线性回归第四周作业

宋歌 2015080086 数52

3/21/2018

1. 对于数据$(X_i, Y_i), 1 \leqslant i \leqslant n$，考虑线性模型

$$Y_i = \beta_1 + \beta_2 X_i + e_i, 1 \leqslant i \leqslant n$$

β_1和β_2是未知参数，X_i是不全相等的固定设计点，随机误差$e_1, \ldots, e_n\ iid \sim N(0, \sigma^2)$，$\sigma^2$未知。设$\hat{\beta}_1$和$\hat{\beta}_2$分别是$\beta_1$和$\beta_2$的最小二乘估计（LS估计）。定义$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$为拟合值，$r_i = Y_i - \hat{Y}_i$为残差，$i = 1, \ldots, n$。
证明：

1) $\sum_{i=1}^n r_i = 0$, $\sum_{i=1}^n X_i r_i = 0$

已知$\hat{\beta}_1$和$\hat{\beta}_2$是最小二乘估计，即它们是

$$L(\beta_1, \beta_2) = \sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_i)^2$$

的极小值点，从而有

$$\frac{\partial L}{\partial \beta_1}(\hat{\beta}_1, \hat{\beta}_2) = -2\sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum_{i=1}^n r_i = 0$$
$$\frac{\partial L}{\partial \beta_2}(\hat{\beta}_1, \hat{\beta}_2) = -2\sum_{i=1}^n X_i(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum_{i=1}^n X_i r_i = 0$$

2) $E(\sum_{i=1}^n r_i^2) = (n-2)\sigma^2$，由此构造出$\sigma^2$的无偏估计.

$$E(\sum_{i=1}^n r_i^2) = \sum_{i=1}^n E\Big[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i + e_i\Big]^2$$
$$= \sum_{i=1}^n E\Big[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i\Big]^2 + n\sigma^2 + 2\sum_{i=1}^n Ee_i\Big[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i\Big]$$

已知

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = \beta_1 + \beta_2\bar{X} + \bar{e} - \hat{\beta}_2\bar{X}$$
$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^n b_i e_i, \quad b_i = \frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

从而

$$\beta_1 - \hat{\beta}_1 = (\hat{\beta}_2 - \beta_2)\bar{X} - \bar{e}, \quad \hat{\beta}_2 - \beta_2 = \sum_{i=1}^n b_i e_i$$

代入第一项得

$$\sum_{i=1}^n E\Big[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i\Big]^2 = \sum_{i=1}^n E\Big[(\hat{\beta}_2 - \beta_2)(\bar{X} - X_i) - \bar{e}\Big]^2$$
$$= \sum_{i=1}^n E\Big[(\hat{\beta}_2 - \beta_2)^2(\bar{X} - X_i)^2 + \bar{e}^2 - 2(\hat{\beta}_2 - \beta_2)(\bar{X} - X_i)\bar{e}\Big]$$
$$= E(\hat{\beta}_2 - \beta_2)^2\sum_{i=1}^n(\bar{X} - X_i)^2 + nE\bar{e}^2 - 2E\Big[(\hat{\beta}_2 - \beta_2)\bar{e}\Big]\sum_{i=1}^n(\bar{X} - X_i)$$
$$= E(\sum_{i=1}^n b_i e_i)^2\sum_{i=1}^n(\bar{X} - X_i)^2 + \sigma^2 - 2E\Big[(\sum_{i=1}^n b_i e_i)\bar{e}\Big]\sum_{i=1}^n(\bar{X} - X_i)$$
$$= \sum_{i=1}^n b_i^2\sigma^2\sum_{i=1}^n(\bar{X} - X_i)^2 + \sigma^2 - 2\frac{\sigma^2}{n}\sum_{i=1}^n b_i\sum_{i=1}^n(\bar{X} - X_i) = 2\sigma^2$$

代入第三项得

$$2\sum_{i=1}^n E\Big[e_i[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i]\Big] = 2\sum_{i=1}^n E\Big[e_i[(\hat{\beta}_2 - \beta_2)(\bar{X} - X_i) - \bar{e}]\Big]$$
$$= 2\sum_{i=1}^n E\Big[e_i[\frac{(\bar{X} - X_i)(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}e_i]\Big] - 2nE\bar{e}^2 = 2\sum_{i=1}^n E\Big[\frac{-(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}e_i^2\Big] - 2\sigma^2$$
$$= -\frac{2}{\sum_{i=1}^n (X_i - \bar{X})^2}\sum_{i=1}^n\Big[(X_i - \bar{X})^2Ee_i^2\Big] - 2\sigma^2 = -4\sigma^2$$

从而

$$E(\sum_{i=1}^{n} r_i^2) = 2\sigma^2 + n\sigma^2 - 4\sigma^2 = (n-2)\sigma^2$$

从而

$$E(\frac{1}{n-2}\sum_{i=1}^{n} r_i^2) = \sigma^2$$

即$\frac{1}{n-2}\sum_{i=1}^{n} r_i^2$是$\sigma^2$的无偏估计。

3) 计算$Cov(\hat{\beta}_1, \hat{\beta}_2)$.

已知

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = \beta_1 + \beta_2\bar{X} + \bar{e} - \hat{\beta}_2\bar{X}$$

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} b_i e_i, \quad b_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

从而

$$\begin{aligned}
&Cov(\hat{\beta}_1, \hat{\beta}_2) = E(\hat{\beta}_1\hat{\beta}_2) - E\hat{\beta}_1 E\hat{\beta}_2 \\
&= E\left(\beta_1\hat{\beta}_2 + \beta_2\hat{\beta}_2\bar{X} + \bar{e}\hat{\beta}_2 - \hat{\beta}_2^2\bar{X}\right) - \beta_1\beta_2 \\
&= \beta_1\beta_2 + \beta_2^2\bar{X} + E(\bar{e}\hat{\beta}_2) - \bar{X}E\hat{\beta}_2^2 - \beta_1\beta_2 \\
&= \beta_2^2\bar{X} + E\left[\bar{e}(\beta_2 + \sum_{i=1}^{n} b_i e_i)\right] - \bar{X}E(\beta_2 + \sum_{i=1}^{n} b_i e_i)^2 \\
&= \beta_2^2\bar{X} + E(\bar{e}\beta_2) + E(\bar{e}\sum_{i=1}^{n} b_i e_i) - \bar{X}E\left[\beta_2^2 + (\sum_{i=1}^{n} b_i e_i)^2 + 2\beta_2\sum_{i=1}^{n} b_i e_i\right] \\
&= \beta_2^2\bar{X} + \sum_{i=1}^{n} b_i E(e_i\bar{e}) - \bar{X}\beta_2^2 - \bar{X}\sum_{i=1}^{n} b_i^2\sigma^2 + 2\beta_2\bar{X}\sum_{i=1}^{n} b_i E e_i \\
&= \frac{1}{n}\sum_{i=1}^{n} b_i\sigma^2 - \bar{X}\sum_{i=1}^{n} b_i^2\sigma^2 = \frac{-\bar{X}\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}
\end{aligned}$$

4) $\hat{\beta}_2$是所有形如$\sum_{i=1}^n a_iY_i$的无偏估计中方差最小的估计.

设$\tilde{\beta}_2=\sum_{i=1}^n a_iY_i$是$\beta_2$的无偏估计，则有

$$E(\tilde{\beta}_2)=E(\sum_{i=1}^n a_iY_i)=\sum_{i=1}^n E\Big[a_i(\beta_1+\beta_2X_i)+e_i\Big]$$
$$=\beta_1\sum_{i=1}^n a_i+\beta_2\sum_{i=1}^n a_iX_i+\sum_{i=1}^n a_iE(e_i)=0$$

从而应有

$$\sum_{i=1}^n a_i=0,\quad \sum_{i=1}^n a_iX_i=1$$

故有

$$\tilde{\beta}_2=\sum_{i=1}^n a_iY_i=\sum_{i=1}^n a_i(\beta_1+\beta_2X_i+e_i)=\beta_2+\sum_{i=1}^n a_ie_i$$

$$Var(\tilde{\beta}_2)=Var(\beta_2+\sum_{i=1}^n a_ie_i)=Var(\sum_{i=1}^n a_ie_i)=\sigma^2\sum_{i=1}^n a_i^2$$

现希望，在约束条件$\begin{cases} sum_{i=1}^n a_i=0 \\ \sum_{i=1}^n a_iX_i=1 \end{cases}$下，找到使得$Var(\tilde{\beta}_2)=\sigma^2\sum_{i=1}^n a_i^2$最小的估计。下用Lagrange乘子法，首先构造Lagrange 函数

$$g\Big((a_1,\ldots,a_n),(\lambda_1,\lambda_2)\Big)=\sigma^2\sum_{i=1}^n a_i^2+\lambda_1\sum_{i=1}^n a_i+\lambda_2(\sum_{i=1}^n a_iX_i-1)$$

由

$$\begin{cases} \frac{\partial g}{\partial\lambda_1}=\sum_{i=1}^n a_i=0 \\ \frac{\partial g}{\partial\lambda_2}=\sum_{i=1}^n a_iX_i=1 \\ \frac{\partial g}{\partial a_i}=2\sigma^2a_+\lambda_1+\lambda_2X_i=0,\ 1\leqslant i\leqslant n \end{cases}$$

解得

$$\begin{cases} \lambda_1=\frac{2\sigma^2\bar{X}}{\sum_{i=1}^n(X_i-\bar{X})^2} \\ \lambda_2=\frac{-2\sigma^2}{\sum_{i=1}^n(X_i-\bar{X})^2} \\ a_i=\frac{X_i-\bar{X}}{\sum_{i=1}^n(X_i-\bar{X})^2}=b_i \end{cases}$$

从而$a_i = b_i, 1 \leqslant i \leqslant n$时，$\tilde{\beta}_2 = \sum_{i=1}^n a_i Y_i = \sum_{i=1}^n b_i Y_i = \hat{\beta}_2$是使得其方差最小的估计。即$\hat{\beta}_2$是UMVUE.

5) $(\hat{\beta}_1, \hat{\beta}_2)$与$\sum_{i=1}^n r_i^2$独立.

已知

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^n b_i e_i$$

$$\hat{\beta}_1 = \beta_1 + (\beta_2 - \hat{\beta}_2)\bar{X} + \bar{e} = \beta_1 + \bar{e} - \bar{X}\sum_{i=1}^n b_i e_i$$

$$r_i = (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_i + e_i = (\sum_{i=1}^n b_i e_i)(\bar{X} - X_i) + e_i - \bar{e}$$

即$\hat{\beta}_1, \hat{\beta}_2, r_i$均为$e_1, \dots, e_n$的线性组合，由$e_1, \dots, e_n \ iid \sim N(0, \sigma^2)$ 知，$\hat{\beta}_1, \hat{\beta}_2, r_i$均服从正态分布。

$$Var(\hat{\beta}_1) = Var(\bar{e} - \bar{X}\sum_{i=1}^n b_i e_i) = \frac{\sigma^2}{n} + \frac{\bar{X}^2\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$Var(\hat{\beta}_2) = Var(\sum_{i=1}^n b_i e_i) = \sum_{i=1}^n b_i^2\sigma^2 = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

从而

$$\begin{aligned}
&Cov(\hat{\beta}_2, r_i) = Cov(\hat{\beta}_2, -\hat{\beta}_1 - \hat{\beta}_2 X_i + e_i) \\
&= -Cov(\hat{\beta}_2, \hat{\beta}_1) - X_i Var(\hat{\beta}_2) + Cov(\hat{\beta}_2, e_i) \\
&= -\frac{-\bar{X}\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} - \frac{X_i\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{(X_i - \bar{X})\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} = 0
\end{aligned}$$

$$\begin{aligned}
&Cov(\hat{\beta}_1, r_i) = Cov(\hat{\beta}_1, -\hat{\beta}_1 - \hat{\beta}_2 X_i + e_i) \\
&= -Var(\hat{\beta}_1) - X_i Cov(\hat{\beta}_1, \hat{\beta}_2) + Cov(\hat{\beta}_1, e_i) \\
&= -\frac{\sigma^2}{n} - \frac{\bar{X}^2\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} - \frac{-X_i\bar{X}\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sigma^2}{n} - \frac{\bar{X}\sigma^2(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = 0
\end{aligned}$$

即$\hat{\beta}_1, \hat{\beta}_2$均与$r_i$独立 $\Rightarrow (\hat{\beta}_1, \hat{\beta}_2)$与$r_i$ 独立 $\Rightarrow (\hat{\beta}_1, \hat{\beta}_2)$与$\sum_{i=1}^n r_i^2$独立.

6) 若将模型中关于误差假定替换为：$e_1,\ldots,e_n\ iid, E(e_i)=0, 0<E(e_i^2)=\sigma^2<\infty$。我们知道$\hat{\beta}_2=\beta_2+\sum_{i=1}^n b_ie_i$，其中$b_i=\frac{X_i-\bar{X}}{\sum_{i=1}^n(X_i-\bar{X})^2}, 1\leqslant i\leqslant n$. 假定

$$\max_{1\leqslant i\leqslant n}\frac{b_i^2}{\sum_{j=1}^n b_j^2}\to 0 \tag{1}$$

试利用概率论中所学的Linderberg-Feller定理，证明

$$\frac{\hat{\beta}_2-\beta_2}{\sigma\sqrt{\sum_{i=1}^n b_i^2}}\xrightarrow{d}N(0,1) \tag{2}$$

进一步，若(2)中的σ用$\hat{\sigma}=\sqrt{(n-2)\sum_{i=1}^n r_i^2}$代替后，(2)的结论仍然成立。将条件(1) 简化后，说明其中蕴含的设计点应该满足的条件。

i) 由条件知

$$\frac{\hat{\beta}_2-\beta_2}{\sigma\sqrt{\sum_{i=1}^n b_i^2}}=\frac{\sum_{i=1}^n b_ie_i}{\sqrt{\sum_{i=1}^n b_i^2\sigma^2}}=\frac{\sum_{i=1}^n b_ie_i-E(\sum_{i=1}^n b_ie_i)}{\sqrt{Var(\sum_{i=1}^n b_ie_i)}}$$

其中b_ie_i是相互独立的随机变量，满足

$$E(b_ie_i)=0, Var(b_ie_i)=b_i^2\sigma^2$$

现只需证明b_ie_i服从CLT，由Linderberg-Feller定理，只需证明

$$\lim_{n\to\infty}\frac{1}{\sum_{i=1}^n b_i^2\sigma^2}\sum_{i=1}^n\int_{|x|\geqslant\tau\sqrt{\sum_{i=1}^n b_i^2\sigma^2}}x^2dF_i(x)=0,\ \forall\tau>0$$

现设e_i的分布函数为F，$|b_i|e_i\ (b_i\neq 0)$的分布函数为F_i，则

$$F_i(x)=P(|b_i|e_i<x)=P(e_i<\frac{x}{|b_i|})=F(\frac{x}{|b_i|})$$

上式化为

$$
\begin{aligned}
&\frac{1}{\sum_{i=1}^{n} b_i^2\sigma^2}\sum_{i=1}^{n} b_i^2 \int_{(x/b_i)^2 \geqslant \tau^2(\sum_{i=1}^{n} b_i^2\sigma^2)/b_i^2} \left(\frac{x}{b_i}\right)^2 dF\left(\frac{x}{|b_i|}\right) \\
&= \frac{1}{\sum_{i=1}^{n} b_i^2\sigma^2}\sum_{i=1}^{n} b_i^2 \int_{y^2 \geqslant \tau^2(\sum_{i=1}^{n} b_i^2\sigma^2)/b_i^2} y^2 dF(y) \\
&\leqslant \frac{1}{\sum_{i=1}^{n} b_i^2\sigma^2}\sum_{i=1}^{n} b_i^2 \int_{y^2 \geqslant \tau^2(\sum_{i=1}^{n} b_i^2\sigma^2)/\max\limits_{1\leqslant i\leqslant n} b_i^2} y^2 dF(y) \\
&= \frac{1}{\sigma^2}\int_{y^2 \geqslant \tau^2(\sum_{i=1}^{n} b_i^2\sigma^2)/\max\limits_{1\leqslant i\leqslant n} b_i^2} y^2 dF(y)
\end{aligned}
$$

由条件知，当$n \to \infty$时

$$
\frac{\sum_{j=1}^{n} b_j^2}{\max\limits_{1\leqslant i\leqslant n} b_i^2} \to \infty
$$

又$0 < E(e_i^2) = \sigma^2 < \infty$，故此时

$$
\frac{1}{\sigma^2}\int_{y^2 \geqslant \tau^2(\sum_{i=1}^{n} b_i^2\sigma^2)/\max\limits_{1\leqslant i\leqslant n} b_i^2} y^2 dF(y) \to 0
$$

即Lindeberg条件成立，从而b_ie_i服从中心极限定理，满足

$$
\frac{\hat{\beta}_2 - \beta_2}{\sigma\sqrt{\sum_{i=1}^{n} b_i^2}} \xrightarrow{d} N(0,1)
$$

ii) 若用$\hat{\sigma}=\sqrt{(n-2)\sum_{i=1}^n r_i^2}$代替(2)中的$\sigma$

$$
\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n-2}\sum_{i=1}^n \left[(\beta_1-\hat{\beta}_1)+(\beta_2-\hat{\beta}_2)X_i+e_i\right]^2 \\
&= \frac{1}{n-2}\sum_{i=1}^n \left[(\sum_{i=1}^n b_i e_i)(\bar{X}-X_i)+e_i-\bar{e}\right]^2 \\
&= \frac{1}{n-2}\sum_{i=1}^n \left[(\sum_{i=1}^n b_i e_i)^2(\bar{X}-X_i)^2+(e_i-\bar{e})^2+2(\sum_{i=1}^n b_i e_i)(e_i-\bar{e})(\bar{X}-X_i)\right] \\
&= \frac{1}{n-2}\left[\sum_{i=1}^n (e_i-\bar{e})^2+\frac{(\sum_{i=1}^n b_i e_i)^2}{\sum_{i=1}^n b_i^2}-2\frac{[\sum_{i=1}^n (X_i-\bar{X})e_i]^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right] \\
&= \frac{1}{n-2}\sum_{i=1}^n (e_i-\bar{e})^2-\frac{1}{n-2}\frac{(\sum_{i=1}^n b_i e_i)^2}{\sum_{i=1}^n b_i^2}
\end{aligned}
$$

其中第一项，由题目假设知

$$
\frac{1}{n-2}\sum_{i=1}^n (e_i-\bar{e})^2=\frac{n-1}{n-2}\frac{1}{n-1}\sum_{i=1}^n (e_i-\bar{e})^2 \to \sigma^2
$$

第二项的一阶矩存在，由Markov不等式知

$$
P\left(\left|\frac{1}{n-2}\frac{(\sum_{i=1}^n b_i e_i)^2}{\sum_{i=1}^n b_i^2}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon}E\left|\frac{1}{n-2}\frac{(\sum_{i=1}^n b_i e_i)^2}{\sum_{i=1}^n b_i^2}\right| = \frac{\sigma^2}{(n-2)\varepsilon} \to 0
$$

从而

$$
\hat{\sigma}_n^2=\hat{\sigma}^2 \xrightarrow{d} \sigma^2 \ \Rightarrow \ \frac{\sigma}{\hat{\sigma}} \xrightarrow{d} 1
$$

由Slutsky定理知

$$
\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}\sqrt{\sum_{i=1}^n b_i^2}}=\frac{\hat{\beta}_2-\beta_2}{\sigma\sqrt{\sum_{i=1}^n b_i^2}}\frac{\sigma}{\hat{\sigma}} \xrightarrow{d} N(0,1)
$$

iii) (1)中条件简化为

$$
\max_{1\leqslant i\leqslant n}\frac{b_i^2}{\sum_{j=1}^n b_j^2}=\max_{1\leqslant i\leqslant n}\frac{(X_i-\bar{X})^2}{\sum_{i=1}^n (X_i-\bar{X})^2} \to 0
$$

该条件意味着，单个设计点X_i与设计点样本均值$\bar{X}$的偏差，是总体偏差的高阶无穷小，即不能出现偏差全部都集中在一个设计点上的情况。

7) 取$b_1=1, b_2=\cdots=b_n=0$，判断(1)和(2)是否成立.

当$b_1=1, b_2=\cdots=b_n=0$时

$$\max_{1\leqslant i\leqslant n}\frac{b_i^2}{\sum_{j=1}^n b_j^2}=1$$

即条件(1)不成立。

$$\frac{\hat{\beta}_2-\beta_2}{\sigma\sqrt{\sum_{i=1}^n b_i^2}}=\frac{\hat{\beta}_2-\beta_2}{\sigma}=\frac{\sum_{i=1}^n b_i e_i}{\sigma}=\frac{e_1}{\sigma}\sim\ N(0,1)$$

即条件(2)成立。

2. 设$X_1,\ldots,X_n\ldots$是$iid\ r.v.s$，则

$$E|X_1|<\infty\ \ iff\ \ \frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\to 0, a.s.$$

Lemma. $\sum_{n=1}^\infty P(|X_1|\geqslant n)\leqslant E(|X_1|)\leqslant 1+\sum_{n=1}^\infty P(|X_1|\geqslant n)$

1) 若$E|X_1|<\infty$，由引理知，对$\forall\varepsilon>0$有

$$\sum_{n=1}^\infty P(\{|X_1|\geqslant n\})<\infty$$

从而对充分小的$\varepsilon>0$，有

$$\sum_{n=1}^\infty P(\{|X_1|\geqslant n\varepsilon\})<\infty \tag{3}$$

已知

$$\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\xrightarrow{a.s.}0\Longleftrightarrow\lim_{k\to\infty}P\left(\bigcup_{n=k}^\infty\{\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\geqslant\varepsilon\}\right)=0$$

现看

$$
\begin{aligned}
&P(\bigcup_{n=k}^{\infty}\{\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\geqslant\varepsilon\})=P(\bigcup_{n=k}^{\infty}\{n\varepsilon\leqslant\max_{1\leqslant i\leqslant n}|X_i|<(n+1)\varepsilon\})\\
&\leqslant\sum_{n=k}^{\infty}P(\{n\varepsilon\leqslant\max_{1\leqslant i\leqslant n}|X_i|<(n+1)\varepsilon\})\\
&=\sum_{n=k}^{\infty}\Bigg[P(\{\max_{1\leqslant i\leqslant n}|X_i|\geqslant n\varepsilon\})-P(\{\max_{1\leqslant i\leqslant n}|X_i|\geqslant(n+1)\varepsilon\})\Bigg]\\
&\leqslant P(\{\max_{1\leqslant i\leqslant k}|X_i|\geqslant k\varepsilon\})+\sum_{n=k}^{\infty}\Bigg[P(\{\max_{1\leqslant i\leqslant n+1}|X_i|\geqslant(n+1)\varepsilon\})-P(\{\max_{1\leqslant i\leqslant n}|X_i|\geqslant(n+1)\varepsilon\})\Bigg]\\
&=P(\{\max_{1\leqslant i\leqslant k}|X_i|\geqslant k\varepsilon\})+\sum_{n=k}^{\infty}P(\{|X_{n+1}|\geqslant(n+1)\varepsilon\})\\
&=1-P(\{\max_{1\leqslant i\leqslant k}|X_i|<k\varepsilon\})+\sum_{n=k}^{\infty}P(\{|X_{n+1}|\geqslant(n+1)\varepsilon\})\\
&=1-\Bigg[P(\{|X_1|<k\varepsilon\})\Bigg]^k+\sum_{n=k}^{\infty}P(\{|X_1|\geqslant(n+1)\varepsilon\})
\end{aligned}
$$

令$k\to\infty$，对第二项有

$$
\begin{aligned}
&\lim_{k\to\infty}\Bigg[P(\{|X_1|<k\varepsilon\})\Bigg]^k=\lim_{k\to\infty}\Bigg[1-P(\{|X_1|\geqslant k\varepsilon\})\Bigg]^k\\
&=\lim_{k\to\infty}\exp\Bigg(kln[1-P(\{|X_1|\geqslant k\varepsilon\})]\Bigg)=\exp\Bigg(\lim_{k\to\infty}kln[1-P(\{|X_1|\geqslant k\varepsilon\})]\Bigg)\\
&=\exp\Bigg(-\lim_{k\to\infty}kP(\{|X_1|\geqslant k\varepsilon\})\Bigg)=1
\end{aligned}
$$

又由(3)知第三项趋于0，故有

$$
\lim_{k\to\infty}P\Bigg(\bigcup_{n=k}^{\infty}\{\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\geqslant\varepsilon\}\Bigg)=0
$$

从而

$$
\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\to 0, a.s.
$$

2) 若$\frac{1}{n}\max\limits_{1\leqslant i\leqslant n}|X_i|\to 0, a.s.$，则$\forall\varepsilon>0$，有

$$0=\lim_{k\to\infty}P(\bigcup_{n=k}^{\infty}\{\frac{1}{n}\max_{1\leqslant i\leqslant n}|X_i|\geqslant\varepsilon\})\geqslant\lim_{k\to\infty}P(\bigcup_{n=k}^{\infty}\{\frac{1}{n}|X_n|\geqslant\varepsilon\})$$

由$X_1,\ldots,X_n\ldots,iid$，上式可化为

$$\lim_{k\to\infty}\sum_{n=k}^{\infty}P(\{\frac{1}{n}|X_n|\geqslant\varepsilon\})=0$$

从而

$$\sum_{n=1}^{\infty}P(\{\frac{1}{n}|X_n|\geqslant\varepsilon\})=\sum_{n=1}^{\infty}P(\{\frac{1}{\varepsilon}|X_1|\geqslant n\})<\infty$$

由引理知

$$E(\frac{1}{\varepsilon}|X_1|)\leqslant 1+\sum_{n=1}^{\infty}P(\{\frac{1}{\varepsilon}|X_1|\geqslant n\})$$

故

$$E|X_1|<\infty$$