Homework 6

Liu Huihang
SA18017026
QQ: 184050473
MAIL: huihang@mail.ustc.edu.cn

Problem 1 *Let $X_1, \ldots, X_m i.i.d \sim F$, $Y_1, \ldots, Y_n i.i.d \sim G$ and i.i.d. within each sample, then*

(1) get the U-statistic U_n with kernel $h(x_1, x_2, y_1, y_2) = I(x_1 < y_1, x_2 < y_2)$,
(2) get the limit distribution of U-statistic U_n with $m + n \to \infty$, $\frac{m}{n+m} \to p \in (0, 1)$,
(3) get the limit distribution of U-statistic U_n under null hypothesis $H_0 : F = G$.

Solution

(1) The kerne is $h(x_1, x_2, y_1, y_2) = I(x_1 < y_1, x_2 < y_2)$, which is of order 2 in both x and y. The corresponding U-statistic is

$$U_n = \frac{1}{\binom{m}{2}\binom{n}{2}} \sum_{i<k}^{m} \sum_{j<l}^{n} I(X_i < Y_j, X_k < Y_l)$$

(2) The projection of $U - \theta$ onto the set of all functions of the form $\sum_{i=1}^{m} k_i(X_i) + \sum_{j=1}^{n} l_j(Y_j)$ is given by

$$\hat{U} = \frac{2}{m} \sum_{i=1}^{m} h_{1,0}(X_i) + \frac{2}{n} \sum_{j=1}^{n} h_{0,1}(Y_j),$$

where the functions $h_{1,0}$ and $h_{0,1}$ are defined by

$$\begin{aligned} h_{1,0}(x) &= Eh(x, X_2, Y_1, Y_2) - \theta, \\ h_{0,1}(y) &= Eh(X_1, X_2, y, Y_2) - \theta. \end{aligned}$$

The sequence $\hat{U}$ is asymtotically normal by the central limit theorem. Then difference between $\hat{U}$ and $U_n - \theta$ is asymtotically negligible.

$$\begin{aligned} \theta &= Eh(X_1, X_2, Y_1, Y_2) \\ &= P(X_1 < Y_1)P(X_2 < Y_2) \\ &= (E[P(X_1 < y | Y_1 = y)])^2 \\ &= (E[F(Y_1)])^2 \\ &= \left(\int F(y) dG(y) \right)^2 \end{aligned}$$

Then calculate $\zeta_{1,0}$ and $\zeta_{0,1}$.

$$
\begin{aligned}
\zeta_{1,0} &= cov(h(X_1, X_2, Y_1, Y_2), h(X_1.X_3, Y_3, Y_4)) \\
&= cov(I(X_1 < Y_1, X_2 < Y_2), I(X_1 < Y_3, X_3 < Y_4)) \\
&= E[I(X_1 < Y_1, X_2 < Y_2)I(X_1 < Y_3, X_3 < Y_4)] \\
&\qquad - E[I(X_1 < Y_1, X_2 < Y_2)]E[I(X_1 < Y_3, X_3 < Y_4)] \\
&= E[I(X_1 < \min(Y_1, Y_3))I(X_2 < Y_2)I(X_3 < Y_4)] \\
&\qquad - E[I(X_1 < Y_1)I(X_2 < Y_2)]E[I(X_1 < Y_3)I(X_3 < Y_4)] \\
&= P(X_1 < \min(Y_1, Y_3))(P(X_1 < Y_1))^2 - (P(X_1 < Y_1))^4 \\
&= \left(\int F(y)dG(y)\right)^2 \int F(z)d(-G^2(z) + 2G(z)) - \left(\int F(y)dG(y)\right)^4
\end{aligned}
$$

and

$$
\begin{aligned}
\zeta_{0,1} &= cov(h(X_1, X_2, Y_1, Y_2), h(X_3.X_4, Y_1, Y_3)) \\
&= cov(I(X_1 < Y_1, X_2 < Y_2), I(X_3 < Y_1, X_4 < Y_3)) \\
&= E[I(X_1 < Y_1, X_2 < Y_2)I(X_3 < Y_1, X_4 < Y_3)] \\
&\qquad - E[I(X_1 < Y_1, X_2 < Y_2)]E[I(X_3 < Y1, X_4 < Y_3)] \\
&= E[I(Y_1 > \max(X_1, X_3))I(Y_2 > X_2)I(Y_3 > X_4)] \\
&\qquad - E[I(X_1 < Y_1)I(X_2 < Y_2)]E[I(X_3 < Y_1)I(X_4 < Y_3)] \\
&= P(Y_1 > \max(X_1, X_3))(P(X_1 < Y_1))^2 - (P(X_1 < Y_1))^4 \\
&= \left(\int F(y)dG(y)\right)^2 \int F^2(z)dG(z) - \left(\int F(y)dG(y)\right)^4
\end{aligned}
$$

Thus,

$$
\sqrt{m+n}(U_n - \theta) \rightsquigarrow N\left(0, \frac{4\zeta_{1,0}}{p} + \frac{4\zeta_{0,1}}{1-p}\right)
$$

where $\theta = \left(\int F(y)dG(y)\right)^2$, $\zeta_{1,0} = \left(\int F(y)dG(y)\right)^2 \int F(z)d(-G^2(z)+2G(z)) - \left(\int F(y)dG(y)\right)^4$ and $\zeta_{0,1} = \left(\int F(y)dG(y)\right)^2 \int F^2(z)dG(z) - \left(\int F(y)dG(y)\right)^4$.

(3) Under $H_0 : F = G$, $\theta = 1/4$, $\zeta_{1,0} = \frac{1}{4*12}$ and $\zeta_{0,1} = \frac{1}{4*12}$.

$$
\sqrt{m+n}\left(U_n - \frac{1}{2}\right) \rightsquigarrow N\left(0, \frac{1}{12p} + \frac{1}{12(1-p)}\right)
$$

■

Problem 2 *Suppose the distribution of X is symmetric about zero with variance $\sigma^2 > 0$ and $EX^4 < \infty$, consider kernel $h(x, y) = xy + (x^2 - \sigma^2)(y^2 - \sigma^2)$, then*

(1) prove that the U-statistic U_n with kernel h is degenerated of order 1,

(2) get λ_1, λ_2 and orthogonal functions $\Phi_1(x), \Phi_2(x)$, such that $h(x, y) = \lambda_1\varphi_1(x)\varphi_1(y) + \lambda_2\varphi_2(x)\varphi_2(y)$,

(3) get the limit distribution of nU_n.

Solution

■

Problem 3 *Prove the Hoeffding decomposition in page 13.*

Solution

■

Problem 4 *Prove the T decomposition in page 12.*

Solution

■