区间估计和方差

内容：

Fisher 信息量。

极大似然估计值的区间估计和方差，

估计值的某些函数的方差。

设 $L(\theta)$ 代表似然函数，$l(\theta)$ 代表对数似然函数。

定理 1 设 $X_1,\cdots,X_n$ 是独立同分布的，其概率密度函数（在离散情况下为概率函数）为 $f(x;\theta)$，θ 为参数. $f(x;\theta)$ 满足以下条件（对于离散情况下面的积分将换作求和）。

（1）参数空间 Ω 是一个开区间；

（2）集合 $A=\{x\,|\,f(x,\theta)>0\}$ 与参数 θ 的取值无关；

（3）对任意 $x\in A$，$f(x,\theta)$ 对 θ 三次可微，且二阶导数是 θ 的连续函数；

（4）积分$\int f(x,\theta)dx$在积分号下二次可微，即$\int \frac{\partial}{\partial\theta} f(x;\theta)dx = 0$，

$\int \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = 0$；

（5）$-\infty < \int f(x;\theta)\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)dx < 0$；

（6）存在函数$H(x)$，使得对任意$x \in A, \theta_0 - c < \theta < \theta_0 + c$，满足

$\int H(x) f(x;\theta_0)dx < \infty$，并且$|\frac{\partial^3}{\partial\theta^3}\ln f(x;\theta)| < H(x)$；

则有如下结论成立：

（a） 当$n\to\infty$时，似然函数方程$[L'(\theta)=0]$有解的概率趋近于 1

（b） 当$n\to\infty$时，极大似然估计值$\hat{\theta}_n$的分布收敛到正态分布，均值为θ、方差满足：$I(\theta)Var(\hat{\theta}_n)\to 1$，其中

$$I(\theta)=-nE[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)]=-n\int f(X;\theta)\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)dx$$

$$=nE[(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2]=n\int f(x;\theta)(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2dx$$

定理的第二个结论表明，对于任意的z，有

$$\lim_{n\to\infty}P(\frac{\hat{\theta}_n-\theta}{[I(\theta)]^{-1/2}}<z)=\Phi(z) \qquad （1）$$

根据定理 1，$[I(\theta)]^{-1}$为$Var(\hat{\theta}_n)$的一个有意义的估计。称$I(\theta)$为 Fisher 信息量。由这个结果我们得到极大似然估计量是渐近无偏和相合的。信息量同时构成了$Crame'r-Rao$下界，也就是说在一般条件下，任何无偏估计量的方差都大于由信息量的倒数给出的方差。因此，至少在渐近意义上，没有哪个无偏估计比极大似然估计更准确。(1)-(6)给出的条件被称为“一般正则性条件”。这些条件通常是成立的，但是难以验证，所以只是假设这些条件都是成立的.这些条件将确保密度函数对参数比较光滑，并且密度函数比较正常.

上面的叙述假设样本由独立同分布的随机变量观测组成。更一般的结果用对数似然函数表示

$$I(\theta)=-E[\frac{\partial^2}{\partial\theta^2}l(\theta)]=E[(\frac{\partial}{\partial\theta}l(\theta))^2] \tag{9.3.2}$$

这里唯一的要求是每个观测的参数值相同.

如果有一个以上的参数，结论只是改为参数向量的极大似然估计服从渐近的多元正态分布。这个分布的协方差阵通过求一个(r,s)处元素，由下式给出的矩阵的逆得到

$$I(\boldsymbol{\theta})_{rs}=-E[\frac{\partial^2}{\partial\theta_s\partial\theta_r}l(\boldsymbol{\theta})]=-nE[\frac{\partial^2}{\partial\theta_s\partial\theta_r}\ln f(X;\boldsymbol{\theta})]$$

$$= E[\frac{\partial}{\partial \theta_r} l(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_s} l(\boldsymbol{\theta})] = nE[\frac{\partial}{\partial \theta_r} \ln f(X;\boldsymbol{\theta}) \frac{\partial}{\partial \theta_s} \ln f(X;\boldsymbol{\theta})]$$

其中第一个和第三个表达式都是在一般情况下成立，第二个和第四个表达式假设似然函数是 n 个相同密度函数的乘积。这个矩阵通常称作信息阵。为保证 $I^{-1}(\boldsymbol{\theta})$ 存在，通常要求信息矩阵 $I(\boldsymbol{\theta})$ 是正定的.

例 8.16 假设用对数正态分布来拟合下面的数据，试用极大似然估计方法计算相应的分布参数，并计算协方差阵。

车险赔付数据

448	2482	2753	3786	3866	3965	6315	6664	6707	7947
839	1454	1544	1559	1560	1602	2212	2387	5512	22087

1	0	5	7	2	0	5	9	2	6

估计对数正态分布的极大似然估计量的方差。

解: 似然函数和对数似然函数为

$$L(u,\sigma)=\prod_{j=1}^{n}\frac{1}{x_j\sigma\sqrt{2\pi}}\exp[-\frac{(\ln x_j-u)^2}{2\sigma^2}]$$

$$l(u,\sigma)=\sum_{j=1}^{n}[-\ln x_j-\ln\sigma-\frac{1}{2}\ln(2\pi)-\frac{1}{2}(\frac{\ln x_j-u}{\sigma})^2]$$

它们的一阶导数为

$$\frac{\partial l}{\partial u}=\sum_{j=1}^{n}\frac{\ln x_j-u}{\sigma^2}\quad 和\ \frac{\partial l}{\partial\sigma}=-\frac{n}{\sigma}+\sum_{j=1}^{n}\frac{(\ln x_j-u)^2}{\sigma^3}$$

二阶导数为

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \sigma \partial \mu} = -2\sum_{j=1}^{n} \frac{\ln x_j - u}{\sigma^3}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - 3\sum_{j=1}^{n} \frac{(\ln x_j - u)^2}{\sigma^4}$$

由于 $\ln X_j$ 服从均值为 u，标准差为 σ 的正态分布，因此

$$E(\frac{\partial^2 l}{\partial u^2}) = -\frac{n}{\sigma^2}, E(\frac{\partial^2 l}{\partial u \partial \sigma}) = 0, E(\frac{\partial^2 l}{\partial \sigma^2}) = -\frac{2n}{\sigma^2}$$

改变符号后求逆可得到协方差阵的估计，为

$$\begin{bmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{\sigma^2}{2n} \end{bmatrix}$$

对于对数正态分布，极大似然估计是下面两个方程的解

$$\sum_{j=1}^{n} \frac{\ln x_j - u}{\sigma^2} = 0 \text{且} -\frac{n}{\sigma} + \sum_{j=1}^{n} \frac{(\ln x_j - u)^2}{\sigma^3} = 0$$

由 第 一 个 等 式 得 到 $\hat{u} = (1/n)\sum_{j=1}^{n} \ln x_j$ ， 由 第 二 个 等 式 有

$\hat{\sigma}^2 = (1/n)\sum_{j=1}^{n} (\ln x_j - \hat{u})^2$ 。 □

对于表中的数据，这些值分别为$\hat{u}$=9.127 和$\hat{\sigma}^2$=1.563。协方差矩阵需要参数的真值，但能够做到的最好方法使用估计值代替真值。于是得到参数μ,σ的极大似然估计值的方差估计值：

$$\mathrm{Var}(\hat{u},\hat{\sigma})=\begin{bmatrix}0.07815 & 0\\ 0 & 0.03908\end{bmatrix}$$

主对角线外的零说明这两个参数估计渐近不相关。在对数正态分布这个特殊情形下，这个结论对任何容量的样本都是成立的。因此可以构造一个置信度为 95%的参数真值的置信区间，在估计值两边的 1.96 个标准差内有

$$u：9.127\pm1.96（0.07815）^{1/2}=9.127\pm0.548$$

σ： $1.25 \pm 1.96(0.03908)^{1/2} = 1.25 \pm 0.3871$

为了得到信息阵需要计算导数和期望值，这通常并不容易。避免这个问题的一个方法是不计算期望值，直接使用观测数据。这个结果称已**观测信息量。**

例 8.16(续) 使用已观测信息量估计前例的协方差阵

解: 将观测值代入二阶导数方程得到

$$\frac{\partial^2 l}{\partial\mu^2}=-\frac{n}{\sigma^2}=--\frac{20}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial\sigma\partial\mu}=-2\sum_{j=1}^{n}\frac{\ln x_j-u}{\sigma^3}=-2\frac{182.54-20u}{\sigma^3}$$

$$\frac{\partial^2 l}{\partial\sigma^2}=\frac{n}{\sigma^2}-3\sum_{j=1}^{n}\frac{(\ln x_j-u)^2}{\sigma^4}=\frac{20}{\sigma^2}-3\frac{1697.2487-365.08u+20u^2}{\sigma^4}$$

带入参数的估计值得到观测信息的负值项

$$\frac{\partial^2 l}{\partial\mu^2}=-12.7959\,,\quad \frac{\partial^2 l}{\partial\sigma\partial\mu}=0\,,\quad \frac{\partial^2 l}{\partial\sigma^2}=-25.5259$$

改变符号求逆得到了和前面相同的结果，这是对数正态分布的特点，对于其他分布并不一定成。

□

极大似然估计量的函数的方差

信息矩阵提供了一种评价参数的极大似然估计质量的方法。但是，人们更关心作为参数函数的某些量的估计，比如，我们希望以帕累托分布的均值作为总体均值的估计。也就是说，以 $\hat{\alpha}\,\hat{\theta}/(\hat{\alpha}-1)$ 作为总体均值的一个估计，其中采用了极大似然估计值。计算这个随机变量的均值和方差是很困难的，因为其中的两个随机变量的分布已经相当的复杂。对于参数函数的方差估计，常用方法为 delta 方法。

delta 方法

令$Y_n = g(X_n)$ ，其中$g(.)$是连续可微的函数， X_n是对EX_n的一致估计。则在适当的正则性条件下，

$$Y_n - EY_n = g(X_n) - Eg(X_n) = g(X_n) - g(EX_n) + g(EX_n) - Eg(X_n)$$
$$= g'(X_n)(X_n - EX_n) + o(X_n - EX_n)$$

因此：$Var(Y_n) \approx [g'(X_n)]^2 Var(X_n)$。

下面定理给出了极大似然函数方差的描述。

定理2 令$X_n = (X_{1n,}\cdots, X_{kn})^T$表示一个容量为$n$的$k$维多元随机变量样本。假设$X$服从渐近正态分布，均值为$\theta$，协方差阵为$\boldsymbol{\Sigma}$/n，其中$\theta$和$\boldsymbol{\Sigma}$都不依赖于$n$。$g$是一个完全可微的$k$元函数，$G_n = g(X_{1n},\cdots,X_{kn})$。则$G_n$也服从均值为$g(\theta)$、方差为$(\partial g)^T \boldsymbol{\Sigma}(\partial g)/n$的渐近正态分布，其中$\partial g$为一阶导数向量$\partial g = (\partial g/\partial\theta_1,\cdots,\partial g/\partial\theta_k)^T$，而且取值于原随机变量参数的真值$\theta$。

对于单参数模型，这个定理的陈述如下：令$\hat{\theta}$为θ的估计量，服从均值为θ、方差为σ^2 / n的渐近正态分布。则$g(\hat{\theta})$也服从渐近正态分布，均值为$g(\theta)$，渐近方差为

$$[g'(\theta)](\sigma^2 / n)[g'(\theta)] = [g'(\theta)]^2 \sigma^2 / n$$

例 8.17 使用 delta 方法给出指数分布超过 4000 的概率值的极大似然估计的方差，然后计算例 8.16 数据的结果。

解:指数分布参数θ的极大似然估计是样本均值。要估计的量为

$$p = P(X > 4000) = \exp(-4000/\theta).$$

其极大似然估计为

$$\hat{p} = \exp(-4000/\hat{\theta}) = \exp(-4000/\bar{x}).$$

计算这个估计量的均值和方差并不容易，但是已知 $Var(\bar{X}) = Var(X)/n = \theta^2/n$.进一步的有

$$g(\theta) = e^{-4000/\theta}, \quad g'(\theta) = 4000\theta^{-2}e^{-4000/\theta}$$

因此由 delta 方法有

$$Var(\hat{p}) = \frac{(4000\theta^{-2}e^{-4000/\theta})^2\theta^2}{n} = \frac{16000000\theta^{-2}e^{-8000/\theta}}{n}$$

对于例 8.16 的数据有

$$\bar{x} = 22626.5$$

$$\hat{p} = \exp(-4000/22626.5) = 0.8380$$

$$\mathrm{Var}(\hat{p}) = \frac{16000000(22626.5)^{-2}e^{-8000/22626.5}}{n} = 0.001097$$

因此，p 的 95%置信区间为 $0.838 \pm 1.96\sqrt{0.001097}$，即（0.773，0.903）。

□

例 8.18 利用例 8.16 的数据构造对数正态总体均值的 95%置信区间。将这个结果与由传统方法样本均值得到的置信区间作比较。

解: 由例 8.16 得到 $\hat{u}$ =9.127 和 $\hat{\sigma}$ =1.2502，两个估计量的协方差矩阵为

$$\frac{\sum}{n}=\begin{bmatrix}0.07815 & 0\\ 0 & 0.03908\end{bmatrix}$$

函数 $g(u,\sigma)=\exp(\mu+\frac{1}{2}\sigma^2)$. 偏导数为

$$\frac{\partial g}{\partial u}=\exp(u+\frac{1}{2}\sigma^2),\quad \frac{\partial g}{\partial \sigma}=\sigma\exp(u+\frac{1}{2}\sigma^2)$$

这些量的估计值分别为 20100.5 和 25129.65.由 *delta* 方法得到下面的估计

$$\begin{aligned}&\operatorname{Var}[g(\hat{u},\hat{\sigma})]\\&=[20100.5\ \ 25129.6]\begin{bmatrix}0.07815 & 0\\ 0 & 0.03908\end{bmatrix}\begin{bmatrix}20100.5\\ 25129.6\end{bmatrix}\\&=56253771.05\end{aligned}$$

置信区间为 20100.5 ± 14700.5，即（5400，34801）。□

例 8.19 已知累积危险率函数 $H(5)$ 的 95%置信区间为(0.283, 1.267)，求 S(5)的 95%的置信区间。

解： 该置信区间的中点为 0.775，即 $\hat{H}=0.775$。由

$$(0.775-1.96\sqrt{\operatorname{Var}(\hat{H}(5)}\ ,0.775+1.96\sqrt{\operatorname{Var}(\hat{H}(5))})=\ (0.283,\ 1.267)$$

知 $\mathrm{Var}(\hat{H}(5))$ 的估计值为 0.063。因为 $S(t)=\exp(-H(t))$，$\dfrac{dS(t)}{dH(t)}=e^{-H(t)}$，

由 delta 方法，

$$\mathrm{Var}(\hat{S}(t))=(e^{-\hat{H}(t)})^2\,\mathrm{Var}(\hat{H}(t))$$

代入 $\hat{H}(5)=0.775$， $Var(\hat{H}(5))$ =0.063 得

$$\mathrm{Var}(\hat{S}(5))=(e^{-0.775})^2\,0.063=0.0134$$

$S(5)$ 的估计值为 $e^{-0.775}=0.4607$，置信区间为 $0.4607\pm1.96\sqrt{0.0134}=(0.23,0.69)$。 □

背景知识

危险率函数

危险率函数是生存分析中的另一个基本函数，它描述被观察个体在某时刻存活的条件下，在以后的单位时间内死亡的（条件）概率。危险率函数也称为条件瞬时死亡率、死亡密度。在人口学中，它也被称为死亡力，在可靠性研究中，也称为条件失效率。

危险率函数的定义为

$$h(t)=\frac{f(t)}{S(t)}$$

显然，$h(t)$是在生存到时刻t的条件下的死亡密度。

我们注意到，密度函数

$$f(t)=-\frac{dS(t)}{dt}$$

这两个式子都是对个体在 t 时刻死亡密度的瞬时度量，但 $f(t)$只需个体在初始时刻生存即可，而$h(t)$却需个体在时刻 t 生存，这也是称 $f(t)$是时刻 t 死亡的无条件密度，而$h(t)$是条

件密度的原因。

将 $f(t)$ 代入 $h(t)$，有

$$h(t) = -\frac{dS(t)/S(t)}{dt} = -\frac{d\ln S(t)}{dt}$$

将上式两边从 0 到 t 积分得

$$\int_0^t h(y)dy = -\ln S(t)$$

所以 $S(t) = e^{-\int_0^t h(y)dy}$ 。

记 $H(t) = \int_0^t h(y)dy = -\ln S(t)$,为累积危险率函数

则 $S(t) = e^{-H(t)}$

277. You are given:

(i) Loss payments for a group health policy follow an exponential distribution with unknown mean.
(ii) A sample of losses is:100 200 400 800 1400 3100
Use the delta method to approximate the variance of the maximum likelihood estimator of $S(1500)$.

(A) 0.019
(B) 0.025
(C) 0.032
(D) 0.039
(E) 0.045
Key:A

极大似然估计值$\bar{\theta}=\bar{x}=1000$，

$$S(\theta)=\exp(-1500/\theta),$$
$$S'(\theta)=1500\theta^{-2}\exp(-1500/\theta)$$

$$Var(\hat{\theta})=\mathrm{Var}(\bar{\mathrm{X}})=\mathrm{Var(X)/n}=\theta^2/n$$

$$\begin{aligned}Var(S(\hat{\theta}))&=[S'(\hat{\theta})]^2 Var(\hat{\theta})\\&=[1500(1000)^{-2}\exp(-1500/1000)]^2(1000^2/6)\\&=0.01867\end{aligned}$$

课堂练习

259. You are given:

(i) A hospital liability policy has experienced the following numbers of claims over a10-year period: 10 2 4 0 6 2 4 5 4 2

(ii) Numbers of claims are independent from year to year.

(iii) You use the method of maximum likelihood to fit a Poisson model.

Determine the estimated coefficient of variation of the estimator of the Poisson parameter.

(A) 0.10

(B) 0.16

(C) 0.22

(D) 0.26

(E) 1.00

Question #259
Key: B

The estimator of the Poisson parameter is the sample mean. Then,

$$E(\hat{\lambda}) = E(\bar{X}) = \lambda$$

$$Var(\hat{\lambda}) = Var(\bar{X}) = \lambda / n$$

$$c.v. = \sqrt{\lambda / n} / \lambda = 1/\sqrt{n\lambda}$$

It is estimated by $1/\sqrt{n\lambda} = 1/\sqrt{39} = 0.1601$.