- 教材[1]: 《Statistical Inference》2nd Edition，George Casella and Roger L. Berger，2002.
- 参考书[0]：《数理统计（第二版）》，韦来生　编著，科学出版社，2015
- 参考书[2]: 《高等数理统计（第二版）》，茆诗松　王静龙　濮小龙　编著，高等教育出版社，2006.
- **主讲内容：【1】第10章　渐近评价**
  1. 第一节　预备知识
  2. 第二节　点估计
  3. 第三节　假设检验
  4. 第四节　区间估计

- 【0】定义2.5.1 当样本容量n趋向无穷时，统计量的分布趋于一确定分布，则后者的分布称为统计量的极限分布，也常称为大样本分布。

注 当样本容量n充分大时，极限分布可作为统计量的近似分布。

- 研究统计量的极限分布的意义
  1. 统计量的精确分布难求，提供了一种近似方法进行估计；
  2. 统计量的精确分布可求但过于复杂，使用极限分布更方便；
  3. 有些统计推断方法的优良性本身在于研究极限性质。

- 【0】定义2.5.2 当样本容量$n \to \infty$时，一个统计量或统计推断方法的性质称为大样本性质（large sample properties）。当样本大小固定时，统计量或统计推断方法的性质称为小样本性质（small sample properties）。

注 大样本性质和小样本性质的差别不在于样本容量的大小，而在于样本容量$n \to \infty$还是n固定时去考虑研究这个性质。

- **概率1收敛**(定义5.5.6) 称随机变量序列$\{X_n\}$以概率1收敛(converge almost surely) 于随机变量X，如果对任意$\epsilon > 0$，都有

$$\mathbb{P}(\lim_{n\to+\infty} |X_n - X| < \epsilon) = 1.$$

记为$X_n \xrightarrow{a.s.} X$.

- **依概率收敛**(定义5.5.1) 称随机变量序列$\{X_n\}$依概率收敛(converge in probability)于随机变量X，如果对任意$\epsilon > 0$，都有

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.$$

记为$X_n \xrightarrow{P} X$ 或者$X_n \xrightarrow{p} X$.

- **依分布收敛**(定义5.5.10) 称随机变量序列$\{X_n\}$依分布收敛(converge in distribution)于随机变量X，如果对$F_X(x)$的任意连续点x，都有

$$\lim_{n\to+\infty} F_{X_n}(x) = F_X(x).$$

记为$X_n \xrightarrow{L} X$ 或者$X_n \xrightarrow{d} X$.

注 三者强弱关系

$$\xrightarrow{a.s.} \quad \Rightarrow \quad \xrightarrow{P} \quad \Rightarrow \quad \xrightarrow{L}.$$

- 设随机变量序列$\{X_n\} i.i.d.$，$\mathbb{E}X_i = \mu$，$Var(X_i) = \sigma^2$ 且$\mu < \infty$, $0 < \sigma^2 < \infty$。关于$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$的
  1. Kolmogorov强大数定律;
  2. 弱大数定律;
  3. Lindeberg中心极限定理.

Theorem (定理5.5.4)

设随机变量序列$\{X_1, X_2, \cdots\}$ 依概率收敛于随机变量X, h是一个连续函数，则$h(X_1), h(X_2), \cdots$ 依概率收敛于随机变量$h(X)$.

Example (1.1)

已知随机变量序列$\{X_1, \cdots, X_n\}$ 同分布，$\mathbb{E}X_i = \mu$，$Var(X_i) = \sigma^2$且$\mu < \infty$, $0 < \sigma^2 < \infty$。设

$$Cov(X_i, X_j) = \begin{cases} c, & |i-j| = 1 \\ 0, & \text{其它} \end{cases},$$

这里$c > 0$。问：$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$是否依概率收敛于$\mu$?

Lemma (Slutsky 定理，定理5.5.17)

令$\{X_n\}$和$\{Y_n\}$是两个随机变量序列，满足当$n \to \infty$时，$X_n \xrightarrow{L} X$, $Y_n \xrightarrow{P} c$ （c 为有限常数），则有

(1) $X_n \pm Y_n \xrightarrow{L} X \pm c$;

(2) $X_n Y_n \xrightarrow{L} cX$;

(3) $X_n / Y_n \xrightarrow{L} X / c (c \neq 0)$.

- **【定义5.5.20】Taylor多项式** 如果函数$g(x)$有r阶导函数，即存在$g^{(r)}(x)=\frac{d^r}{dx^r}g(x)$,则对任意常数$a$，$g(x)$在$a$附近的r 阶Taylor多项式（Taylor Polynomial of order r about a）为

$$T_r(x)=\sum_{j=0}^{r}\frac{g^{(j)}(a)}{j!}(x-a)^j.$$

Theorem (Taylor定理，定理5.5.21)

如果$g^{(r)}(a)=\frac{d^r}{dx^r}g(x)|_{x=a}$存在，，则

$$\lim_{x\to a}\frac{g(x)-T_r(x)}{(x-a)^r}=0.$$

Taylor定理在统计学中的应用

- 设随机向量序列$\mathbf{T}_n=(T_{1,n},\cdots,T_{k,n})$的期望均为$\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)$.
- 若函数g二阶导数存在有限，且当$n\to 0$时$\mathbb{E}_{\boldsymbol{\theta}}|\mathbf{T}_n-\boldsymbol{\theta}|^2\to 0$，则参数估计$g(\mathbf{T_n})$的期望近似估计

$$\mathbb{E}_{\boldsymbol{\theta}}[g(\mathbf{T}_n)]\approx g(\boldsymbol{\theta})+\sum_{j=1}^{k}g_j'(\boldsymbol{\theta})\mathbb{E}_{\boldsymbol{\theta}}(T_{j,n}-\theta_j)=g(\boldsymbol{\theta}).$$

这里$g_j'(\boldsymbol{\theta})=\frac{\partial}{\partial x_j}g(\mathbf{x})|_{\mathbf{x}=\boldsymbol{\theta}}$.

- 若进一步，当$n\to 0$时$\mathbb{E}_{\boldsymbol{\theta}}|\mathbf{T}_n-\boldsymbol{\theta}|^\ell=o\left(\mathbb{E}_{\boldsymbol{\theta}}|\mathbf{T}_n-\boldsymbol{\theta}|^2\right)$, $\ell=3,4$，方差的近似估计

$$Var_{\boldsymbol{\theta}}[g(\mathbf{T}_n)]\approx\sum_{j=1}^{k}\left[g_j'(\boldsymbol{\theta})\right]^2 Var_{\boldsymbol{\theta}}T_{j,n}+2\sum_{j>i}^{k}g_j'(\boldsymbol{\theta})g_i'(\boldsymbol{\theta})Cov_{\boldsymbol{\theta}}(T_{j,n},T_{i,n}).$$

Δ方法

Theorem (Δ方法，定理5.5.24)

设随机变量序列T_n满足：$\sqrt{n}(T_n-\theta)$依分布收敛于$N(0,\sigma^2)$，函数g在指定θ处满足：$g'(\theta)$存在且不为零，则当$n\to\infty$ 时，

$$\sqrt{n}[g(T_n)-g(\theta)] \xrightarrow{L} N(0,\sigma^2[g'(\theta)]^2).$$

Example (1.2)

已知$X_1,\cdots,X_n$ $i.i.d$, $\mathbb{E}X_1=\mu>0$, 且$Var(X_1)=\sigma^2$, $\mathbb{E}|X_1-\mu|^4=\mu_4$均有限。请分别给出总体方差$\sigma^2$已知和未知情形下$g(\overline{X}_n)=\frac{1}{\overline{X}_n}$的的渐近分布。

- 对于随机变量序列T_n，若存在非退化分布F及$a_n(\theta)$及$b_n(\theta)>0$ s.t.

$$\lim_{n\to\infty}\mathbb{P}_\theta\left(\frac{T_n-a_n(\theta)}{b_n(\theta)}\leq x\right)=F(x),\forall x\in\mathbb{R},\theta\in\Theta.$$

则称T_n具有**渐近分布**F 。

Example (1.3)

设随机样本$X_1, \cdots, X_n$ 总体服从$Bernoulli(p)$, $0 < p < 1$. 问：能否给出$h(\overline{X}_n) = \overline{X}_n(1-\overline{X}_n)$的渐近分布？

问题 如果函数$g'(\theta) = 0$，则Δ方法是否仍适用？

Theorem (二阶Δ方法，定理5.5.26)

设随机变量序列T_n满足：$\sqrt{n}(T_n - \theta) \xrightarrow{L} N(0, \sigma^2)$, 函数$g$在指定的$\theta$处满足$g'(\theta) = 0$，$g''(\theta)$存在且不为零，则

$$\frac{2n}{\sigma^2 g''(\theta)}\left[g(T_n) - g(\theta)\right] \xrightarrow{L} \mathcal{X}_1^2.$$

Δ方法的推广

Theorem (多元Δ方法，定理5.5.28)

设随机样本$\mathbf{X}_1, \cdots, \mathbf{X}_n$ 满足：$\mathbb{E}_{\boldsymbol{\mu}} X_{ij} = \mu_i$ 且$Cov(X_{ik}, X_{jk}) = \sigma_{ij}$，函数$g$有连续一阶偏导，且在指定的$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)$处满足：$\tau^2 = \sum\sum \sigma_{ij} \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_i} \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_j} > 0$，则

$$\sqrt{n}[g(\overline{X}_1, \cdots, \overline{X}_p) - g(\mu_1, \cdots, \mu_p)] \xrightarrow{L} N(0, \tau^2).$$

Example (1.4)

设$(X_1, Y_1), \cdots, (X_n, Y_n)$ $i.i.d \sim N_2\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right)$，
μ_x和μ_y均不为零。请给出函数μ_x/μ_y的一个点估计并给出其渐近分布。

作业 习题5: 29, 31, 32；44.