A Appendix

A.1 Proof of Lemma 1

Proof. To simplify the notation, let $\tilde{w}_i = \exp(w_i)$. Then

$$
\begin{aligned}
&P(S_1 \succ \cdots \succ S_M) \\
=& \sum_{(i_1,\cdots,i_N)\in\sigma(S_1\succ\cdots\succ S_M)} \prod_{l=1}^{N} \frac{\tilde{w}_{i_l}}{\sum_{r=l}^{N} \tilde{w}_{i_r}} \\
=& \sum_{\sigma(S_1)} \cdots \sum_{\sigma(S_M)} \prod_{m=1}^{M} \prod_{l=n_{m-1}+1}^{n_m} \frac{\tilde{w}_{i_l}}{\sum_{r=l}^{N} \tilde{w}_{i_r}} \\
=& \prod_{m=1}^{M} \sum_{(j_1,\cdots,j_{n_m})\in\sigma(S_m)} \prod_{l=1}^{n_m} \frac{\tilde{w}_{j_l}}{\sum_{k\in R_m} \tilde{w}_k - \sum_{r=1}^{l-1} \tilde{w}_{j_r}} \\
=& \prod_{m=1}^{M-1} P(S_m \succ R_{m+1}),
\end{aligned}
$$

□

A.2 Proof of Proposition 1

Proof. We first show $P(A \succ B; w) = P(\min_{a\in A} g_{w_a} > \max_{b\in B} g_{w_b})$. If $A \cup B = \lfloor N \rfloor$, then the event of $A \succ B$ is equivalent to the event of $\min_{a\in A} g_{w_a} > \max_{b\in B} g_{w_b}$ so this equality holds true. Otherwise, assume there is a $c \in \lfloor N \rfloor$ but $c \notin A \cup B$.

We introduce a few notations to assist the proof. For any $D \subseteq \lfloor N \rfloor$, let $\mathcal{G}(D) = \{g_{w_i} \mid i \in D\}$. Further let $\Omega(A \succ B; D)$ be the set of all possible permutations of D that are consistent with the partial ranking $A \succ B$, i.e.,

$$
\begin{aligned}
&\Omega(A \succ B; D) \\
=&\{(i_1,\cdots,i_N) \in \sigma(D) \mid k < l, \forall i_k \in A, i_l \in B\}.
\end{aligned}
$$

Then we can write the LHS as

$$
\begin{aligned}
&P(A \succ B; w) \\
=& \sum_{\substack{(i_1,\cdots,i_N)\in \\ \Omega(A\succ B;\lfloor N\rfloor)}} P(g_{w_{i_1}} > g_{w_{i_2}} > \cdots > g_{w_{i_N}}) \\
=& \sum_{\substack{(i_1,\cdots,i_N)\in \\ \Omega(A\succ B;\lfloor N\rfloor)}} \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor)} \mathbb{1}[g_{w_{i_1}} > g_{w_{i_2}} > \cdots > g_{w_{i_N}}] \\
=& \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor)} \sum_{\substack{(i_1,\cdots,i_N)\in \\ \Omega(A\succ B;\lfloor N\rfloor)}} \mathbb{1}[g_{w_{i_1}} > g_{w_{i_2}} > \cdots > g_{w_{i_N}}],
\end{aligned}
$$

where we slightly abused the notation g_{w_i} by using it to refer both the Gumbel random variables in the first line and the corresponding integral variables in the following lines. We have also omitted the integral variables and the probability densities $df(g_{w_i})$ in the derivation. To further ease the notation, we define $g_{w_{j_0}} = +\infty$ and $g_{w_{j_N}} = -\infty$, then

$$
\begin{aligned}
&P(A \succ B; w) \\
=& \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor)} \sum_{k=1}^{N} \sum_{\substack{(j_1,\cdots,j_{N-1})\in \\ \Omega(A\succ B;\lfloor N\rfloor\setminus\{c\})}} \\
&\quad \mathbb{1}[g_{w_{j_1}} > \cdots > g_{w_{j_{k-1}}} > g_{w_c} > g_{w_{j_k}} > \cdots > g_{w_{j_{N-1}}}] \\
=& \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor)} \sum_{\substack{(j_1,\cdots,j_{N-1})\in \\ \Omega(A\succ B;\lfloor N\rfloor\setminus\{c\})}} \mathbb{1}[g_{w_{j_1}} > \cdots > g_{w_{j_{N-1}}}] \\
&\quad \cdot \sum_{k=1}^{N} \mathbb{1}[g_{w_{j_{k-1}}} > g_{w_c} > g_{w_{j_k}}] \\
=& \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor\setminus\{c\})} \sum_{\substack{(j_1,\cdots,j_{N-1})\in \\ \Omega(A\succ B;\lfloor N\rfloor\setminus\{c\})}} \mathbb{1}[g_{w_{j_1}} > \cdots > g_{w_{j_{N-1}}}] \\
&\quad \cdot \int_{g_{w_c}} \sum_{k=1}^{N} \mathbb{1}[g_{w_{j_{k-1}}} > g_{w_c} > g_{w_{j_k}}],
\end{aligned} \tag{9}
$$

where the last equality utilizes the fact that all the Gumbel variables are independent.

Note that, in Eq. (9), given $g_{w_{j_1}} > \cdots > g_{w_{j_{N-1}}}$, $\sum_{k=1}^{N} \mathbb{1}[g_{w_{j_{k-1}}} > g_{w_c} > g_{w_{j_k}}] \equiv 1$ regardless the choice of $(j_1, \cdots, j_{N-1})$. Therefore,

$$
\int_{g_{w_c}} \sum_{k=1}^{N} \mathbb{1}[g_{w_{j_{k-1}}} > g_{w_c} > g_{w_{j_k}}] \equiv 1,
$$

and

$$
\begin{aligned}
&P(A \succ B; w) \\
=& \int\cdots\int_{\mathcal{G}(\lfloor N\rfloor\setminus\{c\})} \sum_{\substack{(j_1,\cdots,j_{N-1})\in \\ \Omega(A\succ B;\lfloor N\rfloor\setminus\{c\})}} \mathbb{1}[g_{w_{j_1}} > \cdots > g_{w_{j_{N-1}}}].
\end{aligned} \tag{10}
$$

By applying Eq. (10) to all the items that do not belong to $A \cup B$, we get

$$
\begin{aligned}
&P(A \succ B; w) \\
=& \int\cdots\int_{\mathcal{G}(A\cup B)} \sum_{\substack{(j_1,\cdots,j_{|A|+|B|})\in \\ \Omega(A\succ B;A\cup B)}} \mathbb{1}[g_{w_{j_1}} > \cdots > g_{w_{j_{|A|+|B|}}}].
\end{aligned} \tag{11}
$$

And note that this reduces to a situation equivalent to the case $A \cup B = \lfloor N \rfloor$. Therefore we have shown $P(A \succ B; w) = P(\min_{a\in A} g_{w_a} > \max_{b\in B} g_{w_b})$.

The proof for

$$P(\min_{a\in A} g_{w_a} > \max_{b\in B} g_{w_b}) = \int_{u=0}^{1} \prod_{a\in A}(1-u^{\exp(w_a - w_B)})du$$

remains the same no matter if $A \cup B = \lfloor N \rfloor$ or not, as the Gumbel variables are independent. We refer the reader to the Appendix B of Kool et al. (2020) for the proof. □

A.3 Proof of Lemma 2

Proof. We first expand the gradients of the log-likelihood w.r.t. $\boldsymbol{w}$ in Eq. (6) below.

$$\begin{aligned}
&\nabla_{\boldsymbol{w}} \log P(S_1 \succ \cdots \succ S_M; \boldsymbol{w}) \\
&= \sum_{m=1}^{M-1} \nabla_{\boldsymbol{w}} \log P(S_m \succ R_{m+1}; \boldsymbol{w}) \\
&= \sum_{m=1}^{M-1} \frac{1}{P(S_m \succ R_{m+1}; \boldsymbol{w})} \nabla_{\boldsymbol{w}} P(S_m \succ R_{m+1}; \boldsymbol{w}) \\
&= \sum_{m=1}^{M-1} \frac{1}{P(S_m \succ R_{m+1}; \boldsymbol{w})} \\
&\quad \cdot \nabla_{\boldsymbol{w}} \int_{u=0}^{1} \prod_{i\in S_m} (1 - u^{\exp(w_i - w_{R_{m+1}})}) du.
\end{aligned} \tag{12}$$

Further note that

$$\begin{aligned}
&\nabla_{\boldsymbol{w}} \prod_{i\in S_m} (1 - u^{\exp(w_i - w_{R_{m+1}})}) \\
&= \sum_{i\in S_m} \Big[\prod_{j\in S_m\setminus\{i\}} (1 - u^{\exp(w_j - w_{R_{m+1}})}) \Big] \\
&\quad \cdot \Big[-\nabla_{\boldsymbol{w}} u^{\exp(w_i - w_{R_{m+1}})} \Big] \\
&= -\sum_{i\in S_m} \Big[\prod_{j\in S_m\setminus\{i\}} (1 - u^{\exp(w_j - w_{R_{m+1}})}) \Big] \\
&\quad \cdot \Big[u^{\exp(w_i - w_{R_{m+1}})} \log u \Big] \nabla_{\boldsymbol{w}} \exp(w_i - w_{R_{m+1}}) \\
&= -\Big[\prod_{j\in S_m} (1 - u^{\exp(w_j - w_{R_{m+1}})}) \Big] \\
&\quad \cdot \sum_{i\in S_m} \frac{u^{\exp(w_i - w_{R_{m+1}})} \log u}{1 - u^{\exp(w_i - w_{R_{m+1}})}} \\
&\quad \cdot \nabla_{\boldsymbol{w}} \exp(w_i - w_{R_{m+1}}).
\end{aligned} \tag{13}$$

Plugging Eq. (13) into the gradients (12), we have

$$\begin{aligned}
&\nabla_{\boldsymbol{w}} \log P(S_1 \succ \cdots \succ S_M; \boldsymbol{w}) \\
&= -\sum_{m=1}^{M-1} \frac{1}{P(S_m \succ R_{m+1}; \boldsymbol{w})} \\
&\quad \cdot \int_{u=0}^{1} \Big[\prod_{j\in S_m} (1 - u^{\exp(w_j - w_{R_{m+1}})}) \Big] \\
&\quad \cdot \sum_{i\in S_m} \frac{u^{\exp(w_i - w_{R_{m+1}})} \log u}{1 - u^{\exp(w_i - w_{R_{m+1}})}} \nabla_{\boldsymbol{w}} \exp(w_i - w_{R_{m+1}}) du \\
&= -\sum_{m=1}^{M-1} \frac{1}{P(S_m \succ R_{m+1}; \boldsymbol{w})} \sum_{i\in S_m} \nabla_{\boldsymbol{w}} \exp(w_i - w_{R_{m+1}}) \\
&\quad \cdot \int_{u=0}^{1} \Big[\prod_{j\in S_m} (1 - u^{\exp(w_j - w_{R_{m+1}})}) \Big] \\
&\quad \cdot \frac{u^{\exp(w_i - w_{R_{m+1}})} \log u}{1 - u^{\exp(w_i - w_{R_{m+1}})}} du.
\end{aligned} \tag{14}$$

□

A.4 Proof of Theorem 1

Before we start our proof of Theorem 1, we first introduce the well-known discretization error bound for the composite mid-point rule of numerical integration in Lemma 3.

Lemma 3 (Discretization Error Bound of the Composite Mid-point Rule.)**.** *Suppose we use the composite mid-point rule with T intervals to approximate the following integral for some $x_1 > x_0$,*

$$\int_{x_0}^{x_1} f(x)dx.$$

Assume $f''(x)$ is continuous for $x \in [x_0, x_1]$ and $M = \sup_{x\in[x_0,x_1]} |f''(x)|$. Then the discretization error is bounded by $\frac{M(x_1-x_0)^3}{24T^2}$.

Proof of the part (a). The sketch of the proof is as follows. We first give an upper bound of the discretization error in terms of the number of intervals. Then we can obtain the number of intervals required for any desired level of error.

In particular, we bound the discretization error in two parts. We first bound the absolute value of the integral on the region $[0, \delta]$ for some sufficiently small $\delta > 0$. We then bound the second derivative of the integrand on $[\delta, 1]$ and apply Lemma 3 to bound the discretization error of the integral on $(\delta, 1]$. The total discretization error is then bounded by the sum of the two parts.

Proof. We first re-write the likelihood as follows,

$$\begin{aligned}
&P(S_1 \succ \cdots \succ S_M; \boldsymbol{w}) \\
=&\prod_{m=1}^{M-1} \int_{u=0}^{1} \prod_{i \in S_m} \left(1 - u^{\exp\left(w_i - w_{R_{m+1}}\right)}\right) du \\
=&\prod_{m=1}^{M-1} \exp(w_{R_{m+1}} + c) \\
&\cdot \int_{v=0}^{1} v^{\exp\left(w_{R_{m+1}}+c\right)-1} \prod_{i \in S_m} \left(1 - v^{\exp(w_i+c)}\right) dv \\
\triangleq& \prod_{m=1}^{M-1} I_m, \qquad (15)
\end{aligned}$$

where in the last second equality we have applied a change of variable $v = u^{\exp\left(-c-w_{R_{m+1}}\right)}$ for each integral.

To simplify the notations, let us define $g_i(v) = 1 - v^{\exp(w_i+c)}$ for any $i \in S_m$, and $g_0(v) = v^{\exp\left(w_{R_{m+1}}+c\right)-1}$. Then I_m can be written as

$$I_m = \exp(w_{R_{m+1}} + c) \int_{v=0}^{1} \prod_{i \in S_m \cup \{0\}} g_i(v) dv.$$

Further let

$$f(v) = \prod_{i \in S_m \cup \{0\}} g_i(v).$$

It remains to investigate the properties of $f(v)$ and its derivatives on $[0, 1]$ to bound the discretization error of I_m.

We first bound the absolute value of the integral on $[0, \delta]$ for some $\delta > 0$. We have

$$\begin{aligned}
&\left| \exp(w_{R_{m+1}} + c) \int_{v=0}^{\delta} \prod_{i \in S_m \cup \{0\}} g_i(v) dv \right| \\
\le& C \int_{v=0}^{\delta} v^{\exp\left(w_{R_{m+1}}+c\right)-1} dv \\
=& \frac{C}{\exp\left(w_{R_{m+1}} + c\right)} \delta^{\exp\left(w_{R_{m+1}}+c\right)} \\
\le& \frac{C}{2}\delta^2.
\end{aligned}$$

For any $\epsilon > 0$, let $\delta = (\frac{\epsilon}{C})^{1/2}$, then

$$\left| \exp(w_{R_{m+1}} + c) \int_{v=0}^{\delta} \prod_{i \in S_m \cup \{0\}} g_i(v) dv \right| \le \epsilon/2.$$

Next we bound the second derivative of the integrand, $f''(v)$, on $[\delta, 1]$. We have

$$\begin{aligned}
f^{''}(v) =& \sum_{i,j \in S_m \cup \{0\}, i \ne j} g_i'(v) g_j'(v) \prod_{k \in S_m \cup \{0\} \setminus \{i,j\})} g_k(v) \\
&+ \sum_{i \in S_m \cup \{0\}} g_i^{''}(v) \prod_{k \in S_m \cup \{0\} \setminus \{i\})} g_k(v).
\end{aligned}$$

For $v \in [\delta, 1]$, and each $i \in S_m$, we know that

$$\begin{aligned}
|g_i'(v)| =& \exp(w_i + c) v^{\exp(w_i+c)-1} \\
\le& \exp(w_i + c) \frac{1}{\delta} \\
\le& \frac{C^{3/2}}{\epsilon^{1/2}},
\end{aligned}$$

and

$$\begin{aligned}
|g_i^{''}(v)| =& |\exp(w_i + c)^2 - \exp(w_i + c)| v^{\exp(w_i+c)-2} \\
\le& |\exp(w_i + c)^2 - \exp(w_i + c)| \frac{1}{\delta^2} \\
\le& \frac{C^3}{\epsilon}.
\end{aligned}$$

Further,

$$\begin{aligned}
|g_0'(v)| =& |\exp(w_{R_{m+1}} + c) - 1| v^{\exp(w_{R_{m+1}}+c)-2} \\
\le& \exp(w_{R_{m+1}} + c) \\
\le& C,
\end{aligned}$$

and

$$\begin{aligned}
|g_0^{''}(v)| =& |\exp(w_{R_{m+1}} + c) - 1| |\exp(w_{R_{m+1}} + c) - 2| \\
&\cdot v^{\exp(w_{R_{m+1}}+c)-3} \\
\le& \exp(w_{R_{m+1}} + c)^2 \frac{1}{\delta} \\
\le& \frac{C^{5/2}}{\epsilon^{1/2}}.
\end{aligned}$$

Therefore, for $v \in [\delta, 1]$, we have

$$\begin{aligned}
|f^{''}(v)| \le& \sum_{i,j \in S_m \cup \{0\}, i \ne j} |g_i'(v)| |g_j'(v)| + \sum_{i \in S_m \cup \{0\}} |g_i^{''}(v)| \\
\le& \frac{C^3 (n_m + 1)^2}{\epsilon}.
\end{aligned}$$

By Lemma 3, we know that the discretization error of the integral on $[\delta, 1]$ is bounded by

$$\frac{C^4 (n_m + 1)^2}{24 T^2 \epsilon}.$$

For the total discretization error of I_m to be smaller than ϵ, it suffices to have

$$\frac{C^4 (n_m + 1)^2}{24 T^2 \epsilon} \le \epsilon/2,$$

which means
$$T \geq \frac{C^2(n_m+1)}{2\sqrt{3}\epsilon}.$$

□

Proof of the part (b). We follow a similar strategy as the proof of part (a). In this case, we bound the discretization error in three parts. We first bound the absolute values of the integral on the region $[0, \delta_1]$ and $[1-\delta_2, 1]$ for some sufficiently small $\delta_1, \delta_2 > 0$. We then bound the second derivative of the integrand on $[\delta_1, 1-\delta_2]$ and apply Lemma 3 to bound the discretization error of the integral on $[\delta_1, 1-\delta_2]$. The total discretization error is then bounded by the sum of the three parts.

Proof. We first denote the (m,i)-th integral in Eq. (14) for each $m = 1, \cdots, M-1$ and each $i \in S_m$ as $I_{m,i}$, i.e.,

$$I_{m,i} = \int_{u=0}^{1} \big[\prod_{j\in S_m}(1-u^{\exp(w_j - w_{R_{m+1}})})\big] \cdot \frac{u^{\exp(w_i - w_{R_{m+1}})}\log u}{1-u^{\exp(w_i - w_{R_{m+1}})}} du.$$

Similarly as what we did in the proof of part (a), by applying a change of variable $v = u^{\exp(-c-w_{R_{m+1}})}$, we can rewrite $I_{m,i}$ as

$$I_{m,i} = \exp(w_{R_{m+1}}+c)^2 \int_{v=0}^{1} \prod_{j\in S_m}(1-v^{\exp(w_j+c)}) \cdot \frac{v^{\exp(w_i+w_{R_{m+1}}+2c)-1}\log v}{1-v^{\exp(w_i+c)}} dv, \tag{16}$$

To simplify the notation, define $g_j(v) = 1 - v^{\exp(w_j+c)}$ for any $j \in S_m$, and

$$g_0(v) = \frac{v^{\exp(w_i+w_{R_{m+1}}+2c)-1}\log v}{1-v^{\exp(w_i+c)}}.$$

Then Eq. (16) becomes

$$I_{m,i} = \exp(w_{R_{m+1}}+c)^2 \int_{v=0}^{1} \prod_{j\in S_m\cup\{0\}} g_j(v) dv.$$

Further let
$$f(v) = \prod_{j\in S_m\cup\{0\}} g_j(v).$$

Recall that we have defined $a = \exp(w_i + w_{R_{m+1}} + 2c)$ and $b = \exp(w_i + c)$ in the statement of the theorem. For ease of notation in the proof, we redefine $a = \exp(w_i + w_{R_{m+1}} + 2c) - 1$, and the assumptions (8) become

$$a > 3, a + 2b > 4, \text{ and } b > C_0.$$

Then we can simplify the notation of $g_0(v)$ as

$$g_0(v) = \frac{v^a \log v}{1-v^b}.$$

It remains to investigate the property of $f(v)$ and its derivatives on $[0,1]$ to bound the discretization error.

We first bound the absolute value of the integral on $[0, \delta_1]$ for some small $\delta_1 > 0$. We have

$$\begin{aligned}
&\left|\exp(w_{R_{m+1}}+c)^2 \int_{v=0}^{\delta_1} \prod_{j\in S_m\cup\{0\}} g_j(v)dv\right| \\
\leq& C^2 \int_{v=0}^{\delta_1} \frac{v^a|\log v|}{1-v^b} \prod_{j\in S_m} g_j(v)dv.
\end{aligned}$$

Observe that $i \in S_m$ so $1 - v^b = g_i(v)$. Further by the fact that $\log v > 1 - \frac{1}{v}$ for all $v > 0$, we know that $|\log v| < |1 - \frac{1}{v}|$ for $v \in (0,1)$. So

$$\begin{aligned}
&C^2 \int_{v=0}^{\delta_1} \frac{v^a|\log v|}{1-v^b} \prod_{j\in S_m} g_j(v)dv \\
\leq& C^2 \int_{v=0}^{\delta_1} (v^{a-1} - v^a) \prod_{j\in S_m\setminus\{i\}} g_j(v)dv \\
\leq& C^2 \int_{v=0}^{\delta_1} v^{a-1}dv \\
=& \frac{C^2}{a}\delta_1^a \\
\leq& \frac{C^2}{3}\delta_1^3
\end{aligned}$$

For any $\epsilon > 0$, let $\delta_1 = (\frac{\epsilon}{C^2})^{1/3}$, then

$$\left|\exp(w_{R_{m+1}}+c)^2 \int_{v=0}^{\delta_1} \prod_{j\in S_m\cup\{0\}} g_j(v)dv\right| \leq \epsilon/3.$$

Similarly as the proof of part (a), for each $j \in S_m$ and $v \in [\delta_1, 1]$, we have

$$|g_j'(v)| \leq C/\delta_1 = C^{5/3}/\epsilon^{1/3} \leq C^2/\epsilon^{1/3},$$

and

$$|g_j''(v)| \leq C^2/\delta_1^2 = C^{10/3}/\epsilon^{2/3} \leq C^4/\epsilon^{2/3}.$$

We note that, however, $g_0(v)$ is not well-defined at $v = 1$. Therefore we instead try to bound the absolute value of the integral on the $[1-\delta_2, 1]$ for some small $\delta_2 > 0$.

When v is close to 1, by L'Hospital's rule, we have

$$\lim_{v\to 1} \frac{\log v}{1-v^b} = \lim_{v\to 1} \frac{1/v}{-bv^{b-1}} = -\frac{1}{b},$$

and hence

$$\lim_{v\to 1} g_0(v) = -\frac{1}{b}.$$

Further, from the assumptions in Eq. (7) and Eq. (8), we can derive that $a > b > 0$. In this case, we can also show that $g_0'(v) < 0$ on $[0, 1]$ so $|g_0(v)| \le \frac{1}{b} \le \frac{1}{C_0}$. Therefore,

$$\begin{aligned}
&\left|\exp(w_{R_{m+1}} + c)^2 \int_{1-\delta_2}^{1} f(v)dv\right| \\
\le& C^2 \int_{1-\delta_2}^{1} \prod_{j\in S_m} g_j(v)|g_0(v)|dv \\
\le& \frac{C^2}{C_0} \int_{1-\delta_2}^{1} \prod_{j\in S_m} (1 - v^{\exp(w_j+c)})dv \\
\le& \frac{C^2}{C_0} \int_{1-\delta_2}^{1} \prod_{j\in S_m} \exp(w_j+c)(1-v)dv \\
\le& \left[\frac{C^2}{C_0} \prod_{j\in S_m} \exp(w_j+c)\right] \delta_2^{n_m+1} \\
\le& \frac{C^2}{C_0} (C\delta_2)^{n_m},
\end{aligned}$$

where for the last third inequality we have used the fact that $1-x^\alpha \le \alpha(1-x)$ when $0 < x < 1$ and $\alpha > 0$.

For any $\epsilon > 0$, let

$$\delta_2 = \frac{1}{C}\left(\frac{C_0\epsilon}{3C^2}\right)^{1/n_m}, \tag{17}$$

then

$$\left|\exp(w_{R_{m+1}} + c)^2 \int_{1-\delta}^{1} f(v)dv\right| \le \frac{\epsilon}{3}.$$

Finally, we seek to bound the first and second derivatives of $g_0(v)$ on $[\delta_1, 1-\delta_2]$ in order to bound $f''(v)$. We write down $g_0'(v)$ and $g_0''(v)$ as follows,

$$g_0'(v) = \frac{av^{a-1}\log v}{1-v^b} + \frac{v^{a-1}}{1-v^b} + \frac{bv^{a+b-1}\log v}{(1-v^b)^2},$$

and

$$\begin{aligned}
g_0''(v) =& \frac{(a^2-a)v^{a-2}\log v}{1-v^b} + \frac{av^{a-2}}{1-v^b} + \frac{abv^{a+b-2}\log v}{(1-v^b)^2} \\
&+ \frac{(a-1)v^{a-2}}{1-v^b} + \frac{bv^{a+b-2}\log v}{(1-v^b)^2} \\
&+ \frac{(b^2+ab-b)v^{a+b-2}\log v}{(1-v^b)^2} + \frac{bv^{a+b-2}}{(1-v^b)^2} \\
&+ \frac{2ab^2v^{a+2b-3}\log v}{(1-v^b)^3}.
\end{aligned}$$

Again we know the fact that $|\log v| < |1-\frac{1}{v}|$ for $v \in (0, 1)$. It is also clear that $a < 2C$ and $b < C$. In combination with the conditions listed in (8), we can bound $g_0(v)$ and its derivatives as follows,

$$|g_0(v)| \le \frac{1}{1-v^b}, |g_0'(v)| \le \frac{4C}{(1-v^b)^2},$$
$$\text{and } |g_0''(v)| \le \frac{16C^3}{(1-v^b)^3}.$$

We can now bound $f''(v)$ on the interval $[\delta_1, 1-\delta_2]$,

$$\begin{aligned}
|f''(v)| \le& \sum_{i,j\in S_m\cup, i\ne j} |g_i'(v)||g_j'(v)||g_0(v)| \\
&+ \sum_{i\in S_m} \Big(|g_i''(v)||g_0(v)| + |g_i'(v)||g_0'(v)|\Big) \\
&+ |g_0''(v)| \\
\le& \frac{C^4 n_m^2}{(1-v^b)\epsilon^{2/3}} + \frac{4C^3 n_m}{(1-v^b)^2\epsilon^{1/3}} + \frac{16C^3}{(1-v^b)^3} \\
\le& \frac{16C^4 n_m^2}{\epsilon^{2/3}(1-(1-\delta_2)^{C_0})^3} \\
\le& \frac{16C^4 n_m^2}{\epsilon^{2/3}(C_0\delta_2)^3}.
\end{aligned} \tag{18}$$

Plugging Eq. (17) into the inequality (18), we have

$$\begin{aligned}
|f''(v)| \le& \frac{16C^7 n_m^2}{\epsilon^{2/3} C_0^{3+3/n_m} \frac{\epsilon}{3C^2}^{3/n_m}} \\
\le& \frac{48C^9 n_m^2}{C_0^4 \epsilon},
\end{aligned}$$

where for the last inequality, we have assumed $n_m \ge 9$ to ease the notation, as the case with small n_m is not very interesting.

Applying the result of Lemma 3, we know the discretization error of the integral on $[\delta_1, 1-\delta_2]$ is bounded by

$$\frac{2C^{11}n_m^2}{T^2 C_0^4 \epsilon}.$$

For the total discretization error of $I_{m,i}$ to be smaller than ϵ, it suffices to have

$$\frac{2C^{11}n_m^2}{T^2 C_0^4 \epsilon} \le \epsilon/3,$$

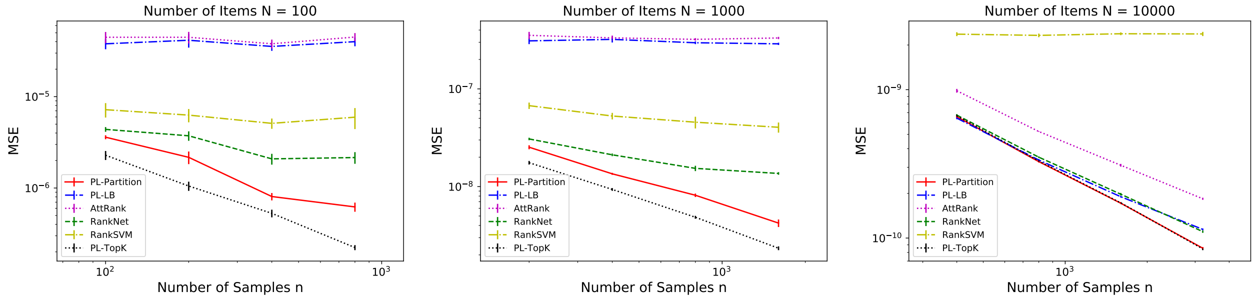

Figure 3: MSE of the estimated PL parameters vs various numbers of items N and number of samples n. Both x-axis and y-axis are in the logarithmic scale with base 10. The results are averaged over 5 different random seeds and error bars indicate the standard error of the mean.

which means

$$T \geq \frac{\sqrt{6}C^{11/2}n_m}{C_0^2\epsilon}.$$

To control the error of the m-th term in Eq. (14), we may want to have discretization error of $I_{m,i}$ smaller than ϵ/n_m. As an immediate corollary, we only need

$$T \geq \frac{\sqrt{6}C^{11/2}n_m^2}{C_0^2\epsilon}.$$

□

A.5 Alleviate the Round-off Error

For each integral in the likelihood (5), the integrand $\prod_{i\in S_m}(1-u^{\exp(w_i-w_{R_{m+1}})})$ is a product of many small numbers and may suffer from round-off errors. We can alleviate such round-off errors by converting the product into summation in the logarithmic space (Kool et al., 2020). In particular, recall the log-likelihood can be written as

$$\begin{aligned}&\log P(S_1 \succ \cdots \succ S_M; \boldsymbol{w})\\ =&\sum_{m=1}^{M-1}\log\left(\int_{u=0}^{1}\prod_{i\in S_m}\left(1-u^{\exp\left(w_i-w_{R_{m+1}}\right)}\right)du\right).\end{aligned}$$

Replacing the integrals with their numerical approximation, we have

$$\begin{aligned}&\log P(S_1 \succ \cdots \succ S_M; \boldsymbol{w})\\ \simeq&\sum_{m=1}^{M-1}\log\left(\sum_{t=1}^{T}\prod_{i\in S_m}\left(1-u_t^{\exp\left(w_i-w_{R_{m+1}}\right)}\right)\right). \quad (19)\end{aligned}$$

We can rewrite Eq. (19) in the form of log-sum-exp as follows

$$\begin{aligned}&\sum_{m=1}^{M-1}\log\left(\sum_{t=1}^{T}\prod_{i\in S_m}\left(1-u_t^{\exp\left(w_i-w_{R_{m+1}}\right)}\right)\right)\\ =&\sum_{m=1}^{M-1}\log\left(\sum_{t=1}^{T}\exp\left(\log\left(\prod_{i\in S_m}\left(1-u_t^{\exp\left(w_i-w_{R_{m+1}}\right)}\right)\right)\right)\right)\\ =&\sum_{m=1}^{M-1}\log\left(\sum_{t=1}^{T}\exp\left(\sum_{i\in S_m}\log\left(1-u_t^{\exp\left(w_i-w_{R_{m+1}}\right)}\right)\right)\right).\end{aligned}$$

Note that both the inner $\log(1-x)$ operations and the outer log-sum-exp operations have numerically stable implementations thus the influence of the round-off errors can be effectively reduced.

A.6 Supplemental Details for Simulation

We also provide additional simulation results with $N = 100, 1000, 10000$ in Figure 3. The trend is similar as what has been shown in Figure 1 in Section 4.1.

A.7 Supplemental Details for Experiments on XML Datasets

We provide the summary statistics of the 4 XML classification datasets in Table 3. We also provide the results of the nDCG-based metrics in Table 4. The nDCG-based metrics are highly correlated with their Precision-based counterparts. We further display in Figure 4 that the proposed PL-Partition is not sensitive in a wide range of the hyper-parameters T and c.

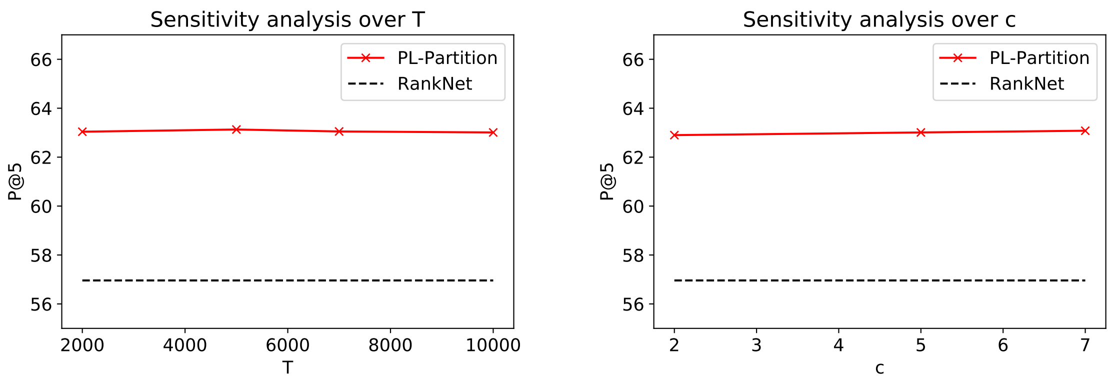

Figure 4: Sensitivity analysis over the hyper-parameters T and c of the proposed PL-Partition on the W-31K dataset. In both plots, the x-axis is the hyper-parameter value, and the y-axis is the P@5 metric. As a reference, we also plot the P@5 of the baseline RankNet, which is the black dashed constant line.

Table 3: Summary Statistics of the XML Classification Datasets

Dataset	#Feature	#Label	#Train	#Test	Avg. #Sample per Label	Avg. #Label per Sample
D-1K	500	983	12920	3185	311.61	19.03
E-4K	5000	3993	15539	3809	25.73	5.31
W-31K	101938	30938	14146	6616	8.52	18.64
D-200K	782585	205443	196606	100095	72.29	75.54

Table 4: nDCG@k and propensity-scored nDCG@k on the real-world XML datasets. Due to space limit, PL-Partition, RankNet, and RankSVM are respectively renamed as PL-P, R-Net, and R-SVM. **Bold numbers** indicate the best performance.

		PL-P	PL-LB	R-Net	R-SVM		PL-P	PL-LB	R-Net	R-SVM
D-1K	nDCG@1	**66.72**	66.12	64.11	61.95	PSnDCG@1	**33.22**	30.36	32.02	31.03
	nDCG@3	**62.78**	61.57	60.00	58.24	PSnDCG@3	**36.01**	33.26	34.34	33.29
	nDCG@5	**59.25**	57.51	56.29	55.01	PSnDCG@5	**37.58**	34.55	35.77	34.87
	nDCG@10	**52.48**	50.70	50.23	49.08	PSnDCG@10	**39.48**	36.09	37.43	36.52
E-4K	nDCG@1	**78.12**	66.46	77.57	76.46	PSnDCG@1	41.93	34.55	**42.81**	42.71
	nDCG@3	66.40	55.16	**66.56**	65.32	PSnDCG@3	47.61	38.18	49.83	**49.87**
	nDCG@5	59.66	48.71	**59.86**	58.61	PSnDCG@5	50.51	39.19	**52.98**	52.50
	nDCG@10	57.34	46.78	**57.78**	56.35	PSnDCG@10	50.76	39.19	**53.33**	52.72
W-31K	nDCG@1	**85.97**	80.73	82.35	80.88	PSnDCG@1	**13.01**	9.24	12.90	12.55
	nDCG@3	**76.08**	59.79	70.88	64.39	PSnDCG@3	**14.41**	8.72	14.07	13.32
	nDCG@5	**68.40**	51.85	62.81	56.67	PSnDCG@5	**15.98**	9.25	15.31	14.33
	nDCG@10	**56.17**	40.24	50.71	45.56	PSnDCG@10	**18.98**	10.22	17.99	16.46
D-200K	nDCG@1	**47.58**	40.38	41.93	41.41	PSnDCG@1	**8.72**	6.79	7.06	7.13
	nDCG@3	**43.32**	37.85	39.68	39.24	PSnDCG@3	**9.98**	8.27	8.67	8.70
	nDCG@5	**41.07**	36.50	38.09	37.77	PSnDCG@5	**10.98**	9.27	9.79	9.79
	nDCG@10	**37.69**	33.99	35.55	35.20	PSnDCG@10	**12.51**	10.83	11.39	11.42