插值与数据拟合建模

插值与数据拟合就是通过一些已知数据去确定某类函数的参数或寻找某个近似函数，使所得的函数与已知数据具有较高的精度，并且能够使用数学分析的工具分析数据所反映的对象的性质.

几种常用的方法：

1、一般插值法

2、样条插值法

3、最小二乘曲线

4、曲面的拟合

一、基本概念

1、插值问题：

不知道某一函数f(x)在待定范围[a,b]上的具体表达式，而只能通过实验测量得到该函数在一系列点$a \leqslant x_1, x_2, ..., x_n \leqslant b$上的值$y_0, y_1, y_2, ..., y_n$，需要找一个简单的函数P(x)来近似地代替f(x)，要求满足：

$$P(x_i)=y_i \quad (i=1,2,...,n)$$

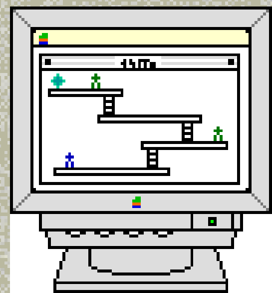
2、插值函数：P(x)，

3、插值法：求插值函数P(x)的方法

二、常用插值函数

1、多项式函数

2、样条函数



1、多项式插值方法

（1）n次代数插值

（2）拉格朗日插值

几点说明：

（1）拉格朗日插值基函数仅与节点有关，而与被插值函数f(x)无关。

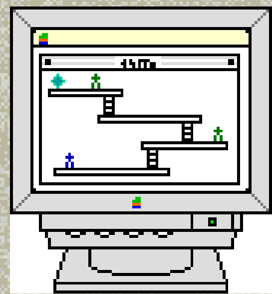
（2）拉格朗日插值多项式仅由数对$(x_i,y_i)(i=1,2,\ldots,n)$确定，而与数对排列次序无关。

（3）多项式插值除了上述插值法外还有其它插值法，如newton插值法、hermite插值法等。

二、常用插值函数

1、多项式函数

2、样条函数



2、样条插值方法

（1）样条函数——m次半截幂函数

$$x_+^0=\begin{cases}1, & x>0;\\ \dfrac{1}{2}, & x=0;\\ 0, & x<0;\end{cases}\qquad x_+^m=\begin{cases}x^m, & x\geqslant 0;\\ 0, & x<0;\end{cases}\qquad (0<m).$$

（2）k次B样条或k次基本样条函数的定义

$$S(x)=\sum_{j=0}^{k}a_jx^j+\frac{1}{k!}\sum_{j=1}^{n-1}c_j(x-x_j)_+^k$$

（一）广泛使用的样条函数

（1）广泛采用：**二次样条、三次样条及B样条**。

（2）力学意义：

A:二次样条在力学上解释为集中力偶作用下的弹性细梁挠度曲线。

B:弹性细梁受集中载荷作用形成的挠度曲线，在小挠度的情况下，恰好表示为三次样条函数，集中载荷的作用点，恰好就是三次样条函数的节点。

（1）二次样条的定义

设[a,b] 的一个划分：$a=x_0<x_1, x_2, ..., x_n=b$，函数f(x)各节点的值分别为：

$f(x_i)=y_i \quad (i=1,2,...,n)$

如果二次样条函数：

$$S(x)=a_0+a_1x+a_2x^2+\sum_{j=1}^{n-1}\frac{c_j(x-x_j)_+^2}{2!}$$

满足：$S(x_i)=y_i \quad (i=1,2,...,n)$

（2）三次样条函数的定义

设[a,b] 的一个划分：$a=x_0<x_1, x_2, \ldots, x_n=b$，

函数f (x)各节点的值分别为：

$f(x_i)=y_i \quad (i=1,2,\ldots,n)$

如果三次样条函数：

$$S(x)=\sum_{j=0}^{3} a_j x^j+\frac{1}{3!}\sum_{j=1}^{N} c_j (x-x_j)_+^3$$

满足：S $(x_i)=y_i \quad (i=1,2,\ldots,n)$

三、曲线拟合最小二乘法

最小二乘法的一般提法是：

对给定的一组数据$(x_i, y_i)$$(i=0，1，2，\ldots，m)$，要求在函数类$\varphi=\{\varphi_1，\varphi_2，\ldots，\varphi_n\}$中找一个函数$y=S^*(x)$，使误差平方和最小：

$$\delta^2=\sum_{j=0}^{m}\delta_j^2=\sum_{j=0}^{m}[S^*(x_j)-y_j]^2=\min_{S(x)\in\varphi}\sum_{j=0}^{m}[S(x_j)-y_j]^2,$$

其中：

$$S(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x),(n<m)$$

例1

在一个SCS系统中，根据实验所得输出信号与时间关系如下表所示，求输出信号y与时间t的拟台曲线y=f (t)。

时间(t)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
输出($y\times10^{-3}$)	4.00	6.40	8.00	8.80	9.22	9.50	9.70	9.86	10.00	10.20	10.32	10.42	10.50	10.55	10.58	10.60

案例1 估计水箱的水流量模型

长度单位：E(＝30．24cm)

容积单位：G(=3.785L(升))

某些镇的用水管理机构需估计公众的用水速度(单位是G/h)和每天总用水量的数据．许多地方没有测量流入或流出水箱流量的设备，而只能测量水箱中的水位(误差不超过5%)．当水箱水位低于某最低水位L时，水泵抽水，灌入水箱，直至水位达到最高水位H为止。但这也无法测量水泵的流量，因此在水泵启动时不易建立水箱中水位和水泵工作时用水量之间关系。水泵一天灌水1~2次，每次约2h。试估计在任意时刻(包括水泵灌水期间)t流出水箱的流量f(t)，并估计一天的总用水量。

案例1数据

表中给出了某镇中某一天的真实用水数据，表中测量时间以s为单位，水位以10^{-2} E为单位．例如3316s以后，水箱中的水深降至31．10 E时，水泵自动启动把水输入水箱；而当水位回升至35.5E时．水泵停止工作。

表 8-2　某镇一天真实用水数据

时间(s)	水位(10^{-2}E)	时间(s)	水位(10^{-2}E)	时间(s)	水位(10^{-2}E)
0	3 175	35 932	泵水	68 535	2 842
3 316	3 110	39 332	泵水	71 854	2 767
6 635	3 054	39 435	3 550	75 021	2 697
10 619	2 994	43 318	3 445	79 254	泵水
13 937	2 947	46 636	3 350	82 649	泵水
17 921	2 892	49 953	3 260	85 968	3 475
21 240	2 850	53 936	3 167	89 953	3 397
25 223	2 795	57 254	3 087	93 270	3 340
28 543	2 752	60 574	3 012		
32 284	2 697	64 554	2 927		

1、模型假设（1）

1）除了问题中特别说明的数据以外，其它给定的数据其测量误差不超过5％.

2）影响水箱流量的唯一因素是该镇公众对水的普通需要．所给的数据反映该镇在通常情况下一天的用水量，不包括任何非常情况．如水箱中水的短缺、水管破裂、自然灾害等.

一个物理定理

1、模型假设（2）

3）水泵的灌水速度为常数，不随时间变化也不是已灌水量的函数，因此假设水泵大约在水位27E时开始灌水，在水位35．5E时停止灌水．同时假设水泵不会损坏或不需要维护．

4）从水箱中流出的最大流速小于水泵的灌水速度．为了满足公众的用水需求不让水棺中的水流尽．

1、模型假设（3）

5）每天的用水量分布都是相似的. 因为公众对水的消耗量是以全天的活动(诸如洗澡、做饭、洗衣服等)为基础的，所以每日用水类型是相似的.

6）水箱的流水速度可用光滑曲线来近似. 每个用户的用水需求量与整个镇的用水需求量相比是微不足道的，而且它与整个镇需求量的增减情况是不相似的.

2、问题分析

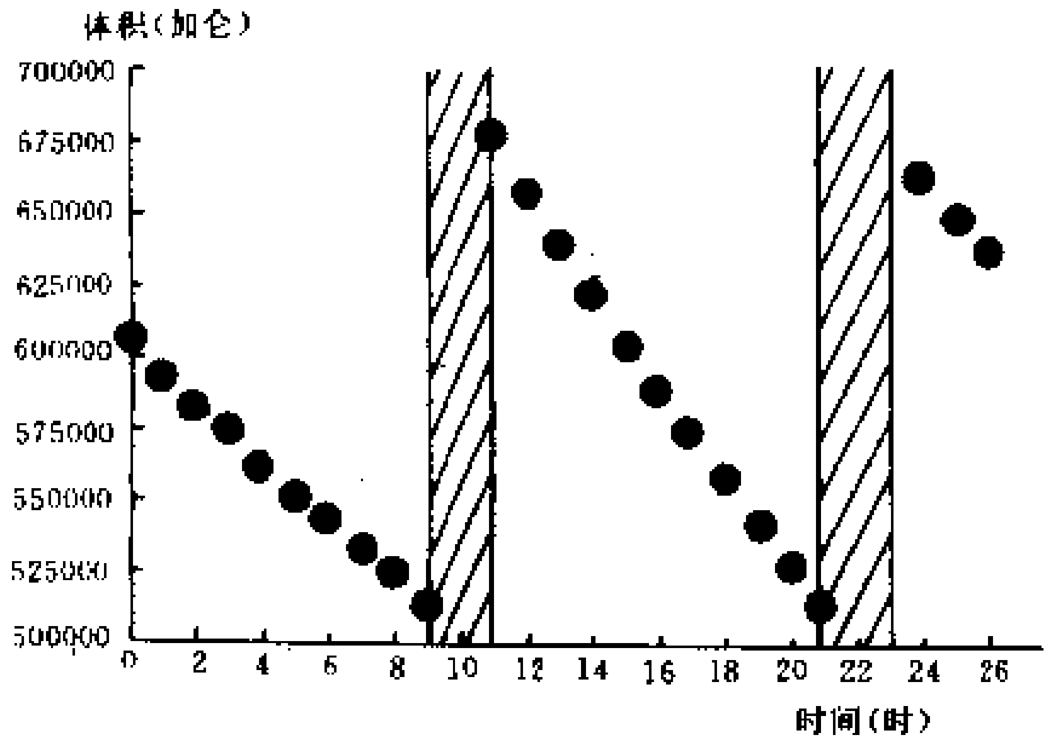


图 8-2　水体积-时间图

3、水流量与时间关系

中心差分公式：

$$f_i := f(t_i) = \left| \frac{-V_{i+2} + 8V_{i+1} - 8V_{i-1} + V_{i-2}}{12(t_{i+1} - t_i)} \right|$$

用来计算每一组的中间数据点的水流量。对每一组数据的前两个和最后两个数据点，采用向前和向后差分公式

4、数值决定的水流量-时间图

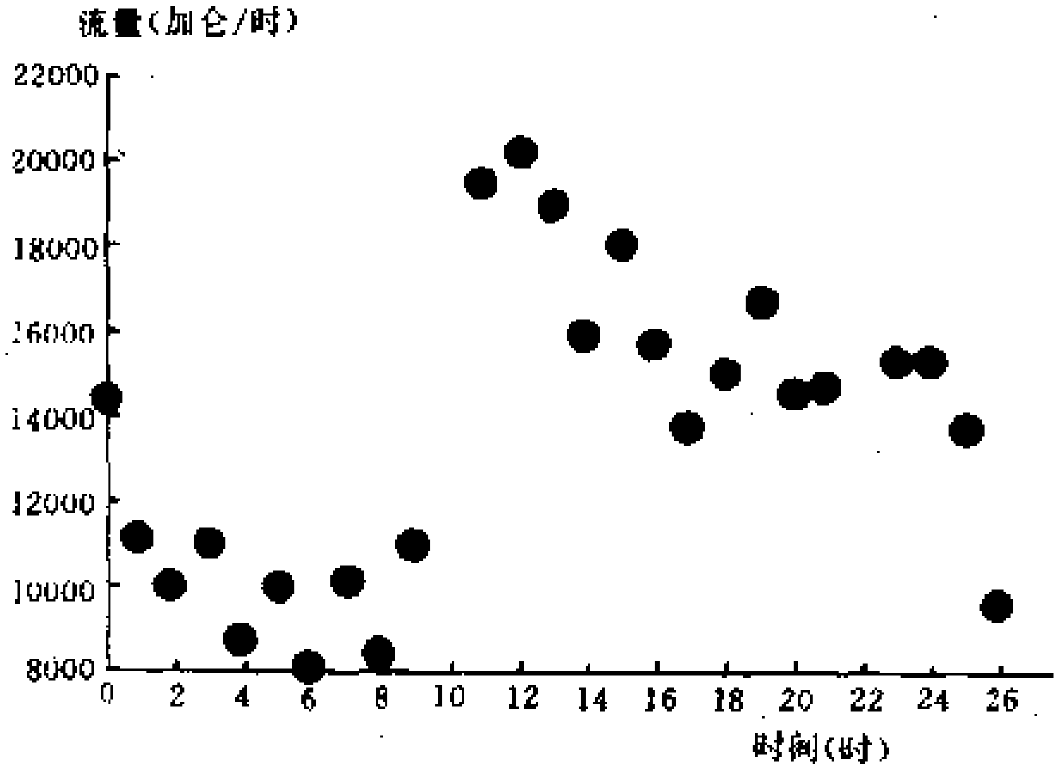


图 8-3 数值决定的水流量-时间图

5、三次样条拟合

$$\begin{cases} a_{0i}=f_i, \\ a_{1i}=\dfrac{f_{i+1}-f_i}{t_{i+1}-t_i}-\dfrac{(t_{i-1}-t_i)(2M_i+M_{i+1})}{6} \\ a_{2i}=\dfrac{1}{2}M_i, \\ a_{3i}=\dfrac{M_{i+1}-M_i}{6(t_{i+1}-t_i)}. \end{cases}$$

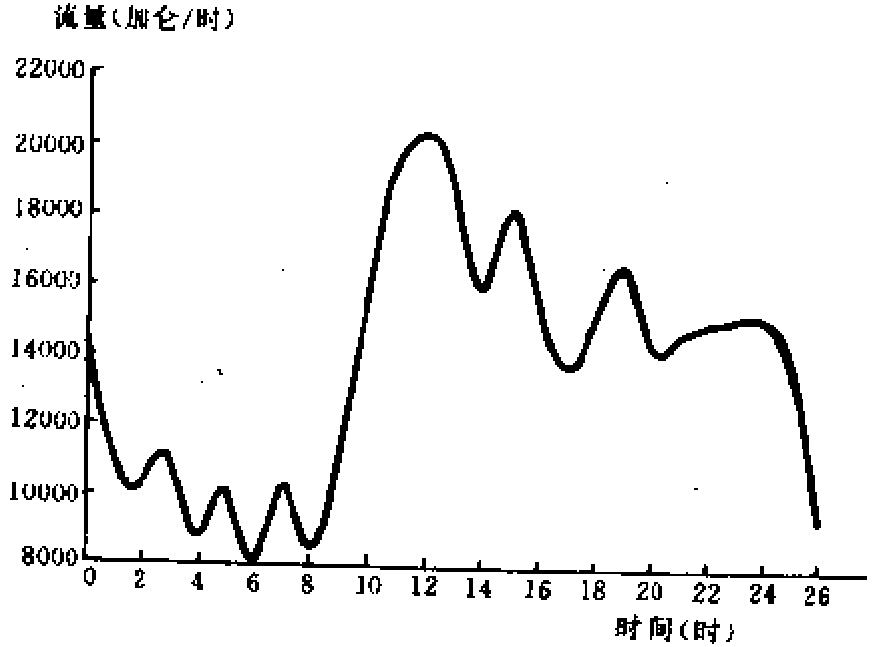


图 8-4 拟合水流量曲线 $f(t)$

6、对水泵两段充水时间的处理

（1）第一次平均水流量：

$$p_1 = \frac{\Delta V_1}{\Delta t_1} + \frac{\int_{8.9678}^{10.9542} f(t)\mathrm{d}t}{\Delta t_1}$$

$$= 81979 + 15597$$

$$= 97576 \text{(加仑/小时)}.$$

（2）第二次平均水流量：

$$p_2 = \frac{\Delta V_2}{\Delta t_2} + \left(\int_{20.8392}^{22.9581} f(t)\mathrm{d}t\right) \Big/ \Delta t_2$$

$$= 76853 + 15057$$

$$= 91910 \text{(加仑/小时)}.$$

（3）平均水流量：

$$p = \frac{1}{2}(p_1 + p_2) = 94743 \text{(加仑/小时)}$$

7、一天总用水量

$$\int_0^{24} f(t)\,\mathrm{d}t \doteq 333189(\text{加仑}).$$

8、检验

以不同的时间为起点得到的一天总用水量相差多少

$$\int_{0.9211}^{24.9211} f(t)\mathrm{d}t \approx 335116(\text{加仑}),$$

$$\int_{1.8431}^{25.8431} f(t)\mathrm{d}t \approx 336782(\text{加仑}),$$

$$\int_{0.5}^{24.5} f(t)\mathrm{d}t \approx 334023(\text{加仑}),$$

$$\int_{1.0}^{25.0} f(t)\mathrm{d}t \approx 335329(\text{加仑}),$$

$$\int_{1.5}^{25.5} f(t)\mathrm{d}t \approx 336480(\text{加仑}),$$

相差只约 1%.

一个物理定理

由托里查里(Torricelli)定律知，从水箱中流出水的最大速度与水位高的平方成正比．对于所给的数据，其水位的最大高度为35．5E，最小高度为27E，因此对两个高度的最大流速比为 $\sqrt{35.5/27}$ ，这个数字已很接近l，所以可以假定水位对流速没有影响．类似地，还假设大气情况、温度变化等对水流速均无直接影响。