Homework 1

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

Problem 1 *Let $X_1, \ldots, X_n \sim F$ and let $F_n(x)$ be the empirical distribution function, for a fixed x, find limiting distribution of $\sqrt{F_n(x)}$.*

Solution

We have $F_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$, and $F_n(x) \sim \frac{1}{n}Binomial(n, F(x))$. By CLT, we get $\frac{nF_n(x)-nF(x)}{\sqrt{n}\sqrt{F(x)(1-F(x))}} = \sqrt{n}\frac{F_n(x)-F(x)}{\sqrt{(F(x)(1-F(x)))}} \rightsquigarrow N(0,1)$. Thus, $\sqrt{n}(F_n(x)-F(x)) \rightsquigarrow N(0, F(x)(1-F(x)))$.

According to *Delta Method*, we have $\sqrt{n}(\sqrt{F_n(x)} - \sqrt{F(x)}) \rightsquigarrow \frac{\partial\sqrt{F(x)}}{\partial F(x)} N(0, F(x)(1-F(x)))$. Finally, we get $\sqrt{F_n} \rightsquigarrow N(\sqrt{F(x)}, \frac{1-F(x)}{4n})$.

■

Problem 2 *Let x, y be two distinct real numbers, find $Cov(F_n(x), F_n(y))$, where F_n be the empirical distribution function.*

Solution

Assume that $x < y$. We have $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$, and $F_n(x) \sim \frac{1}{n}Binomial(n, F(x))$. This gives

$$\begin{aligned} Cov(F_n(x), F_n(y)) &= E[F_n(x) - E(F_n(x))][F_n(y) - E(F_n(y))] \\ &= E[F_n(x)F_n(y)] - E[F_n(x)]E[F_n(y)] \end{aligned}$$

where

$$\begin{aligned} n^2 E[F_n(x)F_n(y)] &= E\left[\sum_{i=1}^{n} I(X_i \le x) \sum_{j=1}^{n} I(X_j \le y)\right] \\ &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} I(X_i \le x) I(X_j \le y)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[I(X_i \le x) I(X_j \le y)\right] \\ &= \sum_{i=1}^{n} E\left[I(X_i \le x) I(X_i \le y)\right] + \sum_{i,j=1; i\ne j}^{n} E\left[I(X_i \le x)\right] E\left[I(X_j \le y)\right] \\ &= nF(x) + n(n-1)F(x)F(y) \end{aligned}$$

and $E[F_n(x)]E[F_n(y)] = F(x)F(y)$.

Thus, $Cov(F_n(x), F_n(y)) = \frac{F(x)+(n-1)F(x)F(y)}{n} - F(x)F(y) = \frac{F(x)-F(x)F(y)}{n}$ ■

Problem 3 *Let* $X_{(1)} \le \cdots \le X_{(n)}$ *be order statistics from continuous popolation* F*, prove that for any* $0 < \beta < 1$

$$\mathbf{P}(F(X_{(n)}) - F(X_{(1)}) > \beta) = 1 - n\beta^{n-1} + (n-1)\beta^n.$$

Solution

We have $\mathbf{P}(F(X_{(n)}) - F(X_{(1)}) > \beta) = 1 - \mathbf{P}(F(X_{(n)}) - F(X_{(1)}) \le \beta)$, $F(X_1), \ldots, F(X_n)$ have *i.i.d.* standard uniform distribution, $F(X_{(1)}), \ldots, F(X_{(n)})$ are ordered statistics, $F(X_{(n)} - F(X_{(1)})) \sim Beta(n-1, 2)$ and

$$\begin{aligned}\mathbf{P}(F(X_{(n)}) - F(X_{(1)}) \le \beta) &= \int_0^1 n(n-1)x^{n-2}(1-x)dx \\ &= n\beta^{n-1} - (n-1)\beta^n.\end{aligned}$$

Thus $\mathbf{P}(F(X_{(n)}) - F(X_{(1)}) > \beta) = 1 - n\beta^{n-1} + (n-1)\beta^n$. ■

Problem 4 *Let* $X_1, \ldots, X_n$ *be simple samples from* $U(0,1)$*, prove that sample median* $\hat{\xi}_{n,1/2}$ *has asymptotic distribution* $N(\frac{1}{2}, \frac{1}{4n})$*.*

Solution

Let F be the distribution function of $U(0,1)$.

$\forall p \in (0,1)$, let $t > 0, p_{nt} = F(\xi_p + \frac{t\sigma_F}{\sqrt{n}}), c_{nt} = \sqrt{p_{nt} - p}/\sqrt{p_{nt}(1-p_{nt})}$ and $Z_{nt} = [B_n(p_{nt}) - np_{nt}]/\sqrt{np_{nt}(1-p_{nt})}$, where $B_n(q)$ denotes a random variable having the binomial distribution $Bi(q, n)$. Then

$$\begin{aligned}\mathbf{P}(\hat{\xi}_{n,p} \le \xi_p + \frac{t\sigma_F}{\sqrt{n}}) &= \mathbf{P}(p \le F_n(\xi_p + \frac{t\sigma_F}{\sqrt{n}})) \\ &= \mathbf{P}(Z_{nt} \ge -c_{nt})\end{aligned}$$

Then $p_{nt} \to p$ and $c_{nt} \to t$. Hence, we get

$$\mathbf{P}(Z_{nt} \le -c_{nt}) - \Phi(-c_{nt}) \to 0$$

by CLT and Polya Throrem. Thus

$$\sqrt{n}(\hat{\xi}_{n,p} - \xi_p) \rightsquigarrow N(0, \sigma_F^2).$$

We have $F(1/2) = 1/2$ and F' exist and is positive at $1/2$. Thus, the result follows from letting $p = 1/2$, which gives $\xi_p = 1/2, \sigma_F = 1/4$. ■