Supplemental materials for "Nonparametric Estimation and Test of Conditional Kendall's Tau under Semi-Competing Risks Data and Truncated Data": Proof of the Theorem

JIN-JIAN HSIEH, WEI-CHENG HUANG

jjhsieh@math.ccu.edu.tw

Department of Mathematics, National Chung Cheng University Chia-Yi, Taiwan, R.O.C.

Appendix: The Proof of Theorem

Without loss of generality, we consider p_{ij} , A_{ij} and $\hat{\tau}_c$ instead of p_{ij}^s , A_{ij}^s and $\hat{\tau}_c^s$ for semi-competing risks data in the Appendix. For convenience, we define

$$D_{ij} = I(A_{ij}) = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij}),$$

and

$$\Delta_{ij} = \begin{cases} 1 & (X_i - X_j)(Y_i - Y_j) > 0 \\ -1 & (X_i - X_j)(Y_i - Y_j) < 0 \end{cases}.$$

Then,

$$\begin{aligned} E\left[\frac{D_{ij}}{p_{ij}}\right] &= E\left[E\left[\frac{I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij})}{P(\tilde{Y}_{ij} < \tilde{C}_{ij}|\tilde{X}_{ij}, \tilde{Y}_{ij})}|\tilde{X}_{ij}, \tilde{Y}_{ij}\right]\right] \\ &= E[I(\tilde{X}_{ij} < \tilde{Y}_{ij})] = P(\tilde{X}_{ij} < \tilde{Y}_{ij}), \end{aligned}$$

and

$$\begin{aligned}
E\left[\frac{D_{ij}\Delta_{ij}}{p_{ij}}\right] &= E\left[E\left[\frac{I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij})\Delta_{ij}}{P(\tilde{Y}_{ij} < \tilde{C}_{ij}|\tilde{X}_{ij}, \tilde{Y}_{ij})}\middle|\tilde{X}_{ij}, \tilde{Y}_{ij}\right]\right] \\
&= E[I(\Delta_{ij} = 1, \tilde{X}_{ij} < \tilde{Y}_{ij}) - I(\Delta_{ij} = -1, \tilde{X}_{ij} < \tilde{Y}_{ij})] \\
&= P(\Delta_{ij} = 1, \tilde{X}_{ij} < \tilde{Y}_{ij}) - P(\Delta_{ij} = -1, \tilde{X}_{ij} < \tilde{Y}_{ij}).
\end{aligned}$$

Note that p_{ij} corrects the selection bias introduced from censoring and makes $E[\frac{D_{ij}}{p_{ij}}]$ and $E[\frac{D_{ij}\Delta_{ij}}{p_{ij}}]$ do not involve $\tilde{C}_{ij}$. Let

$$\tau_c = E[\Delta_{ij}|\tilde{X}_{ij} < \tilde{Y}_{ij}],$$

$$\tau'_c = \left(\sum_{i < j} \frac{D_{ij}}{p_{ij}}\right)^{-1} \sum_{i < j} \frac{D_{ij}\Delta_{ij}}{p_{ij}},$$

and

$$\hat{\tau}_c = \left(\sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}}\right)^{-1} \sum_{i < j} \frac{D_{ij}\Delta_{ij}}{\hat{p}_{ij}}.$$

Now, we consider the following difference

$$\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}\Delta_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}\Delta_{ij}}{p_{ij}} = \frac{1}{\binom{n}{2}} \sum_{i < j} D_{ij}\Delta_{ij} \left(\frac{1}{\hat{p}_{ij}} - \frac{1}{p_{ij}}\right), \quad (\text{A.1})$$

where

$$p_{ij} = P(\tilde{Y}_{ij} < \tilde{C}_{ij}|\tilde{X}_{ij}, \tilde{Y}_{ij}) = (P(C > \tilde{Y}_{ij}))^2 = G^2(\tilde{Y}_{ij}),$$

$$\hat{p}_{ij} = \hat{P}(\tilde{Y}_{ij} < \tilde{C}_{ij}|\tilde{X}_{ij}, \tilde{Y}_{ij}) = (\hat{P}(C > \tilde{Y}_{ij}))^2 = \hat{G}^2(\tilde{Y}_{ij}).$$

Furthermore, we have

$$\begin{aligned}
\frac{1}{\hat{p}_{ij}} - \frac{1}{p_{ij}} &= \frac{1}{\hat{G}^2(\tilde{Y}_{ij})} - \frac{1}{G^2(\tilde{Y}_{ij})} \\
&= \frac{G^2(\tilde{Y}_{ij}) - \hat{G}^2(\tilde{Y}_{ij})}{\hat{G}^2(\tilde{Y}_{ij})G^2(\tilde{Y}_{ij})}
\end{aligned} \quad (\text{A.2})$$

The absolute value of (A.2) is bounded from above by

$$\frac{\sup_{0 \leq t \leq L_Y} |G^2(\tilde{Y}_{ij}) - \hat{G}^2(\tilde{Y}_{ij})|}{\hat{G}^2(L_Y)G^2(L_Y)}. \quad (\text{A.3})$$

By uniform convergence of the Kaplan-Meier estimator on $[0, L_Y]$ and continuity of the mapping: $t \rightarrow t^2$, we have, therefore, (A.3) $\xrightarrow{P} 0$. By Fleming and Harrington (1991), we have

$$\hat{G}(t) - G(t) = -G(t) \int_0^t \frac{\hat{G}(u^-)}{G(u)} \frac{\sum_{k=1}^n dM_k^c(u)}{\sum_{k=1}^n I\{T_k \geq u\}} + o_p(1),$$

where $T_k = \min(Y_k, C_k)$, $M^c(u)$ is the standard martingale associated to censoring variable C , which is defined as $M^c(u) = I\{C \leq u, \delta_Y = 0\} - \int_0^u I\{T \geq s\} d\Lambda^c(s)$ and Λ^c is the cumulative hazard function associated to C . Therefore,

$$\begin{aligned} \sqrt{n} \left[\frac{G^2(t) - \hat{G}^2(t)}{\hat{G}^2(t)G^2(t)} \right] &= \frac{-2\sqrt{n} G(t) [\hat{G}(t) - G(t)]}{\hat{G}^2(t)G^2(t)} + o_p(1) \\ &= \frac{2\sqrt{n} G^2(t) \int_0^t \frac{\hat{G}(u^-)}{G(u)} \frac{\sum_{k=1}^n dM_k^c(u)}{\sum_{k=1}^n I\{T_k \geq u\}}}{\hat{G}^2(t)G^2(t)} + o_p(1). \end{aligned}$$

Thus,

$$\begin{aligned} \sqrt{n} \left[\frac{G^2(\tilde{Y}_{ij}) - \hat{G}^2(\tilde{Y}_{ij})}{\hat{G}^2(\tilde{Y}_{ij})G^2(\tilde{Y}_{ij})} \right] \\ = 2\sqrt{n} \frac{1}{\hat{G}^2(\tilde{Y}_{ij})} \int_0^{L_Y} I\{\tilde{Y}_{ij} \geq u\} \frac{\hat{G}(u^-)}{G(u)} \frac{\sum_{k=1}^n dM_k^c(u)}{\sum_{k=1}^n I\{T_k \geq u\}} + o_p(1). \end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{\sqrt{n}}{\binom{n}{2}} \sum_{i < j} D_{ij} \Delta_{ij} \left(\frac{1}{\hat{p}_{ij}} - \frac{1}{p_{ij}} \right) \\
&= \frac{1}{\binom{n}{2}} \sum_{i < j} D_{ij} \Delta_{ij} 2\sqrt{n} \frac{1}{\hat{G}^2(\tilde{Y}_{ij})} \int_0^{L_Y} I\{\tilde{Y}_{ij} \geq u\} \frac{\hat{G}(u^-)}{G(u)} \frac{\sum_{k=1}^n dM_k^c(u)}{\sum_{k=1}^n I\{T_k \geq u\}} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \int_0^{L_Y} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] 2 \frac{\hat{G}(u^-)}{G(u)} \frac{n}{\sum_{k=1}^n I\{T_k \geq u\}} \sum_{k=1}^n dM_k^c(u) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \int_0^{L_Y} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] \frac{2}{P(T \geq u)} \sum_{k=1}^n dM_k^c(u) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \int_0^{L_Y} E \left[\frac{D_{ij} \Delta_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] \frac{2}{P(T \geq u)} \sum_{k=1}^n dM_k^c(u) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^n \int_0^{L_Y} E \left[\frac{D_{ij} \Delta_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] \frac{2}{P(T \geq u)} dM_k^c(u) + o_p(1), \tag{A.4}
\end{aligned}$$

which is asymptotically equivalent to a sum of zero mean independent and identically distributed terms.

Nextly, one has

$$\begin{aligned}
& \sqrt{n} (\hat{\tau}_c - \tau_c) \\
&= \sqrt{n} \left(\frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{\hat{p}_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}}} - \frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}}} \right) + \sqrt{n} \left(\frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}}} - \right. \\
& \quad \left. \frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}}} \right) + \sqrt{n} \left(\frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}}} - \tau_c \right) \\
&= A + B + C. \\
A &= \frac{1}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}}} \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}} \right] \\
&= \frac{1}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}}} \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}} \right] + o_p(1) \\
&= \frac{1}{P(\tilde{X}_{ij} < \tilde{Y}_{ij})} \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}} \right] + o_p(1) \\
&\stackrel{\text{by (A.4)}}{=} \frac{1}{P(\tilde{X}_{ij} < \tilde{Y}_{ij})} \frac{1}{\sqrt{n}} \sum_{k=1}^n \int_0^{L_Y} E \left[\frac{D_{ij} \Delta_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] \frac{2}{P(T \geq u)} dM_k^c(u) + o_p(1),
\end{aligned}$$

which is equivalent to sum of zero-mean independent and identically distributed terms.

$$\begin{aligned}
B &= \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}} \frac{1}{\left(\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}} \right) \left(\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}} \right)} \times \\
&\quad \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}} \right] \\
&= - \frac{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\left(\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}} \right) \left(\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}} \right)} \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}} \right] \\
&= - \frac{P(\Delta_{ij} = 1, \tilde{X}_{ij} < \tilde{Y}_{ij}) - P(\Delta_{ij} = -1, \tilde{X}_{ij} < \tilde{Y}_{ij})}{(P(\tilde{X}_{ij} < \tilde{Y}_{ij}))^2} \times \\
&\quad \sqrt{n} \left[\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{\hat{p}_{ij}} - \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}} \right] + o_p(1) \\
&\stackrel{\text{by (A.4)}}{=} D_0 \frac{1}{\sqrt{n}} \sum_{k=1}^n \int_0^{L_Y} E \left[\frac{D_{ij} I\{\tilde{Y}_{ij} \geq u\}}{\hat{G}^2(\tilde{Y}_{ij})} \right] \frac{2}{P(T \geq u)} dM_k^c(u) + o_p(1),
\end{aligned}$$

which is also equivalent to a sum of zero-mean independent and identically distributed terms,

where

$$D_0 = - \frac{P(\Delta_{ij} = 1, \tilde{X}_{ij} < \tilde{Y}_{ij}) - P(\Delta_{ij} = -1, \tilde{X}_{ij} < \tilde{Y}_{ij})}{(P(\tilde{X}_{ij} < \tilde{Y}_{ij}))^2}.$$

$$C = \sqrt{n}(\tau'_c - \tau_c)$$

$$\begin{aligned}
& \sqrt{n} \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}} \\
&= \frac{\sqrt{n} \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{\frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij}}{p_{ij}}} - \sqrt{n} \tau_c \\
&= \frac{\sqrt{n} \frac{1}{\binom{n}{2}} \sum_{i < j} \frac{D_{ij} \Delta_{ij}}{p_{ij}}}{P(\tilde{X}_{ij} < \tilde{Y}_{ij})} - \sqrt{n} \tau_c + o_p(1) \\
&= \sqrt{n} \frac{1}{\binom{n}{2}} \sum_{i < j} \left[\frac{\frac{D_{ij} \Delta_{ij}}{p_{ij}}}{P(\tilde{X}_{ij} < \tilde{Y}_{ij})} - \tau_c \right] + o_p(1).
\end{aligned}$$

By the proof of Theorem 1 in Chapter 3 of Lee (1990), the above U-statistic can be written as sum of i.i.d zero-mean. Therefore, we can prove that $\hat{\tau}_c$ is a consistent estimator of τ_c and $\sqrt{n}(\hat{\tau}_c - \tau_c)$ is asymptotic normal with zero-mean.