DOI :10.13203/j.whugis2003.05.015
第 28 卷 第 5 期　　武 汉 大 学 学 报 · 信 息 科 学 版　　Vol. 28 No. 5
2003 年 10 月　　Geomatics and Information Science of Wuhan University　　Oct. 2003

文章编号: 1671-8860(2003)05-0562-04　　文献标识码: A

基于信息扩散的极大似然估计

游扬声[1]　王新洲[2]

(1　武汉大学测绘学院, 武汉市珞喻路 129 号, 430079)
(2　武汉大学科技部, 武汉市珞珈山, 430072)

摘　要: 提出了扩散极大似然估计方法, 利用实际观测值的概率密度函数的信息扩散估计, 代替了对观测值分布的主观假设, 从而具有很强的自适应性。最后设计了两个算例, 说明了扩散极大似然估计的过程, 并考察了扩散极大似然估计的特性。

关键词: 信息扩散; 极大似然估计; 参数估计; 窗宽
中图法分类号: P207.2

1　基于信息扩散的极大似然估计

1.1　极大似然估计

设随机观测向量 l, 其密度函数为 $f(x)$, 它是未知的非随机参数 x 的函数, 此时有似然函数 $f(l;x)=f(x)$。估计值应使 $f(l;x)$ 达到最大, 称为极大似然估值, 并记作 X_{ML}。

为了能够使 $f(l;x)$ 达到最大, 求 $f(l;x)$ 对 x 的偏微分, 并令其等于 0。由于似然函数和密度函数是正的, 因此一般求 $\ln f(l;x)$ 对 x 的偏微分。显然, $\ln f(l;x)$ 是 $f(l;x)$ 的单增函数, 它们具有相同的极值点。这样, 极大似然估计值就是对数似然方程

$$\frac{\partial \ln f(l,x)}{\partial x}=0 \tag{1}$$

的解。

1.2　单参数扩散极大似然估计

假设 l 是某个未知参数 x 的函数, $l=g(x)$, $l_i(i=1,2,\cdots,n)$ 是独立同分布观测值。通常将函数 $l=g(x)$ 线性化, 得:

$$l=kx-\Delta \tag{2}$$

式中, k 是常系数; Δ 是随机误差向量。由于 Δ 的期望为 0, 将式(1)中的 l 用 kx 代替, 得:

$$\hat{f}(l)=\frac{1}{nh}\sum_{i=1}^{n}\mu\left(\frac{kx-l_i}{h}\right)=\hat{f}(l,x) \tag{3}$$

则似然方程为:

$$\frac{\partial \hat{f}(l,x)}{\partial x}=\frac{1}{nh}\sum_{i=1}^{n}\frac{\partial \mu\left(\frac{kx-l_i}{h}\right)}{\partial x}=0 \tag{4}$$

但必须指出, 通常的似然方程里密度是 $f(l,x)$, 而不是 $\hat{f}(l,x)$。但实际上, $f(l,x)$ 很难确切知道, 它们几乎总是假定为已知的, 即使通过对 $f(l,x)$ 作假设检验, 接受了它属于某种类型的假设, 它仍可能与实际分布有相当的差距。在这里, 密度函数是根据实际观测值估计得到的。

选定最优扩散函数[1], 由式(4)容易导出:

$$\sum_i (kx-l_i)\big|_{|kx-l_i|\leqslant\sqrt{5}h}=0 \tag{5}$$

设 $k=1$, 则

$$\sum_i (x-l_i)\big|_{|x-l_i|\leqslant\sqrt{5}h}=0 \tag{6}$$

此时即 l_i 是对 x 的重复观测值的情况。设 l_i 是独立同分布的, 若在区间 $[x-\sqrt{5}h, x+\sqrt{5}h]$ 内, 最多能够包含 $n_0(n_0\leqslant n)$ 个观测值, 则可以根据式(3)、式(4), 得到 x 的极大似然估计值为:

$$x_{DML}=\frac{1}{n_0}\sum_{i=1}^{n_0}l_i \tag{7}$$

1.3　单参数扩散极大似然估计的一点讨论

由式(7), 容易理解单参数扩散极大似然估计是稳健的。由于最优扩散函数在实数域上有紧支撑, 在求解 x_{DML} 时, 给出了一个明确的界限 $\pm\sqrt{5}h$, 落在 $[x-\sqrt{5}h, x+\sqrt{5}h]$ 区域内的观测值被利用, 该区域外的观测值被剔除。从某种意义

收稿日期: 2003-04-15。
项目来源: 国家自然科学基金资助项目(49874002); 国家测绘局测绘科技发展基金资助项目(99010)。

上说, $\pm\sqrt{5}h$ 可看作是粗差和偶然误差的分界。由于这个界限不依赖于研究人员的经验, 因此, 它提供了一个比较客观的标准, 但这个界限决定于窗宽 h。当样本容量很小时, 按文献[2](6.14)式计算, 所得的窗宽又完全依赖于样本的极差, 而极差很可能受粗差影响, 因此, 当样本容量很小时, 粗差还是对估计结果有一定的影响。

1.4　多参数扩散极大似然估计

设 $\boldsymbol{L}$ 是相互独立的 n 组观测值, 第 i ($i=1, 2, \cdots, n$)组包含有 n_i 个独立同分布观测值, 它们是 t 维未知参数 $\boldsymbol{X}$ 的函数: $\boldsymbol{L}=G(\boldsymbol{X})$。为方便计算, 通常取 $\boldsymbol{X}$ 的近似值 $\boldsymbol{X}^0$, 将其线性化后得:

$$\boldsymbol{L} = \boldsymbol{B}\boldsymbol{X}^0 + \boldsymbol{B}\boldsymbol{x} - \Delta \qquad (8)$$

式中, $\boldsymbol{L}$ 是一个 $\sum_i n_i \times 1$ 的观测值阵; $\boldsymbol{B}$ 是一个 $\sum_i n_i \times t$ 的系数矩阵, 其元素记为 $b_{i,j}$ ($j=1, 2, \cdots, t$); Δ 是一个 $\sum_i n_i \times 1$ 的观测误差阵。为方便计算, 将第 i 组观测值看作一个子阵, 则 $\boldsymbol{L}$、Δ 有 n 个子矩阵。

令 $\boldsymbol{l}=\boldsymbol{L}-\boldsymbol{B}\boldsymbol{X}^0$, $\boldsymbol{V}$ 是 Δ 的估计值, $\hat{\boldsymbol{x}}$ 是 $\boldsymbol{x}$ 的估计值, 则

$$\boldsymbol{V} = \boldsymbol{B}\hat{\boldsymbol{x}} - \boldsymbol{l} \qquad (9)$$

选定最优扩散函数, 对每一组观测值作信息扩散估计, 得到各组观测值的密度函数 $f(L_i)$ 的估计为:

$$\hat{f}(L_i) = \frac{1}{n_i h_i}\sum_{j=1}^{n_i} \mu\left(\frac{L_i - L_{i,j}}{h_i}\right) \qquad (10)$$

式中, h_i 是第 i 组观测值的窗宽。

由观测值相互独立, 可得:

$$\hat{f}(\boldsymbol{L}) = \hat{f}(L_1)\hat{f}(L_2)\cdots\hat{f}(L_n) = \prod_i^n \hat{f}(L_i) \qquad (11)$$

容易写出对数似然方程为:

$$\frac{\partial \ln\hat{f}(\boldsymbol{L})}{\partial \boldsymbol{X}} = \frac{\frac{\partial\hat{f}(L_1)}{\partial\boldsymbol{X}}}{\hat{f}(L_1)} + \frac{\frac{\partial\hat{f}(L_2)}{\partial\boldsymbol{X}}}{\hat{f}(L_2)} + \cdots + \frac{\frac{\partial\hat{f}(L_n)}{\partial\boldsymbol{X}}}{\hat{f}(L_n)} = \sum_{i=1}^{n}\frac{\frac{\partial\hat{f}(L_i)}{\partial\boldsymbol{X}}}{\hat{f}(L_i)} = 0 \qquad (12)$$

事实上, 式(12)是由 t 个方程组成的方程组, 其中第一个方程中的第一项为:

$$\frac{\frac{\partial\hat{f}(L_1)}{\partial X_1}}{\hat{f}(L_1)} = \frac{-2b_{1,1}\sum_{j=1}^{n_1}\left(\boldsymbol{B}_1\boldsymbol{X} - L_{1j}\right)}{\left[5h_1^2 n_1 - \sum_{j=1}^{n_1}\left(\boldsymbol{B}_1\boldsymbol{X} - L_{1j}\right)^2\right]} \qquad (13)$$

式中, $\boldsymbol{B}_1$ 是矩阵 $\boldsymbol{B}$ 的第一行。依此可以写出整个对数似然方程。显然, 式(12)是一个非线性方程组, 解此方程组, 即得到未知参数 $\boldsymbol{X}$ 的极大似然估计值 $\boldsymbol{X}_{\mathrm{DML}}$。

为方便计算, 用 $\boldsymbol{x}$ 代替 $\boldsymbol{X}$、$\boldsymbol{l}$ 代替 $\boldsymbol{L}$, 则式(12)的解为 $\hat{\boldsymbol{x}}_{\mathrm{DML}}$。容易理解: $\boldsymbol{X}_{\mathrm{DML}} = \boldsymbol{X}^0 + \hat{\boldsymbol{x}}_{\mathrm{DML}}$。

1.5　多参数扩散极大似然估计的一点讨论

在实际测量中, 通常要进行测站平差和网平差两次平差。测站平差总是对重复观测值取平均, 网平差时仅取用测站平差后的结果。根据信息不增性原理[3]知, 在这种工作模式下, 测站平差很可能损失了一部分信息。在多元扩散极大似然估计中, 在网平差时, 利用了所有原始观测值的信息, 才有可能具有最高的估计效率。

在扩散极大似然估计中, 没有对观测值的理论分布作主观假定, 或者说, 理论分布模式就是实际分布。因此, 这种方法同时兼顾了稳健性与有效性, 有很强的自适应能力, 是一种智能的估计方法。它根据样本, 自动估计具体样本的实际分布, 并根据其分布给出参数的极大似然估值。

这种方法的提出有两方面的理由: ① 随着科技的发展, 要处理的含有误差的数据多种多样, 这些数据的理论分布模式不一定是正态的, 甚至没有对它们的理论分布模式的先验信息, 此时要用极大似然法估计未知参数, 这不可避免地会遇到困难。② 现代的一些测量仪器很容易在短时间内获得大量数据。设GPS 观测历元为 2s, 则仅需 5min 就可获得150 个观测值。在这种情况下, 样本容量很大, 足以运用信息扩散估计得到观测值的具体分布。另外, 由于扩散估计利用了模糊过渡信息, 在小样本的情况下仍能够估计密度函数。

2　算　例

2.1　算例1

图1模拟了一个最小的三角网。设观测了一个三角形的 3 个内角, 每个角观测了 4 个测回, 观测值列于表 1。测站平差总是取各测回的平均值作为网平差的观测值, 其结果列于表 1。

表1　观测值

Tab.1　Observations

	1 ° ′ ″	2 ° ′ ″	3 ° ′ ″	4 ° ′ ″
$L_1(X_1)$	42 30 28	42 30 27	42 30 24	42 30 25
$L_2(X_2)$	40 19 37	40 19 36	40 19 39	40 19 40
L_3	97 09 55	97 09 56	97 09 57	97 09 44

取 L_1、L_2 分别为未知数 X_1、X_2, 得观测方程:

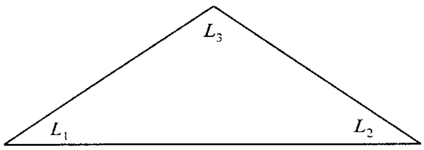

图 1 三角形

Fig. 1 Triangle

$$\begin{cases} L_1 = X_1 - \Delta_1,\ L_2 = X_2 - \Delta_2 \\ L_3 = 180^\circ - X_1 - X_2 - \Delta_3 \end{cases} \tag{14}$$

为方便计算, 对 X 取近似值X^0, 其改正量为 x_1、x_2, $X_1 = X_1{}^0 + x_1$, $X_2 = X_2{}^0 + x_2$。令 v_1 是 Δ_1 的估计值, v_2 是 Δ_2 的估计值, v_3 是$-\Delta_3$ 的估计值, 则相应的误差方程式为:

$$v_1 = x_1, v_2 = x_2, v_3 = x_1 + x_2 - 3'' \tag{15}$$

采用最小二乘估计, 易得 $\hat{x}_{1LS} = \hat{x}_{2LS} = 1.00''$, $X_{1LS} = X_1^0 + \hat{x}_{1LS}$, $X_{2LS} = X_2^0 + \hat{x}_{2LS}$, 其估计结果列于表 2。

表 2 最小二乘估计

Tab. 2 Least Square Estimation

	角度平均值	近似值	LS 估计结果
$L_1(X_1)$	$42^\circ30'26''$	$42^\circ30'26''(X_1^0)$	$42^\circ30'27.00''$
$L_2(X_2)$	$40^\circ19'38''$	$40^\circ19'38''(X_2^0)$	$40^\circ19'39.00''$
L_3	$97^\circ09'53''$	$180^\circ\text{-}X_1^0\text{-}X_2^0$	$97^\circ09'54.00''$

现采用扩散极大似然法来估计角度, 观测方程与式(14)相似, 但每一个原始观测值都有一个观测方程, 仍取近似值, 常数项列于表 3。

相应的误差方程为:

$$\begin{cases} v_{1,1} = x_1 - 2, v_{1,2} = x_1 - 1, v_{1,3} = x_1 + 2, v_{1,4} = x_1 + 1 \\ v_{2,1} = x_2 + 1, v_{2,2} = x_2 + 2, v_{2,3} = x_2 - 1, v_{2,4} = x_2 - 2 \\ v_{3,1} = x_1 + x_2 - 1, v_{3,2} = x_1 + x_2, v_{3,3} = x_1 + x_2 + 1, v_{3,4} = x_1 + x_2 - 12 \end{cases} \tag{16}$$

按照文献[2]的(6.14)式计算各组的窗宽:

$$h_i = 1.274 r_i / (4 - 1)$$

其结果列于表 3。

表 3 扩散极大似然估计/(″)

Tab. 3 Diffusion of Max. Likelihood Estimation/(″)

	1	2	3	4	极差 r	窗宽 h
$l_1(x_1)$	2	1	−2	−1	4	1.699
$l_2(x_2)$	−1	−2	1	2	4	1.699
l_3	1	0	−1	12	13	5.521

按照信息扩散估计式: $\hat{f}(l) = \dfrac{1}{nh}\sum_{i=1}^{n} u\left(\dfrac{l - l_i}{h}\right)$, 选取最优扩散函数, 得到各组观测值的密度函数估计为:

$$\hat{f}(L_i) = \frac{1}{4h_i}\sum_{j=1}^{4}\tilde{u}\left(\frac{L_i - L_{ij}}{h_i}\right)$$

用 l 代替L, 再根据式(11)得:

$$\hat{f}(l_i) = \frac{1}{4h_i}\sum_{j=1}^{4}\tilde{u}\left(\frac{l_i - l_{i,j}}{h_i}\right)$$

按照式(12)、(13), 得对数似然方程为:

$$\begin{cases} -2\sum_j (x_1 - l_{1,j})/[5\times 4\times h_1^2 - \sum_j (x_1 - l_{1,j})^2] + \\ \qquad -2\sum_j (x_1 + x_2 - l_{3,j})/[5\times 4\times h_3^2 - \sum_j (x_1 + x_2 - l_{3,j})^2] = 0 \\ -2\sum_j (x_2 - l_{2,j})/[5\times 4\times h_2^2 - \sum_j (x_2 - l_{2,j})^2] + \\ \qquad -2\sum_j (x_1 + x_2 - l_{3,j})/[5\times 4\times h_3^2 - \sum_j (x_1 + x_2 - l_{3,j})^2] = 0 \end{cases} \tag{17}$$

式(17)的解为: $\hat{x}_{1DML} = \hat{x}_{2DML} = 0.25''$。则平差值为: $X_{1DML} = X_1{}^0 + \hat{x}_{1DML} = 42^\circ30'26.25''$, $X_{2DML} = X_2{}^0 + \hat{x}_{2DML} = 40^\circ19'38.25''$, $\hat{L}_{3DML} = 97^\circ09'55.50''$。

2.2 算例 2

网形及数据均与算例 1 相同, 仅将其中的 $l_{3,4}$改为 $97^\circ09'53''$, 分别采用最小二乘法和扩散极大似然法作估计, 其结果相同, 即 $X_{1LS} = X_{1DML}$, $X_{2LS} = X_{2DML}$。

3 结 语

对算例 1 的原始观测值作简单的分析, 容易发现, 最后一个观测值 $l_{3,4}$比较可疑(可能含有较大误差), 它与近似值的差高达 $12''$, 而其他 11 个观测值与近似值的差最大为 $2''$, 这意味着其他观测值相互之间吻合得较好, 而 $l_{3,4}$离群。根据主观经验, $l_{3,4}$可能含有较大误差, 而由于样本很小, 这种怀疑难以通过假设检验来证实。又因为在测站平差时已经取了平均, 网平差时的观测值完全反映不出 $l_{3,4}$的离群情况, 不宜采用稳健估计。当采用最小二乘法平差时, 每一个角度平差值都受到 $l_{3,4}$的严重影响。事实上, 它对 $\hat{L}_{3LS}$的影响为 $2''$。算例 2 中没有明显离群的观测值, 此时, 采用最小二乘估计是很合理的。

如果采用本文提出的扩散极大似然估计, 算例 1 中可疑的观测值 $l_{3,4}$ 对平差值的影响很小, 事实上, 它对平差值产生的最大影响为 0.5″。这表明, 扩散极大似然估计能够顾及到人们对 $l_{3,4}$ 的怀疑, 合理地降低了它对平差值的影响。显而易见, 扩散极大似然估计不要求知道观测值的密度函数, 从而具有很强的自适应性。

注意到算例 1 和算例 2 的扩散极大似然估计相同(仅相差 0.002″), 而两个例子中有个观测值相差 9″。如果将算例 2 的观测值看作是正常的, 算例 1 的观测值看作是加了 1 个粗差, 则可以发现: ① 在观测值不含粗差时, 扩散极大似然估计结果与最小二乘估计相同; ② 在观测值含有 1 个粗差时, 扩散极大似然估计基本不受粗差的影响(限于篇幅, 在此不讨论多个粗差的情况)。因此可以说, 扩散极大似然估计是稳健的。

观察到在扩散极大似然估计中, 最后的参数估计结果同时利用了所有原始观测值, 所以认为这种方法具有较高的估计效率。

参　考　文　献

1　游扬声. 信息扩散原理及其在测量数据处理中的应用: [学位论文] . 武汉: 武汉大学, 2001

2　Huang C F. Principle of Information Diffusion. Fuzzy Sets and Systems, 1997, 91: 69 ~90

3　金振玉. 信息论. 北京: 北京理工大学出版社, 1991

第一作者简介: 游扬声, 博士生。现从事现代测量数据处理理论及应用研究。已发表论文 4 篇。

E-mail: sheng_yy@sina.com

Maximum Likelihood Estimation Based on the Principle of Information Diffusion

YOU Yangsheng[1]　*WANG Xinzhou*[2]

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

(2 Projects and Consulting Office, Wuhan University, Luojia Hill, Wuhan, China, 430072)

Abstract: This paper introduces the principle of information diffusion and information diffusion estimation (IDE). And with IDE, the observation distribution can be estimated easily. Once the observation distribution is determined, the parameters can be estimated with the maximum likelihood. With IDE as the basis, the diffusion maximum likelihood (DML) estimation is presented. DML is supposed to enjoy priority because of its being free from any supposition of observation distribution. With two simulative persuasive examples, the high self-adapting and robustness of DML are discussed.

Key words: information diffusion; maximum likelihood estimation; parameter estimation; window-width

About the first author: YOU Yangsheng, Ph. D candidate. He is concentrated on the research in the theory and application of surveying data processing. Four of his papers have been published.

E-mail: sheng_yy@sina.com

(责任编辑:　光阳)