Supplementary material for A Learning Theory of Ranking Aggregation

A Optimality

A.0.1 Proof of Remark 6

Suppose P satisfies the *strongly stochastically transitive* condition. According to Theorem 5, there exists $\sigma^* \in \mathfrak{S}_n$ satisfying (9) and (11). We already know that σ^* is a Kemeny consensus since it minimizes the loss with respect to the Kendall's τ distance. Then, Copeland's method order the items by the number of their pairwise victories, which corresponds to sort them according to the mapping s^* and thus σ^* is a Copeland consensus. Finally, the Borda score for an item is: $s(i) = \sum_{\sigma \in \mathfrak{S}_n} \sigma(i) P(\sigma)$. Firstly observe that for any $\sigma \in \mathfrak{S}_n$,

$$\sum_{k \neq i} \mathbb{I}\{\sigma(k) < \sigma(i)\} - \sum_{k \neq i} \mathbb{I}\{\sigma(k) > \sigma(i)\} = \sigma(i) - 1 - (n - \sigma(i)) = 2\sigma(i) - (n+1). \tag{1}$$

According to (1), we have the following calculations:

$$\begin{aligned}
s(i) &= \sum_{\sigma \in \mathfrak{S}_n} \frac{1}{2} \left(n + 1 + \sum_{k \neq i} (2\mathbb{I}\{\sigma(k) < \sigma(i)\} - 1) \right) P(\sigma) \\
&= \frac{n+1}{2} + \frac{1}{2} \left(\sum_{\sigma \in \mathfrak{S}_n} \sum_{k \neq i} 2\mathbb{I}\{\sigma(k) < \sigma(i)\} - (n-1) \right) P(\sigma) \\
&= \frac{n+1}{2} - \frac{n-1}{2} + \sum_{k \neq i} \sum_{\sigma \in \mathfrak{S}_n} \mathbb{I}\{\sigma(k) < \sigma(i)\} P(\sigma) \\
&= 1 + \sum_{k \neq i} p_{k,i}.
\end{aligned}$$

Let i, j such that $p_{i,j} > 1/2$ ($\Leftrightarrow s^*(i) < s^*(j)$ under *stochastic transitivity*).

$$\begin{aligned}
s(j) - s(i) &= \sum_{k \neq j} p_{k,j} - \sum_{k \neq i} p_{k,i} \\
&= \sum_{k \neq i,j} p_{k,j} - \sum_{k \neq i,j} p_{k,i} + p_{i,j} - p_{j,i} \\
&= \sum_{k \neq i,j} p_{k,j} - p_{k,i} + (2p_{i,j} - 1)
\end{aligned}$$

With $(2p_{i,j} - 1) > 0$. Now we focus on the first term, and consider $k \neq i, j$.

(i) First case: $p_{j,k} \geq 1/2$. The strong stochastic transitivity condition implies that :

$$\begin{aligned}
p_{i,k} &\geq \max(p_{i,j}, p_{j,k}) \\
1 - p_{k,i} &\geq \max(p_{i,j}, p_{j,k}) \\
p_{k,j} - p_{k,i} &\geq p_{k,j} - 1 + \max(p_{i,j}, p_{j,k}) \\
p_{k,j} - p_{k,i} &\geq -p_{j,k} + \max(p_{i,j}, p_{j,k}) \\
p_{k,j} - p_{k,i} &\geq \max(p_{i,j} - p_{j,k}, 0) \\
p_{k,j} - p_{k,i} &\geq 0.
\end{aligned}$$

(ii) Second case: $p_{k,j} > 1/2$. If $p_{k,i} \leq 1/2$, $p_{k,j} - p_{k,i} > 0$. Now if $p_{k,i} > 1/2$, having $p_{i,j} > 1/2$, the strong stochastic transitivity condition implies that $p_{k,j} \geq max(p_{k,i}, p_{i,j})$.

Therefore in any case, $\forall k \neq i, j$, $p_{k,j} - p_{k,i} \geq 0$ and $s(j) - s(i) > 0$.

B Empirical consensus

B.1 Universal rates

We can obtain upper bounds using (16) and some calculations on L as follows. First notice that same as in (7) one has for any $\sigma \in \mathfrak{S}_n$:

$$\widehat{L}_N(\sigma) = \binom{n}{2} \mathbb{E}\left[\widehat{p}_{\mathbf{i},\mathbf{j}} \mathbb{I}\{\sigma(\mathbf{i}) > \sigma(\mathbf{j})\} + (1 - \widehat{p}_{\mathbf{i},\mathbf{j}}) \mathbb{I}\{\sigma(\mathbf{i}) < \sigma(\mathbf{j})\}\right]$$

so that

$$\begin{aligned}
|\widehat{L}_N(\sigma) - L(\sigma)| &= \binom{n}{2} \left| \mathbb{E}\left[(\widehat{p}_{\mathbf{i},\mathbf{j}} - p_{\mathbf{i},\mathbf{j}}) \mathbb{I}\{\sigma(\mathbf{i}) > \sigma(\mathbf{j})\} - (\widehat{p}_{\mathbf{i},\mathbf{j}} - p_{\mathbf{i},\mathbf{j}}) \mathbb{I}\{\sigma(\mathbf{i}) < \sigma(\mathbf{j})\}\right] \right| \\
&\leq \binom{n}{2} \mathbb{E}_{\mathbf{i},\mathbf{j}}\left[|p_{\mathbf{i},\mathbf{j}} - \widehat{p}_{\mathbf{i},\mathbf{j}}|\right].
\end{aligned}$$

B.1.1 Proof of Proposition 9

(i) By the Cauchy-Schwartz inequality,

$$\mathbb{E}_{\mathbf{i},\mathbf{j}}\left[|p_{\mathbf{i},\mathbf{j}} - \widehat{p}_{\mathbf{i},\mathbf{j}}|\right] \leq \sqrt{\mathbb{E}_{\mathbf{i},\mathbf{j}}\left[(p_{\mathbf{i},\mathbf{j}} - \widehat{p}_{\mathbf{i},\mathbf{j}})^2\right]} = \sqrt{Var(\widehat{p}_{\mathbf{i},\mathbf{j}})}.$$

Since $\mathbb{E}_{\mathbf{i},\mathbf{j}}\left[p_{\mathbf{i},\mathbf{j}} - \widehat{p}_{\mathbf{i},\mathbf{j}}\right] = 0$. Then, for $i < j$, $N\widehat{p}_{i,j} \sim \mathcal{B}(N, p_{i,j}))$ so $Var(\widehat{p}_{i,j}) = \frac{p_{i,j}(1-p_{i,j})}{N} \leq \frac{1}{4N}$. Finally, we can upper bound the expectation of the excess of risk as follows:

$$\mathbb{E}\left[L(\widehat{\sigma}_N) - L^*\right] \leq 2\mathbb{E}\left[\max_{\sigma \in \mathfrak{S}_n} |\widehat{L}_N(\sigma) - L(\sigma)|\right] \leq 2\binom{n}{2} \frac{1}{\sqrt{4N}} = \frac{n(n-1)}{2\sqrt{N}}.$$

(ii) By (16) one has for any $t > 0$

$$\mathbb{P}\Big\{L(\widehat{\sigma}_N) - L^* > t\Big\} \leq \mathbb{P}\Big\{2\binom{n}{2}\mathbb{E}_{\mathbf{i},\mathbf{j}}\left[|p_{\mathbf{i},\mathbf{j}} - \widehat{p}_{\mathbf{i},\mathbf{j}}|\right] > t\Big\} = \mathbb{P}\Big\{\sum_{1\leq i<j\leq n} |p_{i,j} - \widehat{p}_{i,j}| > \frac{t}{2}\Big\}, \quad (2)$$

and the other hand, it holds that

$$\mathbb{P}\Big\{\sum_{1\leq i<j\leq n} |p_{i,j}-\widehat{p}_{i,j}| > \frac{t}{2}\Big\} \leq \mathbb{P}\Big\{\bigcup_{1\leq i<j\leq n}\Big\{|p_{i,j}-\widehat{p}_{i,j}| > \frac{t}{2\binom{n}{2}}\Big\}\Big\} \leq \sum_{1\leq i<j\leq n} \mathbb{P}\Big\{|p_{i,j}-\widehat{p}_{i,j}| > \frac{t}{2\binom{n}{2}}\Big\}. \quad (3)$$

Now, Hoeffding's inequality to $\widehat{p}_{i,j} = (1/N)\sum_{t=1}^{N} \mathbb{I}\{\Sigma_t(i) < \Sigma_t(j)\}$ gives

$$\mathbb{P}\Big\{|p_{i,j} - \widehat{p}_{i,j}| > \frac{t}{2\binom{n}{2}}\Big\} \leq 2e^{-2N(t/2\binom{n}{2})^2}. \quad (4)$$

Therefore, combining (2), (3) and (4) we get

$$\mathbb{P}\Big\{L(\widehat{\sigma}_N) - L^* > t\Big\} \leq 2\binom{n}{2}e^{-\frac{Nt^2}{2\binom{n}{2}^2}}.$$

Setting $\delta = 2\binom{n}{2}e^{-\frac{Nt^2}{2\binom{n}{2}^2}}$ one obtains that with probability greater than $1 - \delta$,

$$L(\widehat{\sigma}_N) - L^* \leq \binom{n}{2}\sqrt{\frac{2\log(n(n-1)/\delta)}{N}}.$$

B.1.2 Proof of Proposition 11

In the following proof, we follow Le Cam's method, see section 2.3 in Tsybakov (2009).

Consider two Mallows models P_{θ_0} and P_{θ_1} where $\theta_k = (\sigma^*_k, \phi) \in \mathfrak{S}_n \times (0,1)$ and $\sigma^*_0 \neq \sigma^*_1$. We clearly have:

$$\begin{aligned}
\mathcal{R}_N \geq & \inf_{\sigma_N} \max_{k=0,\,1} \mathbb{E}_{P_{\theta_k}}\left[L_{P_{\theta_k}}(\sigma_N) - L^*_{P_{\theta_k}}\right] \\
= & \inf_{\sigma_N} \max_{k=0,\,1} \sum_{i<j} \mathbb{E}_{P_{\theta_k}}\left[2|p_{i,j} - \frac{1}{2}| \times \mathbb{I}\{(\sigma_N(i) - \sigma_N(j)(\sigma^*_k(i) - \sigma^*_k(j)) < 0\}\right] \\
\geq & \inf_{\sigma_N} \frac{|\phi - 1|}{(1+\phi)} \max_{k=0,\,1} \sum_{i<j} \mathbb{E}_{P_{\theta_k}}\left[\mathbb{I}\{(\sigma_N(i) - \sigma_N(j)(\sigma^*_k(i) - \sigma^*_k(j)) < 0\}\right] \\
\geq & \frac{|\phi - 1|}{2} \inf_{\sigma_N} \max_{k=0,\,1} \mathbb{E}_{P_{\theta_k}}\left[d_\tau(\sigma_N, \sigma^*_k)\right],
\end{aligned}$$

using the fact that $|p_{i,j} - \frac{1}{2}| \geq \frac{|\phi-1|}{2(1+\phi)}$ (based on Corollary 3 from Busa-Fekete et al. (2014), see Remark 7). Set $\Delta = d_\tau(\sigma^*_0, \sigma^*_1) \geq 1$, and consider the test statistic related to σ_N:

$$\psi(\Sigma_1,\ \ldots,\ \Sigma_N) = \mathbb{I}\{d_\tau(\sigma_N, \sigma^*_1) \leq d_\tau(\sigma_N, \sigma^*_0)\ \}.$$

If $\psi = 1$, by triangular inequality, we have:

$$\Delta \leq d_\tau(\sigma_N, \sigma_0^*) + d_\tau(\sigma_N, \sigma_1^*) \leq 2 d_\tau(\sigma_N, \sigma_0^*).$$

Hence, we have

$$\mathbb{E}_{P_{\theta_0}}\left[d_\tau(\sigma_N, \sigma_0^*)\right] \geq \mathbb{E}_{P_{\theta_0}}\left[d_\tau(\sigma_N, \sigma_0^*)\mathbb{I}\{\psi = +1\}\right] \geq \frac{\Delta}{2}\mathbb{P}_{\theta_0}\{\psi = +1\}$$

and similarly

$$\mathbb{E}_{P_{\theta_1}}\left[d_\tau(\sigma_N, \sigma_1^*)\right] \geq \mathbb{E}_{P_{\theta_1}}\left[d_\tau(\sigma_N, \sigma_1^*)\mathbb{I}\{\psi = 0\}\right] \geq \frac{\Delta}{2}\mathbb{P}_{\theta_1}\{\psi = 0\}.$$

Bounding by below the maximum by the average, we have:

$$\begin{aligned}\inf_{\sigma_N} \max_{k=0,\,1} \mathbb{E}_{P_{\theta_k}}\left[d_\tau(\sigma_N, \sigma_k^*)\right] &\geq \inf_{\sigma_N} \frac{\Delta}{2}\frac{1}{2}\left\{\mathbb{P}_{\theta_1}\{\psi = 0\} + \mathbb{P}_{\theta_0}\{\psi = 1\}\right\} \\ &\geq \frac{\Delta}{4}\min_{k=0,\,1}\left\{\mathbb{P}_{\theta_1}\{\psi^* = 0\} + \mathbb{P}_{\theta_0}\{\psi^* = 1\}\right\},\end{aligned}$$

where the last inequality follows from a standard Neyman-Pearson argument, denoting by

$$\psi^*(\Sigma_1,\ \ldots,\ \Sigma_N) = \mathbb{I}\left\{\prod_{i=1}^{N} \frac{P_{\theta_1}(\Sigma_i)}{P_{\theta_0}(\Sigma_i)} \geq 1\right\}$$

the likelihood ratio test statistic. We deduce that

$$\mathcal{R}_N \geq \frac{\Delta|\phi_{-}1|}{8} \sum_{\sigma_i \in \mathfrak{S}_N,\ 1 \leq i \leq N} \min\left\{\prod_{i=1}^{N} P_{\theta_0}(\sigma_i),\ \prod_{i=1}^{N} P_{\theta_1}(\sigma_i)\right\},$$

and with Le Cam's inequality that:

$$\mathcal{R}_N \geq \frac{\Delta|\phi_{-}1|}{16} e^{-NK(P_{\theta_0}||P_{\theta_1})},$$

where $K(P_{\theta_0}||P_{\theta_1}) = \sum_{\sigma \in \mathfrak{S}_N} P_{\theta_0}(\sigma)\log(P_{\theta_0}(\sigma)/P_{\theta_1}(\sigma))$ denotes the Kullback-Leibler divergence. In order to establish a minimax lower bound of order $1/\sqrt{N}$, one should choose $\theta_0 = (\phi_0, \sigma_0)$ and $\theta_1 = (\phi_1, \sigma_1)$ so that, for $k \in \{0,\ 1\}$, $\phi_k \to 1$ and $K(P_{\theta_0}||P_{\theta_1}) \to 0$ as $N \to +\infty$ at appropriate rates.

We consider the special case where $\phi_0 = \phi_1 = \phi$, which results in $Z_0 = Z_1 = Z$ for the normalization constant, and we fix $\sigma_0 \in \mathfrak{S}_n$. Let $i < j$ such that $\sigma_0(i) + 1 = \sigma_0(j)$. We consider $\sigma_1 = (i, j)\sigma_0$ the permutation where the adjacent pair (i, j) has been transposed, so that $\sigma_1(i) = \sigma_1(j) + 1$ and $\Delta = 1$. For any $\sigma \in \mathfrak{S}_n$, notice that

$$d_\tau(\sigma_0, \sigma) - d_\tau(\sigma_1, \sigma) = \mathbb{I}\{(\sigma(i) > \sigma(j)\} - \mathbb{I}\{(\sigma(i) < \sigma(j)\} \tag{5}$$

According to (14), the Kullback-Leibler divergence is given by

$$K(P_{\theta_0}||P_{\theta_1}) = \sum_{\sigma \in \mathfrak{S}_n} P_{\theta_0}(\sigma)\log\left(\phi^{d_\tau(\sigma_0,\sigma) - d_\tau(\sigma_1,\sigma)}\right)$$

And combining it with (5) yields

$$K(P_{\theta_0}||P_{\theta_1}) = \log(\phi) \sum_{\sigma \in \mathfrak{S}_n} P_{\theta_0}(\sigma) \left(\mathbb{I}\{(\sigma(i) > \sigma(j)\} - \mathbb{I}\{(\sigma(i) < \sigma(j)\} \right)$$

By denoting $p^0_{j,i} = P_{\theta_0}[\Sigma(i) < \Sigma(j)]$, this gives us

$$K(P_{\theta_0}||P_{\theta_1}) = \log(\phi) \left(p^0_{j,i} - p^0_{i,j} \right) = \log(\frac{1}{\phi}) \left(2p^0_{i,j} - 1 \right) = \log(\frac{1}{\phi}) \frac{1-\phi}{1+\phi} \tag{6}$$

Where the last equality comes from Busa-Fekete et al. (2014) (Corollary 3 for adjacent items in the central permutation, see also Remark 7).

By taking $\phi = 1 - 1/\sqrt{N}$, we firstly have $|\phi - 1| = 1/\sqrt{N}$ and

$$K(P_{\theta_0}||P_{\theta_1}) = -\log(1 - 1/\sqrt{N}) \frac{1/\sqrt{N}}{2 - 1/\sqrt{N}}.$$

Then, since for all $x < 1$, $x \neq 0$, $-log(1-x) > x$ and for all $N \geq 1$, $2 - \frac{1}{\sqrt{N}} \geq 1$, the Kullback-Leibler divergence can be upper bounded as follows:

$$K(P_{\theta_0}||P_{\theta_1}) \leq \frac{1}{\sqrt{N}} . \frac{1}{\sqrt{N}} = \frac{1}{N}$$

and thus the exponential term $e^{-NK(P_{\theta_0}||P_{\theta_1})}$ is lower bounded by e^{-1}. Finally:

$$\mathcal{R}_N \geq \frac{\Delta}{32} \min_{k=0,\,1} |\phi_k - 1| e^{-NK(P_{\theta_0}||P_{\theta_1})} \geq \frac{1}{16e\sqrt{N}}$$

B.2 Fast Rates

B.2.1 Proof of Proposition 14

Let $\mathcal{A}_N = \bigcap_{i<j} \{(p_{i,j} - \frac{1}{2})(\widehat{p}_{i,j} - \frac{1}{2}) > 0\}$. On the event $\mathcal{A}_N$, p and $\widehat{p}$ satisfy the strongly stochastic transitivity property, and agree on each pair, therefore $\widehat{\sigma}_N = \sigma^*$ and $L(\widehat{\sigma}_N) - L^* = 0$. We can suppose without loss of generality that for any $i < j$, $\frac{1}{2} + h \leq p_{i,j} \leq 1$, and we have $N\widehat{p}_{i,j} \sim \mathcal{B}(N, p_{i,j})$. We thus have:

$$\mathbb{P}\left\{ \widehat{p}_{i,j} \leq \frac{1}{2} \right\} = \mathbb{P}\left\{ N\widehat{p}_{i,j} \leq \frac{N}{2} \right\} = \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} p^k_{i,j} (1 - p_{i,j})^{N-k} \tag{7}$$

As $k \mapsto p^k_{i,j}(1 - p_{i,j})^{N-k}$ is an increasing function of k since $p_{i,j} > \frac{1}{2}$, we have

$$\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} p^k_{i,j} (1 - p_{i,j})^{N-k} \leq \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} . p^{\frac{N}{2}}_{i,j} (1 - p_{i,j})^{\frac{N}{2}} \tag{8}$$

Then, since $\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} + \sum_{k=\lfloor \frac{N}{2} \rfloor}^{N} \binom{N}{k} = \sum_{k=0}^{N} \binom{N}{k} = 2^N$ and $p_{i,j} \geq \frac{1}{2} + h$, we obtain

$$\sum_{k=0}^{\frac{N}{2}} \binom{N}{k} . p_{i,j}^{\frac{N}{2}} (1 - p_{i,j})^{\frac{N}{2}} \leq 2^{N-1} . \left(\frac{1}{4} - h^2 \right)^{\frac{N}{2}} = \frac{1}{2} \left(1 - 4h^2\right)^{\frac{N}{2}} = \frac{1}{2} e^{-\frac{N}{2} log\left(\frac{1}{1-4h^2}\right)}, \tag{9}$$

Combining (7), (8) and (9), yields

$$\mathbb{P}\Big\{ \widehat{p}_{i,j} \leq \frac{1}{2} \Big\} \leq \frac{1}{2} e^{-\frac{N}{2} log\left(\frac{1}{1-4h^2}\right)} \tag{10}$$

Since the probability of the complementary of $\mathcal{A}_N$ is

$$\mathbb{P}\Big\{ \mathcal{A}_N^c \Big\} = \mathbb{P}\Big\{ \bigcup_{i<j} \{ (p_{i,j} - \frac{1}{2})(\widehat{p}_{i,j} - \frac{1}{2}) < 0 \} \Big\} = \mathbb{P}\Big\{ \bigcup_{i<j} \{ \widehat{p}_{i,j} \leq \frac{1}{2} \} \Big\}, \tag{11}$$

combining (10) and Boole's inequality on (11) yields

$$\mathbb{P}\Big\{ \mathcal{A}_N^c \Big\} \leq \sum_{i<j} \mathbb{P}\Big\{ \widehat{p}_{i,j} \leq \frac{1}{2} \Big\} \leq \frac{n(n-1)}{4} e^{-\frac{N}{2} log\left(\frac{1}{1-4h^2}\right)}. \tag{12}$$

As the expectation of the excess of risk can be written

$$\mathbb{E}\Big\{ L(\widehat{\sigma}_N) - L^* \Big\} = \mathbb{E}\Big\{ (L(\widehat{\sigma}_N) - L^*) \mathbb{I}\{\mathcal{A}_N\} + \ (L(\widehat{\sigma}_N) - L^*) \mathbb{I}\{\mathcal{A}_N^c\} \Big\},$$

using successively the fact that $L(\widehat{\sigma}_N) - L^* = 0$ on $\mathcal{A}_N$ and (12) we obtain finally

$$\mathbb{E}\Big\{ L(\widehat{\sigma}_N) - L^* \Big\} \leq \frac{n(n-1)}{2} \mathbb{P}\Big\{ \mathcal{A}_N^c \Big\} \leq \frac{n^2(n-1)^2}{8} e^{-\frac{N}{2} log\left(\frac{1}{1-4h^2}\right)}.$$

B.2.2 Remark 12

According to (12) and (20) we have

$$\mathbb{E}[(\widehat{L}^* - L^*)^2] = \mathbb{E}\left[\left(\sum_{i<j} \left\{ \frac{1}{2} - \left| \widehat{p}_{i,j} - \frac{1}{2} \right| \right\} - \sum_{i<j} \left\{ \frac{1}{2} - \left| p_{i,j} - \frac{1}{2} \right| \right\} \right)^2 \right],$$

and pushing further the calculus gives

$$\mathbb{E}[(\widehat{L}^* - L^*)^2] = \mathbb{E}\left[\left(\sum_{i<j} \left| p_{i,j} - \frac{1}{2} \right| - \left| \widehat{p}_{i,j} - \frac{1}{2} \right| \right)^2 \right] = \mathbb{E}\left[\left(\sum_{i<j} |p_{i,j} - \widehat{p}_{i,j}| \right)^2 \right].$$

Firstly, with the bias-variance decomposition we obtain

$$\mathbb{E}[(\widehat{L}^* - L^*)^2] = Var\left(\sum_{i<j} |p_{i,j} - \widehat{p}_{i,j}| \right) + \left(\mathbb{E}\left[\sum_{i<j} |p_{i,j} - \widehat{p}_{i,j}| \right] \right)^2. \tag{13}$$

The bias in (13) can be written as

$$\mathbb{E}\left[\sum_{i<j}|p_{i,j}-\widehat{p}_{i,j}|\right]=\sum_{\substack{i<j\\ p_{i,j}>\widehat{p}_{i,j}}}\mathbb{E}\left[p_{i,j}-\widehat{p}_{i,j}\right]+\sum_{\substack{i<j\\ p_{i,j}<\widehat{p}_{i,j}}}\mathbb{E}\left[\widehat{p}_{i,j}-p_{i,j}\right]=0 \tag{14}$$

And the variance in (13) is

$$Var\left(\sum_{i<j}|p_{i,j}-\widehat{p}_{i,j}|\right)=\sum_{i<j}\sum_{i'<j'}Cov\left(|p_{i,j}-\widehat{p}_{i,j}|,|p_{i',j'}-\widehat{p}_{i',j'}|\right) \tag{15}$$

$$\leq\sum_{i<j}\sum_{i'<j'}\sqrt{Var(|p_{i,j}-\widehat{p}_{i,j}|)Var(|p_{i',j'}-\widehat{p}_{i',j'}|)}. \tag{16}$$

Since for $i<j$, $\widehat{p}_{i,j}\sim\mathcal{B}(N,p_{i,j})$, we have

$$Var(|p_{i,j}-\widehat{p}_{i,j}|)Var(|p_{i',j'}-\widehat{p}_{i',j'}|)=\frac{p_{i,j}(1-p_{i,j})p_{i',j'}(1-p_{i',j'})}{N^2}\leq\frac{1}{16N^2}. \tag{17}$$

Therefore combining (17) with (15) gives

$$Var\left(\sum_{i<j}|p_{i,j}-\widehat{p}_{i,j}|\right)\leq\left(\frac{n(n-1)}{2}\right)^2\frac{1}{4N}. \tag{18}$$

Finally according to (13), (14) and (18) we obtain: $\mathbb{E}[(\widehat{L}^*-L^*)^2]\leq\frac{n^2(n-1)^2}{16N}$.

B.2.3 Proof of Proposition 15

Similarly to Proposition 11, we use Le Cam's method and consider two Mallows models P_{θ_0} and P_{θ_1} where $\theta_k=(\sigma_k^*,\phi)\in\mathfrak{S}_n\times(0,1)$ and $\sigma_0^*\neq\sigma_1^*$. We can lower bound the minimax risk as follows

$$\begin{aligned}\mathcal{R}_N\geq&\inf_{\sigma_N}\max_{k=0,\,1}\mathbb{E}_{P_{\theta_k}}\left[L_{P_{\theta_k}}(\sigma_N)-L^*_{P_{\theta_k}}\right]\\=&\inf_{\sigma_N}\max_{k=0,\,1}\sum_{i<j}\mathbb{E}_{P_{\theta_k}}\left[2|p_{i,j}-\frac{1}{2}|\times\mathbb{I}\{(\sigma_N(i)-\sigma_N(j)(\sigma^*(i)-\sigma_k^*(j))<0\}\right]\\\geq&\inf_{\sigma_N}\max_{k=0,\,1}h\mathbb{E}_{P_{\theta_k}}\left[d_\tau(\sigma_N,\sigma^*)\right]\\\geq&h\frac{\Delta}{4}e^{-NK(P_{\theta_0||\theta_1})}\end{aligned}$$

With $K(P_{\theta_0||\theta_1})=\log(\frac{1}{\phi})\frac{1-\phi}{1+\phi}$ accordig to (6) and $\Delta=1$, choosing σ_0 and σ_1 as in the proof of Proposition 11. Now we take $\phi=\frac{1-2h}{1+2h}$ so that both P_{θ_0} and P_{θ_1} satisfy $\mathbf{NA}(h)$, and we have $K(P_{\theta_0||\theta_1})=2h\log(\frac{1+2h}{1-2h})$, which gives us finally:

$$\mathcal{R}_N\geq\frac{h}{4}e^{-N2h\log(\frac{1+2h}{1-2h})}$$