|S_t|(1 + e^{\tilde{\varepsilon}} - 1)}{k} = \frac{e^{\tilde{\varepsilon}} |S_t|}{k}.$$

From this, we obtain

$$P[1 \in S_t \text{ or } 2 \in S_t] \leq \frac{|S_t|(1 + e^{\tilde{\varepsilon}})}{k} \leq \frac{3|S_t|}{k},$$

where we used the fact that $e^x + 1 \leq 3$ for every $x \in (0, 1/2)$. Together with (31), this allows us to derive our final bound

$$\mathbb{E}_{\nu, \mathbf{u}} [\text{D}_{\text{KL}}(P[\cdot | S_t] || Q[\cdot | S_t])] \leq \frac{6\tilde{\varepsilon}^2}{k}.$$

Combining this with expression (29) yields

$$\begin{aligned} \text{D}_{\text{KL}}(P || Q) &\leq \sum_{t=1}^{k-1} \frac{6\tilde{\varepsilon}^2}{k} \\ &= \frac{6(k-1)\tilde{\varepsilon}^2}{k} \\ &\leq 6\tilde{\varepsilon}^2. \end{aligned}$$

By recalling that $\tilde{\varepsilon} = \varepsilon/\beta$, we finally get

$$\text{D}_{\text{KL}}(P || Q) \leq \frac{6\varepsilon^2}{\beta^2} = \frac{\pi^2 \varepsilon^2}{V}.$$

□

Let $\mathcal{A}^* = (T^*, \hat{\sigma}^*)$ be an (ε, δ) -PAC algorithm, and define the $\mathcal{F}_{T^*}$ -measurable event

$$E = \{\hat{\sigma}^*(1) < \hat{\sigma}^*(2)\}.$$

Then

$$\mathbb{P}_u[E] \geq 1 - \delta \quad \text{and} \quad \mathbb{P}_v[E] \leq \delta.$$

Let us assume that $\delta \in (0, 1/4]$, as the confidence parameter is typically chosen to be very close to zero. For $x \geq 1 - \delta$ and $y \leq \delta$, the binary relative entropy satisfies

$$d(x, y) \geq d(1 - \delta, \delta) \geq \log\left(\frac{1}{3\delta}\right).$$

Using this, a direct application of Lemma 17 yields

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] \text{D}_{\text{KL}}(P || Q) \geq \log\left(\frac{1}{3\delta}\right).$$

Hence, by Lemma 10, we obtain

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] \geq \frac{V}{\pi^2 \varepsilon^2} \log\left(\frac{1}{3\delta}\right).$$

In conclusion, since $\log(1/(3\delta)) \geq \frac{1}{5} \log(1/\delta)$ for every $\delta \in (0, 1/4]$, we have

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] = \Omega\left(\frac{V}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right).$$

D TECHNICAL RESULTS AND PROOFS FOR SECTION 5

D.1 Concentration Bounds

The following result provides anytime-concentration bounds for the unbiased estimators introduced in Section 5. Our inequalities are based on (Maurer and Pontil, 2009). The proof of the lemma below is analogous to the one of Lemma 6.

Lemma 11. *Fix $\delta \in (0, 1)$ and let δ_n be as defined in Section 4.1. Then, with probability at least $1 - \delta$, the following hold simultaneously for all items $i \in [k]$ and every $n \in \mathbb{N} \setminus \{1\}$:*

$$|P_i - P_i^n| \leq \sqrt{\frac{2V_i^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)}, \quad (32)$$

$$\left| \sqrt{V_i^n} - \sqrt{P_i(1 - P_i)} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}}. \quad (33)$$

Proof. For fixed $i \in [k]$ and $n \geq 2$, we define the events

$$A_i^n := \left\{ |P_i - P_i^n| \leq \sqrt{\frac{2V_i^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} \right\}$$

and

$$B_i^n := \left\{ \left| \sqrt{V_i^n} - \sqrt{P_i(1-P_i)} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}} \right\}.$$

By Theorem 4 and Theorem 10 in (Maurer and Pontil, 2009), we have $\mathbb{P}[A_i^n] \geq 1 - \delta_n$ and $\mathbb{P}[B_i^n] \geq 1 - \delta_n$. By taking a union bound, we obtain

$$\begin{aligned} \mathbb{P} \left[\bigcap_{n \in \mathbb{N}} \left(\bigcap_{i \in [k]} (A_i^n \cap B_i^n) \right) \right] &= 1 - \mathbb{P} \left[\bigcup_{n \in \mathbb{N}} \left(\bigcup_{i \in [k]} (A_i^n)^c \cup (B_i^n)^c \right) \right] \\ &\geq 1 - k \sum_{n=1}^{\infty} 2\delta_n \\ &= 1 - \frac{6\delta}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \\ &\stackrel{(i)}{=} 1 - \delta, \end{aligned}$$

where (i) follows from the well-known fact that $\sum_{n \in \mathbb{N}} 1/n^2 = \pi^2/6$. $\square$

D.2 Uniform Lower Bound on the Winner-Only Resolution Factor

We fix $\nu \in \mathcal{N}_V$ and denote its density and CDF by f and F , respectively, and let its standard deviation be $\sigma \in (0, \sqrt{V}]$. Then, using the notation introduced in Section 5, we let $a := \inf \text{Supp}(f)$ and $b := \sup \text{Supp}(f)$, allowing infinite endpoints. By log-concavity, the interior of the support of f is the open interval (a, b) .

Lemma 12. *For almost every $x \in \mathbb{R}$,*

$$f(x) \geq \frac{\min\{F(x), 1 - F(x)\}}{\sqrt{3}\sigma} \geq \frac{F(x)(1 - F(x))}{\sqrt{3}\sigma}. \quad (34)$$

Proof. Let z_0 be the median of ν , so that $F(z_0) = 1/2$. The hazard rate $f/(1 - F)$ is non-decreasing and the reversed hazard rate f/F is non-increasing on (a, b) by log-concavity (Bagnoli and Bergstrom, 2005). Therefore,

$$\frac{f(x)}{F(x)} \geq 2f(z_0) \quad (x \leq z_0), \quad \frac{f(x)}{1 - F(x)} \geq 2f(z_0) \quad (x \geq z_0).$$

The median-density bound $f(z_0) \geq 1/(\sqrt{12}\sigma)$ follows from Bobkov and Ledoux (2019, Proposition B.2, Equation (B.6)). On the two sides of the median, the relevant denominator is respectively $F(x)$ and $1 - F(x)$, which proves the first inequality. The second follows from $\min\{q, 1 - q\} \geq q(1 - q)$ for $q \in [0, 1]$. Outside (a, b) the bounds hold almost everywhere because their right-hand sides vanish. $\square$

The next lemma compares the winning probabilities of any two items and provide a lower bound on their difference in terms of the largest of the two.

Lemma 13. *Let $\nu \in \mathcal{N}_V$ have standard deviation σ , and let $\mathbf{u} \in \mathbb{R}^k$. If $u_i > u_j$, then*

$$P_i - P_j \geq \frac{1 - e^{-(u_i - u_j)/(\sqrt{3}\sigma)}}{2} P_i. \quad (35)$$

Proof. Set $t := u_i - u_j > 0$, $\bar{F} := 1 - F$, and let z_0 be the median of ν . Define

$$D_t(z) := f(z) - f(z+t) \frac{F(z)}{F(z+t)}, \quad c_t := \frac{1 - e^{-t/(\sqrt{3}\sigma)}}{2},$$

assigning the ratio $F(z)/F(z+t)$ the value zero when its denominator is zero. In that case $F(z) = 0$ and $f(z) = 0$ almost everywhere as well. Since F is log-concave, f/F is non-increasing on (a, b) . Thus $D_t(z) \geq 0$ almost everywhere in the support, including when $z+t$ is beyond its upper endpoint. Outside the support, $D_t(z) = 0$ almost everywhere. Also, $F(z) \leq F(z+t)$ implies that the following holds for almost every $z \in \mathbb{R}$:

$$D_t(z) \geq f(z) - f(z+t).$$

We now prove that, for every $\alpha \in \mathbb{R}$:

$$\int_{\alpha}^{\infty} D_t(z) dz \geq c_t \bar{F}(\alpha). \quad (36)$$

To this end, let us first suppose that $\alpha \geq z_0$. If $\bar{F}(\alpha) = 0$, the claim follows from non-negativity. Otherwise, for almost every $z \geq \alpha$ in the support, Lemma 12 gives

$$\frac{f(z)}{\bar{F}(z)} \geq \frac{1}{\sqrt{3}\sigma}.$$

If $\alpha + t < b$, integration over $[\alpha, \alpha + t]$ yields

$$\log \frac{\bar{F}(\alpha + t)}{\bar{F}(\alpha)} = - \int_{\alpha}^{\alpha+t} \frac{f(z)}{\bar{F}(z)} dz \leq - \frac{t}{\sqrt{3}\sigma}.$$

Hence $\bar{F}(\alpha + t) \leq e^{-t/(\sqrt{3}\sigma)} \bar{F}(\alpha)$. The same inequality holds when $\alpha + t \geq b$, because then $\bar{F}(\alpha + t) = 0$. Consequently,

$$\begin{aligned} \int_{\alpha}^{\infty} D_t(z) dz &\geq \bar{F}(\alpha) - \bar{F}(\alpha + t) = \\ &\geq (1 - e^{-t/(\sqrt{3}\sigma)}) \bar{F}(\alpha) = 2c_t \bar{F}(\alpha) \geq c_t \bar{F}(\alpha). \end{aligned}$$

If $\alpha < z_0$, non-negativity and the preceding bound at z_0 give

$$\int_{\alpha}^{\infty} D_t(z) dz \geq \int_{z_0}^{\infty} D_t(z) dz \geq 2c_t \bar{F}(z_0) = c_t \geq c_t \bar{F}(\alpha).$$

This proves Equation (36).

Let w be any bounded, non-negative, non-decreasing function on $\mathbb{R}$, and set $M := \sup_z w(z)$. If we denote $L_y := \{z : w(z) > y\}$, its layer-cake representation is

$$w(z) = \int_0^M \mathbb{1}\{z \in L_y\} dy.$$

Each L_y is an upper interval, possibly empty or all of $\mathbb{R}$. Equation (36) applies to these sets as well, by absolute continuity and monotone convergence. Tonelli's theorem therefore gives

$$\begin{aligned} \int_{\mathbb{R}} w(z) D_t(z) dz &= \int_0^M \int_{L_y} D_t(z) dz dy \geq \\ &\geq c_t \int_0^M \int_{L_y} f(z) dz dy = c_t \int_{\mathbb{R}} w(z) f(z) dz. \end{aligned}$$

Now choose

$$w(z) := F(z+t) \prod_{\ell \in [k] \setminus \{i,j\}} F(z + u_i - u_{\ell}).$$

Expressing both winning probabilities relative to u_i gives

$$P_i = \int_{\mathbb{R}} w(z) f(z) dz, \quad P_i - P_j = \int_{\mathbb{R}} w(z) D_t(z) dz.$$

The second identity remains valid at points where $F(z+t) = 0$: the corresponding integrand for P_j also contains $F(z) = 0$. Applying the preceding inequality yields $P_i - P_j \geq c_t P_i$. $\square$

For an instance satisfying Assumption (WO1), $P_j > 0$. Since $\sigma \leq \sqrt{V}$, the preceding lemma implies

$$\frac{P_i}{P_j} \geq \frac{2}{1 + e^{-t/(\sqrt{3}\sigma)}} \geq \frac{2}{1 + e^{-t/\sqrt{3V}}} = \xi(t),$$

where $t = u_i - u_j$ and ξ is the function given by Equation (11). Accordingly, the resolution factor in Equation (13) is also defined in terms of V , rather than the unknown true variance.

Lemma 14. *Fix $V > 0$. The functions η and η^* in Equations (13) and (14) satisfy*

$$\eta(t) \geq \eta^*(t), \quad t \in (0, b - a).$$

Proof. Set $h(x) := (1 - e^{-x})/(3 + e^{-x})$ for $x \geq 0$. Its derivatives are

$$h'(x) = \frac{4e^{-x}}{(3 + e^{-x})^2} > 0, \quad h''(x) = -\frac{4e^{-x}(3 - e^{-x})}{(3 + e^{-x})^3} < 0.$$

Thus, h is increasing on $[0, \infty)$ and concave on the interval $[0, 1]$, with $h(0) = 0$. By concavity on $[0, 1]$, we have $h(x) \geq xh(1)$, while monotonicity gives $h(x) \geq h(1)$ for $x \geq 1$. Since $h(1) = (e - 1)/(3e + 1)$ and $\eta(t) = h(t/\sqrt{3V})$,

$$\eta(t) \geq \frac{e - 1}{3e + 1} \min \left\{ 1, \frac{t}{\sqrt{3V}} \right\} = \eta^*(t).$$

□

D.3 Consistency of the Winning Probabilities

In the FR case, we showed that the pairwise preference probability P_{ij} is larger than $1/2$ if and only $u_i > u_j$. Here, we rely on an analogous result for the winning probabilities.

Lemma 15. *Let $\nu \in \mathcal{N}_V$ and $\mathbf{u} \in \mathbb{R}^k$ satisfy Assumption (WO1). Then $P_i > 0$ for every $i \in [k]$. Moreover, for any two items $i, j \in [k]$ with $u_i \geq u_j$,*

$$P_i > P_j \iff u_i > u_j.$$

Proof. Let (a, b) be the interior of the support of the density f of ν . Conditioning on the noise of item i gives

$$P_i = \int_{\mathbb{R}} f(z) \prod_{\ell \neq i} F(z + u_i - u_\ell) dz.$$

Since $\Delta_u < b - a$, there is a nonempty open interval of $z \in (a, b)$ on which $z + u_i - u_\ell > a$ for every $\ell \neq i$. The integrand is strictly positive on that interval, so $P_i > 0$. If $u_i = u_j$, exchangeability of their independent noises gives $P_i = P_j$. If $u_i > u_j$, Lemma 13 gives

$$P_i - P_j \geq \frac{1 - e^{-(u_i - u_j)/(\sqrt{3}\sigma)}}{2} P_i > 0,$$

where $\sigma^2 = \mathbb{V}\text{ar}_\nu[N] > 0$. This proves the equivalence. □

D.4 Sample Complexity Upper Bound

In this section, we derive the sample complexity upper bound given by Theorem 3 for the WO-VUR algorithm.

To do so, we let δ_n be defined as in Section 4.1 and define $\mathcal{E}$ as the event consisting of those outcomes for which the following hold simultaneously for all $i \in [k]$ and every $n \in \mathbb{N} \setminus \{1\}$:

$$|P_i - P_i^n| \leq \sqrt{\frac{2V_i^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)}, \quad (37)$$

$$\left| \sqrt{V_i^n} - \sqrt{P_i(1 - P_i)} \right| \leq \sqrt{\frac{2 \log(2/\delta_n)}{n-1}}. \quad (38)$$

Arguing exactly as in Appendix C.5, condition (38) allows us to replace the sample variance V_i^n by the true one, at the price of an extra additive term:

$$\sqrt{\frac{2V_i^n \log(4/\delta_n)}{n}} + \frac{7 \log(4/\delta_n)}{3(n-1)} \leq \sqrt{\frac{2P_i(1-P_i) \log(4/\delta_n)}{n}} + \frac{13 \log(4/\delta_n)}{3(n-1)}. \quad (39)$$

The left-hand side is the quantity appearing both in (37) and in the stopping rule (15), so that, whenever $\mathcal{E}$ occurs, both can be controlled through the right-hand side.

First, we determine a positive integer n_0 such that, for every $n \geq n_0$, the following holds for every $i \in [k]$:

$$\sqrt{\frac{2P_i(1-P_i) \log(4/\delta_n)}{n}} + \frac{13 \log(4/\delta_n)}{3(n-1)} \leq \frac{P_i}{2}. \quad (40)$$

By (39), this guarantees $|P_i - P_i^n| \leq P_i/2$ on $\mathcal{E}$. The inequality is satisfied if the following hold simultaneously:

$$\sqrt{\frac{2P_i(1-P_i) \log(4/\delta_n)}{n}} < \frac{P_i}{4} \quad \text{and} \quad \frac{13 \log(4/\delta_n)}{3(n-1)} < \frac{P_i}{4}.$$

For these conditions to be satisfied, we can enforce the stronger requirement:

$$n > \frac{32 \log(4/\delta_n)}{P_i}.$$

Since we want this to hold for every $i \in [k]$, let us define $P_{\min} := \min_{i \in [k]} P_i$, and let n_0 be the minimum positive integer n such that

$$n \geq \max \left\{ 2, \frac{32 \log(4/\delta_n)}{P_{\min}} \right\}.$$

It follows that

$$n_0 \leq \frac{32 \log(4/\delta_{n_0})}{P_{\min}} + 1 \leq \frac{128}{P_{\min}} \log \left(2\pi n_0 \sqrt{\frac{k}{3\delta}} \right).$$

By Lemma 18, there exists a universal constant $C_1 > 0$ such that

$$n_0 \leq \frac{C_1}{P_{\min}} \log \left(\frac{C_1}{P_{\min}} \cdot \frac{k}{\delta} \right).$$

Now assume that $n > n_0$. If $\mathcal{E}$ occurs, then the condition given by (40) ensures $P_i < 2P_i^n$ for every $i \in [k]$, so that $\frac{\eta^*}{2} P_i^n > \frac{\eta^*}{4} P_i$. In this regime, using (39) once more, the stopping condition of $\mathcal{A}$ is fulfilled if the following hold for every $i \in [k]$:

$$\sqrt{\frac{2P_i(1-P_i) \log(4/\delta_n)}{n}} < \frac{P_i}{8} \eta^* \quad \text{and} \quad \frac{13 \log(4/\delta_n)}{3(n-1)} < \frac{P_i}{8} \eta^*. \quad (41)$$

In particular, these hold simultaneously for every $i \in [k]$ if the following stronger requirement is met:

$$n \geq \frac{256}{P_{\min} \eta^{*2}} \log \left(\frac{4}{\delta_n} \right).$$

If we define n_1 as the minimum positive integer $n > n_0$ such that

$$n \geq \left\lceil \frac{256 \log(4/\delta_n)}{P_{\min} \eta^{*2}} \right\rceil,$$

which exists since the right-hand side grows logarithmically in n , then either $n_1 = n_0 + 1$ or the inequality above fails at $n_1 - 1$, in which case $n_1 \leq 256 \log(4/\delta_{n_1}) / (P_{\min} \eta^{*2}) + 2$. In both cases, we can apply the same reasoning as above to find

$$n_1 \leq \frac{C_2}{P_{\min} \eta^{*2}} \log \left(\frac{C_2}{P_{\min} \eta^{*2}} \cdot \frac{k}{\delta} \right),$$

for some universal constant $C_2 > 0$.

Up to re-tuning C_1 and C_2 , our derivations show that, under event $\mathcal{E}$, the stopping condition is met within a number of steps equal to

$$T_0 := \left\lceil \frac{C_2}{P_{\min} \eta^{\star 2}} \log \left(\frac{C_2}{P_{\min} \eta^{\star 2}} \cdot \frac{k}{\delta} \right) \right\rceil.$$

Therefore, we have

$$\mathbb{E}[T] = \mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}}] + \mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}^c}] \leq T_0 \cdot \mathbb{P}[\mathcal{E}] + \mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}^c}].$$

Using the tail-sum formula for the expectation, we can rewrite

$$\mathbb{E}[T \cdot \mathbb{1}_{\mathcal{E}^c}] = \sum_{n=1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n] \leq T_0 \cdot \mathbb{P}[\mathcal{E}^c] + \sum_{n=T_0+1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n].$$

Now, observe that:

$$T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n \iff T \geq n \text{ and } \mathcal{E}^c \text{ occurs.}$$

However, we showed above that if $n \geq T_0 + 1$ and $T \geq n$, then necessarily $\mathcal{E}^c$ must have occurred, as otherwise we would have found $T \leq T_0$. Thus, for every $n \geq T_0 + 1$, we have

$$\mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n] = \mathbb{P}[T \geq n].$$

If $T \geq n$, then the stopping rule was not satisfied at time $n - 1$. This implies that at least one of the two conditions given by (37) and (38) was violated for at least one item $i \in [k]$. By a union bound over the k items and the two conditions, this occurs with probability at most $2k\delta_{n-1}$. Hence, we have

$$\begin{aligned} \sum_{n=T_0+1}^{\infty} \mathbb{P}[T \cdot \mathbb{1}_{\mathcal{E}^c} \geq n] &\leq 2k \sum_{n=T_0+1}^{\infty} \delta_{n-1} \\ &= \delta \cdot \sum_{n=T_0}^{\infty} \frac{6}{\pi^2 n^2} \\ &\leq C_3, \end{aligned}$$

where $C_3 > 0$ is a universal constant. This implies $\mathbb{E}[T] \leq T_0 + C_3$.

In conclusion, there exists a universal constant $C > 0$ such that

$$\mathbb{E}[T] \leq \frac{C}{P_{\min} \eta^{\star 2}} \log \left(\frac{C}{P_{\min} \eta^{\star 2}} \cdot \frac{k}{\delta} \right).$$

Using the definition of $\eta^{\star}$, when $\sqrt{3V} \geq \varepsilon$, the inequality above can be rewritten as

$$\mathbb{E}[T] \leq \frac{C \cdot V}{P_{\min} \varepsilon^2} \log \left(\frac{C \cdot V}{P_{\min} \varepsilon^2} \cdot \frac{k}{\delta} \right).$$

□

D.5 Information-theoretic Lower Bound

Finally, we prove Theorem 4, thus deriving an information-theoretic lower bound for the ranking problem under WO feedback.

For every $\delta \in (0, 1)$, algorithm $\mathcal{A}^*$ recovers an accurate representation of the utility ranking with probability at least $1 - \delta$, provided that Assumptions (V) and (WO1) are satisfied. In other words, if $\mathbf{u}$ and ν are chosen so that these conditions are met, then:

$$\mathbb{P}[\hat{\sigma}^* \text{ is } \varepsilon\text{-accurate for } \sigma_u \mid \text{RUM}(\mathbf{u}, \nu)] \geq 1 - \delta.$$

We now construct two specific instances of the problem, and we use them to provide a lower bound on the sample complexity of $\mathcal{A}^*$. Throughout, we assume $k \geq 3$. To do so, choose the *minimum-probability level* $p \in (0, 1/k)$

and set an accuracy level $\varepsilon \in (0, b - a)$. As we shall see, the latter must be chosen “sufficiently small”, which, in this case, means $\varepsilon \in (0, \beta)$. However, this restriction does not invalidate our sample complexity result due to its asymptotic nature.

Then, let ν be the centered Gumbel distribution with scale and location parameters given by

$$\beta = \frac{\sqrt{6V}}{\pi} \quad \text{and} \quad \mu = -\beta\gamma,$$

where γ denotes the Euler–Mascheroni constant. In particular, this ensures that ν is centered and that its variance is equal to V . Since the location parameter shifts all sampled utilities by the same constant, the implied winning probabilities do not depend on μ .

Now, consider the utility vectors $\mathbf{u}$ and $\mathbf{v}$ given by:

- $u_1 = \varepsilon/2$, $u_2 = -\varepsilon/2$, $u_3 = \dots = u_{k-1} = 0$, and $u_k = L$.
- $v_1 = -\varepsilon/2$, $v_2 = \varepsilon/2$, $v_3 = \dots = v_{k-1} = 0$, and $v_k = L$.

Here $k \geq 3$ guarantees that item k , which carries the utility L , is distinct from items 1 and 2; for $k = 3$ the intermediate block is empty, and the two instances reduce to $(\varepsilon/2, -\varepsilon/2, L)$ and $(-\varepsilon/2, \varepsilon/2, L)$. Up to choosing ε sufficiently small, the positive constant L is chosen so that the minimum winning probability of both instances is equal to p . We let P and Q be the *winning distributions* corresponding to $\mathbf{u}$ and $\mathbf{v}$, that is,

$$P_i := \frac{e^{u_i/\beta}}{\sum_{\ell=1}^k e^{u_\ell/\beta}} \quad \text{and} \quad Q_i := \frac{e^{v_i/\beta}}{\sum_{\ell=1}^k e^{v_\ell/\beta}} \quad \forall i \in [k].$$

Lemma 16. *Let P and Q be the winning distributions introduced above. Then, the KL divergence between P and Q is upper bounded as*

$$D_{\text{KL}}(P \parallel Q) \leq \frac{\varepsilon^2 \pi^2}{6V} \cdot C \cdot p,$$

where $C > 0$ is a universal constant.

Proof. By direct calculations, we obtain

$$D_{\text{KL}}(P \parallel Q) = \frac{\varepsilon}{\beta} (P_1 - P_2).$$

Up to choosing $\varepsilon \in (0, \beta)$, there exists a universal constant $C > 0$ such that

$$P_1 - P_2 \leq \frac{\varepsilon}{\beta} \cdot C \cdot P_2.$$

Due to our choice of L , $P_2 = P_{\min} = p$, thus implying

$$D_{\text{KL}}(P \parallel Q) \leq \left(\frac{\varepsilon}{\beta} \right)^2 \cdot C \cdot p = \frac{\varepsilon^2 \pi^2}{6V} \cdot C \cdot p.$$

□

In order to employ Lemma 17, we think of our setting as a trivial multi-armed bandit problem, in which the number of arms is $n = 1$. At time t , the only available arm is played. The observed reward is the feedback signal, which consists of the winner of the current round. The two bandit models for the problem are $\text{RUM}(\mathbf{u}, \nu)$ and $\text{RUM}(\mathbf{v}, \nu)$, where $\mathbf{u}$ and $\mathbf{v}$ are the ones introduced above. Since the noise distribution is of the Gumbel type, all winning probabilities are non-zero. Thus, the two reward distributions are mutually absolutely continuous, as needed in order to apply Lemma 17.

Now, let $\mathcal{E}$ be the event that item 1 is ranked before item 2. Since the output ranking is an $\mathcal{F}_T$ -measurable function, $\mathcal{E} \in \mathcal{F}_T$. If we define $\mathbb{P}_u$ and $\mathbb{P}_v$ as the probability measures induced by $\text{RUM}(\mathbf{u}, \nu)$ and $\text{RUM}(\mathbf{v}, \nu)$, respectively, a direct application of Lemma 17 yields

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] D_{\text{KL}}(P \parallel Q) \geq d(\mathbb{P}_u(\mathcal{E}), \mathbb{P}_v(\mathcal{E})).$$

Let us assume that $\delta \in (0, 1/4]$. Algorithm $\mathcal{A}^*$ is (ε, δ) -PAC, so

$$\mathbb{P}_u(\mathcal{E}) \geq 1 - \delta \quad \text{and} \quad \mathbb{P}_v(\mathcal{E}) \leq \delta.$$

Due to the properties of the binary relative entropy, we have

$$d(\mathbb{P}_u(\mathcal{E}), \mathbb{P}_v(\mathcal{E})) \geq d(1 - \delta, \delta) \geq \log\left(\frac{1}{3\delta}\right).$$

It follows that

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] \geq \frac{1}{D_{\text{KL}}(P \parallel Q)} \log\left(\frac{1}{3\delta}\right).$$

Owing to Lemma 16, we conclude that

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] \geq \frac{6V}{Cp\pi^2\varepsilon^2} \log\left(\frac{1}{3\delta}\right).$$

In conclusion, since $\log(1/(3\delta)) \geq \frac{1}{5} \log(1/\delta)$ for every $\delta \in (0, 1/4]$, we have shown that

$$\mathbb{E}_{\nu, \mathbf{u}}[T^*] = \Omega\left(\frac{V}{p\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right).$$

E AUXILIARY RESULTS

E.1 Change-of-measure Lemma for Multi-armed Bandit Problems

The next result restates Lemma 1 of (Kaufmann et al., 2016). As this is typically used in the context of multi-armed bandit problems, we need to introduce some preliminary notation. The number of arms is any positive integer n , and we identify the arm set with $[n]$. At time t , let A_t and Z_t be the (possibly random) arm being chosen and the corresponding reward, respectively. Let $\mathcal{F}_t = \sigma(A_1, Z_1, \dots, A_t, Z_t)$ be the sigma-algebra generated by the trajectory followed up to time t . For every $i \in [n]$, we define $N_i(t)$ as the number of times that arm i was played up to time t . We use the subscripts $\mathbb{P}_\nu$ and $\mathbb{E}_\nu$ to refer to the probability measure induced by a given bandit model ν and the implied expectation functional.

Lemma 17. *Let $\nu = (\nu_i)_{i=1}^n$ and $\nu' = (\nu'_i)_{i=1}^n$ be two bandit models, where ν_i (resp. ν'_i) is the reward distribution of arm $i \in [n]$. Assume that, for every $i \in [n]$, the probability distributions ν_i and ν'_i are equivalent, i.e., they are mutually absolutely continuous. For any almost-surely finite stopping time τ with respect to the filtration $(\mathcal{F}_t)_{t \in \mathbb{N}}$, the following holds:*

$$\sum_{i=1}^n \mathbb{E}_\nu[N_i(\tau)] D_{\text{KL}}(\nu_i \parallel \nu'_i) \geq \sup_{E \in \mathcal{F}_\tau} d(\mathbb{P}_\nu[E], \mathbb{P}_{\nu'}[E]),$$

where $d(x, y) := x \log(x/y) + (1-x) \log((1-x)/(1-y))$ is the binary relative entropy, with the convention that $d(0, 0) = d(1, 1) = 0$.

As we shall motivate later, the setting considered in this lemma encompasses ours by choosing the number of arms equal to one.

E.2 Technical Result for Bounding the Sample Complexity

The next result is a technical lemma used to derive the sample complexity upper bounds given by Theorem 1 and Theorem 3. A more general version can be found as Lemma 12 in (Jonsson et al., 2020). The slightly modified version presented below is sufficient for our purposes.

Lemma 18. *Let $n \in \mathbb{N}$ and $a, b > 0$. Suppose that $n \leq a \log(bn)$. Then*

$$n \leq 2a \log(ab).$$

Proof. For every $x > 0$, $\log(x) \leq \sqrt{x}$. Therefore, we have

$$n \leq a \log(bn) \leq a\sqrt{bn} = a\sqrt{b}\sqrt{n},$$

from which we obtain

$$\sqrt{n} \leq a\sqrt{b}.$$

Equivalently, $n \leq a^2b$. Since the logarithm is increasing and $b > 0$, we have

$$n \leq a \log(bn) \leq a \log(b(a^2b)) = 2a \log(ab).$$

□

E.3 Deviation Bound for Symmetric Unimodal Distributions

The following is a standard result providing useful deviation bounds for symmetric unimodal distributions. The version for standardized random variables can be found as Theorem 7.1 in (Ion et al., 2022).

Lemma 19. *Let Z be a symmetric unimodal random variable with $\mathbb{V}ar[Z] = \sigma^2 > 0$. Then*

$$\mathbb{P}[Z \geq t] \leq \begin{cases} \frac{1}{2}(1 - \frac{t}{\sqrt{3}\sigma}) & \text{if } 0 \leq t < 2\sigma/\sqrt{3} \\ \frac{2\sigma^2}{9t^2} & \text{if } t \geq 2\sigma/\sqrt{3} \end{cases}. \quad (42)$$