Supplementary Material: Parameter Estimation for Generalized Thurstone Choice Models

Milan Vojnovic MILANV@MICROSOFT.COM
Microsoft Research, Cambridge, UK

Se-Young Yun YUNSEYOUNG@GMAIL.COM
Microsoft Research, Cambridge, UK

A. Some Basic Facts

A.1. Cramér-Rao Inequality

Let $\mathrm{cov}[Y]$ denote the covariance matrix of a multivariate random variable Y, i.e., $\mathrm{cov}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])(Y - \mathbf{E}[Y])^\top]$.

Proposition 7 (Cramér-Rao inequality)**.** *Suppose that X is a multivariate random variable with distribution $p(x;\theta)$, for parameter $\theta \in \Theta_n$, and let $\mathbf{T}(X) = (T_1(X), \ldots, T_r(X))^\top$ be any unbiased estimator of $\psi(\theta) = (\psi_1(\theta), \ldots, \psi_r(\theta))^\top$, i.e., $\psi(\theta) = \mathbf{E}[\mathbf{T}(X)]$. Then, we have*

$$\mathrm{cov}[\mathbf{T}(X)] \geq \frac{\partial \boldsymbol{\psi}(\theta)}{\partial \theta} F^{-1}(\theta) \frac{\partial \boldsymbol{\psi}(\theta)}{\partial \theta}^\top$$

where $\frac{\partial \psi(\theta)}{\partial \theta}$ is the Jacobian matrix of ψ and $F(\theta)$ is the Fisher information matrix with (i,j) element defined by

$$F_{i,j}(\theta) = \mathbf{E}\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j}(-\log(p(X;\theta)))\right].$$

A.2. Azuma-Hoeffding's Inequality for Vectors

The inequality is known as the Azuma-Hoeffding's inequality for multivariate random Variables, which was established in Theorem 1.8 (Hayes).

Proposition 8 (Azuma-Hoeffding's inequality)**.** *Suppose that $S_m = \sum_{t=1}^m X_t$ is a martingale where $X_1, X_2, \ldots, X_m$ take values in $\mathbf{R}^n$ and are such that $\mathbf{E}[X_t] = \mathbf{0}$ and $\|X_t\|_2 \leq D$ for all t, for $D > 0$. Then, for every $x > 0$,*

$$\mathbf{P}\left[\|S_m\|_2 \geq x\right] \leq 2e^2 e^{-\frac{x^2}{2mD^2}}.$$

A.3. Chernoff's Inequality for Matrices

The inequality is known as the Chernoff's inequalities for random matrices; e.g. stated as Theorem 5.1.1 in (Tropp, 2015).

Proposition 9 (Matrix Chernoff's inequality)**.** *Let $X_1, X_2, \ldots, X_m$ be a finite sequence of independent, random, Hermitian matrices with dimension d. Assume that*

$$0 \leq \lambda_1(X_i) \quad \text{and} \quad \|X_i\|_2 \leq \alpha \quad \text{for all} \quad i.$$

Let

$$\beta_{\min} = \lambda_1\left(\sum_{i=1}^m \mathbf{E}[X_i]\right)$$

and

$$\beta_{\max} = \lambda_d\left(\sum_{i=1}^m \mathbf{E}[X_i]\right).$$

Then, for $\varepsilon \geq 0$,

$$\mathbf{P}\left[\lambda_d\left(\sum_{i=1}^m X_i\right) \geq (1+\varepsilon)\beta_{\max}\right] \leq d\left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\beta_{\max}/\alpha} \quad \text{for } \varepsilon \geq 0 \tag{16}$$

and, for $\epsilon \in [0,1)$,

$$\mathbf{P}\left[\lambda_1\left(\sum_{i=1}^m X_i\right) \leq (1-\varepsilon)\beta_{\min}\right] \leq d\left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\beta_{\min}/\alpha} \quad \text{for } \varepsilon \in [0,1). \tag{17}$$

We have the following corollary:

Corollary 10. *Under the assumptions of Proposition 9, for $\varepsilon \in [0,1)$,*

$$\mathbf{P}\left[\lambda_1\left(\sum_{i=1}^m X_i\right) \leq (1-\varepsilon)\beta_{\min}\right] \leq de^{-\frac{\varepsilon^2 \beta_{min}}{2\alpha}}$$

Proof. This follows from (17) and the following fact

$$\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}} \leq e^{-\frac{\varepsilon^2}{2}}, \text{ for all } \varepsilon \in (0,1].$$

□

A.4. A Chernoff's Tail Bound

The following tail bound follows from the Chernoff's bound and is proved in Appendix L.3.

Proposition 11. *Suppose that X is a sum of m independent Bernoulli random variables each with mean p, then if $q \leq p \leq 2q$,*

$$\mathbf{P}[X \leq qm] \leq \exp\left(-\frac{(q-p)^2}{4q}m\right) \quad (18)$$

and, if $p \leq q$,

$$\mathbf{P}[X \geq qm] \leq \exp\left(-\frac{(q-p)^2}{4q}m\right). \quad (19)$$

A.5. Properties of Laplacian Matrices

If $\mathbf{A}$ is a symmetric non-negative matrix and the diagonal of $\mathbf{A}$ is zero, we have the following properties (Boyd, 2006):

$$0 = \lambda_1(\Lambda_{\mathbf{A}}) \leq \cdots \leq \lambda_n(\Lambda_{\mathbf{A}})$$

and

$$\lambda_{i+1}(\Lambda_{\mathbf{A}}) = \lambda_i(\mathbf{Q_1}^\top \Lambda_{\mathbf{A}} \mathbf{Q_1}) \text{ for } i = 1, 2, \ldots, n-1 \quad (20)$$

where $\mathbf{Q_1} \in \mathbf{R}^{n \times n-1}$ denotes a matrix whose columns are orthonormal to the all-one vector $\mathbf{1}$.

From (20), we have for all symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ with zero diagonals, it holds that

$$\Lambda_{\mathbf{A}} \succeq \Lambda_{\mathbf{B}} \quad \text{when} \quad \mathbf{A} \succeq \mathbf{B}, \quad (21)$$

where $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semi-definite matrix.

B. $\mathcal{T}_F$ Log-likelihoods

Let $H_{i,j}(\theta, S)$ be defined for $\theta \in \Theta_n$, $S \subseteq N$ and $i \in S$, $j \in S$, as follows

$$H_{i,j}(\theta, S) = \sum_{y \in S} p_{y,S}(\theta) \frac{\partial^2}{\partial\theta_i \partial\theta_j}(-\log(p_{y,S}(\theta))).$$

Lemma 12. *For every comparison set $S \subseteq N$, we have*

$$H_{i,j}(\mathbf{0}, S) = \begin{cases} k^2(k-1)\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 & \text{if} \quad i = j \\ -k^2\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 & \text{if} \quad i \neq j. \end{cases} \quad (22)$$

Proof of the last above lemma is provided in Appendix L.1.

Lemma 13. *Let $S \subseteq N$ and $y \in S$ and let k be the cardinality of set S. Then, it holds*

1. $\mathbf{1}^\top \nabla^2(-\log(p_{y,S}(\mathbf{0}))) = 0$, *and*
2. $\frac{1}{k}\sum_{y \in S} \nabla^2(-\log(p_{y,S}(\mathbf{0}))) = \Lambda_{\mathbf{M}_S} k^2 \left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2$

where $\mathbf{M}_S$ denotes a matrix that has all (i,j) elements such that $\{i,j\} \subseteq S$ equal to 1, and all other elements equal to 0.

Proof of the last above lemma follows easily from that of Lemma 12.

Lemma 14. *If for a comparison set $S \subseteq N$ of cardinality k, $\nabla^2(-\log(p_{y,S}(\mathbf{0})))$ is a positive semi-definite matrix, then it holds that*

$$\left\|\nabla^2(-\log(p_{y,S}(\mathbf{0})))\right\|_2 \leq \frac{2}{\gamma_{F,k}}.$$

Proof of the last above lemma is given in Appendix L.2.

C. Proof of Theorem 1

Let $\Delta = \hat{\theta} - \theta^\star$. By the Taylor expansion, we have

$$\begin{aligned} \ell(\hat{\theta}) \leq & \ell(\theta^\star) + \nabla\ell(\theta^\star)^\top \Delta \\ & + \frac{1}{2} \max_{\alpha \in [0,1]} \Delta^\top \nabla^2 \ell(\theta^\star + \alpha\Delta)\Delta. \end{aligned} \quad (23)$$

Note that Δ is orthogonal to the all-one vector, i.e., $\sum_{i=1}^n \Delta_i = 0$.

By the Cauchy-Schwartz inequality, we have

$$\nabla\ell(\theta^\star)^\top \Delta \leq \|\nabla\ell(\theta^\star)\|_2 \|\Delta\|_2. \quad (24)$$

Since $\hat{\theta}$ is a maximum likelihood estimator, we have

$$\ell(\hat{\theta}) - \ell(\theta^\star) \geq 0. \quad (25)$$

From (23), (24) and (25),

$$-\max_{\alpha \in [0,1]} \Delta^\top \nabla^2 \ell(\theta^\star + \alpha\Delta)\Delta \leq 2\|\nabla\ell(\theta^\star)\|_2 \|\Delta\|_2. \quad (26)$$

Now, note that for every $\theta \in \mathbf{R}^n$ and $i, j \in N$,

$$\begin{aligned} \frac{d^2}{dx^2}\log(p_2(\theta_i - \theta_j)) &= \frac{\partial^2}{\partial\theta_i^2}\log(p_2(\theta_i - \theta_j)) \\ &= \frac{\partial^2}{\partial\theta_j^2}\log(p_2(\theta_i - \theta_j)) \\ &= -\frac{\partial^2}{\partial\theta_i \partial\theta_j}\log(p_2(\theta_i - \theta_j)). \end{aligned}$$

Hence, for every $\theta \in \Theta_n$ and $\mathbf{x} \in \mathbf{R}^n$, we have

$$\begin{aligned}
&\mathbf{x}^\top \nabla^2 \ell(\theta) \mathbf{x} \\
&= \sum_{i=1}^{n} \sum_{j \neq i} w_{i,j} (x_i - x_j)^2 \frac{d^2}{dx^2} \log(p_2(\theta_i - \theta_j)) \\
&\leq - \sum_{i=1}^{n} \sum_{j \neq i} w_{i,j} B (x_i - x_j)^2 \\
&= - B \frac{4m}{n} \mathbf{x}^\top \Lambda_{\mathbf{M}} \mathbf{x} \\
&\leq - B \frac{4m}{n} \|\mathbf{x}\|_2^2 \lambda_2(\Lambda_{\mathbf{M}}), \qquad (27)
\end{aligned}$$

where we deduce the last inequality from (20).

From (26) and (27), we obtain

$$\frac{2Bm}{n} \|\Delta\|_2^2 \lambda_2(\Lambda_{\mathbf{M}}) \leq \|\nabla \ell(\theta^\star)\|_2 \|\Delta\|_2. \qquad (28)$$

We bound $\|\nabla \ell(\theta^\star)\|_2$ using the Azuma-Hoeffding's inequality for multivariate random variables in Proposition 8. Note that $\nabla \ell(\theta^\star)$ is a sum of m independent random vectors having zero-mean, where each comparison of a pair of items (i, j) results in a vector of value $\nabla \log(p_2(\theta_i^\star - \theta_j^\star))$ with probability $p_2(\theta_i^\star - \theta_j^\star)$ and of value $\nabla \log(p_2(\theta_j^\star - \theta_i^\star))$ with probability $p_2(\theta_j^\star - \theta_i^\star)$. Note that for every pair of items (i,j), $\|\nabla \log(p_2(\theta_i^\star - \theta_j^\star))\|_2 \leq A\sqrt{2}$.

By the Azuma-Hoeffding's inequality in Proposition 8, it follows that

$$\mathbf{P}\left[\|\nabla \ell(\theta^\star)\|_2 \geq 2A\sqrt{m(\log(n)+2)}\right] \leq \frac{2}{n}. \qquad (29)$$

Finally, from (28), (24) and (29), with probability $1 - 2/n$, it holds

$$\|\Delta\|_2 \leq \frac{An\sqrt{(\log(n)+2)}}{B\lambda_2(\Lambda_{\mathbf{M}})\sqrt{m}}.$$

D. Proof of Theorem 2

This proof follows the main steps of the proof of Theorem 1. Let $\Delta = \hat{\theta} - \theta^\star$. By the same arguments as in the proof of Theorem 1, we have that equation (26) holds, i.e.,

$$- \max_{\alpha \in [0,1]} \Delta^\top \nabla^2 \ell(\theta^\star + \alpha\Delta)\Delta \leq 2\|\nabla \ell(\theta^\star)\|_2 \|\Delta\|_2. \quad (30)$$

Since $\nabla^2(-\ell(\theta))$ is a Laplacian matrix, from assumption **A1** and (21), we have

$$\nabla^2(-\ell(\theta)) \succeq \underline{A}_{F,b} \nabla^2(-\ell(\mathbf{0})) \text{ for all } \theta \in [-b,b]^n. \quad (31)$$

From (30) and (31), we obtain

$$\begin{aligned}
&\lambda_1\left(\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\ell(\mathbf{0}))\mathbf{Q}_\mathbf{1}\right) \underline{A}_{F,b} \|\Delta\|_2 \\
&\leq 2\|\nabla \ell(\theta^\star)\|_2, \qquad (32)
\end{aligned}$$

which follows by the fact that $\hat{\theta}$ is orthogonal to $\mathbf{1}$.

We state two lemmas whose proofs are given at the end of this section, in Appendix D.1 and D.2.

Lemma 15. *Suppose that*

$$m \geq 32 \frac{\sigma_{F,K}}{\underline{B}_{F,b} \lambda_2(\Lambda_{\overline{\mathbf{M}}_F})} n \log(n)$$

then, with probability at least $1 - 1/n$*,*

$$\lambda_1\left(\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\ell(\mathbf{0}))\mathbf{Q}_\mathbf{1}\right) \geq \frac{\underline{B}_{F,b} m}{2n} \lambda_2(\Lambda_{\overline{\mathbf{M}}_F}).$$

and

Lemma 16. *With probability at least* $1 - 2/n$*, it holds that*

$$\|\nabla \ell(\theta^\star)\|_2 \leq \overline{C}_{F,b} \sqrt{\sigma_{F,K}} \sqrt{2m(\log(n)+2)}.$$

From (32) and the bounds in Lemma 15 and Lemma 16, it follows that if

$$m \geq 32 \frac{\sigma_{F,K}}{\underline{B}_{F,b} \lambda_2(\Lambda_{\overline{\mathbf{M}}_F})} n \log(n),$$

then, with probability at least $1 - 3/n$,

$$\|\Delta\|_2 \leq 32 \left(\frac{\overline{C}_{F,b}}{\underline{A}_{F,b} \underline{B}_{F,b}}\right)^2 \sigma_{F,K} \frac{n(\log(n)+2)}{\lambda_2(\Lambda_{\overline{\mathbf{M}}_F})^2} \frac{1}{m}.$$

D.1. Proof of Lemma 15

From the definition of the log-likelihood function $\ell(\theta)$, $\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\ell(\mathbf{0}))\mathbf{Q}_\mathbf{1}$ is a sum of a sequence of random matrices $\{\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\log(p_{y_t,S_t}(\mathbf{0})))\mathbf{Q}_\mathbf{1}\}_{1 \leq t \leq m}$, i.e.,

$$\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\ell(\mathbf{0}))\mathbf{Q}_\mathbf{1} = \sum_{t=1}^{m} \mathbf{Q}_\mathbf{1}^\top \nabla^2(-\log(p_{y_t,S_t}(\mathbf{0})))\mathbf{Q}_\mathbf{1}.$$

From assumption **A1** and (20), for every observation t,

$$\lambda_1\left(\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\log(p_{y_t,S_t}(\mathbf{0})))\mathbf{Q}_\mathbf{1}\right) \geq 0.$$

We can thus apply the matrix Chernoff's inequality, given in Proposition 9, once we find a lower bound for $\lambda_1\left(\mathbf{E}\left[\mathbf{Q}_\mathbf{1}^\top \nabla^2(-\ell(\mathbf{0}))\mathbf{Q}_\mathbf{1}\right]\right)$ and an upper bound for $\left\|\mathbf{Q}_\mathbf{1}^\top \left(\nabla^2 \log(p_{y_t,S_t}(\mathbf{0}))\right)\mathbf{Q}_\mathbf{1}\right\|_2$ for every observation t.

We have the following sequence of relations

$$\begin{aligned}
&\mathbf{E}_{\theta^\star}\left[\nabla^2(-\log(\ell(\mathbf{0})))\right] \\
&= \sum_{t=1}^{m} \mathbf{E}_{\theta^\star}\left[\nabla^2(-\log(p_{y_t,S_t}(\mathbf{0})))\right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^{m}\sum_{y\in S_t} p_{y,S_t}(\theta^\star)\nabla^2(-\log(p_{y,S_t}(\mathbf{0}))) \\
&\succeq \underline{B}_{F,b}\sum_{t=1}^{m}\sum_{y\in S_t}\frac{1}{|S_t|}\nabla^2(-\log(p_{y,S_t}(\mathbf{0}))) \\
&= \underline{B}_{F,b}\sum_{t=1}^{m}\sum_{y\in S_t}\frac{1}{|S_t|}\Lambda_{\mathbf{M}_{S_t}}|S_t|^2\left(\frac{\partial p_{|S_t|}(\mathbf{0})}{\partial x_1}\right)^2 \quad (33) \\
&= \underline{B}_{F,b}\frac{m}{n}\Lambda_{\overline{\mathbf{M}}_F} \quad (34)
\end{aligned}$$

where (33) follows Lemma 13 and $\mathbf{M}_S$ denotes a matrix that has all (i,j) elements such that $\{i,j\}\subseteq S$ equal to 1, and all other elements equal to 0.

From (34), we have

$$\begin{aligned}
&\lambda_1\left(\mathbf{E}\left[\mathbf{Q_1}^\top\nabla^2(-\ell(\mathbf{0}))\mathbf{Q_1}\right]\right) \\
\geq\ & \underline{B}_{F,b}\frac{m}{n}\lambda_1\left(\mathbf{Q_1}^\top\Lambda_{\overline{\mathbf{M}}_F}\mathbf{Q_1}\right) \\
=\ & \underline{B}_{F,b}\frac{m}{n}\lambda_2(\Lambda_{\overline{\mathbf{M}}_F}), \quad (35)
\end{aligned}$$

where the last equality holds by (20).

From Lemma 14, for every observation t,

$$\left\|\nabla^2\log(p_{y_t,S_t}(\mathbf{0}))\right\|_2 \leq \frac{2}{\gamma_{F,|S_t|}} \leq 2\sigma_{F,K}. \quad (36)$$

Using the matrix Chernoff's inequality in Corollary 10 with $\varepsilon = 1/2$, $\beta_{min}\geq \underline{B}_{F,b}\frac{m}{n}\lambda_2(\Lambda_{\overline{\mathbf{M}}_F})$ by (35) and $\alpha\leq\sigma_{F,K}$ by (36), we obtain the assertion of the lemma.

D.2. Proof of Lemma 16

For every comparison set $S\subseteq N$ and $i\in S$, we have

$$\frac{\partial\log p_{i,S}(\theta)}{\partial\theta_i} = -\frac{1}{p_{i,S}(\theta)}\sum_{v\in S\setminus\{i\}}\frac{\partial p_{v,S}(\theta)}{\partial\theta_i} \quad (37)$$

and, for all $j\in S\setminus\{i\}$,

$$\frac{\partial\log p_{j,S}(\theta)}{\partial\theta_i} = \frac{1}{p_{j,S}(\theta)}\frac{\partial p_{j,S}(\theta)}{\partial\theta_i}. \quad (38)$$

From (37) and (38), we have

$$\mathbf{E}\left[\nabla\log p_{y,S}(\theta^\star)\right] = \mathbf{0} \quad (39)$$

and

$$\|\nabla\log p_{y,S}(\mathbf{0})\|_2^2 = k^3(k-1)\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 = \frac{1}{\gamma_{F,k}}.$$

By assumption **A3**, every $S\subseteq N$ such that $|S|\in K$,

$$\begin{aligned}
\|\nabla\log p_{y,S}(\theta^\star)\|_2^2 &\leq \overline{C}_{F,b}^2\|\nabla\log p_{y,S}(\mathbf{0})\|_2^2 \\
&\leq \overline{C}_{F,b}^2\sigma_{F,K}. \quad (40)
\end{aligned}$$

Using (39) and (40) with the Azuma-Hoeffding inequality for multivariate random variables in Proposition 8, we obtain that with probability at least $1-2/n$,

$$\|\nabla\ell(\theta^\star)\|_2 \leq \overline{C}_{F,b}\sqrt{\sigma_{F,K}}\sqrt{2m(\log(n)+2)}.$$

E. Remark for Theorem 2

For the special case of noise according to the double-exponential distribution with parameter β, we have

$$p_k(\mathbf{x}) = \frac{1}{1+\sum_{i=1}^{k-1}e^{-x_i/\beta}}.$$

For every $\theta\in\theta_n$ and every $S\subseteq N$ of cardinality k and $i,j,y\in S$, we can easily check that

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_{y,S}(\theta))) = -\frac{1}{\beta^2}p_{i,S}(\theta)p_{j,S}(\theta).$$

Furthermore, the following two relations hold

$$\frac{k}{\beta(k-1)}\left(1-p_{y,S}(\theta)\right)^2 \leq \|\nabla p_{y,S}(\theta)\|_2 \leq \frac{2}{\beta}\left(1-p_{y,S}(\theta)\right)^2.$$

Since

$$\begin{aligned}
\min_{y\in S,\theta\in[-b,b]^n} p_{y,S}(\theta) &= \frac{1}{1+(k-1)e^{2b/\beta}} \\
&\geq p_{y,S}(\mathbf{0})e^{-2b/\beta}
\end{aligned}$$

and

$$\begin{aligned}
\max_{y\in S,\theta\in[-b,b]^n} p_{y,S}(\theta) &= \frac{1}{1+(k-1)e^{-2b/\beta}} \\
&\leq p_{y,S}(\mathbf{0})e^{2b/\beta}
\end{aligned}$$

we have that

$$\sigma_{F,K}\leq\frac{1}{\beta^2}$$

and

$$\begin{aligned}
e^{-4b/\beta} \leq \underline{A}_{F,b} \leq \overline{A}_{F,b} \leq e^{4b/\beta}, \quad (41) \\
e^{-2b/\beta} \leq \underline{B}_{F,b} \leq \overline{B}_{F,b} \leq e^{2b/\beta}, \quad (42) \\
e^{-4b/\beta} \leq \underline{C}_{F,b} \leq \overline{C}_{F,b} \leq 4. \quad (43)
\end{aligned}$$

F. Proof of Theorem 3

The proof of the theorem follows from the well-known Cramér-Rao inequality, which is given in Proposition 7.

Since $\sum_{i=1}^n\theta_i = 0$, we define $\psi_i(\theta) = \theta_i - \frac{1}{n}\sum_{l=1}^n\theta_l$. Note that $\sum_{i=1}^n\psi_i(\theta)=0$. Then,

$$\frac{\partial\psi(\theta)}{\partial\theta} = \mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^\top. \quad (44)$$

Let $F(\theta)$ be the Fisher information matrix of the random vector $X = (\mathbf{S}, \mathbf{y})$ where $\mathbf{S} = (S_1, S_2, \dots, S_m)$ are the comparison sets and $\mathbf{y} = (y_1, y_2, \dots, y_m)$ are the choices of comparisons.

Then, we have that the (i,j) element of matrix $F(\theta)$ is given by

$$F(\theta) = \sum_{t=1}^{m} \mathbf{E}\left[\nabla^2(-\log(p_{y_t,S_t}(\theta)))\right]. \quad (45)$$

From the assumptions **A1**, **A2**, and Lemma 13, we have

$$\begin{aligned}
&\mathbf{E}\left[\nabla^2(-\log(p_{y_t,S_t}(\theta)))\middle| S_t = S\right] \\
&= \sum_{y\in S} p_{y,S}(\theta)\nabla^2(-\log(p_{y,S}(\theta))) \\
&\preceq \sum_{y\in S} \frac{\overline{B}_{F,b}}{|S|}\nabla^2(-\log(p_{y,S}(\theta))) \\
&\preceq \sum_{y\in S} \frac{\overline{A}_{F,b}\overline{B}_{F,b}}{|S|}\nabla^2(-\log(p_{y,S}(\mathbf{0}))) \\
&= \overline{A}_{F,b}\overline{B}_{F,b}\left(|S|\frac{\partial p_{|S|}(\mathbf{0})}{\partial x_1}\right)^2 \Lambda_{\mathbf{M}_S} \quad (46)
\end{aligned}$$

where we use (21) for the two inequalities and $\mathbf{M}_S$ that has each element (i,j) such that $\{i,j\} \subseteq S$ equal to 1 and all other elements equal to 0.

From (45) and (46),

$$F(\theta) \preceq \overline{A}_{F,b}\overline{B}_{F,b}\frac{m}{n}\Lambda_{\overline{\mathbf{M}}_F}. \quad (47)$$

For a $n \times n$ matrix $\mathbf{A} = [a_{i,j}]$, let $\mathrm{tr}(\mathbf{A})$ denote its trace, i.e. $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{i,i}$. Note that

$$\begin{aligned}
\mathbf{E}[\|\hat{\theta} - \theta\|_2^2] &= \mathrm{tr}(\mathrm{cov}[\mathbf{T}(X)]) \\
&= \sum_{i=1}^{n} \lambda_i\left(\mathrm{cov}[\mathbf{T}(X)]\right).
\end{aligned}$$

By the Cramér-Rao bound and (47), we have

$$\begin{aligned}
\frac{1}{n}\mathbf{E}[\|\hat{\theta} - \theta\|_2^2] &\geq \frac{1}{n}\sum_{i=1}^{n}\lambda_i\left(\frac{\partial\boldsymbol{\psi}(\theta)}{\partial\theta}F^{-1}(\theta)\frac{\partial\boldsymbol{\psi}(\theta)}{\partial\theta}^\top\right) \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\lambda_i(\mathbf{Q_1}^\top F^{-1}(\mathbf{0})\mathbf{Q_1}) \\
&= \frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\lambda_i(\mathbf{Q_1}^\top F(\mathbf{0})\mathbf{Q_1})} \\
&\geq \frac{1}{\overline{A}_{F,b}\overline{B}_{F,b}m}\sum_{i=1}^{n-1}\frac{1}{\lambda_i(\mathbf{Q_1}^\top\Lambda_{\overline{\mathbf{M}}_F}\mathbf{Q_1})} \\
&= \frac{1}{\overline{A}_{F,b}\overline{B}_{F,b}m}\sum_{i=2}^{n}\frac{1}{\lambda_i(\Lambda_{\overline{\mathbf{M}}_F})},
\end{aligned}$$

where the last equality is obtained from (20).

G. Proof of Theorem 4

Let p^e denote the probability that the point score ranking method incorrectly classifies at least one item:

$$p^e = \mathbf{P}\left[\bigcup_{l\in N_1}\{l \in \hat{N}_1\} \cup \bigcup_{l\in N_2}\{l \in \hat{N}_2\}\right]$$

Let R_i denote the point score of item $i \in N$. If the point scores are such that $R_l > m/n$ for every $l \in N_1$ and $R_l < m/n$ for every $l \in N_2$, then this implies a correct classification. Hence, it must be that in the event of a misclassification of an item, $R_l \leq m/n$ for some $l \in N_1$ or $R_l \geq m/n$ for some $l \in N_2$. Combining this with the union bound, we have

$$\begin{aligned}
p^e &\leq \mathbf{P}\left[\bigcup_{l\in N_1}\left\{R_l \leq \frac{m}{n}\right\} \cup \bigcup_{l\in N_2}\left\{R_l \geq \frac{m}{n}\right\}\right] \\
&\leq \sum_{l\in N_1}\mathbf{P}\left[R_l \leq \frac{m}{n}\right] + \sum_{l\in N_2}\mathbf{P}\left[R_l \geq \frac{m}{n}\right]. \quad (48)
\end{aligned}$$

Let i and j be arbitrarily fixed items such that $i \in N_1$ and $j \in N_2$. We will show that for every observation t,

$$\mathbf{P}[y_t = i] \geq \frac{1}{n} + \frac{bk^2}{4n}\frac{\partial p_k(\mathbf{0})}{\partial x_1} \quad (49)$$

and

$$\mathbf{P}[y_t = j] \leq \frac{1}{n} - \frac{bk^2}{4n}\frac{\partial p_k(\mathbf{0})}{\partial x_1}. \quad (50)$$

From the Chernoff's bound in Lemma 11, we have the following bounds.

Using (18) for the random variable R_i, we obtain

$$\begin{aligned}
\mathbf{P}\left[R_i \leq \frac{m}{n}\right] &\leq \exp\left(-\frac{1}{4}n\left(\frac{1}{n} - \mathbf{E}[y_1 = i]\right)^2 m\right) \\
&\leq \exp\left(-\frac{1}{4}\left(\frac{bk^2}{4n}\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 m\right) \\
&\leq \exp\left(-\log(n/\delta)\right) \\
&= \frac{\delta}{n}.
\end{aligned}$$

Using (19) and using the same arguments, we obtain

$$\mathbf{P}\left[R_j \geq \frac{m}{n}\right] \leq \frac{\delta}{n}.$$

Combining with (48), it follows that

$$p^e \leq \delta.$$

In the remainder of the proof we show that inequalities (49) and (50) hold.

Let $\mathcal{A}$ be the set of all $A \subseteq N$ such that $|A| = k-1$ and $A \cap \{i,j\} = \emptyset$ and $\mathcal{B}$ be the set of all $B \subseteq N$ such that $|B| = k-2$ and $B \cap \{i,j\} = \emptyset$. Then, we have

$$\begin{aligned} & \mathbf{P}[y_t = i] - \mathbf{P}[y_t = j] \\ = & \sum_{A\in\mathcal{A}} \mathbf{P}[S_t = A \cup \{i\}] D_{i,j}(A) \\ & + \sum_{B\in\mathcal{B}} \mathbf{P}[S_t = B \cup \{i,j\}] D_{i,j}(B) \end{aligned} \quad (51)$$

where

$$\begin{aligned} D_{i,j}(A) &= \mathbf{P}[y_t = i | S_t = A\cup\{i\}] \\ & -\mathbf{P}[y_t = j | S_t = A \cup \{j\}] \end{aligned}$$

and

$$\begin{aligned} D_{i,j}(B) &= \mathbf{P}[y_t = i | S_t = B\cup\{i,j\}] \\ & -\mathbf{P}[y_t = j | S_t = B \cup \{i,j\}]. \end{aligned}$$

Let $\mathbf{b}$ be a $k-1$-dimensional vector with all elements equal to b. Then, note that

$$D_{i,j}(A) = p_k(\mathbf{b} - \theta_A) - p_k(-\mathbf{b} - \theta_A).$$

By limited Taylor series development, we have

$$p_k(\mathbf{x}) \geq p_k(\mathbf{0}) + \nabla p_k(\mathbf{0})^\top \mathbf{x} - \frac{1}{2}\beta \|\mathbf{x}\|_2^2 \quad (52)$$

$$p_k(\mathbf{x}) \leq p_k(\mathbf{0}) + \nabla p_k(\mathbf{0})^\top \mathbf{x} + \frac{1}{2}\beta \|\mathbf{x}\|_2^2 \quad (53)$$

where

$$\beta = \max_{\mathbf{x}\in[-2b,2b]^{k-1}} \|\nabla^2 p_k(\mathbf{x})\|_2. \quad (54)$$

Hence, it follows that for every $\theta_A \in \{-b, b\}^{k-1}$,

$$D_{i,j}(A) \geq 2(k-1)b\frac{\partial p_k(\mathbf{0})}{\partial x_1} - 4(k-1)b^2\beta. \quad (55)$$

Under the condition of the theorem, we have

$$\beta \leq \frac{1}{4b}\frac{\partial p_k(\mathbf{0})}{\partial x_1}.$$

Hence, combining with (55), for every $\theta_A \in \{-b,b\}^{k-1}$,

$$\begin{aligned} D_{i,j}(A) &\geq (k-1)b\frac{\partial p_k(\mathbf{0})}{\partial x_1} \\ &\geq \frac{kb}{2}\frac{\partial p_k(\mathbf{0})}{\partial x_1}. \end{aligned} \quad (56)$$

By the same arguments, we can show that

$$\begin{aligned} D_{i,j}(B) &= p_k(\boldsymbol{b} - \theta_B^{(-b)}) - p_k(-\boldsymbol{b} - \theta_B^{(b)}) \\ &\geq \frac{kb}{2}\frac{\partial p_k(\mathbf{0})}{\partial x_1} \end{aligned} \quad (57)$$

where $\theta_B^{(b)} \in \{-b,b\}^{k-1}$ and $\theta_B^{(-b)} \in \{-b,b\}^{k-1}$ are $(k-1)$-dimensional with the first elements equal to b and $-b$, respectively, and other elements equal to the parameters of items B.

Since the comparison sets are sampled uniformly at random without replacement, note that

$$\mathbf{P}[S_t = A\cup\{i\}] = \frac{\binom{n-1}{k-1}}{\binom{n}{k}}, \text{ for all } A \in \mathcal{A} \quad (58)$$

and

$$\mathbf{P}[S_t = B\cup\{i,j\}] = \frac{\binom{n-2}{k-2}}{\binom{n}{k}}, \text{ for all } B \in \mathcal{B}. \quad (59)$$

From (51), (56), (57), (58) and (59), we have

$$\mathbf{P}[y_t = i] - \mathbf{P}[y_t = j] \geq \frac{k^2 b}{2n}\frac{\partial p_k(\mathbf{0})}{\partial x_1}.$$

Using this inequality together with the following facts (i) $\mathbf{P}[y_t = l] = \mathbf{P}[y_t = i]$ for every $l \in N_1$, (ii) $\mathbf{P}[y_t = l] = \mathbf{P}[y_t = j]$ for every $l \in N_2$, (iii) $\sum_{l\in N}\mathbf{P}[y_t = l] = 1$, and (iv) $|N_1| = |N_2| = n/2$, it can be readily shown that

$$\mathbf{P}[y_t = i] \geq \frac{1}{n} + \frac{k^2 b}{4n}\frac{\partial p_k(\mathbf{0})}{\partial x_1},$$

which establishes (49). By the same arguments one can establish (50).

H. Proof of Theorem 5

Suppose that n is a positive even integer and θ is the parameter vector such that $\theta_i = b$ for $i \in N_1$ and $\theta_i = -b$ for $i \in N_2$, where $N_1 = \{1, 2, \dots, n/2\}$ and $N_2 = \{n/2+1, \dots, n\}$. Let θ' be the parameter vector that is identical to θ except for swapping the first and the last item, i.e. $\theta'_i = b$ for $i \in N'_1$ and $\theta'_i = -b$ for $i \in N'_2$, where $N'_1 = \{n, 2, \dots, n/2\}$ and $N'_2 = \{n/2+1, \dots, n-1, 1\}$.

We denote with $\mathbf{P}_\theta[A]$ and $\mathbf{P}_{\theta'}[A]$ the probabilities of an event A under hypothesis that the generalized Thurstone model is according to parameter θ and θ', respectively. We denote with $\mathbf{E}_\theta$ and $\mathbf{E}_{\theta'}$ the expectations under the two respective distributions.

Given observed data $(\mathbf{S}, \mathbf{y}) = (S_1, y_1), \dots, (S_m, y_m)$, we denote the log-likelihood ratio statistic $L(\mathbf{S}, \mathbf{y})$ as follows

$$L(\mathbf{S},\mathbf{y}) = \sum_{t=1}^m \log\left(\frac{p_{y_t,S_t}(\theta')\rho_t(S_t)}{p_{y_t,S_t}(\theta)\rho_t(S_t)}\right), \quad (60)$$

where $\rho_t(S)$ is the probability that S is drawn at time t.

The proof follows the following two steps:

Step 1: We show that for given $\delta \in [0, 1]$, for the existence of an algorithm that correctly classifies all the items with probability at least $1 - \delta$, it is necessary that the following condition holds

$$\mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \geq \log(n/\delta)] \geq \frac{1}{2}. \tag{61}$$

Step 2: We show that

$$\begin{aligned}\mathbf{E}_{\theta'}[L(\mathbf{S}, \mathbf{y})] &\leq 36\frac{m}{n}\left(k^2 b\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 &(62)\\ \sigma^2_{\theta'}[L(\mathbf{S}, \mathbf{y})] &\leq 144\frac{m}{n}\left(k^2 b\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 &(63)\end{aligned}$$

where $\sigma^2_{\theta'}[L(\mathbf{S}, \mathbf{y})]$ denotes the variance of random variable $L(\mathbf{S}, \mathbf{y})$ under a generalized Thurstone model with parameter θ'.

By Chebyshev's inequality, for every $g \in \mathbf{R}$,

$$\mathbf{P}_{\theta'}[|L(\mathbf{S}, \mathbf{y}) - \mathbf{E}_{\theta'}[L(\mathbf{S}, \mathbf{y})]| \geq |g|] \leq \frac{\sigma^2_{\theta'}[L(\mathbf{S}, \mathbf{y})]}{g^2}.$$

Using this for $g = \log(n/\delta) - \mathbf{E}_{\theta'}[L(\mathbf{S}, \mathbf{y})]$, it follows that (61) implies the following condition:

$$\begin{aligned}\log(n/\delta) - \mathbf{E}_{\theta'}[L(\mathbf{S}, \mathbf{y})] &\leq |\log(n/\delta) - \mathbf{E}_{\theta'}[L(\mathbf{S}, \mathbf{y})]|\\ &\leq \sqrt{2}\sigma_{\theta'}[L(\mathbf{S}, \mathbf{y})].\end{aligned}$$

Further combining with (62) and (63), we obtain

$$m \geq \frac{1}{62}\frac{1}{b^2 k^4 (\partial p_k(\mathbf{0})/\partial x_1)^2} n(\log(n) + \log(1/\delta))$$

which is the condition asserted in the theorem.

Proof of Step 1. Let us define the following two events

$$A = \{|N_1 \setminus \hat{N}_1| = 1\} \cap \{|N_2 \setminus \hat{N}_2| = 1\}$$

and

$$B = \{\hat{N}_1 = N'_1\} \cap \{\hat{N}_2 = N'_2\}.$$

Let B^c denote the complement of the event B.

Note that

$$\begin{aligned}\mathbf{P}_\theta[B] &= \mathbf{P}_\theta[B|A]\mathbf{P}_\theta[A]\\ &= \left(\frac{2}{n}\right)^2 \mathbf{P}_\theta[A]\\ &\leq \frac{4}{n^2}\delta\end{aligned}$$

where the second equation holds because $B \subseteq A$ and every possible partition in A has the same probability under θ.

For every $g \in \mathbf{R}$, we have

$$\begin{aligned}\mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq g] &= \mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq g, B]\\ &\quad + \mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq g, B^c].\end{aligned}$$

Now, note

$$\begin{aligned}\mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq g, B] &= \mathbf{E}_{\theta'}[\mathbf{1}(L(\mathbf{S}, \mathbf{y}) \leq g, B)]\\ &= \mathbf{E}_{\theta}[e^{L(\mathbf{S}, \mathbf{y})}\mathbf{1}(L(\mathbf{S}, \mathbf{y}) \leq g, B)]\\ &\leq \mathbf{E}_{\theta}[e^{g}\mathbf{1}(L(\mathbf{S}, \mathbf{y}) \leq g, B)]\\ &= e^g \mathbf{P}_\theta[L(\mathbf{S}, \mathbf{y}) \leq g, B]\\ &\leq e^g \mathbf{P}_\theta[B]\\ &\leq e^g \frac{4}{n^2}\delta \end{aligned} \tag{64}$$

where in the second equation we make use of the standard change of measure argument.

Since the algorithm correctly classifies all the items with probability at least $1 - \delta$, we have

$$\mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq g, B^c] \leq \mathbf{P}_{\theta'}[B^c] \leq \delta. \tag{65}$$

For $g = \log(n/\delta)$, from (64) and (65), it follows that

$$\mathbf{P}_{\theta'}[L(\mathbf{S}, \mathbf{y}) \leq \log(n/\delta)] \leq \delta + \frac{4}{n} \leq \frac{1}{2}$$

where the last inequality is by the conditions of the theorem.

Proof of Step 2. If the observed comparison sets $S_1, S_2, \ldots, S_m$ are such that $S_t \cap \{1, n\} = \emptyset$, for every observation t, then we obviously have

$$\log\left(\frac{p_{y_t, S_t}(\theta')}{p_{y_t, S_t}(\theta)}\right) = 0, \text{ for all } t.$$

We therefore consider the case when $S_t \cap \{1, n\} \neq \emptyset$.

Using (52), (53), and (54), we have for every S and $i \in S$,

$$\begin{aligned}&|p_{i,S}(\theta') - p_{i,S}(\theta)|\\ &\leq 2kb\frac{\partial p_k(\mathbf{0})}{\partial x_1} + 4\beta bk\\ &\leq 3kb\frac{\partial p_k(\mathbf{0})}{\partial x_1},\end{aligned} \tag{66}$$

where the last inequality is obtained from the condition of this theorem.

From (66), for every comparison set S such that $S \cap \{1,n\} \neq \emptyset$, we have

$$\begin{aligned}
&\sum_{i\in S} \left(p_{i,S}(\theta') - p_{i,S}(\theta)\right)^2 \\
&\leq \sum_{i\in\{1,n\}\cap S} \left(p_{i,S}(\theta') - p_{i,S}(\theta)\right)^2 + \\
&\quad \left(\sum_{i\in S\setminus\{1,n\}} p_{i,S}(\theta') - p_{i,S}(\theta)\right)^2 \\
&\leq 2\left(3kb\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2, \qquad (67)
\end{aligned}$$

which is because for every comparison set S such that $1 \in S$,

$$\begin{aligned}
&p_{1,S}(\theta') \leq \frac{1}{k} \leq p_{1,S}(\theta) \quad \text{and} \\
&p_{i,S}(\theta') \geq p_{i,S}(\theta) \quad \forall i \neq 1;
\end{aligned}$$

for every comparison set S such that $n \in S$,

$$\begin{aligned}
&p_{n,S}(\theta') \geq \frac{1}{k} \geq p_{n,S}(\theta) \quad \text{and} \\
&p_{i,S}(\theta') \leq p_{i,S}(\theta) \quad \forall i \neq n.
\end{aligned}$$

From (66) and the assumption of the theorem, we have

$$\begin{aligned}
\min_S \min_{i\in S} p_{i,S}(\theta) = \min_{S:n\in S} p_{n,S}(\theta) \\
\geq \frac{1}{k} - 3kb\frac{\partial p_k(\mathbf{0})}{\partial x_1} \\
\geq \frac{1}{2k}. \qquad (68)
\end{aligned}$$

For simplicity of notation, let

$$D = 3kb\frac{\partial p_k(\mathbf{0})}{\partial x_1}. \qquad (69)$$

Then, for all S such that $S \cap \{1,n\} \neq \emptyset$, we have

$$\sum_{i\in S} p_{i,S}(\theta') \log\left(\frac{p_{i,S}(\theta')}{p_{i,S}(\theta)}\right) \leq 2kD^2 \qquad (70)$$

which is obtained from

(i) $p_{i,S}(\theta) \geq 1/(2k)$ for all $i \in S$ that holds by (68),

(ii) $\sum_{i\in S}(p_{i,S}(\theta') - p_{i,S}(\theta))^2 = 2D^2$ from (67),

(iii) $a\log\frac{a}{b} \leq \frac{(a-b)^2}{2b} + a - b$.

Similarly to (70), from (i) and (ii) and $a\left(\log\frac{a}{b}\right)^2 \leq \frac{(a-b)^2}{a\wedge b}\left(1 + \frac{|a-b|}{3(a\wedge b)}\right)$, we have

$$\sum_{i\in S} p_{i,S}(\theta')\left(\log\left(\frac{p_{i,S}(\theta')}{p_{i,S}(\theta)}\right)\right)^2 \leq 8kD^2. \qquad (71)$$

Since

$$\mathbf{P}_{\theta'}[\{S_t \cap \{1,n\} \neq \emptyset\}] = 1 - \frac{\binom{n-2}{k}}{\binom{n}{k}} \leq 2\frac{k}{n}$$

and according to the model, the input observations are independent, from (70) and (71), we have

$$\begin{aligned}
&\mathbf{E}_{\theta'}[L(\mathbf{S},\mathbf{y})] \\
&= m\mathbf{E}_{\theta'}\left[\log\left(\frac{p_{y_1,S_1}(\theta')}{p_{y_1,S_1}(\theta)}\right)\right] \\
&= m\sum_{S:S\cap\{1,n\}\neq\emptyset} \mathbf{P}_{\theta'}[S_1 = S] \\
&\quad \sum_{y\in S} p_{y,S}(\theta')\left[\log\left(\frac{p_{y,S}(\theta')}{p_{y,S}(\theta)}\right)\right] \\
&\leq 4\frac{m}{n}k^2D^2 \qquad (72)
\end{aligned}$$

and

$$\begin{aligned}
&\sigma^2_{\theta'}[L(\mathbf{S},\mathbf{y})] \\
&= m\sigma^2_{\theta'}\left[\log\left(\frac{p_{y_1,S_1}(\theta')}{p_{y_1,S_1}(\theta)}\right)\right] \\
&\leq m\mathbf{E}_{\theta'}\left[\left(\log\left(\frac{p_{y_1,S_1}(\theta')}{p_{y_1,S_1}(\theta)}\right)\right)^2\right] \\
&= m\sum_{S:S\cap\{1,n\}\neq\emptyset} \mathbf{P}_{\theta'}[S_1 = S] \\
&\quad \sum_{y\in S} p_{y,S}(\theta')\left[\left(\log\left(\frac{p_{y,S}(\theta')}{p_{y,S}(\theta)}\right)\right)^2\right] \\
&\leq 16\frac{m}{n}k^2D^2. \qquad (73)
\end{aligned}$$

I. Characterizations of $\partial p_k(\mathbf{0})/\partial x_1$

In this section, we note several different representations of the parameter $\partial p_k(\mathbf{0})/\partial x_1$.

First, note that

$$\frac{\partial p_k(\mathbf{0})}{\partial x_1} = \frac{1}{k-1}\int_{\mathbf{R}} f(x)dF(x)^{k-1}. \qquad (74)$$

The integral corresponds to $\mathbf{E}[f(X)]$ where X is a random variable whose distribution is equal to that of a maximum of $k-1$ independent and identically distributed random variables with cumulative distribution F.

Second, suppose that F is a cumulative distribution function with its support contained in $[-a,a]$, and that has a differentiable density function f. Then, we have

$$\frac{\partial p_k(\mathbf{0})}{\partial x_1} = A_{F,k} + B_{F,k} \quad (75)$$

where

$$A_{F,k} = \frac{1}{k-1} f(a)$$

and

$$B_{F,k} = \frac{1}{k(k-1)} \int_{-a}^{a} (-f'(x)) dF(x)^k.$$

The identity (75) is shown to hold as follows. Note that

$$\begin{aligned} & \frac{d^2}{dx^2} F(x)^k \\ = & \frac{d}{dx}(kF(x)^{k-1} f(x)) \\ = & k(k-1)F(x)^{k-2} f(x)^2 + kF(x)^{k-1} f'(x). \end{aligned}$$

By integrating over $[-a, a]$, we obtain

$$\begin{aligned} \frac{d}{dx} F(x)^k |_{-a}^{a} &= k(k-1) \int_{-a}^{a} f(x)^2 F(x)^{k-2} dx \\ &\quad + k \int_{-a}^{a} f'(x) F(x)^{k-1} dx. \end{aligned}$$

Combining with the fact

$$\frac{d}{dx} F(x)^k |_{-a}^{a} = k f(x) F^{k-1}(x) |_{-a}^{a} = k f(a),$$

we obtain (75).

Note that $B_{F,k} = \mathbf{E}[-f'(X)]/(k(k-1))$ where X is a random variable with distribution that corresponds to that of a maximum of k independent samples from the cumulative distribution function F. Note also that if, in addition, f is an even function, then (i) $B_{F,k} \geq 0$ and (ii) $B_{F,k}$ is increasing in k.

Third, for any cumulative distribution function F with an even density function f, we have $F(-x) = 1 - F(x)$ for all $x \in \mathbf{R}$. In this case, we have the identity

$$\frac{\partial p_k(\mathbf{0})}{\partial x_1} = \int_0^\infty f(x)^2 (F(x)^{k-2} + (1 - F(x))^{k-2}) dx. \quad (76)$$

J. Proof of Proposition 6

The upper bound follows by noting that that $B_{F,k}$ in (75) is such that $B_{F,k} = \Omega(1/k^2)$. Hence, it follows that

$$\gamma_{F,k} = O(1).$$

The lower bound follows by noting that for every cumulative distribution function F such that there exists a constant $C > 0$ such that $f(x) \leq C$ for all $x \in \mathbf{R}$,

$$\begin{aligned} \frac{\partial p_k(\mathbf{0})}{\partial x_1} &= \int_{\mathbf{R}} f(x)^2 F(x)^{k-2} dx \\ &\leq C \int_{\mathbf{R}} f(x) F(x)^{k-2} dx \\ &= C \frac{1}{k-1}. \end{aligned}$$

Hence, $\gamma_{F,k} \geq (1/C)(k-1)/k^3 = \Omega(1/k^2)$.

K. Derivations of parameter $\gamma_{F,k}$

We derive explicit expressions for parameter $\gamma_{F,k}$ for our example generalized Thurstone choice models introduced in Section 2

Recall from (7) that we have that

$$\gamma_{F,k} = \frac{1}{(k-1)k^3} \frac{1}{(\partial p_k(\mathbf{0})/\partial x_1)^2}$$

where

$$\frac{\partial p_k(\mathbf{0})}{\partial x_1} = \int_{\mathbf{R}} f(x)^2 F(x)^{k-2} dx$$

Gaussian distribution A cumulative distribution function F is said to have a type-3 domain of maximum attraction if the maximum of r independent and identically distributed random variables with cumulative distribution function F has as a limit a double-exponential cumulative distribution function:

$$e^{-e^{-\frac{x-a_r}{b_r}}}$$

where

$$a_r = F^{-1}\left(1 - \frac{1}{r}\right)$$

and

$$b_r = F^{-1}\left(1 - \frac{1}{er}\right) - F^{-1}\left(1 - \frac{1}{r}\right).$$

It is a well known fact that any Gaussian cumulative distribution function has a type-3 domain of maximum attraction. Let Φ denote the cumulative distribution function of a standard normal random variable, and let ϕ denotes its density.

Note that

$$\begin{aligned}
& \int_{\mathbf{R}} \phi(x) d\Phi(x)^r \\
\sim\ & \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} d(e^{-e^{-\frac{x-a_r}{b_r}}}) \\
=\ & \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(a_r + b_r \log(1/z))^2} e^{-z} dz \\
=\ & \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_r^2} \int_0^\infty z^{a_r b_r} e^{-\frac{1}{2}b_r^2 \log(1/z)^2} e^{-z} dz \\
\leq\ & \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_r^2} \int_0^\infty z^{a_r b_r} e^{-z} dz \\
=\ & \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a_r^2} \Gamma(a_r b_r + 1).
\end{aligned}$$

Now, note that

$$a_r \sim \sqrt{2\log(r)} \text{ and } b_r = \Theta(1), \text{ for large } r.$$

It is readily checked that $e^{-a_r^2/2} \sim 1/r$ and $\Gamma(a_r b_r + 1) = O(r^\epsilon)$ for every constant $\epsilon > 0$. Hence, we have that

$$\int_{\mathbf{R}} \phi(x) d\Phi(x)^r = O(1/r^{1-\epsilon})$$

and thus, $\partial p_k(\mathbf{0})/\partial x_1 = O(1/k^{2-\epsilon})$. Hence,

$$\gamma_{F,k} = \Omega(1/k^{2\epsilon}).$$

Double-exponential distribution Note that $f(x) = \frac{1}{\beta} e^{-\frac{x+\beta\gamma}{\beta}} F(x)$. Hence, we have

$$\begin{aligned}
\frac{\partial p_k(\mathbf{0})}{\partial x_1} &= \int_{\mathbf{R}} f(x)^2 F(x)^{k-2} dx \\
&= \frac{1}{\beta^2} \int_{\mathbf{R}} e^{-2\frac{x+\beta\gamma}{\beta}} F(x)^k dx \\
&= \frac{1}{\beta} \int_0^\infty z e^{-kz} dz \\
&= \frac{1}{\beta k^2}.
\end{aligned}$$

Laplace distribution Let $\beta = \sigma/\sqrt{2}$. Note that

$$F(x) = 1 - \frac{1}{2} e^{-x/\beta} \text{ and } f(x) = \frac{1}{2\beta} e^{-x/\beta}, \text{ for } x \in \mathbf{R}_+.$$

$$\begin{aligned}
A &= \int_0^\infty f(x)^2 F(x)^{k-2} dx \\
&= \int_0^\infty \left(\frac{1}{2\beta}\right)^2 e^{-2x/\beta} \left(1 - \frac{1}{2} e^{-x/\beta}\right)^{k-2} dx \\
&= \frac{1}{2\beta} \int_{1/2}^1 2(1-z) z^{k-2} dz \\
&= \frac{1}{\beta} \left(\frac{1}{k-1} \left(1 - \frac{1}{2^{k-1}}\right) - \frac{1}{k} \left(1 - \frac{1}{2^k}\right) \right) \\
&= \frac{1}{\beta k(k-1)} \left(1 - \frac{k}{2^{k-1}} + \frac{k-1}{2^k}\right)
\end{aligned}$$

and

$$\begin{aligned}
B &= \int_0^\infty f(x)^2 (1 - F(x))^{k-2} dx \\
&= \int_0^\infty \left(\frac{1}{2\beta}\right)^2 e^{-2x/\beta} \frac{1}{2^{k-2}} e^{-(k-2)x/\beta} dx \\
&= \frac{1}{\beta^2 2^k} \int_0^\infty e^{-kx/\beta} dx \\
&= \frac{1}{\beta k 2^k}.
\end{aligned}$$

Combining with (76), we obtain

$$\frac{\partial p_k(\mathbf{0})}{\partial x_1} = A + B = \frac{1}{\beta k(k-1)} \left(1 - \frac{1}{2^{k-1}}\right).$$

Uniform distribution Note that

$$\begin{aligned}
\frac{\partial p_k(\mathbf{0})}{\partial x_1} &= \int_{\mathbf{R}} f(x)^2 F(x)^{k-2} dx \\
&= \frac{1}{(2a)^2} \int_{-a}^a \left(\frac{x+a}{2a}\right)^{k-2} dx \\
&= \frac{1}{2a} \int_0^1 z^{k-2} dz \\
&= \frac{1}{2a(k-1)}.
\end{aligned}$$

L. Some Remaining Proofs

L.1. Proof of Lemma 12

Consider a set $S \subseteq N$ such that $|S| = k$, for an arbitrary integer $2 \leq k \leq n$. Without loss of generality, consider $S = \{1, 2, \ldots, k\}$. Let $\mathbf{x}_l(\theta) = \theta_i - \theta_{S \setminus \{l\}}$, for $l \in S$. For simplicity, with a slight abuse of notation, we write x_l in lie of $x_i(\theta)$, for $l \in S$. We first consider the case $i \neq j$. By straightforward derivation, we have

$$\begin{aligned}
& \frac{\partial^2}{\partial\theta_i \partial\theta_j} (-\log(p_k(\mathbf{x}_l))) \\
=\ & -\frac{1}{p_k(\mathbf{x}_l)} \frac{\partial^2 p_k(\mathbf{x}_l)}{\partial\theta_i \partial\theta_j} + \frac{1}{p_k(\mathbf{x}_l)^2} \frac{\partial p_k(\mathbf{x}_l)}{\partial\theta_i} \frac{\partial p_k(\mathbf{x}_l)}{\partial\theta_j}.
\end{aligned}$$

We separately consider three different cases.

Case 1: i, j, l are all distinct. Note that

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_l)))|_{\theta=\mathbf{0}} = I_1 \tag{77}$$

where

$$I_1 = -k\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1\partial x_2} + k^2\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2.$$

Case 2: $i \neq l$ and $j = l$. In this case, we characterize the following quantity for $\theta = \mathbf{0}$,

$$\begin{aligned}&\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_j)))\\ = \quad & -\frac{1}{p_k(\mathbf{x}_j)}\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i\partial\theta_j} + \frac{1}{p_k(\mathbf{x}_j)^2}\frac{\partial p_k(\mathbf{x}_j)}{\partial\theta_i}\frac{\partial p_k(\mathbf{x}_j)}{\partial\theta_j}\end{aligned} \tag{78}$$

For every $u \in S$, $p_k(\mathbf{x}_u)$ does not change its value by changing the parameter θ to value $\theta + \Delta\theta$, for every constant $\Delta\theta \in \mathbf{R}$. Hence, by the full differential, we have

$$\frac{\partial p_k(\mathbf{x}_u)}{\partial\theta_j} = -\sum_{v\in S\setminus\{j\}}\frac{\partial p_k(\mathbf{x}_u)}{\partial\theta_v}. \tag{79}$$

Using (79), we have

$$\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i\partial\theta_j} = -\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i^2} - \sum_{v\in S\setminus\{i,j\}}\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i\partial\theta_v}. \tag{80}$$

Note that

$$\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i^2} = \int_{\mathbf{R}} f(z)f'(x_i+z)\prod_{v\in S\setminus\{i,j\}}F(x_v+z)dz.$$

Hence, we can derive

$$\begin{aligned}&\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i^2}|_{\theta=\mathbf{0}}\\ = \quad & \int_{\mathbf{R}} f(z)f'(z)\prod_{v\in S\setminus\{i,j\}}F(z)^{k-2}dz\\ = \quad & f(z)^2F(z)^{k-1}|_{-\infty}^{\infty} - \int_{\mathbf{R}} f(z)(f(z)F(z)^{k-2})'dz\\ = \quad & -\int_{\mathbf{R}} f(z)f'(z)F(z)^{k-1} - (k-2)\int_{\mathbf{R}} f(z)^2F(z)^{k-3}dz\\ = \quad & -\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i^2}|_{\theta=\mathbf{0}} - (k-2)\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1\partial x_2}\end{aligned}$$

from which it follows that

$$\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i^2}|_{\theta=\mathbf{0}} = -\frac{k-2}{2}\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1\partial x_2}. \tag{81}$$

From (80) and (81), we obtain

$$\frac{\partial^2 p_k(\mathbf{x}_j)}{\partial\theta_i\partial\theta_j}|_{\theta=\mathbf{0}} = -\frac{k-2}{2}\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1\partial x_2}. \tag{82}$$

Using (79), we have

$$\frac{\partial p_k(\mathbf{x}_j)}{\partial\theta_j}|_{\theta=\mathbf{0}} = (k-1)\frac{\partial p_k(\mathbf{0})}{\partial x_1}. \tag{83}$$

Combining (78), (82) and (83), we have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_j)))|_{\theta=\mathbf{0}} = I_2 \tag{84}$$

where

$$I_2 = \frac{k(k-2)}{2}\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1\partial x_2} - k^2(k-1)\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2.$$

Case 3: $i = l$ and $j \neq l$. By symmetry, from Case 2, we have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_i)))|_{\theta=\mathbf{0}} = I_2. \tag{85}$$

Final step Putting the pieces together, from (77), (84), and (85), we have for $\theta = \mathbf{0}$,

$$\begin{aligned}H_{i,j}(\theta,S) \quad = \quad & \sum_{l\in S\setminus\{i,j\}} p_k(\mathbf{x}_l)\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_l)))\\ & +p_k(\mathbf{x}_i)\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_i)))\\ & +p_k(\mathbf{x}_l)\frac{\partial^2}{\partial\theta_i\partial\theta_j}(-\log(p_k(\mathbf{x}_j)))\\ = \quad & \frac{k-2}{k}I_1 + \frac{1}{k}I_2 + \frac{1}{k}I_2\\ = \quad & -k^2\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2.\end{aligned} \tag{86}$$

Now, we consider the case $i = j$. Using same argument as in (79), we have

$$\frac{\partial^2(-\log(p_k(\mathbf{x}_l)))}{\partial\theta_i^2} = -\sum_{v\in S\setminus\{i\}}\frac{\partial^2(-\log(p_k(\mathbf{x}_l)))}{\partial\theta_i\partial\theta_v}.$$

Hence,

$$H_{i,i}(\theta,S) = -\sum_{v\in S\setminus\{i\}}\sum_{l\in S}p_k(\mathbf{x_l})\frac{\partial^2(-\log(p_k(\mathbf{x}_l)))}{\partial\theta_i\partial\theta_v}.$$

Combining with $H_{i,i}(\theta,S) = -\sum_{v\in S\setminus\{i\}}H_{i,v}(\theta,S)$ and the result established in (86), we have for $\theta = \mathbf{0}$,

$$H_{i,i}(\theta,S) = k^2(k-1)\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2.$$

L.2. Proof of Lemma 14

Without loss of generality, let $y = 1$ and $S = \{1, \dots, k\}$. Then, we have

$$\frac{1}{k^2}\frac{\partial^2(-\log(p_{1,S}(\mathbf{0})))}{\partial\theta_1\partial\theta_2} = -(k-1)\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 + \frac{k-2}{2k}\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1 \partial x_2} \quad (87)$$

and for $i \neq 1$ and $j \neq i$,

$$\frac{1}{k^2}\frac{\partial^2(-\log(p_{1,S}(\mathbf{0})))}{\partial\theta_i\partial\theta_j} = -\frac{1}{k}\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1 \partial x_2} + \left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2. \quad (88)$$

From assumption **A1** and (87),

$$\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1 \partial x_2} \le \frac{2k(k-1)}{k-2}\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2. \quad (89)$$

Note that it holds that

$$\frac{\partial^2 p_k(\mathbf{0})}{\partial x_1 \partial x_2} \ge 0. \quad (90)$$

Combining (87), (88), (89), and (90), we have

$$\frac{\partial^2(-\log(p_{y,S}(\mathbf{0})))}{\partial\theta_i\partial\theta_j} \ge -k^3\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 \quad \forall \quad i \neq j.$$

From the above inequality and (21),

$$k^3\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2 \Lambda_{\mathbf{M}_S} \succeq \nabla(-\log(p_{y,S}(\mathbf{0}))).$$

Therefore, we conclude

$$\|\nabla^2(-\log(p_{y,S}(\mathbf{0})))\|_2 \le k^4\left(\frac{\partial p_k(\mathbf{0})}{\partial x_1}\right)^2,$$

which holds because $\|\Lambda_{\mathbf{M}_S}\|_2 = k$.

L.3. Proof of Proposition 11

We prove only (19) as the proof of (18) follows by similar arguments.

By Chernoff's bound, for every $s > 0$,

$$\begin{aligned}\mathbf{P}[X \ge qm] &\le e^{-sqm}\mathbf{E}[e^{sX}] \\ &= e^{-sqm}(1-p+pe^s)^m \\ &= e^{-mh(s)}\end{aligned}$$

where

$$h(s) = qs - \log(1-p+pe^s)$$

Now, using the elementary fact $\log(1-x) \le -x$, we have

$$h(s) \ge qs + p - pe^s.$$

Take $s = s^* := \log(q/p)$, then,

$$h(s^*) \ge q\log\left(\frac{q}{p}\right) + p - q.$$

Now, let $\epsilon = q - p$, and note that

$$q\log\left(\frac{q}{p}\right) + p - q \quad := \quad g(\epsilon)$$

where

$$g(\epsilon) = q\log\left(\frac{q}{q-\epsilon}\right) - \epsilon.$$

Since

$$g'(\epsilon) = \frac{q}{q-\epsilon} - 1 = \frac{\epsilon}{q-\epsilon} \ge \frac{1}{2q}\epsilon$$

we have

$$g(\epsilon) = \int_0^\epsilon g'(x)dx \ge \frac{1}{4q}\epsilon^2$$

Hence, it follows that

$$h(s^*) \ge \frac{1}{4q}(p-q)^2$$

and, thus,

$$\mathbf{P}[X \ge qm] \le \exp\left(-\frac{1}{4q}(p-q)^2\right).$$

M. Experimental Results: Fiedler Values of Pair-weight Matrices

We found that Fiedler value of a pair-weight matrix is an important factor that determines the mean square error in Section 3.1 and Section 3.2. In the supplementary material, we evaluate Fiedler value for different pair-weight matrices of different schedules of comparisons. Throughout this section, we use the definition of a pair-weight matrix in (5) with the weight function $w(k) = 1/k^2$. Our first two examples are representative of schedules in sport competitions, which are typically carefully designed by sport associations and exhibit a large degree of regularity. Our second two examples are representative of comparisons that are induced by user choices in the context of online services, which exhibit much more irregularity. Please refer to the supplementary material for more details.

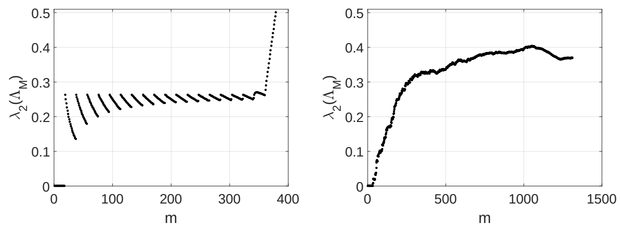

Figure 1. Fiedler value of the pair-weight matrices for the game fixtures of two sports in the season 2014-2015: (left) football Barclays premier league, and (right) basketball NBA league.

Sport competitions We consider the fixtures of games for the season 2014-2015 for (i) football Barclays premier league and (ii) basketball NBA league. In the Barclays premier league, there are 20 teams, each team plays a home and an away game with each other team; thus there are 380 games in total. In the NBA league, there are 30 teams, 1,230 regular games, and 81 playoff games.[1] We evaluate Fiedler value of pair-weight matrices defined for first m matches of each season; see Figure 1.

For the Barclays premier league dataset, at the end of the season, the Fiedler value of the pair-weight matrix is of value $n/[2(n-1)] \approx 1/2$. The schedule of matches is such that at the middle of the season, each team played against each other team exactly once, at which point the Fiedler value is $n/[4(n-1)] \approx 1/4$. The Fiedler value is of a strictly positive value after the first round of matches. For most part of the season, its value is near to $1/4$ and it grows to the highest value of approximately $1/2$ in the last round of the matches.

For the NBA league dataset, at the end of the season, the Fiedler value of the pair-weight matrix is approximately 0.375. It grows more slowly with the number of games played than for the Barclays premier league; this is intuitive as the schedule of games is more irregular, with each team not playing against each other team the same number of times.

Crowdsourcing contests We consider participation of users in contests of two competition-based online labour platforms: (i) online platform for software development TopCoder and (ii) online platform for various kinds of tasks Tackcn. We refer to coders in TopCoder and workers in Taskcn as users. We consider contests of different categories observed in year 2012; more information about datasets is provided in Appendix. We present results only for one category of tasks for each system, which are representative. In both these systems, the participation in contests is according to choices made by users.

[1]The NBA league consists of two conferences, each with three divisions, and the fixture of games has to obey constraints on the number of games played between teams from different divisions.

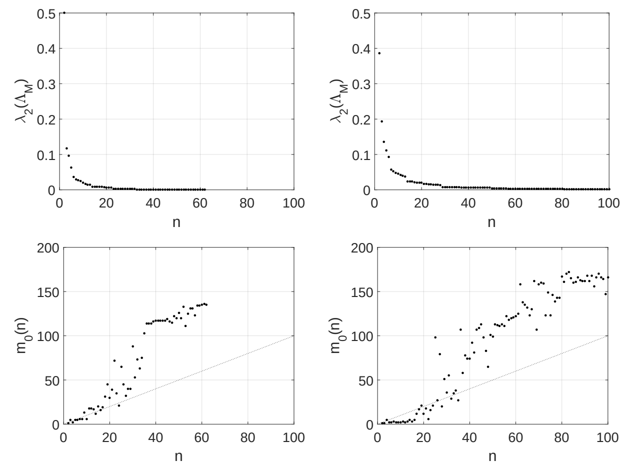

Figure 2. (Left) Topcoder data restricted to top-n coders and (Right) same as left but for Taskcn, for Design and Website task categories, respectively. The top plots show the Fiedler value and the bottom plots show the minimum number of contests to observe a strictly positive Fiedler value.

For each set of tasks of given category, we conduct the following analysis. We consider a thinned dataset that consists only of a set of top-n users with respect to the number of contests they participated in given year, and of all contests attended by at least two users from this set. We then evaluate Fiedler value of the pair-weight matrix for parameter n ranging from 2 to the smaller of 100 or the total number of users. Our analysis reveals that the Fiedler value tends to decrease with n. This indicates that the larger the number of users included, the less connected the pair-weight matrix is. See the top plots in Figure 2.

We also evaluated the smallest number of contests from the beginning of the year that is needed for the Fiedler value of the pair-weight matrix to assume a strictly positive value. See the bottom plots in Figure 2. We observe that this threshold number of contests tends to increase with the number of top users considered. There are instances for which this threshold substantially increases for some number of the top users. This, again, indicates that the algebraic connectivity of the pair-weight matrices tends to decrease with the number of top users considered.

References

Boyd, Stephen. Convex optimization of graph Laplacian eigenvalues. In *Proceedings of the International Congress of Mathematicians*, pp. 1311–1319, 2006.

Hayes, Thomas P. A large-deviation inequality for vector-valued martingales. URL http://www.cs.unm.edu/~hayes/papers/VectorAzuma/VectorAzuma20030207.pdf.

Table 1. Summary statistics for TopCoder and Taskcn datasets. The rightmost two columns contain, repectively, mean and median values of comparison sets' cardinalities.

Category	# contests	# workers	mean	median
TopCoder				
Design	209	62	1.99	2
Development	198	171	3.07	2
Specification	75	39	2.39	2
Architecture	238	55	1.75	2
Taskcn				
Website	131	636	9.87	6
Design	1,967	6,891	27.3	18
Coding	31	284	27.1	18
Writing	420	15,575	46.11	19

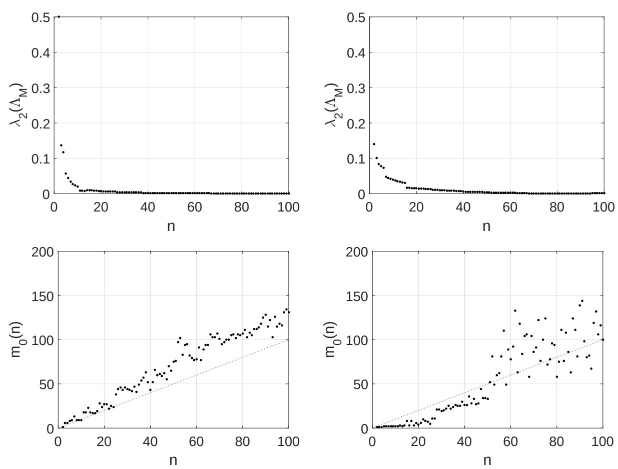

Figure 3. Same as in Figure 3 but for different categories (Development and Writing).

Tropp, Joel A. An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*, 2015.