第二周

《高等概率论》
《高等统计学》
《统计机器学习》

概率论 $\to$ 研究随机性，减少不确定性给决策带来的困难

课程设置：

1. 极大似然估计 $\to$ 证明方法（优化方法引进）
2. 渐进有效估计
3. 交换方法 $\to$ 给一类估计量算...，只用一个公式
4. 总体分布的估计
5. 非参数回归系数的局部常数和局部线性统计
6. 分位数估计与Bahadur表示
7. M-估计
8. Jackknife 1
9. Jackknife 2
10. Bootstrap
11. 多重检验
12. 统计机器学习方法

Chapter 1

模型 $\to$ 机器学习三要素之一

概率模型 $\begin{cases} \text{参数模型} \quad \text{分布确定，参数不确定} \quad \text{能用有限个参数来确定Y的分布} \quad \text{可识别性} \\ \text{非参数模型} \end{cases}$

可识别性：分布确定，则参数就确定下来

例 $N(\mu,\sigma^2)$ 参数为(μ,σ^2)

$$\mu = 0$$
$$\sigma^2 = 1$$

则标准正态分布 $\mu = 1, \sigma^2 = 1$ 唯一确定

若参数为 (μ,σ)

则标准正态分布 $\mu = 1, \sigma = \pm 1$ 不唯一确定

对于指数分布 $\varepsilon(\lambda\mu) = \varepsilon(\eta)$

以η 为参数的指数分布 $\to$ 可识别

以(λ, μ) 为参数的指数分布 $\rightarrow$ 不可识别

可识别性： 若 $\theta_1 \neq \theta_2$ ，则 $F_{\theta_1} \neq F_{\theta_2}$

不能确定参数，则不好解释，则一般要求参数模型可识别；若不能识别，则加限制条件，去掉冗余，使之可识别

非参数模型

例如：

- 所有连续分布
- 所有对称分布

有很多的可能性，不能确定某一类，则要考虑一个很大的非参模型

子集

$\Downarrow$

半参数模型： 一部分特征可以参数化

例：线性模型 $Y = X^T\beta + \varepsilon \quad E(\varepsilon) = 0 \quad Var(\varepsilon) < +\infty$

Y 为响应变量

则Y 的分布 $F(X^T\beta, \sigma^2)$ 为一类分布，不知道其具体的，但知道其均值方差

优缺点：

参数模型：

- 优点：简单
- 缺点：可能犯模型设定错误

非参数模型：

- 稳健 $\Rightarrow$ 神经网络
- 难算

指数分布族

$$f_\theta(X)=\exp\left\{\sum_{j=1}^{k}\eta_j(\theta)T_j(X)-B(\theta)\right\}h(x)$$

参数根统计量可分离 $\Rightarrow$ 充分统计量 $\to$ 完备统计量 $\to$ 数据压缩（最好用不超过参数个数的统计量）

$$B(\theta)=\ln\{\cdots\}$$

自由的是 $\eta_1(\theta),\cdots,\eta_k(\theta)$ ，确定之后 $B(\theta)$ 随之确定

Canonical form

$$f_\eta(X)=\exp\left\{\sum\eta_jT_j(X)-A(\eta)\right\}h(X)$$

参数变为η

例： $X\sim N(\mu,\sigma^2)\qquad\theta=(\mu,\sigma^2)$

$$\begin{aligned}
f_\theta(X)&=\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\\
&\doteq\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\sigma^2\\
&\text{后面的部分移入}B(\theta)\text{ 或}H(x)\\
&\doteq\exp\left\{-\frac{1}{2}\frac{x^2-2\mu x+\mu^2}{\sigma^2}\right\}\\
&=\exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\cdot x^2+\frac{\mu}{\sigma^2}\cdot x-\frac{1}{2}\frac{1}{\sigma^2}\mu^2\right\}\\
&\text{最后一项移入}B(\theta)\\
&\doteq\exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\cdot x^2+\frac{\mu}{\sigma^2}x\right\}\\
&\doteq\exp\left\{\eta_1(\theta)x^2+\eta_2(\theta)x\right\}
\end{aligned}$$

$\eta=(\eta_1,\eta_2)$

$(\mu,\sigma^2)\mapsto(\eta_1,\eta_2)$

$f_\eta(x)=\exp\left\{\eta_1x^2+\eta_2x\right\}\cdots$

D 1.1.4 Ex ... of full rank

- T,η 都不是满足一个线性约束
- 参数空间包含k 维矩形

  例子： $f_\eta(X)=\exp\left\{\eta_1x^2+\eta_2x\right\}$

Ex 1.1.3 $X_1,\ldots,X_n\sim N(\mu,\theta^2)$ 不是 **full rank**

参数空间 $\eta_\theta=(\frac{1}{\theta},-\frac{1}{2\theta^2})$ 不包含一个长方形

1.1.2 指数分布具有封闭性

- X 是k 参数指数分布
- Y 是k 参数指数分布

- X, Y 独立

  $\Rightarrow (X, Y)$ 仍然是 k 参数指数分布

1.1.3 X 是k 参数指数分布，则其 k 个统计量的分布函数 $T = (T_1(x), \cdots, T_k(x))$

$f_T(t) = \cdots = \exp\{\sum \eta_j(\theta)t_j - B(\theta)\}h_\theta(t)$ 不是 k 参数指数分布

1.1.4 $T(x)$ 的各界矩 $\to$ 只跟 $A(\eta)$ 有关系

m.g.f 矩母函数 $M(s) = \exp\{A(s+\eta) - A(\eta)\}$

矩母函数 $M(s) = Ee^{s\cdot Y} = \int e^{sY} f(y)dy < \infty$

若 $s = it$，则转化为特征函数， $s \cdot Y$ 为内积

可以求导来算Y 的矩

$$\begin{aligned} E(Y) = M'(s)|_{s=0} = (EYe^{sY})|_{s=0} \\ Var(Y) = E(Y^2) - (EY)^2 \end{aligned}$$

Ber(p) 伯努利分布 $x = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$

$$\begin{aligned} f_p(x) = p^x(1-p)^{1-x} &= \exp\{x\ln p + (1-x)\ln(1-p)\} \\ &= \exp\left\{x \ln\frac{p}{1-p} + \ln(1-p)\right\} \\ \eta = \ln\frac{p}{1-p} \to p = \frac{e^\eta}{1+e^\eta} \qquad &= \exp\{x\cdot\eta - \ln(1+e^\eta)\} \\ &= \exp\{x\cdot\eta - A(\eta)\} \end{aligned}$$

$$p = EX = E(T(x)) = A'(\eta) = \frac{e^\eta}{1+e^\eta} = \frac{e^{x^T\beta}}{1+e^{x^T\beta}}$$

其中，$A'(\eta) = \frac{e^\eta}{1+e^\eta}$ 是连接函数

$Y \quad EY = X^T\beta$ 线性模型 $\to$ 逻辑回归

- 两点分布对于逻辑回归
- 泊松分布对于泊松回归 $\to$ 指数分布族

参数与统计量分不开的反例 $[0, \theta]$ 上的均匀分布

不可能写成指数分布

$X \sim U(0, \theta) \qquad f_\theta(x) = \frac{1}{\theta} I_{0<x<\theta} \qquad \neq \eta(\theta)\cdot T(X)$

最大次序统计量

尺度变换族

Chapter 2

1. 依分布收敛 $X_n \to^d X \qquad F_n(x) = P(X_n \leq x) \to P(X \leq x) = F(X) \quad n \to \infty$
2. 依概率收敛 $X_n \to^p X \qquad \forall \varepsilon > 0. \quad P\{\mid X_n - X \mid > \varepsilon\} \to 0 \quad n \to \infty$
3. 几乎既然收敛 $X_n \to^{a.s.} X \qquad \forall \varepsilon > 0 \quad \forall n > 0 \quad p\{\cup_{k \geq n} [\mid X_k - X \mid > \varepsilon]\} \to 0$
4. Lr 收敛 $X_n \to^{lr} X \qquad E \mid X_n \to X \mid^r \to 0$

这几个收敛的关系（推导写作作业）

- 3 $\Rightarrow$ 2
- 4 $\Rightarrow$ 2 马尔可夫不等式 $P\{\mid X_n - X \mid < \varepsilon\} \leq \frac{E|X_n - X|}{\varepsilon^r}$
- 2 $\Rightarrow$ 1
- 若 $X = c$ 为常数，则 1 $\Rightarrow$ 2

解答：

- 依概率收敛 $\rightarrow$ 依分布收敛

依概率收敛：

$$X_n \rightarrow^p X \qquad \forall \epsilon > 0 \qquad P\{\mid X_n - X \mid > \epsilon\} \rightarrow 0 \qquad n \rightarrow \infty$$

依分布收敛：

$$X_n \rightarrow^d X \qquad F_n(x) = P\{X_n \leq x\} \rightarrow P\{X \leq x\} = F(x) \qquad n \rightarrow \infty$$

证明：

设随机变量的分布函数为 $F(x), F_1(x), F_2(x), \cdots$

令 $x' < x$ ，则

$$\begin{aligned}\{X \leq x'\} &= \{X_n \leq x, X \leq x'\} \cup \{X_n > x, X \leq x'\} \\ &\subset \{X_n \leq x\} \cup \{\mid X_n - X \mid \geq x - x'\}\end{aligned}$$

由 $X_n \rightarrow^p X$ ，有$P\{\mid X_n - X \mid \geq x - x'\} \rightarrow_0 \qquad n \rightarrow \infty$

所以 $F(x') \leq \lim_{n\rightarrow\infty} F_n(x)$

再令 $x' \rightarrow x$ ，即可得 $F(x-0) = \lim_{x'\rightarrow x^-} F(x') \leq F_n(x)$

同理可得 ，当$x'' > x$ 时，有 $\lim_{n\rightarrow\infty} F_n(x) \leq F(x'')$

令 $x'' \rightarrow x$ ，即得 $\lim_{n\rightarrow\infty} F_n(x) \leq F(x+0)$

所以，对所有的x ，有

$$F(x-0) \leq \lim_{n\rightarrow\infty} F_n(x) \leq F(x+0)$$

当x 是 $F(x)$ 的连续点时，有 $F(x-0) = F(x+0)$ ，定理得证

- 若 $x=c$ 是常数，则 依分布收敛 $\rightarrow$ 依概率收敛

依分布收敛：

$$X_n \rightarrow^d X \qquad F_n(x)=P\{X_n \leq x\} \rightarrow P\{X \leq x\}=F(x) \qquad n \rightarrow \infty$$

依概率收敛：

$$X_n \rightarrow^p X \qquad \forall \epsilon>0 \qquad P\{\mid X_n-X \mid>\epsilon\} \rightarrow 0 \qquad n \rightarrow \infty$$

证明：

记 X_n 的分布函数为 $F_n(x) \quad n=1,2,\cdots$

常数 c 的分布函数为退化分布

$$F(x)=\begin{cases}0 & x<c \\ 1 & x \geq c\end{cases}$$

对任意的 $\epsilon>0$ ，有

$$\begin{aligned} P\{\mid X_n-c \mid \geq \epsilon\} &= P\{X_n \geq c+\epsilon\}+P\{X_n \leq c-\epsilon\} \\ &\leq P\{X_n>c+\epsilon/2\}+P\{X_n \leq c-\epsilon\} \\ &=1-F_n(c+\epsilon/2)+F_n(c-\epsilon)\end{aligned}$$

由于 $x=c+\epsilon/2$ 和 $x=c-\epsilon$ 均为 $F(x)$ 的连续点，且 $F_n(x) \rightarrow^d F(x)$

所以当 $n \rightarrow \infty$ 时，有

$$F_n(c+\epsilon/2) \rightarrow F(c+\epsilon/2)=1 \qquad F_n(c-\epsilon) \rightarrow F(c-\epsilon)=0$$

由此可得

$$P\{\mid X_n-c \mid \geq \epsilon\} \rightarrow 0$$

定理得证

- 几乎必然收敛 $\rightarrow$ 依概率收敛

几乎必然收敛：

$$X_n \rightarrow^{a.s.} X \qquad \forall \epsilon > 0 \quad \forall n > 0 \quad P\left(\cup_{k \geq n}\{\mid X_k - X \mid > \epsilon\}\right) \rightarrow 0$$

依概率收敛：

$$X_n \rightarrow^{p} X \qquad \forall \epsilon > 0 \qquad P\{\mid X_n - X \mid > \epsilon\} \rightarrow 0 \qquad n \rightarrow \infty$$

任取一个正实数 $\epsilon > 0$

$$A_n = \cup_{k > n}\{x : \mid X_m(x) - X(x) \mid > \epsilon\}$$

因为 $A_{n+1} \subseteq A_n$ ，所以 $A_\infty = \cap_{n \geq 1} A_n$

令 $O = \{x : \lim_{n \rightarrow \infty} X_n(x) = X(x)\}$

因为几乎必然收敛 $P(O) = 1$

对于 O 中任意元素 x ，能找到一个正整数 N ，使得当 $n > N$ 时，有

$$|X_n(x) - X(x) \mid < \epsilon$$

则 O 中任意元素不属于 A_∞

所以 $P(A_\infty) < P(\overline{O}) = 0$

$$P\{\mid X_n - X \mid > \epsilon\} < P(A_n) \rightarrow 0$$

定理得证

- Lr收敛 $\rightarrow$ 依概率收敛

Lr收敛：

$$X_n \rightarrow^{Lr} X \qquad E \mid X_n - X \mid^r \rightarrow 0 \qquad n \rightarrow \infty$$

依概率收敛：

$$X_n \rightarrow^{p} X \qquad \forall \epsilon > 0 \qquad P\{\mid X_n - X \mid > \epsilon\} \rightarrow 0 \qquad n \rightarrow \infty$$

证明：

```
由Chebyshev不等式：
```

$$\forall \epsilon > 0 \qquad P\{\mid X_n - X \mid > \epsilon\} \leq \frac{E \mid X_n - X|^r}{\epsilon^r}$$

$$\begin{aligned} P\{\mid X_n - X \mid > \epsilon\} &= \int_{|X_n - X| > \epsilon} dF(x) \\ &\leq \int_{|X_n - X| > \epsilon} \frac{|X_n - X|^r}{\epsilon^r} dF(x) \\ &= \frac{1}{\epsilon^r} \int_{|X_n - X| > \epsilon} \mid X_n - X \mid^r \, dF(x) \\ &= \frac{1}{\epsilon^r} E \mid X_n - X|^r \end{aligned}$$

```
可得：
```

$$\begin{aligned} P\{\mid X_n - X \mid > \epsilon\} &\leq \frac{E \mid X_n - X|^r}{\epsilon^r} \\ &= \frac{1}{\epsilon^r} E \mid X_n - X|^r \\ \text{由Lr收敛} &\rightarrow 0 \qquad n \rightarrow \infty \end{aligned}$$

```
定理可证
```