EXERCISE 5

WEIYU LI

1. Suppose $X, X_1, \ldots, X_n \sim F, i.i.d.$, solve the kernel $h(x_1, x_2, x_3)$ such that $E_F h(X_1, X_2, X_3) = E(X - E_F X)^3$.

Solve. For abbreviation, we omit F in the arguments. Notice that

$$E(X - EX)^3 = EX^3 - 3EX^2EX + 3EX(EX)^2 - (EX)^3 = EX^3 - 3EX^2EX + 2(EX)^3.$$

We check the following symmetric homogeneous terms to solve the kernel:

$$E(X_1^3 + X_2^3 + X_3^3) = 3EX^3, \tag{1}$$

$$E(X_1X_2X_3) = (EX)^3, \tag{2}$$

$$E(X_1 + X_2 + X_3)^3 = 3EX^3 + 18EX^2EX + 6(EX)^3. \tag{3}$$

Computing $\frac{1}{2} \times (1) + 3 \times (2) - \frac{1}{6} \times (3)$ gives that

$$E[\frac{1}{2}(X_1^3 + X_2^3 + X_3^3) + 3X_1X_2X_3 - \frac{1}{6}(X_1 + X_2 + X_3)^3] = E(X - EX)^3,$$

thus we derive the kernel

$$h(x_1, x_2, x_3) = \frac{1}{2}(x_1^3 + x_2^3 + x_3^3) + 3x_1x_2x_3 - \frac{1}{6}(x_1 + x_2 + x_3)^3.$$

□

2. Prove $\zeta_1 = 1/9$ in Page 25 in the slides.

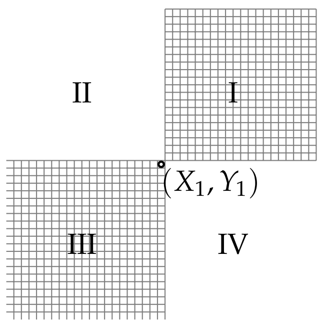

FIGURE 1. Conditioned on P_1, P_2 in the gray part gives $h(P_1, P_2) = 1$, otherwise $h(P_1, P_2) = -1$.

Proof. Since X, Y are continuous, almost surely the line P_1P_2 cannot be parallel to the axes, so we don't care about the parallel case. For (X_2, Y_2) independent with (X_1, Y_1),

Date: 2019/10/14.
liweiyu@mail.ustc.edu.cn.

conditioned on (X_1, Y_1), we have that $h(P_1, P_2) = 1$ if P_2 lies in the gray part, else $h(P_1, P_2) = -1$. Denote $F(\cdot), G(\cdot)$ the cdf of X and Y, respectively. Then

$$\begin{aligned}
P(P_2 \in \mathrm{I}|P_1) =& \big(1 - F(X_1)\big)\big(1 - G(Y_1)\big),\\
P(P_2 \in \mathrm{II}|P_1) =& F(X_1)\big(1 - G(Y_1)\big),\\
P(P_2 \in \mathrm{III}|P_1) =& F(X_1)G(Y_1),\\
P(P_2 \in \mathrm{IV}|P_1) =& \big(1 - F(X_1)\big)G(Y_1).
\end{aligned}$$

Consequently,

$$\begin{aligned}
P\big(h(P_1, P_2) = 1\big) =& E\big[P(P_2 \in \text{I or III}|P_1)\big)\big]\\
=& \big(1 - E[F(X_1)]\big)\big(1 - E[G(Y_1)]\big) + E[F(X_1)]E[G(Y_1)] = \frac{1}{2},
\end{aligned}$$

where the last equality comes from that $[F(X_1), G(Y_1) \sim U[0, 1]$, and

$$\begin{aligned}
&P\big(h(P_1, P_2)h(P_1, P_3) = 1\big)\\
=& E\big[P(P_2, P_3 \text{ in the same color part}|P_1)\big)\big]\\
=& E\Big[\big[\big(1 - F(X_1)\big)\big(1 - G(Y_1)\big) + F(X_1)G(Y_1)\big]^2 + \big[\big(1 - F(X_1)\big)G(Y_1) + F(X_1)\big(1 - G(Y_1)\big)\big]^2\Big]\\
=& E\Big[\big(1 - F(X_1)\big)^2\big(1 - G(Y_1)\big)^2 + F(X_1)^2G(Y_1)^2 + \big(1 - F(X_1)\big)^2G(Y_1)^2 + F(X_1)^2\big(1 - G(Y_1)\big)^2\\
&+ 4F(X_1)\big(1 - F(X_1)\big)G(Y_1)\big(1 - G(Y_1)\big)\Big]\\
=& 4 \times \frac{1}{3} \times \frac{1}{3} + 4 \times \frac{1}{6} \times \frac{1}{6} = \frac{5}{9}.
\end{aligned}$$

Now we conclude that

$$\begin{aligned}
\zeta_1 =& Cov\big(h(P_1, P_2), h(P_1, P_3)\big)\\
=& E\big(h(P_1, P_2)h(P_1, P_3)\big) - E\big(h(P_1, P_2)\big)E\big(h(P_1, P_3)\big)\\
=& \big(\frac{5}{9} - \frac{4}{9}\big) - \big(\frac{1}{2} - \frac{1}{2}\big)^2 = \frac{1}{9}.
\end{aligned}$$

□

3. Suppose samples $X_1, \ldots, X_n \sim U(0, \tau), i.i.d.$, for the kernel $h(x, y) = |x - y|$, its corresponding U statistics is $G_n = \frac{1}{\binom{n}{2}} \sum_{i<j} |X_i - X_j|$. Solve the asymptotic distribution of G_n.

Solve. For the kernel h, G_n estimates

$$\theta = E|X_1 - X_2| = \int_0^\tau \int_0^\tau \frac{1}{\tau^2}|x - y|dydx = \frac{2}{\tau^2}\int_0^\tau \int_0^x (x - y)dydx = \frac{\tau}{3}.$$

From the notations in the slides, we know that

$$\sqrt{n}(G_n - \theta) \to N(0, r^2\zeta_1) = N(0, 4\zeta_1),$$

where $\zeta_1 = Var(E(h(X, Y)|X))$.

It remains to solve ζ_1. First we have

$$E(h(X,Y)|X) = \int_0^\tau \frac{1}{\tau}|X-y|dy = \frac{1}{2\tau}\left(X^2+(\tau-X)^2\right) = \frac{X^2}{\tau} - X + \frac{\tau}{2}.$$

Since $EX = \frac{\tau}{2}$, $EX^2 = \frac{\tau^2}{3}$, $EX^3 = \frac{\tau^3}{4}$ and $EX^4 = \frac{\tau^4}{5}$,

$$\begin{aligned}
\zeta_1 &= Var(\frac{X^2}{\tau} - X + \frac{\tau}{2}) = Var(\frac{X^2}{\tau} - X) \\
&= E(\frac{X^2}{\tau} - X)^2 - (\frac{EX^2}{\tau} - EX)^2 \\
&= \frac{EX^4}{\tau^2} - \frac{2EX^3}{\tau} + EX^2 - (\frac{EX^2}{\tau} - EX)^2 \\
&= \frac{\tau^2}{5} - \frac{\tau^2}{2} + \frac{\tau^2}{3} - (\frac{\tau}{3} - \frac{\tau}{2})^2 = \frac{\tau^2}{180}.
\end{aligned}$$

Therefore, we conclude that

$$\sqrt{n}(G_n - \frac{\tau}{3}) \to N(0, \frac{\tau^2}{45}).$$

□