

ch10: 渐进评价

迄今，我们考虑的一直是在有限样本条件下的情形。与之对照，我们可以考虑渐近性质，这些性质描述的是当样本量变成无穷时的一个方法。在这一章我们将着眼于某些渐近性质，并分别考虑点估计，假设检验和区间估计。我们将特别强调极大似然方法的渐近性。

渐近评价的力量在于，当样本量变成无穷时，计算简化了。在有限样本情形不可能做的评价变成了常规。这种简化还允许我们去检查其他的技术（例如自助法和M-估计），这些技术的典型特点是对其只能做渐近评价。

令样本量无限制增加（有时称其为“asymptopia”）不应仅仅被嘲笑为一个想像的练习。反之，渐近性揭露出一个方法最基本的性质，并且给予我们一个非常强有力而应用广泛的评价工具。

10.1 点估计的渐近评价

10.1.1 相合性

相合性似乎是一个相当基本的性质，它要求当样本量变成无穷时估计量收敛到正确的值。相合性非常重要，一个非相合的估计量的价值是值得怀疑的（或者至少是值得严格考查的）。

虽然人们常讲相合估计量，但相合性（也包括所有的渐近性质）关注的是一个估计量序列，而非一个单独的估计量。如果我们按照一个分布 $f(x|\theta)$ 去观测 X_1, X_2, \dots ，我们只要通过对于每个样本量 n 执行相同的估计过程，就可以构造出一个估计量序列 $W_n = W_n(X_1, \dots, X_n)$ 。例如， $\bar{X}_1 = X_1$ ， $\bar{X}_2 = (X_1 + X_2)/2$ ， $\bar{X}_3 = (X_1 + X_2 + X_3)/3$ ， \dots 。我们现在可以给相合序列下定义。

定义10.1.1

定义 10.1.1 一个估计量序列 $W_n = W_n(X_1, \dots, X_n)$ 是参数 θ 的一个相合估计量序列（consistent sequence of estimators），如果对于每个 $\epsilon > 0$ 和每个 $\theta \in \Theta$ ，

$$(10.1.1) \quad \lim_{n \rightarrow \infty} P_\theta(|W_n - \theta| < \epsilon) = 1.$$

通俗地讲，方程 (10.1.1) 就是说当样本量变成无穷（而且样本信息变得越来越好）时，估计量将以高概率任意接近于参数 θ ，这是人们特别渴望的一个性质。或者把事情转换一下，我们可以说一个相合估计量序列未能达到真实参数的概率很小。方程 (10.1.1) 的一个等价说法是，一个相合序列 W_n 将满足：对于每个 $\epsilon > 0$ 和每个 $\theta \in \Theta$ ，

$$(10.1.2) \quad \lim_{n \rightarrow \infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0.$$

定义 10.1.1 可以和定义 5.5.1 即依概率收敛的定义，来比较。定义 10.1.1 讲的就是一个相合估计量序列依概率收敛于被估计的参数 θ 。定义 5.5.1 是以一个概率结构来处理一个随机变量序列，而定义 10.1.1 是以用 θ 为指标的一整族概率结构做处理。对于每个不同的 θ 值，与序列 W_n 关联的概率结构是不同的。而且该定义说对于各个 θ 值，相应概率结构都使得序列依概率收敛到真实 θ 。这就是一个概率的定义与一个统计学定义通常的区别。概率的定义处理一个概率结构，而统计学定义处理一整族。

例10.1.2

例 10.1.2 (\bar{X} 的相合性) 设 X_1, X_2, \dots 是 iid $n(\theta, 1)$ 的, 来考虑序列

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

回忆有 $\bar{X}_n \sim n(\theta, 1/n)$, 所以

$$P_\theta(|\bar{X}_n - \theta| < \epsilon) = \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}(\bar{x}_n - \theta)^2} d\bar{x}_n \quad (\text{根据定义})$$

$$= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}y^2} dy \quad (\text{变量替换 } y = \bar{x}_n - \theta)$$

$$= \int_{-\epsilon/\sqrt{n}}^{\epsilon/\sqrt{n}} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-(1/2)t^2} dt \quad (\text{变量替换 } t = y\sqrt{n})$$

$$= P(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \quad (Z \sim n(0, 1))$$

$$\rightarrow 1 \quad \text{当 } n \rightarrow \infty$$

因此 \bar{X}_n 是 θ 的一个相合估计量序列. ||

补充定义

Definition

Let $\{T_n = T_n(X_1, \dots, X_n)\}$ be a sequence of estimators of a parametric function $\tau(\theta)$.

i) $\{T_n\}$ is said to be *consistent* estimator of $\tau(\theta)$ if for all $\epsilon > 0$,

$$P_\theta\{|T_n - \tau(\theta)| > \epsilon\} \rightarrow 0$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$. $[T_n \xrightarrow{P} \tau(\theta)]$.

ii) $\{T_n\}$ is said to be *MSE consistent (MSEC)* to $\tau(\theta)$ if

$$E_\theta\{[T_n - \tau(\theta)]^2\} \rightarrow 0$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$.

◁ Note: MSEC implies consistency

Definition (- Continued)

iii) $\{T_n\}$ is said to be strong consistent to $\tau(\theta)$ if for all $\epsilon > 0$,

$$T_n \xrightarrow{a.s.} \tau(\theta)$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$

iv) $\{T_n\}$ is said to be *Asymptotically Unbiased (AU)* to $\tau(\theta)$ if

$$E_{\theta}\{T_n\} \rightarrow \tau(\theta)$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$.

如何验证相合性-定理10.1.3

一般情况下, 为验证相合性不必做像上面那样详细的计算. 对于一个估计量 W_n , 回想一下 Chebychev 不等式的陈述

$$P_{\theta}(|W_n - \theta| \geq \epsilon) \leq \frac{E_{\theta}[(W_n - \theta)^2]}{\epsilon^2},$$

这样, 如果对于每个 $\theta \in \Theta$ 有

$$\lim_{n \rightarrow \infty} E_{\theta}[(W_n - \theta)^2] = 0,$$

则这个估计量序列就是相合的. 此外, 根据式 (7.3.1) 就有

$$(10.1.3) \quad E_{\theta}[(W_n - \theta)^2] = \text{Var}_{\theta} W_n + [\text{Bias}_{\theta} W_n]^2.$$

把其总括在一起, 我们可以给出以下的定理.

定理 10.1.3 如果 W_n 是参数 θ 的一个估计量序列, 它对于每个 $\theta \in \Theta$ 都满足

$$(i) \lim_{n \rightarrow \infty} \text{Var}_{\theta} W_n = 0,$$

$$(ii) \lim_{n \rightarrow \infty} \text{Bias}_{\theta} W_n = 0,$$

则 W_n 是参数 θ 的一个相合估计量序列.

非相合估计量序列的价值为何应该被质疑

在本节开始, 我们谈到一个非相合估计量序列时认为它的价值应该受到质疑. 这样评论的部分根据就是如以下定理所揭示的, 存在如此之多的相合序列. 这个定理的证明留给习题 10.2.

定理 10.1.5 设 W_n 是参数 θ 的一个相合估计量序列. 设 a_1, a_2, \dots 和 b_1, b_2, \dots 是常数序列, 满足

$$(i) \lim_{n \rightarrow \infty} a_n = 1,$$

$$(ii) \lim_{n \rightarrow \infty} b_n = 0.$$

则序列 $U_n = a_n W_n + b_n$ 是参数 θ 的一个相合估计量序列.

MLE的相合性-定理10.1.6

作为本节的结束，我们概要给出关于极大似然估计量相合性的一个更一般的结果。这个结果表明极大似然估计量是其参数的相合估计量，而且是我们见到的求估计量的方法能保证一种最优性质的第一例。

为了使 MLE 具有相合性，其基础密度（似然函数）必须满足一定的“正则性条件”，这里我们将不深入探究，不过可以参见 10.6 节 10.6.2 了解详情。

定理 10.1.6 (MLE 的相合性) 设 X_1, X_2, \dots 是 iid $f(x|\theta)$ 的， $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ 是似然函数，而 $\hat{\theta}$ 表示 θ 的 MLE. 设 $\tau(\theta)$ 是 θ 的一个连续函数。那么在杂录 10.6.2 的关于 $f(x|\theta)$ ，从而也就是对 $L(\theta|\mathbf{x})$ 的正则性条件之下，对于每个 $\epsilon > 0$ 和每个 $\theta \in \Theta$ ，有

$$\lim_{n \rightarrow \infty} P_{\theta}(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0.$$

这就是说，此 $\tau(\hat{\theta})$ 是 $\tau(\theta)$ 的一个相合估计量。

10.1.2 有效性

相合性考虑的是一个估计量的渐近精确性：它是收敛到要估计的参数吗？本节我们来看一个有关的性质，即有效性，这个性质关心的是一个估计量的渐近方差。

在计算渐近方差时，或许我们想如下进行。给出一个基于样本量 n 的估计量 T_n ，我们计算有限样本方差 $\text{Var} T_n$ ，然后计算 $\lim_{n \rightarrow \infty} k_n \text{Var} T_n$ ，其中 k_n 是某规格化常数。（注：在很多情况中，当 $n \rightarrow \infty$ 时， $\text{Var} T_n \rightarrow 0$ ，所以我们需要一个因子 k_n 以迫使它趋向一个极限。）

定义 10.1.7

定义 10.1.7 对于一个估计量 T_n ，如果 $\lim_{n \rightarrow \infty} k_n \text{Var} T_n = \tau^2 < \infty$ ，其中 $\{k_n\}$ 是一个常数序列，则 τ^2 叫做极限方差（limiting variance）或方差的极限。

例 10.1.8

例 10.1.8 (极限方差) 关于 n 个 iid 的具有 $E(X) = \mu$ 和 $\text{Var}(X) = \sigma^2$ 的正态观测的平均值 \bar{X}_n ，如果我们取 $T_n = \bar{X}_n$ ，则 $\lim_{n \rightarrow \infty} n \text{Var} \bar{X}_n = \sigma^2$ 是 T_n 的极限方差。

但是，若我们要用 $1/\bar{X}_n$ 估计 $1/\mu$ ，麻烦事就发生了。如果我们现在取 $T_n = 1/\bar{X}_n$ ，就发现方差 $\text{Var} T_n = \infty$ ，那么方差的极限是无穷。然而回忆例 5.5.23，那里我们讲过 $1/\bar{X}_n$ 的近似均值和方差是

$$E\left(\frac{1}{\bar{X}_n}\right) \approx \frac{1}{\mu},$$

$$\text{Var}\left(\frac{1}{\bar{X}_n}\right) \approx \left(\frac{1}{\mu}\right)^4 \text{Var} \bar{X}_n,$$

于是根据这第二次的计算，方差是 $\text{Var} T_n \approx \frac{\sigma^2}{n\mu^4} < \infty$ 。

定义 10.1.9

||

定义 10.1.9 对于一个估计量 T_n , 假定有依分布收敛 $k_n(T_n - \tau(\theta)) \rightarrow n(0, \sigma^2)$, 则参数 σ^2 叫做 T_n 的渐近方差或 T_n 的极限分布的方差.

对于计算样本均值或者其他类型平均的方差, 典型情况是极限方差和渐近方差有相同值. 但是在更复杂的情况, 有时极限方差将会令我们失望. 注意总是有这样的情况, 渐近方差小于极限方差, 这也是有趣的 (见 Lehmann and Casella 1998, 6.1 节). 这里举个例子来说明.

例10.1.10

例 10.1.10 (大样本混合方差) 分层模型

$$Y_n | W_n = w_n \sim n(0, w_n + (1-w_n)\sigma_n^2), \\ W_n \sim \text{Bernoulli}(p_n),$$

可以展示出渐近方差和极限方差间的重大差别. (这个模型有时也描述为一个混合模型, 在其中我们以 p_n 的概率观测 $Y_n \sim n(0, 1)$ 而以 $1-p_n$ 的概率观测 $Y_n \sim n(0, \sigma_n^2)$.)

首先利用定理 4.4.7, 我们就有

$$\text{Var}(Y_n) = p_n + (1-p_n)\sigma_n^2.$$

由此就得到, 只有在 $\lim_{n \rightarrow \infty} (1-p_n)\sigma_n^2 < \infty$ 时, Y_n 的极限方差有限.

另一方面, Y_n 的渐近方差可以利用

$$P(Y_n < a) = p_n P(Z < a) + (1-p_n) P(Z < a/\sigma_n).$$

直接计算出来.

假定现在我们让 $p_n \rightarrow 1$ 及 $\sigma_n \rightarrow \infty$, 并且使得 $(1-p_n)\sigma_n^2 \rightarrow \infty$. 这样就得到 $P(Y_n < a) \rightarrow P(Z < a)$, 就是说 $Y_n \rightarrow n(0, 1)$, 并且我们有

$$\text{极限方差} = \lim_{n \rightarrow \infty} p_n + (1-p_n)\sigma_n^2 = \infty$$

$$\text{渐近方差} = 1$$

更多的细节参见习题 10.6.

||

定义10.1.11-有效性

定义 10.1.11 一个估计量序列 W_n 关于一个参数 $\tau(\theta)$ 是渐近有效的, 如果 $\sqrt{n}[W_n - \tau(\theta)] \xrightarrow{L} n(0, \nu(\theta))$ 而且

$$\nu(\theta) = \frac{[\tau'(\theta)]^2}{E_\theta \left(\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right)},$$

就是说, W_n 的渐近方差达到了 Cramér-Rao 下界.

定理10.1.12-MLE的渐近有效性

定理 10.1.12 (MLE 的渐近有效性) 设 X_1, X_2, \dots 是 iid $f(x|\theta)$, θ 的 MLE 记作 $\hat{\theta}$, 设 $\tau(\theta)$ 是 θ 的一个连续函数. 那么在 10.6 节 10.6.2 的关于 $f(x|\theta)$, 从而也就是对 $L(\theta|x)$ 的正则性条件之下,

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \xrightarrow{L} n(0, \nu(\theta)),$$

其中 $\nu(\theta)$ 是 Cramér-Rao 下界. 就是说, $\tau(\hat{\theta})$ 是 $\tau(\theta)$ 的一个相合且渐近有效的估计量.

10.1.3 计算与比较

前面几节展示的渐近公式可以提供给我们用在大样本的近似方差. 当然, 我们必须要考虑正则性条件 (10.6 节 10.6.2), 但是这些条件是相当一般的而且几乎在通常情况下总能得到满足. 不过, 有一个条件应当特别提及, 就像我们在例 7.3.13

已经见过的那样, 违背了它就会导致混乱. 为了使下面的近似成为正当, 概率密度函数或概率质量函数的支撑集, 也就是似然函数的支撑集必须与参数无关.

如果一个 MLE 是渐近有效的, 定理 10.1.6 中的渐近方差就是定理 5.5.24 中 (去掉 $1/n$ 项) 的 Δ 方法方差. 这样, 我们就可以把 Cramér-Rao 下界用作 MLE 的真实方差的一个近似. 设 X_1, X_2, \dots 是 iid $f(x|\theta)$ 的, $\hat{\theta}$ 是 θ 的 MLE, 而 $I_n(\theta) = E_\theta \left(\frac{\partial}{\partial \theta} \log L(\theta | \mathbf{X}) \right)^2$ 是样本的信息数. 根据 Δ 方法与 MLE 的渐近有效性, $h(\hat{\theta})$ 的方差可以由以下来近似

$$\begin{aligned} (10.1.7) \quad \text{Var}(h(\hat{\theta})|\theta) &\approx \frac{[h'(\theta)]^2}{I_n(\theta)} \\ &= \frac{[h'(\theta)]^2}{E_\theta \left(-\frac{\partial^2}{\partial \theta^2} \log L(\theta | \mathbf{X}) \right)} && \text{(利用引理 7.3.11 的恒等式)} \\ &\approx \frac{[h'(\theta)]^2|_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial \theta^2} \log L(\theta | \mathbf{X})|_{\theta=\hat{\theta}}} && \text{(分母是 } \hat{I}_n(\hat{\theta}), \text{ 即观测信息数)} \end{aligned}$$

此外, 已被证明 (Efron and Hinkley 1978), 使用观测信息数胜于使用出现在 Cramér-Rao 下界中的期望信息数.

注意方差估计的步骤是一个两步的过程, 这个事实或多或少被式 (10.1.7) 掩盖了. 为估计 $\text{Var}_\theta h(\hat{\theta})$, 首先我们近似 $\text{Var}_\theta h(\hat{\theta})$, 然后再估计这个近似结果, 而这通常是用 $\hat{\theta}$ 替换 θ . 作为结果的估计, 可以记作 $\text{Var}_{\hat{\theta}} h(\hat{\theta})$ 或 $\widehat{\text{Var}}_\theta h(\hat{\theta})$.

从定理 10.1.6 得出 $-\frac{1}{n} \frac{\partial^2}{\partial \theta^2} \log L(\theta | \mathbf{X}) |_{\theta=\hat{\theta}}$ 是 $I(\theta)$ 的一个相合估计量, 所以就得到 $\text{Var}_{\hat{\theta}} h(\hat{\theta})$ 是 $\text{Var}_\theta h(\hat{\theta})$ 的一个相合估计量.

渐近相对效率-定义 10.1.16

渐近有效的性质给予我们一个在求渐近方差时希望达到的基准点 (见 10.6 节 10.6.1). 通过渐近相对效率 (asymptotic relative efficiency) 的概念, 我们还能将渐近方差当作比较估计量的一个工具.

定义 10.1.16 如果两个估计量 W_n 和 V_n 满足

$$\begin{aligned} \sqrt{n}[W_n - \tau(\theta)] &\xrightarrow{L} N[0, \sigma_W^2], \\ \sqrt{n}[V_n - \tau(\theta)] &\xrightarrow{L} N[0, \sigma_V^2], \end{aligned}$$

V_n 关于 W_n 的渐近相对效率 (ARE) 是

$$\text{ARE}(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}$$

例 10.1.17

例 10.1.17 (Poisson 估计量的 ARE) 设 X_1, X_2, \dots 是 iid Poisson (λ) 的, 而我们对估计 0 概率感兴趣. 例如, 在一个给定时间段内进入一家银行的顾客数有时用一个泊松随机变量来建模, 而 0 概率就是在这段时间段将没有一个顾客进入此银行的概率. 如果 $X \sim \text{Poisson}(\lambda)$, 则 $P(X=0) = e^{-\lambda}$, 而一个自然的 (然而有些朴素的) 估计量 $\hat{\tau}$ 是

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

其中 $Y_i = I(X_i = 0)$. 这些 Y_i 服从分布 Bernoulli ($e^{-\lambda}$), 于是由此就可得到

$$E(\hat{\tau}) = e^{-\lambda} \text{ 和 } \text{Var}(\hat{\tau}) = \frac{e^{-\lambda}(1-e^{-\lambda})}{n}.$$

另外一种方法, $e^{-\lambda}$ 的 MLE 是 $e^{-\hat{\lambda}}$, 其中 $\hat{\lambda} = \sum_i X_i/n$ 是 λ 的 MLE. 使用 Δ 方

法近似, 我们就有

$$E(e^{-\hat{\lambda}}) \approx e^{-\lambda} \text{ 和 } \text{Var}(e^{-\hat{\lambda}}) \approx \frac{\lambda e^{-2\lambda}}{n}.$$

因为

$$\begin{aligned} \sqrt{n}[\hat{\tau} - e^{-\lambda}] &\xrightarrow{L} N[0, e^{-\lambda}(1-e^{-\lambda})] \\ \sqrt{n}[e^{-\hat{\lambda}} - e^{-\lambda}] &\xrightarrow{L} N[0, \lambda e^{-2\lambda}], \end{aligned}$$

所以 $\hat{\tau}$ 关于 MLE $e^{-\hat{\lambda}}$ 的渐近相对效率 (ARE) 是

$$\text{ARE}(\hat{\tau}, e^{-\hat{\lambda}}) = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1-e^{-\lambda})} = \frac{\lambda}{e^{\lambda}-1}.$$

对这个函数的考察表明它是严格减的, 在 $\lambda=0$ 达到最大值 1 (这是 $\hat{\tau}$ 所能达到的最好状态), 而且随 $\lambda \rightarrow \infty$ 快速地逐渐变小 (在 $\lambda=4$ 小于 10%) 趋向渐近线 0. (参见习题 10.9) ||

注意

因为在典型的情况下 MLE 是渐近有效的, 所以不可能期望另外的估计量有比它更小的渐近方差. 然而其他估计量也许具有其他的令人满意的性质 (易于计算, 对基础假定稳健) 使其合乎愿望. 在这种情形下, 如果我们采用别的估计量, 那么 MLE 的有效性对于校准我们将会放弃什么就变得很重要.

我们来看最后一个例, 其中要得到最佳方差计算可没那么容易. 下一节将讨论稳健性.

例 10.1.18

例 10.1.18 (估计一个伽玛分布均值) 眼见为实, 伽玛分布均值的估计不是一件容易的任务. 回忆伽玛分布的概率密度函数 $f(x|\alpha, \beta)$ 由下式给出

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}.$$

这个分布的均值是 $\alpha\beta$, 为计算极大似然估计量我们必须处理 Γ 函数的导数 (称为双 Γ 函数 digamma function), 这种计算很困难. 与之对比, 矩法给予我们一个易于计算的估计.

具体来说, 设我们有来自上面伽玛密度的一随机样本 X_1, \dots, X_n , 此密度中参数已经改设, 把均值作为显参数, 记为 $\mu = \alpha\beta$, 从而密度成为

$$f(x|\mu, \beta) = \frac{1}{\Gamma(\mu/\beta)\beta^{\mu/\beta}} x^{\mu/\beta-1} e^{-x/\beta},$$

而且 μ 的矩估计量是 \bar{X} , 具有方差 $\beta\mu/n$.

为计算 MLE, 我们利用对数似然

$$l(\mu, \beta|\mathbf{x}) = \sum_{i=1}^n \log f(x_i|\mu, \beta).$$

为了计算简便, 假定 β 已知, 于是我们求解 $\frac{d}{d\mu} l(\mu, \beta|\mathbf{x}) = 0$ 以得到 MLE $\hat{\mu}$. 不存在

显式解, 所以我们用数值方法.

根据定理 10.1.6 我们知道 $\hat{\mu}$ 是渐近有效的. 让我们有兴趣的问题是使用易于计算的矩估计量我们会有多少损失. 为了做比较, 我们计算渐近相对效率

$$\text{ARE}(\hat{\mu}, \bar{X}) = [\beta\mu] E\left[-\frac{d^2}{d\mu^2} l(\mu, \beta|\mathbf{X})\right],$$

并且选择了几个 β 值把它们的图显示在图 10.1.1 中. 当然, 我们知道此 ARE 必定大于 1, 但是从图我们看到对于较大的 β 值, 进行较复杂计算和使用 MLE 是值得的. (参见习题 10.11 做的扩充, 计算细节参见例 A.0.7.)

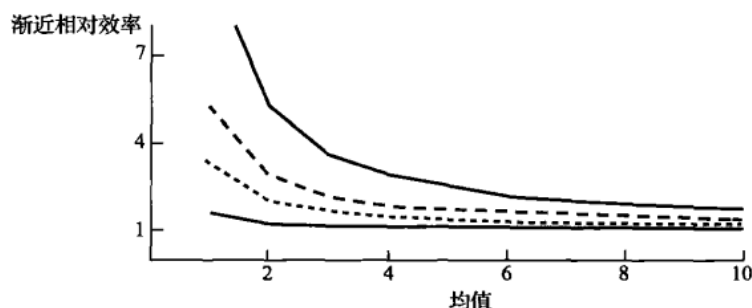


图 10.1.1 Γ 均值矩法估计量对 MLE 估计量的渐近相对效率. 四条曲线相应于尺度参数的四个值 (1, 3, 5, 10), 其中较高的曲线相应于较大的尺度参数的值

补充定义

Definition

Relative efficiency of an unbiased estimator T of $\tau(\theta)$ to another unbiased estimator T^* is given by

$$RE(T, T^*) = \frac{Var(T^*)}{Var(T)}.$$

Definition The efficiency of an unbiased estimator T of $\tau(\theta)$

is given by

$$\text{eff}(T) = \frac{[\tau'(\theta)]^2 / n I_1(\theta)}{Var_{\theta}(T)}$$

The unbiased estimator T is efficient to $\tau(\theta)$ if $\text{eff}(T)=1$.

Definition: For an estimator W_n , if $k_n(W_n - \tau(\theta)) \rightarrow N(0, \sigma^2)$ in distribution, then σ^2 is called the asymptotic variance of the limiting distribution of W_n .

Definition: W_n is *asymptotically efficient* for $\tau(\theta)$ if

$$\sqrt{n}(W_n - \tau(\theta)) \rightarrow N(0, v(\theta))$$

with

$$v(\theta) = \frac{[\tau'(\theta)]^2}{E_{\theta} \left[\left\{ \frac{\partial}{\partial \theta} \log f(x|\theta) \right\}^2 \right]} = \frac{[\tau'(\theta)]^2}{I_1(\theta)}.$$

that is the asymptotic variance of W_n achieves the Cramer-Rao Lower Bound.

注意

- 两个不一样的估计量，取方差较小的那个估计量
- 而有些统计量的方差并不好求，则使用渐近相对效率

10.1.4 自助法标准误差

我们在例 1.2.20 中曾首次看到自助法 (bootstrap). 自助法提供了计算标准误差的一个替代方法. (它还能够提供更多, 参见 10.6 节 10.6.3)

自助法建立在一个简单然而强有力的思想上 (它可能涉及相当多的数学)^①. 在统计学上, 我们通过提取样本获悉总体特征. 因为样本代表总体, 相似的样本特征就应该给予我们关于总体特征的信息. 自助法通过提取再抽样样本 (resample) 帮助我们了解样本特征 (即我们从原始样本中再提取样本), 并且利用这些信息去推断总体. 自助法是由 Efron 在 1970 年代后期发展起来的, 其最初的想法出现在 Efron (1970a, b), 并且有 Efron (1982) 的专著. 另外参看 Efron (1998) 有关近期更多的想法和发展.

例 10.1.19

例 10.1.19 (一个方差的自助法) 在例 1.2.20 中我们计算过选自

2, 4, 9, 12

的所有可能的四个数的平均, 那里我们是有放回的取数. 这是最简单的自助法形式, 有时叫做非参数自助法. 图 1.2.2 在一个直方图中显示了这些值.

我们所建立的是样本均值可能值的再抽样. 我们看到有 $\binom{4+4-1}{4} = 35$ 个不同的可能值, 但是这些值不是等概率的 (这样, 就不能视为随机样本). 而 $4^4 = 256$ 个 (非不同的) 再抽样样本都是等可能的, 所以它们可以被视为随机样本. 对于第 i 个再抽样样本, 我们设 \bar{x}_i^* 是该再抽样样本的平均数. 则我们可以用下式估计样本均值 \bar{X} 的方差

$$(10.1.9) \quad \text{Var}^*(\bar{X}) = \frac{1}{n^n - 1} \sum_{i=1}^{n^n} (\bar{x}_i^* - \bar{\bar{x}}^*)^2,$$

其中 $\bar{\bar{x}}^* = \frac{1}{n^n - 1} \sum_{i=1}^{n^n} \bar{x}_i^*$ 为再抽样样本的平均数. (这里一律用*表示自助的, 或者说再抽样的值)

在我们这个例中, 自助法均值和方差为 $\bar{\bar{x}}^* = 6.75$ 和 $\text{Var}^*(\bar{X}) = 3.94$. 结果表明, 就所关注的均值和方差而言, 自助法估计与通常估计几乎是相同的 (见习题 10.13). ||

现在我们已经看到怎样计算一个自助法标准误差, 但在上述问题中并不真正需要这样做. 然而自助法的真正优越性, 就像 Δ 方法, 公式 (10.1.9) 几乎可以应用于任何的估计量. 这样, 对任何的估计量 $\hat{\theta}(x) = \hat{\theta}$, 我们可以写成

$$(10.1.10) \quad \text{Var}^*(\hat{\theta}) = \frac{1}{n^n - 1} \sum_{i=1}^{n^n} (\hat{\theta}_i^* - \bar{\hat{\theta}}^*)^2,$$

其中 $\hat{\theta}_i^*$ 是从第 i 个再抽样样本计算出的估计量, 而 $\bar{\hat{\theta}}^* = \frac{1}{n^n - 1} \sum_{i=1}^{n^n} \hat{\theta}_i^*$ 是再抽样估计值的平均数.

10.2 稳健性

到现在为止，我们在基础模型正确的假定下已经评价了估计量的性能。在这个假定下，我们已经推导出在某种意义下的最佳估计量。然而，如果基础模型不正确，则就不能保证我们的估计量最佳。

我们无法预防所有可能的情形，此外，如果我们的模型是经过仔细考虑做出的，则也不必这样预防。但是我们关心对于假定的模型有小的或中等偏离的情形。这样，这就引导我们去考虑稳健估计量 (robust estimator)。这种估计量放弃在假定模型上的最佳性，而换以当假定模型不是真实模型时的合理表现。这样，我们就有一个平衡得失问题，最佳性或稳健性哪个标准更重要，这或许最好根据具体情况来决定。

术语“稳健性”可以有多种解释，但是也许 Huber (1981, 1.2 节) 概括得最好，他指出^①：

任何统计方法应该具有如下特性：

- (1) 在假定模型下它应当具有一个合理的好（最佳或接近最佳）效率。
- (2) 应该在这样的意义下是稳健的：对于假定模型的微小偏离应该仅引起性能的轻微损伤
- (3) 对模型大一些的偏离也不应导致灾难性后果。

10.2.1 均值和中位数

例10.2.1

例 10.2.1 (样本均值的稳健性) 设 X_1, X_2, \dots, X_n 是 iid $n(\mu, \sigma^2)$ 的。我们知道 \bar{X} 具有方差 $\text{Var}(\bar{X}) = \sigma^2/n$ ，而该方差是 Cramér-Rao 下界，这样在假定模型下它达到最好的方差因此 \bar{X} 满足 (1)。

为研究 (2)，即研究对模型微小的偏离之下 \bar{X} 的性能，我们首先需要决定这意味着什么。一种通常的解释就是使用一个 δ -污染模型 (δ -contamination model)，就是说，对于一个小 δ ，假定我们的观测

$$X_i \sim \begin{cases} n(\mu, \sigma^2) & \text{以 } 1-\delta \text{ 的概率} \\ f(x) & \text{以 } \delta \text{ 的概率} \end{cases},$$

其中 $f(x)$ 是某个其他分布。

设我们把 $f(x)$ 取为任何一个均值是 θ 而方差是 τ^2 的密度。则

$$\text{Var}(\bar{X}) = (1-\delta)\frac{\sigma^2}{n} + \delta\frac{\tau^2}{n} + \frac{\delta(1-\delta)(\theta-\mu)^2}{n}.$$

这看上去真是不错，因为如果 $\theta \approx \mu$ 且 $\sigma \approx \tau$ ， \bar{X} 将接近最佳。然而，我们可以对模型做更大一点的扰动，并把事情搞得相当糟。考虑如果 $f(x)$ 是一个 Cauchy 概率密度函数将会发生什么。这时立即就得出 $\text{Var}(\bar{X}) = \infty$ 。（细节参见习题 10.18 而习题 10.19 讨论的是另一个情况。）

定义10.2.2

||

定义 10.2.2 设 $X_{(1)} < \dots < X_{(n)}$ 是容量为 n 的顺序样本, 而设 T_n 是一个基于这个样本的统计量. T_n 具有崩溃值 (breakdown value) b , $0 \leq b \leq 1$, 如果对于每一个 $\epsilon > 0$, 都有

$$\lim_{X_{(\lfloor (1-b)n \rfloor)} \rightarrow \infty} T_n < \infty \text{ 和 } \lim_{X_{(\lfloor (1-b+\epsilon)n \rfloor)} \rightarrow \infty} T_n = \infty.$$

(关于百分位数记号回忆定义 5.4.2.)

容易看出 \bar{X} 的崩溃值是 0; 就是说, 如果这个样本中任何比例的样本值趋向无穷, 则 \bar{X} 的值也趋向无穷. 与此鲜明对照的是, 样本中位数在样本值的这种变化下是不变的. 这种对于极端观测值的不敏感性有的时候被认为是样本中位数的一个优点, 它的崩溃值为 50%. (关于崩溃值的更多内容参见习题 10.20.)

由于中位数在稳健性方面对于均值有改善, 我们就可以问, 转而使用一个更加稳健的估计量 (当然我们必须这样做!) 是否会失去什么. 例如在例 10.2.1 的简单正态模型中, 如果模型正确, 则样本均值是最优的无偏估计量. 因此就可以推出, 对这个正态模型 (以及它的附近), 样本均值是一个良好的估计量. 但关键问题是对这个正态模型样本均值到底比中位数好多少? 如果我们能够回答这个问题, 在做出使用哪个估计量以及侧重考虑哪个准则 (最佳性或者稳健性) 的选择时, 就有了更丰富的信息. 为了在某种普遍意义下回答这个问题, 我们号召用渐近相对效率准则.

为了计算中位数对均值的 ARE, 首先必须建立中位数的渐近正态性并计算其渐近分布的方差.

例 10.2.3

例 10.2.3 (中位数的渐近正态性) 为了求中位数的极限分布, 我们采用一种类似于在定理 5.4.3 和 5.4.4 证明当中的论证, 即基于二项分布的论证.

设 X_1, X_2, \dots, X_n 是来自具有概率密度函数为 f 和 cdf 为 F (设是可微的) 的总体的样本, 并且 $P(X_i \leq \mu) = 1/2$, 所以 μ 是总体中位数. 设 M_n 是样本中位数, 并考虑对于某 a 计算

$$\lim_{n \rightarrow \infty} P(\sqrt{n}(M_n - \mu) \leq a)$$

如果我们通过

$$Y_i = \begin{cases} 1 & \text{如果 } X_i \leq \mu + a/\sqrt{n} \\ 0 & \text{其他} \end{cases}$$

定义随机变量 Y_i , 由此就可推出 Y_i 是一个成功概率为 $p_n = F(\mu + a/\sqrt{n})$ 的 Bernoulli 随机变量. 为避免复杂, 我们将假定 n 是奇数, 这样事件 $\{M_n \leq \mu + a/\sqrt{n}\}$ 就等价于事件 $\{\sum_i Y_i \geq (n+1)/2\}$.

经过一些代数计算就得到

$$P(\sqrt{n}(M_n - \mu) \leq a) = P\left(\frac{\sum_i Y_i - np_n}{\sqrt{np_n(1-p_n)}} \geq \frac{(n+1)/2 - np_n}{\sqrt{np_n(1-p_n)}}\right).$$

现在 $p_n \rightarrow p = F(\mu) = 1/2$, 于是我们能应用中心极限定理证明 $\frac{\sum_i Y_i - np_n}{\sqrt{np_n(1-p_n)}}$ 依分布收敛到一个标准正态随机变量 Z . 简单的极限计算又证明有

$$\frac{(n+1)/2 - np_n}{\sqrt{np_n(1-p_n)}} \rightarrow -2aF'(\mu) = -2af(\mu).$$

把这些都放在一起就得到

$$P(\sqrt{n}(M_n - \mu) \leq a) \rightarrow P(Z \geq -2af(\mu)),$$

而且因此 $\sqrt{n}(M_n - \mu)$ 的渐近分布是均值 0 而方差为 $1/[2f(\mu)]^2$ 的正态分布. (关于细节, 见习题 10.22, 而严谨推导以及对于这个结果发展的更一般结果, 参见 Shao, 1999, 5.3 节.)

例10.2.4

例 10.2.4 (中位数对均值的渐近相对效率) 由于对于均值和中位数的渐近方差有简单的表达式, 所以 ARE 是易于计算的. 下面的表给出相应三种对称分布的渐近相对效率. 我们发现, 如所料当分布的尾部越重则得到的 ARE 越大. 这就是说, 在重尾分布情况, 用中位数性能会改善. 更多的比较见习题 10.23.

中位数/均值的渐近相对效率			
正态分布	罗吉斯蒂克分布	双指数分布	
0.64	0.82	2	

10.2.2 M-估计量

我们所使用的很多统计量是最小化一个特别的准则的结果. 例如, 如果 X_1, X_2, \dots, X_n 是来自 $f(x|\theta)$ 的, 那么, 可能的估计量有: 样本均值, 它是使 $\sum (x_i - a)^2$ 最小的量; 样本中位数, 它是使 $\sum |x_i - a|$ 最小的量; 再就是 MLE, 它是使 $\prod_{i=1}^n f(x_i|\theta)$ 最大 (或者使负的对数似然最小) 的量. 作为获得一个稳健估计量的系统方法, 我们应当试图写下一个准则函数, 它的最小值导致一个具有令人满意的稳健性质的估计量.

在试图定义一个稳健准则时, Huber (1964) 曾考虑一种均值和中位数间的折中方案. 均值的准则是一个平方, 它使之具有敏感性, 但是在“尾部”平方对大的观测值给出太多的权重. 与之相反, 中位数的绝对值准则不偏重大的或者小的观测值. 折中方案就是最小化准则函数

$$(10.2.1) \quad \sum_{i=1}^n \rho(x_i - a)$$

其中函数 ρ 是由

$$(10.2.2) \quad \rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{若 } |x| \leq k \\ k|x| - \frac{1}{2}k^2 & \text{若 } |x| > k \end{cases}.$$

函数 $\rho(x)$ 的性态对于 $|x| \leq k$ 像 x^2 而对于 $|x| > k$ 像 $|x|$. 此外, 因为 $\frac{1}{2}k^2 = k|k| - \frac{1}{2}k^2$, 所以这个函数连续 (见习题 10.28). 事实上 ρ 是可微的. 常数 k 可被称为一个调节参数, 它控制着混合, 对于较小的 k 值, 则产生一个更“像中位数”的估计量.

表 10.2.1 Huber 估计量

k	0	1	2	3	4	5	6	8	10
估计值	-2.1	0.03	-0.04	0.29	0.41	0.52	0.87	0.97	1.33

最小化 (10.2.2) 的估计量是 Huber 所研究的估计量的一个特例. 对于一般的函数 ρ , 我们把使 $\sum_i \rho(x_i - \theta)$ 达最小的估计量叫做一个 M-估计量 (M-estimator), 这个名字使我们联想起它们是极大似然类型的估计量. 注意到如果把 ρ 选成负的对数似然 $-l(\theta|x)$, 则 M-估计量就是通常的 MLE. 但是更灵活地选择这个欲最小化的函数, 可以推演出具有各种不同性质的估计量.

由于最小化一个函数的典型做法是通过解出其导数的零点 (指我们能够求导数的时候) 而进行, 定义 $\psi = \rho'$, 我们看到, M-估计就是

$$(10.2.3) \quad \sum_{i=1}^n \psi(x_i - \theta) = 0$$

的解. 把估计量刻画为一个方程的根对于获取估计量的性质是特别有用的, 这是由于那些在极大似然估计量中使用过的论证方法能够扩展. 特别地, 看 10.1.2 节, 尤其是定理 10.1.12 的证明. 我们假定函数 $\rho(x)$ 是对称的, 而它的导数 $\psi(x)$ 是单调增的 (这保证 (10.2.3) 的根是唯一的最小点). 于是, 就像定理 10.1.12 的证明, 我们写出 ψ 的 Taylor 展开式为

$$\sum_{i=1}^n \psi(x_i - \theta) = \sum_{i=1}^n \psi(x_i - \theta_0) + (\theta - \theta_0) \sum_{i=1}^n \psi'(x_i - \theta_0) + \dots,$$

其中 θ_0 是真值, 而且我们忽略高阶项. 设 $\hat{\theta}_M$ 是方程 (10.2.3) 的解并且用它替换 θ 就得到

$$0 = \sum_{i=1}^n \psi(x_i - \theta_0) + (\hat{\theta}_M - \theta_0) \sum_{i=1}^n \psi'(x_i - \theta_0) + \dots,$$

其中左侧为 0 是因为 $\hat{\theta}_M$ 是方程 (10.2.3) 的解. 现在, 再次类似于定理 10.1.12 的证明, 我们重排这些项, 然后除以 \sqrt{n} , 并且忽略余项就得到

$$\sqrt{n}(\hat{\theta}_M - \theta_0) = \frac{-\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(x_i - \theta_0)}{\frac{1}{n} \sum_{i=1}^n \psi'(x_i - \theta_0)}.$$

现在我们假定 θ_0 满足 $E_{\theta_0} \psi(X - \theta_0) = 0$ (这通常被当作 θ_0 的定义). 于是就可得到

$$(10.2.4) \quad \frac{-1}{\sqrt{n}} \sum_{i=1}^n \psi(x_i - \theta_0) = \sqrt{n} \left[\frac{-1}{n} \sum_{i=1}^n \psi(x_i - \theta_0) \right] \xrightarrow{L} N(0, E_{\theta_0} \psi(X - \theta_0)^2),$$

而且根据大数定律得到

$$(10.2.5) \quad \frac{1}{n} \sum_{i=1}^n \psi'(x_i - \theta_0) \xrightarrow{P} E_{\theta_0} \psi'(X - \theta_0).$$

把这些放在一起, 我们就得到

$$(10.2.6) \quad \sqrt{n}(\hat{\theta}_M - \theta_0) \rightarrow N\left(0, \frac{E_{\theta_0} \psi(X - \theta_0)^2}{[E_{\theta_0} \psi'(X - \theta_0)]^2}\right).$$

例子

例 10.2.5 (Huber 估计量) 定义使 (10.2.1) 和 (10.2.2) 达最小的估计量叫做 Huber 估计量. 为了解这个估计量怎样工作以及 k 的选择如何重要, 考虑以下含有 8 个标准正态偏离值与 3 个“离群值”的数据集合:

$x = -1.28, -0.96, -0.46, -0.44, -0.26, -0.21, -0.063, 0.39, 3, 6, 9$

对于这些数据, 均值是 1.33 而中位数是 -0.21 . 当 k 变化时, 我们得到一系列 Huber 估计的值, 列在表 10.2.1 中. 我们看到, 当 k 增大时 Huber 估计值在中位数与均值间变动, 因而我们解释为随着 k 的增大, 则对离群值的稳健性下降. ||

例 10.2.6 (Huber 估计量的极限分布) 设 X_1, X_2, \dots, X_n 是 iid 的, 来自概率密度函数 $f(x - \theta)$, 其中 f 关于 0 对称, 则对于由式 (10.2.2) 给出的 ρ , 我们有

$$(10.2.7) \quad \psi(x) = \begin{cases} x & \text{若 } |x| \leq k \\ k & \text{若 } x > k \\ -k & \text{若 } x < -k \end{cases}$$

而且因此有

$$(10.2.8) \quad \begin{aligned} E_{\theta} \psi(X - \theta) &= \int_{\theta-k}^{\theta+k} (x - \theta) f(x - \theta) dx - \\ &\quad k \int_{-\infty}^{\theta-k} f(x - \theta) dx + k \int_{\theta+k}^{\infty} f(x - \theta) dx \\ &= \int_{-k}^k y f(y) dy - k \int_{-\infty}^{-k} f(y) dy + k \int_k^{\infty} f(y) dy = 0, \end{aligned}$$

其中我们做了替换 $y = x - \theta$. 几个积分相加得 0 是由于 f 的对称. 因此, Huber 估计量具有正确的均值 (见习题 10.25).

为了计算方差, 我们需要 ψ' 的期望值. 虽然 ψ 是不可微的, 但是越过不可微点 ($x = \pm k$), ψ' 将为 0. 因此我们只需要处理对于 $|x| \leq k$ 的期望, 从而我们有

$$\begin{aligned} E_{\theta} \psi'(X - \theta) &= \int_{\theta-k}^{\theta+k} f(x - \theta) dx = P_0(|X| \leq k), \\ E_{\theta} \psi(X - \theta)^2 &= \int_{\theta-k}^{\theta+k} (x - \theta)^2 f(x - \theta) dx + k^2 \int_{\theta+k}^{\infty} f(x - \theta) dx + k^2 \int_{-\infty}^{\theta-k} f(x - \theta) dx \\ &= \int_{-k}^k x^2 f(x) dx + 2k^2 \int_k^{\infty} f(x) dx. \end{aligned}$$

这样, 我们可以得出结论, Huber 统计量是渐近正态的, 具有均值 θ 和渐近方差

$$\frac{\int_{-k}^k x^2 f(x) dx + 2k^2 P_0(|X| > k)}{[P_0(|X| \leq k)]^2}. \quad \parallel$$

例 10.2.7 (Huber 估计量的 ARE) 因为 Huber 估计量在某种意义上是均值与中位数的折中, 我们将考察它相对于这两个估计量的相对效率.

Huber 估计量的渐近相对效率, $k=1.5$

	正态分布	罗吉斯蒂克分布	双指数分布
与均值比较	0.96	1.08	1.37
与中位数比较	1.51	1.31	0.68

对于正态及罗吉斯蒂克分布, Huber 估计量的表现类似于均值而比中位数有改进. 对于双指数分布, Huber 估计量比均值有改进但是不如中位数好. 回忆均值是关于正态的 MLE, 而中位数是关于双指数分布的 MLE (所以 $ARE < 1$ 是料想之中的). 对于这些分布, Huber 估计量具有类似于 MLE 的性能, 但是在其他情况似乎也保持合理性. ||

我们看到 M-估计量是稳健性和效率的一个折衷. 现在我们更仔细地分析为了得到稳健性, 在效率方面我们可能放弃了什么.

式 (10.2.6) 中的渐近方差的分母含有 $E_{\theta_0} \psi'(X - \theta_0)$ 项, 我们可以把它写成

$$E_{\theta} \psi'(X - \theta) = \int \psi'(x - \theta) f(x - \theta) dx = - \int \left[\frac{\partial}{\partial \theta} \psi(x - \theta) \right] f(x - \theta) dx.$$

现在我们运用乘积微分法则就得到

$$\begin{aligned} &\frac{d}{d\theta} \int \psi(x - \theta) f(x - \theta) dx \\ &= \int \left[\frac{d}{d\theta} \psi(x - \theta) \right] f(x - \theta) dx + \int \psi(x - \theta) \left[\frac{d}{d\theta} f(x - \theta) \right] dx. \end{aligned}$$

因为 $E_{\theta} \psi(X - \theta) = 0$, 所以上式的左侧是 0, 于是我们有

$$- \int \left[\frac{d}{d\theta} \psi(x - \theta) \right] f(x - \theta) dx = \int \psi(x - \theta) \left[\frac{d}{d\theta} f(x - \theta) \right] dx$$

$$= \int \psi(x-\theta) \left[\frac{d}{d\theta} \log f(x-\theta) \right] f(x-\theta) dx,$$

这里我们用到 $\frac{d}{dy} g(y)/g(y) = \frac{d}{dy} \log g(y)$ 这个事实. 最后一个表达式可以写成 $E_{\theta}[\psi(X-\theta)l'(\theta|X)]$, 其中 $l(\theta|x)$ 是对数似然函数, 这样就产生出恒等式

$$E_{\theta}\psi'(X-\theta) = -E_{\theta}\left[\frac{d}{d\theta}\psi(x-\theta)\right] = E_{\theta}[\psi(X-\theta)l'(\theta|X)]$$

(这里, 当我们取 $\psi = l'$, 就得出 (我们希望的) 熟悉的等式 $-E_{\theta}[l''(\theta|X)] = E_{\theta}l'(\theta|X)^2$; 见引理 7.3.11).

现在比较一个 M-估计量和 MLE 的渐近方差就是一件简单的事情了, 回忆 MLE $\hat{\theta}$ 的渐近方差是 $1/E_{\theta}l'(\theta|X)^2$, 所以借助于 Cauchy-Schwarz 不等式我们有

$$(10.2.9) \quad \text{ARE}(\hat{\theta}_M, \hat{\theta}) = \frac{[E_{\theta}\psi(X-\theta_0)l'(\theta|X)]^2}{E_{\theta}\psi(X-\theta)^2 E_{\theta}l'(\theta|X)^2} \leq 1$$

因此, 一个 M-估计量的效率总比 MLE 低, 只有当 ψ 和 l' 成比例时它的效率才能与 MLE 相匹敌 (见习题 10.29).

10.3 假设检验的渐近评价

与在 10.1 节一样, 本节描述几种在复杂的问题中获得某些检验的方法. 在这些问题当中, 不存在或不知道有像以前几节中所定义的那些最佳 (例如 UMP 无偏) 检验. 在这种情况下, 任何合理的检验的推导都可能有用. 下面我们将用两小节讨论似然比检验的大样本性质和其他近似的大样本检验.

10.3.1 LRT 的渐近分布

对于复杂模型最有用的方法之一就是构造检验的似然比方法, 因为它给出检验统计量的一个显式的定义

$$\lambda(x) = \frac{\sup_{\theta_0} L(\theta|x)}{\sup_{\theta} L(\theta|x)},$$

而且给出了拒绝区域的一个显式的形式, 即 $\{x: \lambda(x) \leq c\}$. 在观测到数据 $X=x$ 之后, 似然函数 $L(\theta|x)$ 就是变量 θ 的一个完全被定义了的函数. 即使 $L(\theta|x)$ 在集合 Θ_0 和集合 Θ 上的两个上确界不能被解析地得到, 它们一般也可以用数值方法计算. 因此, 即使 $\lambda(x)$ 没有方便的定义式可用, 对于观测的数据点, 检验统计

量 $\lambda(x)$ 值仍可以得到.

为了定义一个水平 α 检验, 必须选择常数 c 以使得

$$(10.3.1) \quad \sup_{\theta_0} P_{\theta}(\lambda(X) \leq c) \leq \alpha.$$

如果我们不能得到 $\lambda(x)$ 的一个简单公式, 似乎就没有希望得出 $\lambda(X)$ 的样本分布也就不知如何挑选 c 以使方程 (10.3.1) 成立. 然而, 如果我们借助于渐近分布, 我们就能够得到一个近似答案.

定理 10.3.1

定理 10.3.1 (LRT 的渐近分布简单 H_0) 关于检验 $H_0: \theta = \theta_0$ 对 $H_1: \theta \neq \theta_0$, 设 X_1, \dots, X_n 是 iid $f(x|\theta)$, $\hat{\theta}$ 是 θ 的 MLE, 并且 $f(x|\theta)$ 满足在杂录 10.6.2 中的正则性条件. 则在 H_0 之下, 当 $n \rightarrow \infty$,

$$-2\log \lambda(\mathbf{X}) \xrightarrow{L} \chi_1^2,$$

其中 χ_1^2 是一个具有自由度 1 的 χ^2 分布随机变量.

证明: 首先在 $\hat{\theta}$ 的邻域展开 $\log L(\theta|\mathbf{x}) = l(\theta|\mathbf{x})$ 为 Taylor 级数, 有

$$l(\theta|\mathbf{x}) = l(\hat{\theta}|\mathbf{x}) + l'(\hat{\theta}|\mathbf{x})(\theta - \hat{\theta}) + l''(\hat{\theta}|\mathbf{x})\frac{(\theta - \hat{\theta})^2}{2} + \dots$$

现在把 $l(\theta_0|\mathbf{x})$ 的展开式代入 $-2\log \lambda(\mathbf{x}) = -2l(\theta_0|\mathbf{x}) + 2l(\hat{\theta}|\mathbf{x})$ 中, 得到

$$-2\log \lambda(\mathbf{x}) \approx \frac{(\theta - \hat{\theta})^2}{-l''(\hat{\theta}|\mathbf{x})},$$

这里我们用到 $l'(\hat{\theta}|\mathbf{x}) = 0$ 这个事实. 因为分母就是观测信息数 $\hat{I}_n(\hat{\theta})$ 并且 $\hat{I}_n(\hat{\theta}) \rightarrow I(\theta_0)$, 于是根据定理 10.1.12 和 Slutsky 定理 (定理 5.5.17) 就推断出 $-2\log \lambda(\mathbf{X}) \rightarrow \chi_1^2$. ||

定理 10.3.3

定理 10.3.3 设 X_1, \dots, X_n 是来自一个概率密度函数或概率质量函数 $f(x|\theta)$ 的随机样本. 在杂录 10.6.2 中的正则性条件之下, 如果 $\theta \in \Theta_0$, 则统计量 $-2\log \lambda(\mathbf{X})$ 的分布在样本容量 $n \rightarrow \infty$ 时收敛到一个 χ^2 分布. 这个极限分布的自由度是由 $\theta \in \Theta$ 指明的自由参数个数与由 $\theta \in \Theta_0$ 指明的自由参数个数之差.

对于 $\lambda(\mathbf{X})$ 过小的值拒绝 $H_0: \theta \in \Theta_0$ 等价于对于 $-2\log \lambda(\mathbf{X})$ 过大的值作出拒绝. 因此

$$H_0 \text{ 被拒绝, 当且仅当 } -2\log \lambda(\mathbf{X}) \geq \chi_{\nu, \alpha}^2$$

其中 ν 是定理 10.3.3 中指出的自由度. 如果 $\theta \in \Theta_0$ 且样本量很大, 犯第一类错误的概率将近似为 α . 这样, 对于大的样本量, 方程 (10.3.1) 将近似地得到满足, 从而也定义了一个渐近的真实水平 α 检验. 注意定理实际上仅仅蕴涵

$$\lim_{n \rightarrow \infty} P_\theta(\text{拒绝 } H_0) = \alpha, \text{ 对于每个 } \theta \in \Theta_0,$$

而不是 $\sup_{\theta \in \Theta_0} P_\theta(\text{拒绝 } H_0)$ 收敛到 α . 渐近的真实水平 α 检验情况通常是这样.

检验统计量自由度的计算通常是直接的. 最经常的是, Θ 可以表示为 q -维欧氏空间的一个子集合, 它包含 \mathbf{R}^q 中的一个开子集, 而 Θ_0 可以表示为 p -维欧氏空间的一个子集合, 它包含 \mathbf{R}^p 中的一个开子集, 其中 $p < q$. 则 $q - p = \nu$ 就是这个检验统计量的自由度.

例 10.3.2

例 10.3.2 (Poisson LRT) 考虑基于 iid Poisson (λ) 的样本 X_1, \dots, X_n 的检验 $H_0: \lambda = \lambda_0$ 对 $H_1: \lambda \neq \lambda_0$, 我们有

$$-2\log\lambda(\mathbf{x}) = -2\log\left[\frac{e^{-n\lambda_0}\lambda_0^{\sum x_i}}{e^{-n\hat{\lambda}}\hat{\lambda}^{\sum x_i}}\right] = 2n[(\lambda_0 - \hat{\lambda}) - \hat{\lambda}\log(\lambda_0/\hat{\lambda})],$$

其中 $\hat{\lambda} = \sum x_i/n$ 是 λ 的 MLE. 运用定理 10.3.1, 如果 $-2\log\lambda(\mathbf{x}) > \chi^2_{1,\alpha}$, 我们就将在水平 α 上拒绝 H_0 .

为了对这个渐近的准确度有些认识, 这里给出这个检验的一个小型模拟. 设 $\lambda_0 = 5$ 和 $n=25$, 图 10.3.1. 显示的是把 $-2\log\lambda(\mathbf{x})$ 的 10000 个值做成的直方图与 χ^2_1 的概率密度函数图放在一起. 从图看起来, 符合是较好的. 此外, 在下面的表中给出了模拟的 (确切的) 和 χ^2_1 (近似的) 分界点的比较值, 它表明两者的分位点非常近似.

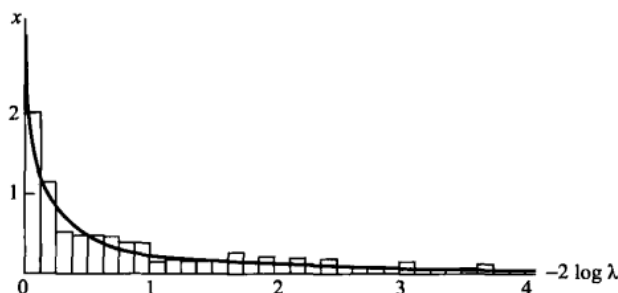


图 10.3.1 $-2\log\lambda(\mathbf{x})$ 的 10000 个值做成的直方图与 χ^2_1 的概率密度函数图, $\lambda_0 = 5$ 和 $n=25$

Poisson LRT 统计量的模拟的 (确切的) 和近似的百分位数

百分位数	0.80	0.90	0.95	0.99
模拟的	1.630	2.726	3.744	6.304
χ^2_1	1.642	2.706	3.841	6.635

||

10.3.2 其他大样本检验

另外一个构造大样本检验统计量的通用方法是建立在一个具有渐近正态分布的

估计量之上的. 设我们要检验关于一个实值参数 θ 的假设, 而 $W_n = W(X_1, \dots, X_n)$ 是一个基于样本容量 n 的通过某种方法得到的 θ 的点估计量. 例如, W_n 可能是 θ 的 MLE. 于是一个基于正态近似的近似检验可以通过下面途径证明它的合理性. 如果把 W_n 的方差记作 σ_n^2 , 而且如果我们能够用某种形式的中心极限定理证明当 $n \rightarrow \infty$ 时, $(W_n - \theta) / \sigma_n$ 依分布收敛到一个标准正态随机变量, 则 $(W_n - \theta) / \sigma_n$ 就可以比作一个 $N(0, 1)$ 分布. 我们因此就有了一个近似检验的基础.

当然, 在前段论述中有很多细节要验证, 但是这些想法的确应用于很多情形. 例如, 如果 W_n 是一个 MLE, 则可以使用定理 10.1.12 来证实上面论述的正确性. 要注意, W_n 的分布, 也许还有 σ_n 的值依赖于 θ 的值. 因此, 关于收敛更正式的说法是, 对于每个固定的 $\theta \in \Theta$, 如果我们对 W_n 使用其相应的分布而对 σ_n 使用其相应的值, 则 $(W_n - \theta) / \sigma_n$ 收敛到一个标准正态分布. 如果对于每个 n , σ_n 是一个可计算的常数 (它可能依赖于 θ 但不依赖于其他未知参数), 那么就可以推导出一个基于 $(W_n - \theta) / \sigma_n$ 的检验.

在某些情况中, σ_n 还依赖于未知的参数. 在这种情况下, 我们寻找 σ_n 的一个的估计 S_n , 满足 σ_n/S_n 依概率收敛到 1. 然后运用 Slutsky 定理 (如例 5.5.18) 我们就可以推出 $(W_n - \theta)/S_n$ 同样依分布收敛到一个标准正态分布. 这样就可以建立一个大样本检验.

设我们想检验双侧假设 $H_0: \theta = \theta_0$ 对 $H_1: \theta \neq \theta_0$. 一个近似检验就可以建立在统计量 $Z_n = (W_n - \theta_0)/S_n$ 之上, 并且当且仅当 $Z_n < -z_{\alpha/2}$ 或 $Z_n > z_{\alpha/2}$ 时拒绝 H_0 . 如果 H_0 为真, 则 $\theta = \theta_0$ 并且 Z_n 依分布收敛到 $Z \sim n(0, 1)$. 这样, 犯第一类错误的概率

$$P_{\theta_0}(Z_n < -z_{\alpha/2} \text{ 或 } Z_n > z_{\alpha/2}) \rightarrow P_{\theta_0}(Z < -z_{\alpha/2} \text{ 或 } Z > z_{\alpha/2}),$$

从而这是一个渐近的真实水平 α 检验.

现在考虑另一个 $\theta \neq \theta_0$ 的参数值. 我们可以写成

$$(10.3.3) \quad Z_n = \frac{W_n - \theta_0}{S_n} = \frac{W_n - \theta}{S_n} + \frac{\theta - \theta_0}{S_n}.$$

不管 θ 的值是什么, 都有 $(W_n - \theta)/S_n \rightarrow n(0, 1)$. 典型情况下还有当 $n \rightarrow \infty$ 时, $\sigma_n \rightarrow 0$. (回忆, $\sigma_n = \text{Var}W_n$, 而典型情况下当 $n \rightarrow \infty$ 时估计量变得愈渐精确.) 这样, S_n 将依概率收敛到 0, $(\theta - \theta_0)/S_n$ 将依概率收敛到 $+\infty$ 或 $-\infty$, 依赖于 $(\theta - \theta_0)$ 为正或负. 所以 Z_n 将依概率收敛到 $+\infty$ 或 $-\infty$, 而且

$$P_\theta(\text{拒绝 } H_0) = P_\theta(Z < -z_{\alpha/2} \text{ 或 } Z > z_{\alpha/2}) \rightarrow 1, \text{ 当 } n \rightarrow \infty.$$

这样, 就可以构造出一个具有渐近真实水平 α 和渐近功效 1 的检验.

我们想检验单侧假设 $H_0: \theta \leq \theta_0$ 对 $H_1: \theta > \theta_0$, 可以构造一个类似的检验. 这时将再次利用检验统计量 $Z_n = (W_n - \theta_0)/S_n$, 而且当且仅当 $Z_n > z_\alpha$ 时拒绝 H_0 . 运用与前面类似的推理, 我们可以断言这个检验的功效函数根据 $\theta < \theta_0$, $\theta = \theta_0$ 或

$\theta > \theta_0$ 收敛到 0, α 或 1. 因此, 这个检验也具有合理的渐近功效性质.

一般而言, 一个 Wald 检验 (Wald test) 是一个基于形式为

$$Z_n = \frac{W_n - \theta_0}{S_n}$$

的统计量的检验, 其中 θ_0 是参数 θ 的一个假设值, W_n 是 θ 的一个估计量, 而 S_n 是 W_n 的标准误差, 即 W_n 标准差的一个估计. 如果 W_n 是 θ 的 MLE, 那么就如 10.1.3 节所讨论, $1/\sqrt{I_n(W_n)}$ 是 W_n 的一个合理的标准误差. 也经常用 $1/\sqrt{\hat{I}_n(W_n)}$ 替换它, 其中

$$\hat{I}_n(W_n) = -\frac{\partial^2}{\partial \theta^2} \log L(\theta | \mathbf{X}) \Big|_{\theta = W_n}$$

是观测信息数 [参见 (10.1.7)].

例10.3.5

例 10.3.5 (大样本二项检验) 设 X_1, \dots, X_n 是来自总体 Bernoulli (p) 的随机样本. 考虑检验 $H_0: p \leq p_0$ 对 $H_1: p > p_0$, 其中 $0 < p_0 < 1$ 是一个指定的值. p 的基于样本容量 n 的 MLE 是 $\hat{p}_n = \sum_{i=1}^n X_i/n$. 因为 \hat{p}_n 正好是样本均值, 所以中心极限定理适用, 而且说明对于任何的 p , $0 < p < 1$, $(\hat{p}_n - p)/\sigma_n$ 收敛到一个标准正态随机变量, 其中 $\sigma_n = \sqrt{p(1-p)/n}$ 是一个依赖于未知参数 p 的值. σ_n 的一个合理的估计是 $S_n = \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$, 而且可以证明 (见习题 5.32) σ_n/S_n 依概率收敛到 1. 这样, 对于任何的 p , $0 < p < 1$,

$$\frac{\hat{p}_n - p}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} \rightarrow_{n(0,1)}.$$

Wald 检验统计量 Z_n 是在上式中把 p 换成 p_0 , 而大样本 Wald 检验当 $Z_n > z_\alpha$ 时拒绝 H_0 . 作为 σ_n 的一种替换的估计, 容易验证 $1/I_n(\hat{p}_n) = \hat{p}_n(1-\hat{p}_n)/n$. 所以, 如果我们使用信息数去推导 \hat{p}_n 的标准误差, 则得到相同的统计量 Z_n .

如果对于双侧检验 $H_0: p = p_0$ 对 $H_1: p \neq p_0$ 感兴趣, 其中 $0 < p_0 < 1$ 是一个指定的值, 可以再次应用上面的策略. 然而在这种情况下, 有另外一个近似检验. 根据中心极限定理, 对于任何的 p , $0 < p < 1$,

$$\frac{\hat{p}_n - p}{\sqrt{p(1-p)/n}} \rightarrow_{n(0,1)}.$$

因此, 如果原假设 H_0 为真, 则统计量

$$(10.3.4) \quad Z'_n = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim_n (0, 1) \quad (\text{近似地})$$

这个近似的水平 α 检验在 $|Z'_n| > z_{\alpha/2}$ 时拒绝 H_0 .

在两个检验都适用的情况, 例如, 当检验假设 $H_0: p = p_0$ 时, 不清楚选择哪

个. 它们的功效函数 (指实际的, 而不是近似的) 互相交叉, 所以每一个检验都是在一部分的参数空间上功效更强. (Ghosh 1979 对这个问题给出一些启示. Robbins 1977 及 Eberhardt and Fligner 1977 讨论了关于两样本二项分布问题的论证. 习题 10.31 给出了关于这个问题的两个不同的检验.)

当然, 任何对于功效函数的比较都被如下的事实所混扰, 即这些检验是近似的而不必保持水平 α . 利用连续性校正 (见例 3.3.2) 有助于这个问题. 在很多情况中, 使用连续性校正的近似方法是保守的, 就是说, 它们保持其名义 α 水平 (见例 10.4.6). ||

式 (10.3.4) 是另一个有用的大样本检验, 即记分检验 (score test) 的一个特例, 记分统计量 (score statistics) 的定义是

$$S(\theta) = \frac{\partial}{\partial \theta} \log f(\mathbf{X} | \theta) = \frac{\partial}{\partial \theta} \log L(\theta | \mathbf{X}).$$

根据式 (7.3.8) 我们知道, 对于所有的 θ , $E_{\theta} S(\theta) = 0$. 特别地, 如果我们在检验 $H_0: \theta = \theta_0$ 并且 H_0 为真, 则 $S(\theta_0)$ 的均值是 0. 进一步, 根据式 (7.3.10),

$$\text{Var}_{\theta} S(\theta) = E_{\theta} \left(\left(\frac{\partial}{\partial \theta} \log L(\theta | \mathbf{X}) \right)^2 \right) = -E_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \log L(\theta | \mathbf{X}) \right) = I_n(\theta);$$

这里的信息数是记分统计量的方差. 记分检验的检验统计量是

$$Z_S = S(\theta_0) / \sqrt{I_n(\theta_0)}.$$

如果 H_0 为真, Z_S 具有 0 均值和 1 方差. 根据定理 10.1.12 就可推出如果 H_0 为真, 则 Z_S 收敛到一个标准正态随机变量. 这样, 近似的水平 α 记分检验当 $|Z_S| > z_{\alpha/2}$ 时拒绝 H_0 . 如果 H_0 是复合假设, $\hat{\theta}_0$ 是假定 H_0 真时 θ 的估计, 则把 Z_S 中的 θ_0 替换成 $\hat{\theta}_0$. 如果 $\hat{\theta}_0$ 是限制的 MLE, 限制极大化可利用拉格朗日乘数法实现. 因此, 这个记分检验有时称为拉格朗日乘数检验 (Lagrange multiplier test).

10.4 区间估计的渐近评价

像我们在前面两节已做的那样, 现在来探索几种近似和渐近置信集合形式. 就像以往, 我们的目的是用例子说明一些将被用于更加复杂情况的方法, 将得到某种解答的方法. 这里得到的解答几乎一定不是最好的, 但是一定不是最坏的. 在很多情况下, 它们却是我们所能做到的最好的.

仍像过去, 我们从基于 MLE 的近似开始.

10.4.1 近似极大似然区间

根据 10.1.2 节的讨论, 并且运用定理 10.1.12, 我们就有了求 MLE 渐近分布的一般方法, 从而就有了构造一个置信区间的一般方法.

如果 X_1, \dots, X_n 是 iid $f(x|\theta)$ 的而 $\hat{\theta}$ 是 θ 的 MLE, 则根据式 (10.1.7), $\hat{\theta}$ 的一个函数 $h(\hat{\theta})$ 的方差可以由

$$\widehat{\text{Var}}(h(\hat{\theta})|\theta) \approx \frac{[h'(\theta)]^2|_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial \theta^2} \log L(\theta|\mathbf{x})|_{\theta=\hat{\theta}}}$$

近似. 现在, 对于一个固定的但是任意的 θ 值, 我们对

$$\frac{h(\hat{\theta}) - h(\theta)}{\sqrt{\widehat{\text{Var}}(h(\hat{\theta})|\theta)}}$$

的渐近分布感兴趣. 从定理 10.1.12 和 Slutsky 定理 (定理 5.5.17) (见习题 10.33) 就可推出

$$\frac{h(\hat{\theta}) - h(\theta)}{\sqrt{\widehat{\text{Var}}(h(\hat{\theta})|\theta)}} \rightarrow N(0, 1),$$

于是给出近似的置信区间

$$h(\hat{\theta}) - z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h(\hat{\theta})|\theta)} \leq h(\theta) \leq h(\hat{\theta}) + z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h(\hat{\theta})|\theta)}.$$

例 10.4.2

例 10.4.2 (二项记分区间) 仍旧利用一个二项分布的例, 如果 $Y = \sum_{i=1}^n X_i$,

其中每个 X_i 是一个独立的 Bernoulli (p) 随机变量, 我们有

$$\begin{aligned} Q(\mathbf{X} | p) &= \frac{-\frac{\partial}{\partial p} \log L(p | \mathbf{X})}{\sqrt{-E_p \left(\frac{\partial^2}{\partial p^2} \log L(p | \mathbf{X}) \right)}} \\ &= \frac{\frac{y}{p} - \frac{n-y}{1-p}}{\sqrt{\frac{n}{p(1-p)}}} \\ &= \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}, \end{aligned}$$

其中 $\hat{p} = y/n$. 由式 (10.4.2), 就给出了一个近似的 $1-\alpha$ 置信区间

$$(10.4.4) \quad \left\{ p : \left| \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \right| \leq z_{\alpha/2} \right\}.$$

这就是由反转记分统计量 (见例 10.3.6) 得到的区间. 为了计算这个区间我们需要解一个关于 p 的二次方程; 关于细节参见例 10.4.6. ||

在 10.3 节我们曾推出另外一个基于 $-2 \log \lambda(\mathbf{X})$ 具有渐近 χ^2 分布这个事实的似然检验. 这就表明如果 X_1, \dots, X_n 是 iid $f(x|\theta)$ 的并且 $\hat{\theta}$ 是 θ 的 MLE, 则集合

$$(10.4.5) \quad \left\{ \theta : -2 \log \left(\frac{L(\theta|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})} \right) \leq \chi_{1,\alpha}^2 \right\}$$

是一个近似的 $1-\alpha$ 置信区间. 情况就是如此, 并且给了我们另一个近似的似然区间.

当然, 式 (10.4.5) 恰好就是我们最初通过反转 LRT 统计量推出的最高似然区域 (9.2.7). 然而, 现在我们有了一条自动附加上近似置信水平的途径.

10.4.2 其他大样本区间

大多数近似置信区间是基于求近似的（或渐近的）枢轴或者反转近似的水平 α 检验统计量. 如果有任何的统计量 W 与 V 和一个参数 θ 使得当 $n \rightarrow \infty$, 有

$$\frac{W-\theta}{V} \rightarrow n(0,1),$$

则我们就可以通过

$$W - z_{\alpha/2}V \leq \theta \leq W + z_{\alpha/2}V$$

构成一个关于 θ 的近似的置信区间, 它本质上是一个 Wald-型区间. 直接应用中心极限定理连同 Slutsky 定理, 我们通常将给出一个近似的置信区间. (注意, 前一节的近似极大似然区间都反映了这个策略.)

例10.4.5

例 10.4.5 (近似的 Poisson 区间) 如果 X_1, \dots, X_n 是 iid Poisson (p) 的, 则我们知道有

$$\frac{\bar{X}-\lambda}{S/\sqrt{n}} \rightarrow n(0,1).$$

但是, 这即使在我们不从 Poisson 总体抽取样本的情况下也是对的. 利用 Poisson 假定, 我们就知道 $\text{Var}(X) = \lambda = E(\bar{X})$ 以及 \bar{X} 是 λ 的一个好的估计量 (见例 7.3.12). 因此, 根据 Poisson 假定, 我们也可以由

$$\frac{\bar{X}-\lambda}{\sqrt{\bar{X}/n}} \rightarrow n(0,1)$$

这个事实得到一个近似的置信区间, 它就是从反转 Wald 检验所得到的区间. 我们可以从另外一条路径利用 Poisson 假定. 因为 $\text{Var}(X) = \lambda$; 由此就可以得出

$$\frac{\bar{X}-\lambda}{\sqrt{\lambda/n}} \rightarrow n(0,1).$$

由此导致相应于记分检验的区间, 它也就是式 (10.4.2) 的似然区间并且根据 Wilks (1938) 的结论, 它是最好的 (见习题 10.40). ||

一般而言, 一个合理的经验法则是, 在近似中要尽可能少用估计多用参数. 其道理很简单, 参数被固定且不把任何附加的变动引入近似当中, 而每个统计量都代入更多的变动.

10.5 习题

1

10.1 X_1, \dots, X_n 是摘自概率密度函数为

$$f(x|\theta) = \frac{1}{2}(1+\theta x), \quad -1 < x < 1, \quad -1 < \theta < 1$$

的总体的随机样本. 求 θ 的一个相合估计量并证明其相合性.

2

10.3 X_1, \dots, X_n 是取自总体 $n(\theta, \theta)$ 的随机样本, 其中 $\theta > 0$.

- (a) 证明 θ 的 MLE, 即 $\hat{\theta}$ 是二次方程 $\theta^2 + \theta - W = 0$ 的一个根, 其中 $W = (1/n) \sum_{i=1}^n X_i^2$, 并确定哪个根是 MLE.
- (b) 用 10.1.3 节中的技术求 $\hat{\theta}$ 的近似方差.

3

10.5 就例 10.1.8 中的情形, 对 $T_n = \sqrt{n}/\bar{X}_n$ 证明:

- (a) $\text{Var}(T_n) = \infty$.
- (b) 如果 $\mu \neq 0$, 并且从样本空间中删去 $(-\delta, \delta)$, 则 $\text{Var}(T_n) < \infty$.
- (c) 如果 $\mu \neq 0$, 区间 $(-\delta, \delta)$ 中的概率当 $n \rightarrow \infty$ 时趋近于 0.

4

10.9 假定 X_1, \dots, X_n 是 iid Poisson (λ) 的. 求下列量的最佳无偏估计量:

- (a) $e^{-\lambda}$, 这是 $X=0$ 的概率.
- (b) $\lambda e^{-\lambda}$, 这是 $X=1$ 的概率.
- (c) 对 (a) 和 (b) 中的这些最佳无偏估计量, 计算它们相对于 MLE 的渐近相对效率. 你喜欢哪个估计? 为什么?
- (d) 对于可能致癌的化合物, 可以通过测量暴露于这种化合物下的微生物的突

统计推断

变率来进行基本检测. 试验人员把这种化合物放在 15 个皮氏培养皿中, 记录到下列数目的突变群体:

10, 7, 8, 13, 8, 9, 5, 7, 6, 8, 3, 6, 6, 3, 5.

估计 $e^{-\lambda}$, 即没有突变群体出现的概率, 以及 $\lambda e^{-\lambda}$, 也就是只有一个突变群体出现的概率. 计算最佳无偏估计和 MLE.

5

10.32 (a) 设 $(X_1, \dots, X_n) \sim \text{multinomial}(m, p_1, \dots, p_n)$. 考虑检验 $H_0: p_1 = p_2$ 对 $H_1: p_1 \neq p_2$. 一个常用的检验是所谓 **McNemar 检验**, 当

$$\frac{(X_1 - X_2)^2}{X_1 + X_2} > \chi^2_{1, \alpha}$$

时拒绝 H_0 . 证明这个检验统计量有形式

$$\sum_1^n \frac{(\text{观测频数} - \text{期望频数})^2}{\text{期望频数}},$$

其中 X_i 为观测到的格子频数, 期望格子频数为在假设 $p_1 = p_2$ 下 $m p_i$ 的 MLE.

(b) McNemar 检验经常用在下列类型的问题中. 问调查对象是否同意某个说法, 然后让他们读到一些关于这个说法的信息, 再问他们是否同意. 把每种情况的响应数量总结在下面的 2×2 表中:

		前	
		同意	不同意
后	同意	X_3	X_2
	不同意	X_1	X_4

假设 $H_0: p_1 = p_2$ 是说从同意变到不同意的人在所有人中的比例与从不同意变到同意的人在所有人中的比例相同. 可能检验的另一个假设是原来同意的人中改变态度的人的比例与原来不同意的人中改变态度的人的比例相同. 用条件概率表述这个假设, 并说明它不同于上述的 H_0 . (这个假设可以用习题 10.31 中的 χ^2 检验来进行.)

6

10.37 设 X_1, \dots, X_n 为来自总体 $N(\mu, \sigma^2)$ 的随机样本.

(a) 如果 μ 未知而 σ 已知, 证明 $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma$ 是检验 $H_0: \mu = \mu_0$ 的记分统计量.

(b) 如果 σ 未知而 μ 已知, 求检验 $H_0: \sigma = \sigma_0$ 的一个记分统计量.

7

10.41 设 X_1, \dots, X_n 是 iid 的, 服从负二项分布 $NB(r, p)$. 我们要构造负二项分布参数的近似置信区间.

(a) 计算 Wilks 近似 (10.4.1), 并说明如何用这个表达式形成置信区间.

(b) 求负二项分布的均值的近似 $1 - \alpha$ 置信区间, 并说明如何对求出的区间做连续性校正.

(c) 用负二项分布作为习题 9.23 中蚜虫数据的模型, 用 (b) 中的结果构造近似 90% 的置信区间. 把这个区间与习题 9.23 中基于 Poisson 分布的置信区间进行比较.