《高等统计》总复习

何志坚

2020-12-23

提纲

- 第一章：概率论基础
- 第二章：统计学基础
- 第三章：无偏估计
- 第六章：假设检验

模型设定

统计学研究随机样本: X 为概率空间$(\Omega, \mathcal{F}, P)$ 上的随机向量.

- 样本: $X = (X_1, \ldots, X_n)$
- 总体: P (未知或部分未知)
- 统计模型: $X \sim P \in \mathcal{P}$, $\mathcal{P}$ 为分布族

参数分布族: $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$, θ 未知, Θ 参数空间

非参分布族: $\mathcal{P} = \{$所有连续型分布$\}$

概率密度

如果$P \ll \nu$ (绝对连续), 则 Radon-Nikodym 导数定义为P 关于σ-有限测度ν 的概率密度 (PDF)

$$f(x) = \frac{dP}{d\nu}(x) \quad \text{a.e. } \nu$$

- 如果X 取值可列, ν 为计数测度, 则$f(x)$ 为离散分布的概率质量函数 (PMF)
- 如果ν 为 Lebesgue 测度, 则$f(x)$ 为连续型分布的概率密度
- 此外, ν 可以是上述两种测度的混合, 如截断分布 (例 1.12)

注意: 统计模型往往假设分布族关于某σ-有限测度ν 绝对连续, i.e., $\mathcal{P} \ll \nu$, 于是存在ν 下的密度$\frac{dP}{d\nu}(x)$. 参数分布族有时表示为

$$\mathcal{P} = \left\{ f_\theta(x) = \frac{dP_\theta}{d\nu}(x), \theta \in \Theta \right\}$$

例子: IID 数据

假设$X_1, \ldots, X_n$ 是独立同分布的样本. 此时, X_i 的边际分布唯一决定X 的分布.

- $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$
- $X_i \overset{iid}{\sim} E(a, \theta)$
- $X_i \overset{iid}{\sim} U(a, b)$
- $X_i \overset{iid}{\sim} Bi(p, 1)$

更多常见分布参见表 1.2.

例子: 独立但不同分布 (IND) 数据

$$X_i \overset{ind}{\sim} P_i,\ i = 1, \ldots, n$$

- 线性回归模型:

$$X_i = Z_i^\top \beta + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \Leftrightarrow X_i \overset{ind}{\sim} N(Z_i^\top \beta, \sigma^2)$$

指数型分布族

$$\frac{dP_\theta}{d\nu}(\omega) = \exp\{\eta(\theta)^\top T(\omega) - \xi(\theta)\}h(\omega)$$

- $\eta : \Theta \mapsto \mathbb{R}^p$, p 是该指数型分布族的维数
- $T(\omega) : \Omega \mapsto \mathbb{R}^p$
- $T(X)$ 为 "重要的" 统计量

自然指数型分布族:

$$\frac{dP_\theta}{d\nu}(\omega) = \exp\{\eta^\top T(\omega) - \zeta(\eta)\}h(\omega)$$

- 自然参数$\eta \in \Xi$
- 自然参数空间$\Xi = \{\eta(\theta)|\theta \in \Theta\}$
- 若Ξ 存在一个内点, 则该分布族满秩

指数型分布族的一些性质

- P_θ 的支持集$\mathcal{S}=\{\omega\in\Omega|\frac{dP_\theta}{d\nu}(\omega)>0\}$ 与θ 无关
- 排除均匀分布族$U(a,b)$
- P_{θ_1} 与P_{θ_2} 等价 (互相绝对连续)
- 指数型分布族表示不唯一
- $T(X)$ 是充分统计量
- 若分布族满秩, 则$T(X)$ 是充分完备和最小充分统计量.

位置 -尺度分布族

$$P_{(\mu,\Sigma)}(B) = P(\Sigma^{-1/2}(B-\mu)),\ B \in \mathcal{B}^k$$

- 位置参数μ, 尺度参数$\Sigma^{1/2}$
- 若$X \sim P_{(\mu,\Sigma)}$, 则$\Sigma^{-1/2}(X-\mu)$ 分布确定, 即与μ 和$\Sigma^{1/2}$ 无关
- 任何一种给定分布都可以通过这种方式得到一个位置 -尺度分布族

例子:

1. $N(\mu,\sigma^2) \stackrel{d}{=} \mu + \sigma N(0,1),\ \mu \in \mathbb{R},\ \sigma > 0$
2. $E(a,\theta) \stackrel{d}{=} a + \theta E(0,1),\ a \in \mathbb{R},\ \theta > 0$

统计量$T(X)$ 是样本信息的汇总，广泛用于统计推断.

1. 点估计：估计量$T(X) \in \mathbb{R}$
2. 区间估计：两个估计量$[T_1(X), T_2(X)] \subset \mathbb{R}$
3. 假设检验：检验统计量$T(X) \in [0, 1]$

几类重要的统计量

1. 充分统计量
2. 最小充分统计量
3. 辅助统计量
4. 完备统计量
5. 充分完备统计量

定义: 条件分布$X|T(X)$ 与总体分布P 无关 (或与θ 无关)

直观理解: $T(X)$ 包含所有X 中有关P 的信息

比较两种信息:

1. 提供X 的观测值x
2. 仅提供$T(x)$: "山寨" 样本$\tilde{X} \sim X|T(X) = T(x)$ 的观测$\tilde{x}$

容易证明: $X \overset{d}{=} \tilde{X}$

- 如果$\dim(T(X)) \ll \dim(X)$, 则$T(X)$ 在不损失信息的基础上能够很好地压缩数据.
- 什么情况$\dim(T(X))$ 与$\dim(X)$ 无关? 这对大数据降维非常重要!

充分统计量的充要条件 -因子分解定理

一般形式:

$$\frac{dP}{d\nu}(\omega) = g_P(T(\omega))h(\omega)$$

参数形式:

$$\frac{dP_\theta}{d\nu}(\omega) = g_\theta(T(\omega))h(\omega)$$

指数型分布族:

$$\frac{dP_\theta}{d\nu}(\omega) = \exp\{\eta(\theta)^\top T(\omega) - \xi(\theta)\}h(\omega)$$

- $T(X)$ 是充分统计量
- 更重要的是$\dim(T(X)) = p$ 与X 的维数n(样本量) 无关

两个事实:

- 完全样本X 是一个平凡的充分统计量
- $T=\psi(S)$ 是充分统计量$\Rightarrow$ S 是充分统计量

如何找到一个“精简”的充分统计量?— 最小充分统计量

定义: T 是最小充分统计量$\Leftrightarrow$ 对任意充分统计量S, 存在可测函数$\psi(\cdot)$ 使得$T=\psi(S)$ a.s. $\mathcal{P}$.

直观含义: 所有的充分统计量都可以“压缩”成最精简的充分统计量T.

最小充分统计量存在性以及判断条件

存在性: Bahadur (1957) 证明了只要$\mathcal{P}$ 关于某一σ 有限测度绝对连续, 则最小充分统计量存在.

判断条件 (定理 2.3): 在子分布族$\mathcal{P}_0 \subset \mathcal{P}$ 讨论最小性即可, 但前提 a.s. $\mathcal{P}_0 \Rightarrow$ a.s. $\mathcal{P}$.

例 2.14 给出一个重要的结果: 在一定条件下 (例如满秩), 指数型分布族中的$T(X)$ 是最小充分的.

辅助统计量与完备统计量

辅助统计量$V(X)$: 分布确定与P 无关

一阶辅助统计量$V(X)$: $E[V(X)]$ 与P 无关

完备统计量$T(X)$: $E[f(T)] = 0\ \forall P \in \mathcal{P} \Leftrightarrow f(T) = 0$ a.s. $\mathcal{P}$.

命题 2.1: 对满秩的指数分布族, $T(X)$ 是充分完备的.

注: 一般来讲, 通过定义来判断完备性 (可在子分布族$\mathcal{P}_0$ 中证明)

- 例 2.16: $X_i \overset{iid}{\sim} U(0,\theta)$, $X_{(n)}$ 是充分完备的
- 例 2.17: 对所有连续型分布构成的分布族, $T = (X_{(1)}, \ldots, X_{(n)})$ 是充分完备的

Basu 定理: 如果$V(X)$ 是辅助统计量, $T(X)$ 是充分完备统计量, 则两者独立.

一些联系

- 完备充分统计量$\Rightarrow$ 最小充分统计量, 反之不成立
- 如果T 是完备统计量, 则$S=\psi(T)$ 也是完备的
- 如果$T=\psi(S)$ 是充分, 则S 也是充分的
- 如果T 与S 可相互表示 (等价), 则

$$T\text{充分完备} \Leftrightarrow S\text{充分完备}$$

$$T\text{最小充分} \Leftrightarrow S\text{最小充分}$$

例子

正态总体: $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

- μ 未知, σ^2 已知, $\bar{X}$ 充分完备, 最小充分
- σ 未知, μ 已知, $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ 充分完备, 最小充分
- μ, σ^2 均未知, $(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2)$ 充分完备, 最小充分

指数总体: $X_i \overset{iid}{\sim} E(a, \theta)$

- θ 未知, a 已知, $\bar{X}$ 充分完备, 最小充分
- a 未知, θ 已知, $X_{(1)}$ 充分完备, 最小充分
- a, θ 均未知, $(\bar{X}, X_{(1)})$ 充分完备, 最小充分

- 决策规则$T(X) : (\mathcal{X}, \mathcal{F}_{\mathcal{X}}) \mapsto (\mathbb{A}, \mathcal{F}_{\mathbb{A}})$, 样本空间到决策空间的可测函数
- 损失函数: $L(P, a)$ or $L(\theta, a)$
- 风险: $R_T(P) = E[L(P, T(X))]$ or $R_T(\theta) = E[L(\theta, T(X))]$

参数估计: 平方损失下风险为

$$R_T(\theta) = E[(T(X) - \theta)^2] = MSE$$

假设检验: 0-1 损失: 如果做出正确判断$L(p, a) = 0$, 如果做出错误判断$L(p, a) = 1$, 则风险为

$$R_T(P) = \begin{cases} E[T(X)], & P \in \mathcal{P}_0 \\ 1 - E[T(X)], & P \in \mathcal{P}_1 \end{cases}$$

随机化决策规则

给定样本观测$X = x$, 随机决策服从分布: $T(x) \sim \delta(x, \cdot)$

- 损失函数$L(P, \delta, x) = E_{a \sim \delta(x,\cdot)}[L(P, a)]$
- 风险: $R_\delta(P) = E_{X \sim P}[L(P, \delta, X)]$

设$\mathcal{T}$ 为一些决策规则的集合.

- T 是$\mathcal{T}$-最优: T 不次于其它任何$\mathcal{T}$ 中的决策规则, i.e., $R_T(P) \le R_{T'}(P)\ \forall P \in \mathcal{P}, T' \in \mathcal{T}$
- T 是$\mathcal{T}$-可容: $\mathcal{T}$ 中不存在比T 更好的
- $\mathcal{T}$-最优$\Rightarrow\mathcal{T}$-可容, 反之不成立

Rao-Blackwell 定理 (定理 2.5): 设T 是充分统计量, T_0 为任意决策规则. 如果$L(P,a)$ 是凸函数, 则$T_1 = E[T_0|T]$ 不次于T_0. 如果$L(P,a)$ 是严格凸函数且T_0 不是T 的函数, 则$T_1 = E[T_0|T]$ 优于T_0(即说明T_0 不是容许的).

注意: $\mathcal{T}$-可容是最起码的要求, 但$\mathcal{T}$-最优要求太高, 往往不存在.

贝叶斯规则与 minmax 规则

贝叶斯风险 (平均风险):

$$r_T(\Pi) = E_{P\sim\Pi}[R_T(P)]$$

- T 是$\mathcal{T}$-贝叶斯规则: T 的贝叶斯风险不大于其它任何$\mathcal{T}$ 中的贝叶斯风险

最坏情况下的风险:

$$w_T := \sup_{P\in\mathcal{P}} R_T(P)$$

- T 是$\mathcal{T}$-minmax 规则: T 的最坏情况下的风险不大于其它任何$\mathcal{T}$ 中的最坏情况下的风险

无偏估计与 UMVUE

- $X \sim P \in \mathcal{P}$
- $T(X)$ 是参数ϑ 的估计量
- 无偏估计: $E[T(X)] = \vartheta$ 对任意$P \in \mathcal{P}$ 成立
- 若存在参数ϑ 的一个无偏估计量, 则称ϑ 是可估的

UMVUE: 一致最小方差无偏估计, 即$\mathcal{T}$-最优, 其中$\mathcal{T}$ 为所有无偏估计

定理 3.1 (Lehmann-Scheffe 定理): 设$T(X)$ 是充分完备统计量. 如果ϑ 是可估的, 则存在 (a.s. $\mathcal{P}$) 唯一的 UMVUE 且可表示为$h(T)$ 的形式.

UMVUE 的求解

思路一: 假设$h(T)$ 为 UMVUE, 利用$E[h(T)] = \vartheta$ 对任意$P \in \mathcal{P}$ 成立求解函数$h(\cdot)$ 的形式.

- 例 3.1 和例 3.2

思路二: 任取一无偏估计T_0, 计算$E[T_0|T]$

- 例 3.3 和例 3.4

注: 无论哪种思路, 首先需要找出充分完备统计量T, 这一步与待估的参数ϑ 无关.

定理 3.2: $T(X)$ 是 UMVUE 的充要条件是$T(X)$ 与所有 “零” 的无偏估计$U(X)$ 不相关, i.e., $E[T(X)U(X)] = 0 \ \forall P \in \mathcal{P}$.

- 例 3.7 和例 3.8

设$T(X)$ 是$g(\theta)$ 的无偏估计，在一些正则性条件下 (积分和求导可交换顺序)，有

$$\operatorname{Var}(T) \geq [\nabla g(\theta)]^\top I(\theta)^{-1} \nabla g(\theta),$$

其中 Fisher 信息矩阵

$$I(\theta) = E[\nabla \log f_\theta(X) \nabla \log f_\theta(X)^\top] = -E[\nabla^2 \log f_\theta(X)].$$

单参数时，

$$\operatorname{Var}(T) \geq \frac{g'(\theta)^2}{E[(\partial_\theta \log f_\theta(X))^2]}.$$

- 例 3.10

$$X = Z\beta + \epsilon$$

- Z 给定
- ϵ 的假设: (A1) $\epsilon \sim N(0, \sigma^2 I_n)$; (A2) $\epsilon \sim (0, \sigma^2 I_n)$; (A3) $\epsilon \sim (0, \Sigma)$

最小二乘估计 (LSE):

$$\hat{\beta} = \arg\min_{\beta} ||X - Z\beta||_2^2 = (Z^\top Z)^- Z^\top X$$

- LSE 一定存在, 但可能不唯一
- LSE 对应线性方程组$Z^\top Z\hat{\beta} = Z^\top X$ 的解 (见例 3.12,3.13)
- LSE 唯一$\Leftrightarrow$ Gram 矩阵$Z^\top Z$ 非奇异$\Leftrightarrow$ Z 列满秩

LSE 统计性质

考虑两类参数估计问题

1. $\ell^\top \beta$, 其中$\ell \in \mathcal{R}(Z)$
2. σ^2

估计量及其性质 (定理 3.6,3.7,3.9):

1. $\ell^\top \hat{\beta}$: 在 (A3) 下无偏且唯一; 在 (A1) 下 UMVUE; 在 (A2) 下 BLUE
2. $\hat{\sigma}^2 = ||X - Z\hat{\beta}||_2^2/(n-r)$, $r = \mathrm{rank}(Z) < n$: 在 (A2) 下无偏, 在 (A1) 下 UMVUE

抽样分布定理 (定理 3.8): 在 (A1) 下, $\ell^\top \hat{\beta}$ 与$\hat{\sigma}^2$ 独立, 且

$$\ell^\top \hat{\beta} \sim N(\ell^\top \beta, \sigma^2 \ell^\top (Z^\top Z)^- \ell), \ (n-r)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-r}.$$

数学形式:

- $H_0 : P \in \mathcal{P}_0 \ v.s. P \in \mathcal{P}_1$
- $H_0 : \theta \in \Theta_0 \ v.s. P \in \Theta_1$

检验统计量$T(X) \in [0,1]$ 表示给定X 拒绝H_0 的概率

功效函数: $\beta_T(P) = E[T(X)]$ 表示在P 下作出拒绝H_0 的概率

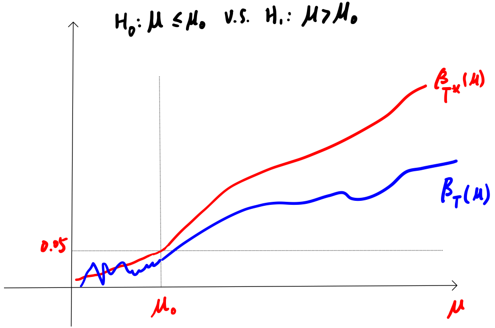
检验准则 (控制第一类错误的概率):

$$\sup_{P \in \mathcal{P}_0} \beta_T(P) \le \alpha,$$

左边为T 的检验大小, 右边为显著性水平.

UMP 定义: T^* 是大小为α 的 UMP 当且仅当

- $\sup_{P\in\mathcal{P}_0}\beta_{T^*}(P)=\alpha$
- 对任意检验T 满足$\sup_{P\in\mathcal{P}_0}\beta_T(P)\le\alpha$, 则有$\beta_{T^*}(P)\le\beta_T(P)\ \forall P\in\mathcal{P}_1$.



考虑简单检验

$$H_0 : P = P_0 \ v.s. \ H_1 : P = P_1$$

UMP 检验为:

$$T_*(X) = \begin{cases} 1 & f_1(X) > cf_0(X) \\ \gamma & f_1(X) = cf_0(X) \\ 0 & f_1(X) < cf_0(X) \end{cases}$$

其中$c \geq 0, \gamma \in [0,1]$ 满足$E_{P_0}[T_*(X)] = \alpha$.

- 应用: 例 6.1

单调似然比 (单参数分布族)

单调似然比定义：对任意$\theta_2 > \theta_1$，似然比

$$\frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = h_{\theta_1,\theta_2}(Y(X))$$

关于统计量$Y(X) \in \mathbb{R}$ 单调不减.

$$H_0 : \theta = \theta_0 \ v.s. \ H_1 : \theta = \theta_1 > \theta_0$$
$$H_0 : \theta = \theta_0 \ v.s. \ H_1 : \theta > \theta_0$$
$$H_0 : \theta \leq \theta_0 \ v.s. \ H_1 : \theta > \theta_0$$

的 UMP 均为

$$T_*(X) = \begin{cases} 1 & f_{\theta_1}(X) > cf_{\theta_0}(X) \\ \gamma & f_{\theta_1}(X) = cf_{\theta_0}(X) \\ 0 & f_{\theta_1}(X) < cf_{\theta_0}(X) \end{cases} = \begin{cases} 1 & Y(X) > \tilde{c} \\ \gamma & Y(X) = \tilde{c} \\ 0 & Y(X) < \tilde{c} \end{cases}$$

- 应用：例 6.6

考试安排

- 时间: 2020/12/30 (周三) 下午 3:00-5:00
- 地点: 四号楼 4141
- 考试范围: 除了第一章
- 考试重点: 作业和上课讲过的例子 (包括第六章)
- 注意事项: 雨课堂请用真实姓名以便准确登记平时成绩

Questions?