三点插值参数估计法及其在生产中的应用

李　　宁

(石油勘探开发科学研究院研究生部)

一、问题的提出

已知可控测量条件 x 与随机测量结果 y 之间有如下形式的指数统计规律：

$$y = a_1 e^{a_2 x} + a_3, \tag{1}$$

试在下列实验条件下估计参数 a_1, a_2, a_3. 实验条件是，在现有实验设备、实验对象或实验费用的限制下，只可能提供三个互异实验点(即能够估计 a_1, a_2, a_3 的最少实验点数目)：(x_1, y_1), (x_2, y_2), (x_3, y_3), 且 $x_i (i = 1, 2, 3)$ 之间不等距，但不排除重复实验的可能.

显然，在上述实验条件约束下，不能直接用回归方法估计参数 a_1, a_2, a_3，亦不能用一般代数方法求解方程组

$$y_i = a_1 e^{a_2 x_i} + a_3 \qquad (i = 1, 2, 3), \tag{2}$$

来确定 a_1, a_2, a_3. 本文讨论在这种情况下估计参数 a_1, a_2, a_3 的一种方法. 这种方法的剩余离差平方和 $\Sigma(y_i - \hat{y}_i)^2$ (y_i——实测值，$\hat{y}_i$——估计值) 不小于回归方法的剩余离差平方和. 因此，当不受上述实验条件限制(即可以用回归方法估计参数 a_1, a_2, a_3)时，本方法不宜采用.

二、方法的基本思想

设 $x_3 > x_2 > x_1$, 由于 $x_2 - x_1 \neq x_3 - x_2$, 取 $x^* = \frac{(x_1 + x_3)}{2}$ ($\neq x_2$), 如果能找到与 x^* 对应的 y^*，则方程组(2)中的

$$a_3 = \frac{y_1 y_3 - (y^*)^2}{y_1 + y_3 - 2y^*}. \tag{3}$$

注意，当实验可重复时，这里及下面的 $y_i (i = 1, 2, 3)$ 取作平均值. 现在利用已知的三个实验点 (x_i, y_i) $(i = 1, 2, 3)$, 用插值方法求 y^*. 过三点做插值的方法有多种，通常采用二次插值(抛物插值)，即令

$$P_{01} = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1), \tag{4}$$

$$P_{02} = y_1 + \frac{y_3 - y_1}{x_3 - x_1}(x - x_1), \tag{5}$$

于是插值函数

$$f(x) \approx P_{01} + \frac{P_{02} - P_{01}}{x_3 - x_2}(x - x_2)$$
$$(x_1 \leqslant x \leqslant x_3). \tag{6}$$

那么，

$$y^* = f(x^*). \tag{7}$$

a_3 确定后，由(2)式中第一、二两方程解出

$$a_{21} = \frac{1}{x_1 - x_2} \ln \frac{y_1 - a_3}{y_2 - a_3}, \tag{8}$$

再由一、三两方程解出

$$a_{22} = \frac{1}{x_1 - x_3} \ln \frac{y_1 - a_3}{y_3 - a_3}, \tag{9}$$

最后由二、三两方程解出

$$a_{23} = \frac{1}{x_2 - x_3} \ln \frac{y_2 - a_3}{y_3 - a_3}, \tag{10}$$

从而

$$a_2 = \frac{a_{21} + a_{22} + a_{23}}{3}. \tag{11}$$

再次由方程组(2)、并考虑到(11)式，分别有

$$a_{11} = \frac{y_1 - a_3}{e^{a_2 x_1}}, \tag{12}$$

$$a_{12} = \frac{y_2 - a_3}{e^{a_2 x_2}}, \tag{13}$$

$$a_{13} = \frac{y_3 - a_3}{e^{a_2 x_3}}, \tag{14}$$

从而

$$a_1 = \frac{a_{11} + a_{12} + a_{13}}{3}. \qquad (15)$$

(15), (11)和(3)式分别给出了 a_1, a_2, a_3 的估计值. 由估计值 a_1, a_2, a_3 确定的(1)式称为估计方程. 由估计方程计算得出的 y 值称为 y 的估计值 $\hat{y}$. 实测值 y 与估计值 $\hat{y}$ 的离差平方和

$$Q = \sum_{i=1}^{3} (y_i - \hat{y}_i)^2 \qquad (16)$$

称为剩余离差平方和，是检验估计值与实测值符合程度的一个标准. Q 越小，表明符合程度越高；$Q = 0$，是理想情况. 根据最小二乘法原理，回归方程具有最小的 Q 值，因而由本方法建立的估计方程其 Q 值不小于回归方程，所以它只能作为无法确定回归方程时的一种替代方法. 不难看出，影响 Q 值的因素有二：一是三个实验点对于所求曲线的离散程度. 在已知 x, y 之间有指数统计规律的前提下，重复实验的平均值可以降低三个点的统计起伏，从而缩小这种离散程度，降低 Q 值；二是插值点偏离实际值造成的误差. 需要指出的是，这里的插值与通常数值逼近理论中所讲的插值，尽管方法相同但意义有较大差别. 在数值逼近理论中，是构造插值函数在插值区间上近似代替已知函数，插值点上自变量与因变量的关系是确定的函数关系，因此可以估计截断误差；而在我们这里，"已知函数"恰恰是待求的，并且插值点上自变量与因变量的函数关系具有统计起伏性，因此无法估计截断误差，但插值效果的好坏却能通过 Q 值反映出来. 这种反映又体现在两方面：一是重复实验得到的平均值可以减小三个实验点的统计起伏，有利于提高插值精度，降低 Q 值；二是采用不同的三点插值方法，可以得到不同 Q 值的估计方程，通过比较，取 Q 值最低者为所求. 但是，究竟采用那种插值方法为好，不能一概而论，因为这与插值区间 $[x_1, x_3]$ 和三个实验点的分布有关，应该根据具体情况具体分析.

三、应用实例

地球物理测井的任务之一，是确定井下储集层的含油饱和度. 在均匀含油地层中，由于钻井泥浆的侵入，造成从井壁向地层纵深有一个含油饱和度由小到大的渐变过程. 如以 r 表示地层中的点到井壁的距离，S_o 表示含油饱和度，则 r, S_o 之间有下列统计规律：

$$S_o(r) = S_{o_{max}}(1 - e^{-\beta r}) + S_{o_0} e^{-\beta r}$$

式中 $S_{o_{max}}$ 为 $r \to \infty$ 时的含油饱和度(即未受泥浆侵入影响地层的原始含油饱和度)，S_{o_0} 为 $r \to 0$ 时的含油饱和度，称残余油饱和度. $S_{o_{max}}$, S_{o_0} 和系数 β 都随地层不同而变化. 已知某次测井时测得某均匀含油地层三个不同径向深度处的 $S_o(r)$ 值如下：

r(m)	0.20	0.45	1.50
$S_o(r)$	0.190	0.246	0.322

试确定该含油地层的 $S_{o_{max}}$, S_{o_0} 和 β (目前的双侧向-微球形聚焦组合测井仪只能提供上述三个径向深度处的测量值，且测井费用是以万元为单位的，无特殊情况一般不允许重测).

令 $x = r$, $y = S_o(r)$, $a_1 = S_{o_0} - S_{o_{max}}$, $a_2 = -\beta$, $a_3 = S_{o_{max}}$, 则上述问题等价于估计方程

$$y = a_1 e^{a_2 x} + a_3$$

的参数 a_1, a_2, a_3. 由于这时 $x_1 = 0.20$, $y_1 = 0.190$; $x_2 = 0.45$, $y_2 = 0.246$; $x_3 = 1.50$, $y_3 = 0.322$; 且 $x^* = \frac{x_1 + x_3}{2} = 0.85 (\neq x_2)$, 故由(4)—(7)式，得

$$f(x) \approx 0.1347 + 0.2998x - 0.1167x^2$$
$$(0.20 \leqslant x \leqslant 1.50),$$
$$y^* = f(0.85) = 0.305.$$

再由(3)式，得 $a_3 = 0.325$. 进而由(8)—(11)式，分别有：$a_{21} = -2.143$, $a_{22} = -2.928$, $a_{23} = -3.115$; $a_2 = -2.729$; 由(12)—(15)式，分别有：$a_{11} = -0.233$, $a_{12} = -0.270$, $a_{13} =$

(下转第62页)