第二章　点估计

- 数理统计的任务：用样本去推断总体的分布。
- 参数统计模型：分布依赖于一个或几个参数，参数估计就是拟合概率分布。
- 参数估计：点估计和区间估计
- 点估计：（【0】定义3.1.1）设$\vec{X} = (X_1, \ldots, X_n)$为从某个总体中抽取的样本，$\widehat{g}(\vec{X})$ 是样本函数，用$\widehat{g}(\vec{X})$作为$g(\theta)$的估计，称为点估计（Point estimation）.
- 主讲内容：
  1. 求估计量的方法
  2. 估计量的评价方法
- 第一节的主讲内容
  1. 矩估计
  2. 极大似然估计
  3. 贝叶斯估计

1.1 矩估计(Method of Moments, 简记MoM)

约定 样本$\vec{X} = (X_1, \ldots, X_n)$ $i.i.d. \sim X$，k有限正整数，总体矩和样本矩

1. 总体k阶原点矩：$\alpha_k = \mathbb{E}(X^k)$;
2. 样本k阶原点矩：$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$;
3. 总体k阶中心矩：$\mu_k = \mathbb{E}(X - \mathbb{E}X)^k$;
4. 样本k阶中心矩：$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^k$.

- **矩估计方法** 用样本矩估计总体矩。具体而言：

Step1 计算低阶总体矩α_k 或μ_k，它们应是关于参数$\vec{\theta}$的函数表达式；

Step2 建立方程，反解得到参数$\vec{\theta}$关于总体矩α_k、μ_k的函数表达式，记为$\vec{\theta}(\alpha_k, \mu_k)$；

Step3 将函数表达式中的总体矩α_k、μ_k分别替换为样本矩$\hat{\alpha}_k$、$\hat{\mu}_k$，从而得到关于样本矩函数表达式的参数估计：$\hat{\vec{\theta}} = \vec{\theta}(\hat{\alpha}_k, \hat{\mu}_k)$.

注1 在Step1中，通常需要的低阶矩个数等于参数个数；

注2 记$\vec{\theta}$的矩估计$\hat{\vec{\theta}}_{MoM}$, 则可测函数$g(\vec{\theta})$的矩估计$\hat{g}_{MoM} = g\left(\hat{\vec{\theta}}_{MoM}\right)$.

注3 上述定义与结果可参考【0】定义3.2.1.

举例说明

Example (1.1)

设$X_1, \ldots, X_n$ $i.i.d. \sim Poisson(\lambda)$, $\lambda > 0$未知。求λ和$g(\lambda) = \lambda^2 - \lambda$的矩估计。

注 计算$Var(X) = \lambda$，得λ的另一估计

$$\widehat{\lambda}^*_{MoM} = \widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2.$$

问：哪个估计更好？（待讨论，见本章2.2小节）

Example (1.2)

设$X_1, \ldots, X_n$ $i.i.d. \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$ 均未知。求：

1. μ, σ^2 以及$g(\mu, \sigma^2) = \frac{\mu}{\sigma}$的矩估计；
2. 总体X的变异系数（coefficient of variation）$\nu = \frac{\sqrt{\mu_2}}{\alpha_1}$，偏度（skewness）$\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$和峰度（kurtosis）$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3$ 的矩估计。（定义参考【0】p20）

举例说明

Example (1.3)

设$X_1, \ldots, X_n$ $i.i.d. \sim Gamma(\alpha, \beta)$, $\alpha, \beta > 0$ 均未知。求α, β的矩估计。

问题 参数估计是随机变量/向量，其分布如何？

Example (1.4)

[【0】例3.2.7] 设$(X_i, Y_i), i = 1, 2, \ldots, n$是从一个2维总体中抽取的简单随机样本，求总体分布的协方差σ_{XY}与相关系数ρ的矩估计。

练习 【0】例3.2.1-3.2.6

作业 习题3：Ex. 7, 8, 9.

- **定义1.1** （【0】定义3.3.1）设随机变量$X_1,\dots,X_n$ 具有联合概率密度函数$f(x_1,\dots,x_n|\vec{\theta})$, $\vec{\theta}\in\Theta$, 则$\vec{\theta}$ 的**似然函数**（Likelihood function）定义为

$$L(\vec{\theta})=f(x_1,\dots,x_n|\vec{\theta}).$$

- **定义1.2** （【0】定义3.3.2）**极大似然估计**（Maximum Likelihood Estimator）是使似然函数$L(\vec{\theta})$在Θ上（极限）达到唯一极大的$\vec{\theta}$值，简记$\hat{\vec{\theta}}_{MLE}$, i.e.

$$\hat{\vec{\theta}}_{MLE}=\arg\max_{\vec{\theta}\in\Theta}L(\vec{\theta}).$$

对数似然函数（log likelihood）定义为

$$\ell(\vec{\theta})=\log L(\vec{\theta}).$$

注1 $\hat{\vec{\theta}}_{MLE}$亦是使对数似然$\ell(\vec{\theta})$达到最大的参数值。

注2 这里$L(\vec{\theta})$的表达式虽与$\vec{X}=(X_1,\ldots,X_n)$的联合p.d.f./p.m.f.一样，但意义不同。

1. 联合p.d.f.：$\vec{\theta}$固定，$\vec{x}=(x_1,\ldots,x_n)$是变量；
2. $L(\vec{\theta})$：$\vec{x}$固定，$\vec{\theta}$是变量；也即，$L(\vec{\theta})$是给定$\vec{X}$的一个观测值$\vec{x}$后关于$\vec{\theta}$的函数，i.e. $L(\vec{\theta})=L(\vec{\theta}\,|\,\vec{x})$.

注3 $\hat{\vec{\theta}}_{MLE}=\hat{\vec{\theta}}_{MLE}(\vec{x})$，即$\hat{\vec{\theta}}_{MLE}$依赖于观测值$\vec{x}$，是使观测值$\vec{x}$"最有可能"出现的$\vec{\theta}$参数值。

注4 设g为定义在Θ上的某一可测函数，则$\hat{g}_{MLE}=g\left(\hat{\vec{\theta}}_{MLE}\right)$（参考【1】定理7.2.10）.

注5 对于某一特定样本观测值$\vec{x}$，若$\hat{\vec{\theta}}_{MLE}(\vec{x})$在$\Theta$ 上取不到，但存在$\{\hat{\vec{\theta}}_{MLE}(\vec{x_n})\}_{n=1}^{\infty}\subset\Theta$, $s.t.$ $\lim_{n\to\infty}\hat{\vec{\theta}}_{MLE}(\vec{x_n})=\hat{\vec{\theta}}_{MLE}(\vec{x})$，则$\hat{\vec{\theta}}_{MLE}(\vec{x})$也是$\vec{\theta}$的MLE.

Example (1.5)

[【0】例3.3.1] 单个样本$X \sim Binomial(n,p)$，已知$p=\frac{1}{4}$或$\frac{3}{4}$，如何由观测值判断p取$\frac{1}{4}$还是$\frac{3}{4}$?

问题 设$X_1,\ldots,X_n$ $i.i.d. \sim X$，总体X的p.d.f.为$f(x|\vec{\theta})$，$\vec{\theta}\in\Theta$，
则$X_1,\ldots,X_n$的联合$p.d.f.$为$f(x_1,\cdots,x_n|\vec{\theta})=\prod_{i=1}^n f(x_i|\vec{\theta})$，从而得似然函数

$$L(\vec{\theta})=\prod_{i=1}^n f(x_i|\vec{\theta}).$$

对数似然

$$\ell(\vec{\theta})=\sum_{i=1}^n \log f(x_i|\vec{\theta}).$$

如何求$\widehat{\vec{\theta}}_{MLE}=\arg\max_{\vec{\theta}\in\Theta} L(\vec{\theta})=\arg\max_{\vec{\theta}\in\Theta}\ell(\vec{\theta})$?

如何寻找MLE 1

- 一般情形： $\ell(\vec{\theta})$在Θ内是$\vec{\theta}=(\theta_1,\ldots,\theta_k)$的连续可微函数。

Step1 求解

$$\frac{\partial \ell}{\partial \theta_i}(\vec{\theta})=0, i=1,\cdots,k,$$

得$\hat{\theta}^*$；

Step2 证明$\ell(\vec{\theta})$的Hessian矩阵（二阶导矩阵）在$\hat{\theta}^*$处负定，且这样的$\hat{\theta}^*$唯一。这一条件也可换为“唯一极大值在Θ内点上达到”。

- 特殊情形：指数族

Theorem (1.1)

[【0】定理3.3.1] 设$X_1, \ldots, X_n$ *i.i.d.* 服从具有自然参数形式

$$f(x|\overrightarrow{\eta}) = C(\overrightarrow{\eta}) \exp\left\{\sum_{i=1}^{k} \eta_i T_i(x)\right\} h(x)$$

的指数族总体，$\overrightarrow{\eta} \in \Theta^*$. 对于其对数似然

$$\ell(\overrightarrow{\eta}) = n \log C(\overrightarrow{\eta}) + \sum_{i=1}^{k} \eta_i \sum_{j=1}^{n} T_i(x_j) + \sum_{j=1}^{n} \log h(x_j)$$

若一阶导$\frac{\partial \ell}{\partial \eta_i}(\overrightarrow{\eta}) = 0, i = 1, \cdots, k$ 方程组的解$\widehat{\overrightarrow{\eta}}^*$ 存在且是参数空间Θ^*的内点，则$\widehat{\overrightarrow{\eta}}^*$ 必唯一且是$\overrightarrow{\eta}$的*MLE*。

举例说明

注　回忆定义1.2注4，对于具有自然参数形式的指数族的一般表达式，

$$f(x|\vec{\theta}) = c(\vec{\theta})\exp\left\{\sum_{i=1}^{k} Q_i(\vec{\theta})T_i(x)\right\}h(x),$$

若存在函数$\vec{R}$, s.t. $\vec{\theta} = \vec{R}(\vec{\eta})$，则$\hat{\vec{\theta}}_{MLE} = \vec{R}(\hat{\vec{\eta}}_{MLE})$。

Example (1.6)

分别求如下分布族中未知参数的MLE估计：

1. $Poisson(\lambda), \lambda > 0$;
2. $N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$.

问题1　回忆各个参数的矩估计，与MLE相同吗？各个参数MLE的分布?

问题2　考虑$g(\lambda) = \lambda^2 - \lambda$和$h(\mu, \sigma^2) = \frac{\mu}{\sigma}$的MLE?

练习　【0】例3.3.2 ∼ 3.3.5, 3.3.8.

如何寻找MLE 2

- 当求导方法不适用时，可以从定义1.1出发求参数的MLE.

Example (1.7)

[【0】例3.3.6] 设$X_1, \ldots, X_n$ $i.i.d. \sim U(0, \theta)$, $\theta > 0$ 未知。求θ的MLE。

Example (1.8)

[【0】例3.3.7] 设$X_1, \ldots, X_n$ $i.i.d. \sim U(\theta, \theta + 1)$, $\theta \in \mathbb{R}$ 未知。求θ的MLE。

练习 【0】例3.3.6, 3.3.7.

作业 习题3：Ex. 13, 15, 17, 21.

- 矩估计和极大似然估计方法：参数$\vec{\theta}$是一个未知的固定值/向量。
- 贝叶斯方法：参数$\vec{\theta}$是一个具有概率分布的随机变量。
- 基本模型 假设样本$\vec{X} = (X_1, \ldots, X_n)$具有联合概率密度/质量函数$f(\vec{x}|\vartheta)$,
  1. 设定先验分布(Prior distribution)(【0】定义7.1.1)：参数ϑ是一个随机变量，具有概率密度/质量函数$\pi(\theta)$, $\theta \in \Theta$;
  2. 求后验分布（Posterior distribution）(【0】定义7.1.2)：给定样本观测值$\vec{x}$，求ϑ的条件分布$\pi(\theta|\vec{x})$;
- **定义1.3** （【0】7.3.1） 参数ϑ的**贝叶斯估计**(Bayes Estimator)是其后验分布的期望：

$$\hat{\vartheta}_B = \mathbb{E}(\vartheta|\vec{X}),$$

也称为ϑ的平方损失意义下的贝叶斯估计。

举例说明

- 后验分布的求解步骤
  1. 求$(\vec{X}, \vartheta)$的联合p.d.f/p.m.f.，$f(\vec{x}, \theta) = f(\vec{x}|\theta)\pi(\theta)$;
  2. 求$\vec{X}$的边缘p.d.f/p.m.f.，$f_X(\vec{x}) = \int_\Theta f(\vec{x}, \theta)d\theta$;
  3. 后验分布

$$\pi(\theta|\vec{x}) = \frac{f(\vec{x}, \theta)}{f_X(\vec{x})} = \frac{f(\vec{x}|\theta)\pi(\theta)}{\int_\Theta f(\vec{x}|\nu)\pi(\nu)d\nu} \tag{1}$$

Example (1.9)

（【1】例7.2.14） 设$X_1, \ldots, X_n$ $i.i.d. \sim Bernoulli(p)$, $p \in (0, 1)$未知,

$$P_\theta(X_i = 1) = p, \quad P_\theta(X_i = 0) = 1 - p, \quad i = 1, \ldots, n.$$

分别求如下两种情形下p的后验分布及其贝叶斯估计：

1. $p \sim U(0, 1)$;
2. $p \sim Beta(a, b)$, $a > 0, b > 0$ 已知。

注1 由第一章第6节例6.1可知，$S=\sum_{i=1}^{n}X_i$是p的充分统计量，因此我们可考虑贝叶斯估计$\widehat{p}_B=\mathbb{E}(p|S)$.

注2 $U(0,1)\overset{d}{=}Beta(1,1)$，因此例1.9中(1)可看成是(2)的特例。

- 求后验分布的更简方法: 注意到(1)中分母与θ无关，因此可看成（关于θ）是常数；同时，如果只关心包含θ的因子部分，（i）令"核"表示一个函数中与参数θ有关的因子，（ii）令"$\propto$"表示"正比于"，则求后验分布的过程可简化为:
  1. 分别写出似然函数$L(\theta)=f(\vec{x}|\theta)$和先验分布p.d.f./p.m.f.$\pi(\theta)$的核；
  2. 后验分布p.d.f./p.m.f.的核
  $$\pi(\theta|\vec{x})\propto\{L(\theta)\text{的核}\}\cdot\{\pi(\theta)\text{的核}\} \tag{2}$$
  3. 将(2)式添加一个正则化常数因子（可与$\vec{x}$有关），即可得后验分布的概率密度/质量函数$\pi(\theta|\vec{x})$.

举例说明

Example (1.10)

设$X_1, \ldots, X_n$ $i.i.d. \sim N(\mu, \sigma^2)$. 令$\xi = \frac{1}{\sigma^2}$，

(1)若μ已知，$\sigma^2 > 0$未知，假设ξ的先验分布$\xi \sim Gamma(\alpha, \beta)$, 求$\xi$的后验分布及其贝叶斯估计$\widehat{\xi}_B$. （参考【0】例7.2.13逆Gamma分布）

(2)若$\mu \in \mathbb{R}, \sigma^2 > 0$均未知，假设$\mu, \xi$独立，它们的先验分布分别为$\mu \sim N(\mu_0, \tau^2)$, $\xi \sim Gamma(\alpha, \beta)$，求$(\mu, \xi)$ 的后验分布.

注1 μ未知，$\sigma^2 > 0$已知情形，参考【0】例7.2.4；另自行练习【0】例7.2.10，7.2.11，7.2.12，7.4.2，7.4.3。

注2 观察在例1.9和1.10(1)中，均有先验分布和后验分布同属同一分布族的特点。

共轭先验分布

- **定义1.4** [【0】定义7.2.3, 【1】定义7.2.15] 假设先验分布$\pi(\theta)$取自分布族$\mathcal{F}$，如果对任一样本观测值$\vec{x}$，后验分布$\pi(\theta|\vec{x})$仍属于$\mathcal{F}$，则称$\mathcal{F}$是一个**共轭先验分布族**（Conjugate Prior Distribution Family）.

Example (1.11)

设$X_1, \ldots, X_n$ $i.i.d. \sim Poisson(\lambda)$. 证明参数$\lambda$的共轭先验分布族为Gamma分布族。

注1 λ的贝叶斯估计$\widehat{\lambda}_B = \frac{\sum_{i=1}^n X_i + \alpha}{n+\beta}$;

注2 自行练习【0】例7.2.9.

作业 习题7：Ex. 3, 4, 5, 11, 14(后验期望估计改为贝叶斯估计).