

| STAT201B Lecture 7 MoM and MLE

| 1 Parametric Inference 概述

| 1.1 Definition: Location scale family

🔗 Logic ▾

关于 location scale family 的更多描述, 见 [STA3020 Lecture 3](#)

令:

- Y 为服从分布 F 的随机变量
- F_μ 为 $Y + \mu$ 的 distribution function
- F_σ 为 σY 的 distribution function
- $F_{\mu,\sigma}$ 为 $\sigma Y + \mu$ 的 distribution function

则

- Family $\{F_\mu : -\infty < \mu < \infty\}$ 被称为 location family (e.g. $\mathcal{N}(\mu, 1)$)
- Family $\{F_\sigma : \sigma > 0\}$ 被称为 scale family (e.g. $\mathcal{N}(0, \sigma^2)$)
- Family $\{F_{\mu,\sigma} : -\infty < \mu < \infty, \sigma > 0\}$ 被称为 location scale family (e.g. $\mathcal{N}(\mu, \sigma^2)$)

⚠ Remark ▾

WLOG, 我们通常假设 $\mathbb{E}[Y] = 0, \text{Var}[Y] = 1$

| 1.2 Definition: Parametric model

一个 parametric model 通常有以下形式:

$$\mathcal{F} = \{F(x, \theta) : \theta \in \Theta\}$$

其中 $\Theta \subset \mathbb{R}^k$ 为 parametric space

⚠ Remark ▾

Class \mathcal{F} 的选取通常基于我们对于特定问题的 knowledge (如 data generating mechanism), 需要特别注意是否存在违背这些 assumptions 的情形

🔗 Logic ▾

接下来我们将介绍两种 parametric estimation methods: Method of Moment 和 Maximum Likelihood Estimation

| 2 Method of Moments 的定义

| 2.1 Definition: Method of Moments

🔗 Logic ▾

关于 method of moments 的更多论述, 见 [STA2004 Lecture 4](#)

令:

- parameter of interest 为 $\theta = (\theta_1, \dots, \theta_k)$
- j^{th} (population) moment 为

$$\alpha_j := \alpha_j(\theta) = \mathbb{E}_\theta[X^j] = \int x^j dF_\theta(x), \quad \text{for } j = 1, \dots, k$$

- j^{th} sample moment 为

$$\hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^n X_i^j$$

则 **method of moments (MOM) estimator** $\hat{\theta}_n$ 满足:

$$\begin{aligned} \alpha_1(\hat{\theta}_n) &= \hat{\alpha}_1 \\ \alpha_2(\hat{\theta}_n) &= \hat{\alpha}_2 \\ &\dots \\ \alpha_k(\hat{\theta}_n) &= \hat{\alpha}_k \end{aligned}$$

⚠ Remark: MOM generalization ▾

除了考虑 $\alpha_j(\theta) = \mathbb{E}_\theta[X^j]$, 我们还可以转而去考虑 $\alpha_j(\theta) = \mathbb{E}_\theta[g(X)^j]$, 并且令 $\hat{\theta}_n$ 满足:

$$\alpha_j(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n g(X_i)^j, \quad \text{for } j = 1, \dots, k$$

3 Maximum likelihood estimator 的定义

🔗 Logic ▾

关于 maximum likelihood estimator 的更多描述, 见 [STA3020 Lecture 7](#), 包括:

- MLE 的定义
- MLE 的 consistency 及其 conditions
- MLE 的 CAN property 及其 conditions
- MLE 的 invariance property

关于 likelihood function 的变种, 见 [STA3020 Lecture 8](#), [STA3020 Lecture 9](#), [STA3020 Lecture 10](#), [STA3020 Lecture 11](#), 包括:

- Composite likelihood
- Quasi likelihood
- Profile likelihood
- Generalized profile likelihood

3.1 Definition: Maximum likelihood estimator

令 sample $X = \{X_1, \dots, X_n\}$ 有 distribution function $f(x|\theta), \theta \in \Theta$, 则

- **likelihood function** 被定义为:

$$\mathcal{L}_n(\theta) = f(X_1, \dots, X_n; \theta) = \prod_{i=1}^n f(X_i; \theta) \quad \text{for indenpendet data}$$

⚠ Remark ▾

换言之, likelihood 为 data 的 joint density, 但是被视作一个关于 θ 的函数

- **log-likelihood function** 被定义为:

$$l_n(\theta) = \log \mathcal{L}_n(\theta) = \sum_{i=1}^n \log f(X_i; \theta)$$

- **maximum likelihood estimator (MLE)** 被定义为

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \mathcal{L}_n(\theta|x) = \arg \max_{\theta \in \Theta} l_n(\theta|x)$$

Remark

- 若 log-likelihood 关于 θ differentiable, 则 MLE 的 candidates (in the interior of Θ) 满足:

$$\frac{\partial}{\partial \theta_j} l_n(\theta) = 0, j = 1, \dots, k$$

需要特别注意:

- 是否为 **global maximum** (检查 second derivative)
- maximum** 是否位于 Θ 的 **boundary** (first derivative 可能不为 0)
- 可能会出现无法求出解析解的情况, 此时需要使用 numerical maximization methods
- 该定义下的 MLE 不一定存在, 例如: $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} Unif(0, \theta)$, 则

$$\mathcal{L}_n(\theta|x) = \frac{1}{\theta^n} \cdot \mathbf{1}\{\max_{1 \leq i \leq n} \{x_i\} < \theta\}$$

注意到 $\theta \in \Theta = (x_{(n)}, \infty)$, 因此实际上 θ 取不到 $x_{(n)}$; 为了避免这种情况, 可以将 MLE 定义在 Θ 的 closure 上:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \bar{\Theta}} \mathcal{L}_n(\theta|x) = \arg \max_{\theta \in \bar{\Theta}} l_n(\theta|x)$$

Example

问题:

令 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$, 求 θ 的 MLE

求解:

log-likelihood 为:

$$l(\theta) = \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}} \right) + \left(-\frac{\sum_{i=1}^n (X_i - \theta)^2}{2} \right), \theta \in \mathbb{R}$$

求导, 可以得到:

$$\frac{\partial l(\theta)}{\partial \theta} = -\frac{\sum_{i=1}^n -2(X_i - \theta)}{2} = \sum_{i=1}^n X_i - n\theta$$

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = -n < 0$$

因此 $\hat{\theta}_{n,MLE} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ 为 θ 的 MLE

Example

问题:

考虑同样的例子, 但是限制 $\Theta = [0, \infty)$

求解:

类似的, 可以得到:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}} \right) + \left(-\frac{\sum_{i=1}^n (X_i - \theta)^2}{2} \right) \\ &:= C - \frac{\sum_{i=1}^n X_i^2 - 2(\sum_{i=1}^n X_i)\theta + n\theta^2}{2} \\ &= C - \frac{n(\theta^2 - 2\bar{X}\theta) + \sum_{i=1}^n X_i^2}{2} \\ &= C - \frac{n(\theta^2 - \bar{X}_n)^2 + \sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2}{2} \\ &:= C' - \frac{n(\theta - \bar{X}_n)^2}{2} \end{aligned}$$

因此可以得到:

$$\hat{\theta}_n = \begin{cases} \bar{X} & \text{if } \bar{X} \geq 0 \\ 0 & \text{if } \bar{X} < 0 \end{cases} = \max\{\bar{X}_n, 0\}$$

Example ▾

问题:

令 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \text{Unif}(0, \theta)$, 求 θ 的 MLE

求解:

Likelihood 为:

$$\mathcal{L}(\theta) = \prod_{i=1}^n \left[\frac{1}{\theta} \cdot \mathbf{1}(X_i \geq 0) \mathbf{1}(X_i \leq \theta) \right] = \left(\frac{1}{\theta} \right)^n \cdot \mathbf{1}(X_{(1)} \geq 0) \cdot \mathbf{1}(X_{(n)} \leq \theta)$$

若 $\hat{\theta}_n$ maximizes $\mathcal{L}_n(\theta)$, 则必须满足 $\mathbf{1}(X_{(n)} \leq \hat{\theta}_n) = 1$, 因此 $\hat{\theta}_n \geq X_{(n)}$;

与此同时, 为了 maximize $(\frac{1}{\theta})^n$, 我们希望 $\hat{\theta}_n$ 尽可能的小, 因此 $\hat{\theta}_{n,MLE} = X_{(n)}$

⚠ Remark: θ 的 MoM ▾

X 的 first moment 为: $\alpha_1(\theta) = \mathbb{E}[X_1] = \frac{\theta}{2}$,

X 的 first moment 的 plug-in estimator 为: $\hat{\alpha}_1 = \sum_{i=1}^n X_i/n$;

为了得到 $\hat{\theta}_{MoM}$, 需要求解: $\alpha_1(\hat{\theta}_{MoM}) = \hat{\alpha}_1$, 即

$$\frac{\hat{\theta}_{MoM}}{2} = \sum_{i=1}^n X_i/n$$

因此,

$$\hat{\theta}_{MoM} = 2\bar{X}_n$$

Example ▸ ▾

问题:

令 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \text{Unif}(\theta, \theta+1)$, 求 θ 的 MLE

求解:

Likelihood 为:

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{i=1}^n [1 \cdot \mathbf{1}(X_i \geq \theta) \cdot \mathbf{1}(X_i \leq \theta+1)] \\ &= \mathbf{1}(X_i \geq \theta) \cdot \mathbf{1}(X_i \leq \theta+1) \\ &= \begin{cases} 1 & \text{if } X_{(n)} - 1 \leq \hat{\theta}_n \leq X_{(1)} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

因此 $\hat{\theta}_n$ 可以为 any value in $[X_{(n)} - 1, X_{(1)}]$

⚠ Remark ▾

由此可以得出: MLE 不一定唯一

⚠ Remark: MLE 是关于 sufficient statistics 的函数 ▾

根据 Neyman factorization theorem:

$$T \text{ is sufficient for } \theta \iff \exists \text{ functions } h \text{ and } g, \text{ such that } p(x; \theta) = h(x) \cdot g(T(x), \theta)$$

MLE 可以写作:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \mathcal{L}_n(\theta|x) = \arg \max_{\theta \in \Theta} h(x) \cdot g(T(x), \theta) = \arg \max_{\theta \in \Theta} g(T(x), \theta)$$

仅与 sufficient statistics $T(x)$ 有关