

随机误差传递与合成的蒙特卡罗模拟

张大康^①

冀东水泥厂 唐山 063031

摘要 对分析测试过程中随机误差的传递与合成进行了蒙特卡罗模拟,并编制了 BASIC 程序。该方法解决了当被合成误差不相互独立时泰勒级数公式所遇到的困难;当被合成误差相互独立时,与泰勒公式计算结果一致。

关键词 蒙特卡罗模拟 随机误差 BASIC 程序

多数情况下,分析测试结果是一个间接测定的量,它是一个或多个直接测定量的函数。直接测定量的误差依一定规律传递给测定结果,合成后成为测定结果的误差。在各直接测定量的误差是相互独立的随机变量的情形下,资料^[1~3]应用将误差展开为泰勒级数的方法讨论了随机误差传递与合成的计算。当各直接测定量的随机误差不能相互独立时,需知道各直接测定量的相关系数,才能应用该方法。但是求得相关系数相当困难,许多场合根本无法做到,因而也就无法根据资料^[1~3]中的公式处理随机误差的传递与合成。

本文应用蒙特卡罗方法^[4]对随机误差传递过程进行了间接模拟,通过统计方法求出它们在测定结果中的合成。当各直接测定量的随机误差相互独立时,与资料^[1~3]中公式计算结果一致;当各直接测定量的随机误差不相互独立时,亦能得到正确结果。

1 方法原理

设间接测定量 y 是各直接测定量 $x_1, x_2, \dots, x_n$ 的函数:

$$y = f(x_1, x_2, \dots, x_n) \quad (1)$$

间接测定量是随各直接测定量连续变化

的,测定误差(指随机误差,下同)相对于被测量是一个微小量。因此,可将式(1)在各变量的邻近区域内按泰勒级数展开,且作为近似只取误差的一阶项,略去更高阶项。则有:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2 + R + \dots \quad (2)$$

$$R = 2\left(\frac{\partial f}{\partial x_1}\right)\left(\frac{\partial f}{\partial x_2}\right)\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + \dots + 2\left(\frac{\partial f}{\partial x_1}\right)\left(\frac{\partial f}{\partial x_n}\right) \cdot \rho_{x_1x_n}\sigma_{x_1}\sigma_{x_n} + \dots + 2\left(\frac{\partial f}{\partial x_2}\right)\left(\frac{\partial f}{\partial x_n}\right) \cdot \rho_{x_2x_n}\sigma_{x_2}\sigma_{x_n} + \dots \quad (3)$$

$$\text{由于 } \rho_{x_i x_j} = \frac{\sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum (x_i - \bar{x}_i)^2 \sum (x_j - \bar{x}_j)^2}} \quad (4)$$

(4)式是随机变量 x_i 与 x_j 的相关系数。当各直接测定量 $x_1, x_2, \dots, x_n$ 相互独立时 $\rho_{x_i x_j} = 0$, 式(2)中 $R = 0$, 可变为:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2 \quad (5)$$

此为式(2)的一个特例,在计算时可化为更简单的算式^[1,2]。但是,许多分析测试过程中,直接测定量不是相互独立的,其相关系数 $\rho_{x_i x_j}$ 又很难求出,此时上述方法遇到了困难。

应用蒙特卡罗方法,不必知道各直接测定量的相关系数,即可研究误差的传递规律

① 张大康 男,工程师,研究兴趣在数值方法、微机应用技术。

完成测定结果的误差合成。

蒙特卡罗方法是以随机变量按一定概率分布抽样为主要手段,通过统计分析求解实际问题近似解的一种数值方法。应用蒙特卡罗方法求解问题的一般步骤是:

① 建立一个简单而又便于实际应用的概型;

② 确定概型中随机变量的抽样方法;

③ 通过统计的方法求解。

在分析测试过程中,各种随机误差都服从或近似服从于正态分布,可以采用服从于正态分布的随机变量抽样。对于极少数不服从于正态分布的随机误差,需先确定其分布概型再按概型抽样。

对误差按正态分布模拟,需要产生一组服从于正态分布的独立的随机变量 x ,即 $x \sim N(\mu, \sigma^2)$ 。为此,可以在计算机上使用随机函数 $RND(x)$ 产生一组均匀分布在 $(0, 1)$ 上的随机数,再通过下列变换:

$$R = \sqrt{-2 \ln r_1} \cos(2\pi r_2) \quad (6)$$

$$\text{或: } R = \sqrt{-2 \ln r_1} \sin(2\pi r_2) \quad (7)$$

得到一组相互独立的服从标准正态分布的随机变量 R ,即 $R \sim N(0, 1)$ 。其中 r_1 和 r_2 是两个独立的、均匀分布在 $(0, 1)$ 上的随机数。任意均值为 μ ,方差为 σ^2 的正态分布随机变量 x 可通过下列变换得到:

$$x = \mu + \sigma R \quad (8)$$

即 $x \sim N(\mu, \sigma^2)$ 。

在计算机上产生 n 个均匀分布在 $(0, 1)$ 上的随机数,通过(6)~(8)式即可产生 $n/2$ 个服从于正态分布的随机数,再按一定算式运算,求其标准差即为误差合成结果。

2 产生正态分布随机数的子程序

应用蒙特卡罗方法模拟误差的传递与合成,对于不同的分析测试方法合成运算过程各异,很难确定一个通用的计算程序。这里给

出了程序共用的主体部分,即产生正态分布随机变量的子程序,由两段相对独立的程序构成。该程序用 BASIC 语言编写。

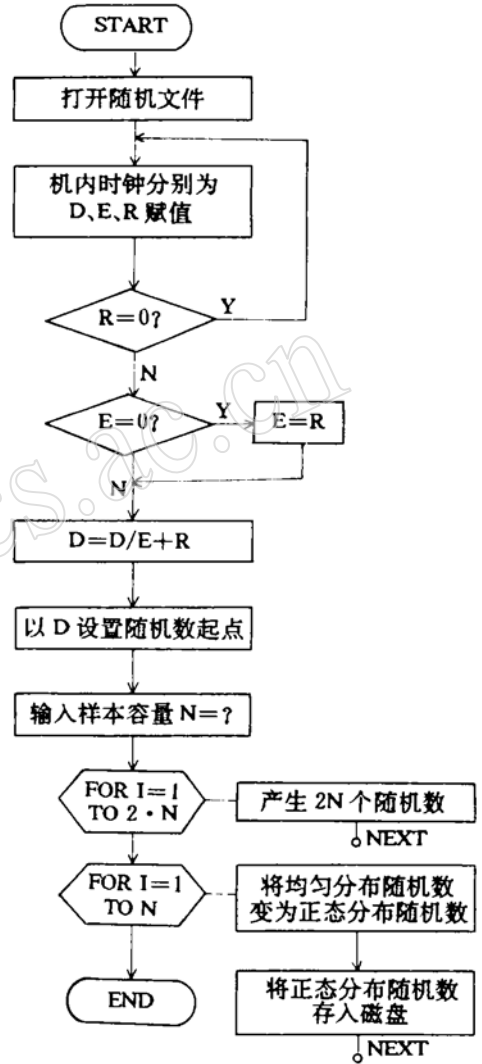


图 1 产生和储存正态分布随机数程序框图

2.1 产生和储存正态分布随机数子程序

具有较多变量的算式进行较大样本容量(如 $n > 1000$)模拟时,数组的定义受到微机内存的限制,本子程序(图 1)利用了外存设备磁盘做为缓冲,其功能是产生足够多的服从正态分布的随机数并存入磁盘。

微机的随机函数 $RND(x)$ 产生的随机数, 实际上是具有固定起点和一定周期的伪随机数, 它使得对同一问题相同样本容量的多次模拟产生相同的计算结果。本子程序采用 RANDOMIZE 语句为随机数设置起点, RANDOMIZE 语句的参数由机内时钟自动生成。将机内时钟的时、分、秒 3 个串变量分别转变为实型变量, 再通过一定的算式将它们合成为一个重复出现概率极小的数值, 作为 RANDOMIZE 语句的参数。运行程序的时

间不同, 产生随机数的起点就不同。这样产生的随机数避免了伪随机数的重现性, 使得对同一误差合成算式的多次模拟运算, 产生不同的相近结果以利参照对比。

2.2 读出正态分布随机数及 X^2 检验子程序

本子程序(图 2)的功能是从磁盘中读出指定样本容量的上一子程序存入的服从正态分布的随机数, 并对其进行 X^2 检验自动剔除偏离正态分布较大的样本, 为模拟计算提供比较严格地服从于正态分布的随机数样本。

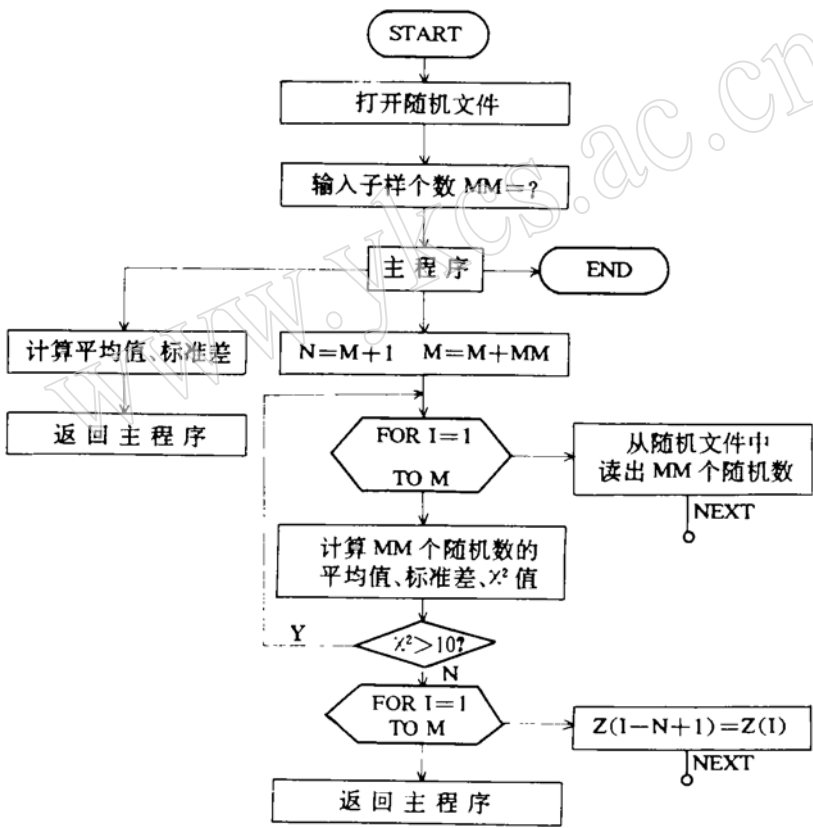


图 2 读出正态分布随机数并进行 X^2 检验程序框图

3 结果与讨论

由子程序 1 产生的 3 组不同样本容量的正态分布随机数的 X^2 值见表 1。

表 1 数据表明, 随样本容量的增大, X^2 值呈减小趋势, 且出现较大值(如 $X^2 > 10$)的

概率减小。参考表 1 数据, 根据所模拟问题的精度要求, 可以选择出被剔除正态分布随机数样本的 X^2 临界值。

对下式的误差传递与合成进行了蒙特卡罗模拟, 结果见表 2。

$$\frac{(4.10 \pm 0.02) \times (0.0050 \pm 0.0001)}{1.97 \pm 0.04} = 0.0104 \pm 0.000302 \quad (9)$$

表 1 不同样本容量正态分布随机数 X^2 值

样本容量		正态分布随机数的 X^2 值									
100	5.42	1.84	4.18	7.47	8.40	8.79	15.17	9.40	18.78	6.32	
	5.35	11.42	13.90	4.45	1.73	16.14	9.45	5.31	1.55	6.22	
500	2.24	7.94	8.02	7.16	11.78	4.94	9.72	5.41	5.44	10.85	
	8.46	8.94	14.92	7.70	2.96	3.44	7.28	7.70	8.16	6.32	
1000	5.61	10.21	4.86	1.13	7.20	8.23	8.71	8.75	10.61	4.85	
	7.82	5.00	4.94	9.45	2.43	3.20	9.12	3.00	6.45	10.19	

表 2 误差传递与合成的模拟结果^①

样本容量	模拟结果			
	平均值	标准差	平均值	标准差
100	0.01042	0.00031	0.01037	0.00030
	0.01038	0.00032	0.01042	0.00031
	0.01044	0.00030	0.01036	0.00028
	0.01045	0.00030	0.01040	0.00028
	0.01046	0.00031	0.01041	0.00029
	0.01040	0.00027	0.01036	0.00029
	0.01042	0.00026	0.01040	0.00027
500	0.01041	0.00031	0.01040	0.00028
	0.01041	0.00031	0.01041	0.00030
	0.01040	0.00030	0.01041	0.00030
	0.01041	0.00031	0.01040	0.00030
	0.01040	0.00029	0.01042	0.00030
	0.01043	0.00031	0.01041	0.00031
	0.01040	0.00031	0.01040	0.00029
1000	0.01040	0.00029	0.01040	0.00030
	0.01041	0.00031	0.01041	0.00031
	0.01042	0.00030	0.01042	0.00030
	0.01041	0.00031	0.01041	0.00030
	0.01042	0.00029	0.01040	0.00031
	0.01041	0.00030	0.01040	0.00030
	0.01041	0.00030	0.01041	0.00029

① 本表数据模拟时,正态分布随机数样本未进行 X^2 检验和剔除,以便观察样本容量对模拟结果的影响。

表 2 数据表明,模拟结果与资料^[1,2]中按泰勒级数展开的算式计算结果十分接近,是非常理想的估计值。对资料^[1]中的 4 个例题进行了同样的模拟,均与资料^[1]的结果吻合(数据从略)。表 2 中的标准差实际上是或然误差,按绝对值来说比或然误差大或小的误差出现的概率相等。

资料^[5]成功地研究了 ISE 多次标准加入法误差传递与合成的蒙特卡罗模拟。在溶液条件恒定时,多次标准加入法的能斯特方程为:

$$E_i = K + S \ln \frac{C_x V_x \sum_{k=1}^i C_s V_k}{V_k + \sum_{k=1}^i V_k} \quad (10)$$

在一个测量系统中,待求量由 n 个参数决定,如式(1)。只要明确知道式(1)的函数解析式,如式(9)、式(10),即可由蒙特卡罗模拟求出待求量 y 的误差估计值,无论参量 $x_1, x_2, \dots, x_n$ 之间是否相互独立,亦可单一或同时改变参量 $x_1, x_2, \dots, x_n$ 以观察各参量单独或综合作用对待求量的影响,确认各测定条件是否合理,选择出最佳测定条件。这种情况在实际的测量系统中经常遇到,如在 AAS、AES 分析中,为了消除基体效应可将待测溶液稀释,但这样同时需要增大仪器对电信号

的放大倍数,致使信噪比降低。究竟怎样确定稀释比和放大倍数以保证测定结果误差最小,只要确定出不同水平参量的标准差,此类问题即可由蒙特卡罗模拟得到解决。

应用本方法研究了 AES 法测定水泥中 K_2O 、 Na_2O 主要测定条件对结果的影响,验证了资料^[6]方法的测定条件。AES 法测定水泥中 K_2O 、 Na_2O 的主要误差来源是水泥中 Ca^{2+} 、 SiO_3^{2-} 、 Al^{3+} 、 Fe^{3+} 、 Mg^{2+} 等杂质离子和试样前处理过程中引入的 NH_4^+ 、 SO_4^{2-} 、 Cl^- 等离子所构成的基体效应。这里仅考虑基体效应和信噪比的影响,而对其它影响因素忽略。首先以实验的方法确定各实验条件的标准差并以相对标准偏差表示。参考资料^[6]的测定条件,以 K_2O 为例确定以下模拟参数: $C_1=5$; $C_2=10$; $C_3=20$; $C_4=40$; $C_5=60$; $C_6=80$, C 是待测离子的浓度。

固定信噪比相对标准偏差,得到基体效应对测定结果的影响示于图 3 中 1。

固定基体效应相对标准偏差,得到信噪比对测定结果影响示于图 3 中 2。

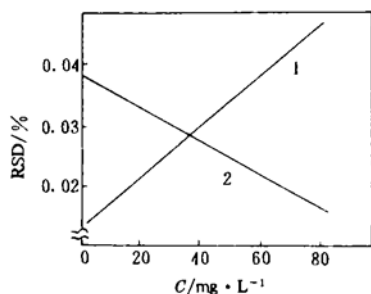


图 3 测定条件对结果的影响

- 1—基体效应对结果的影响;
2—信噪比对结果的影响。

综合图 3 的结果确定的最佳工作条件是待测离子 $[K^+]$ 稀释为 15~20mg/L 为宜,与

资料^[6]方法的条件一致。

现代仪器分析迅速发展,随着微机的普及各种数学方法大量地应用于仪器分析的数据处理,显著提高了仪器分析的信噪比、灵敏度和应用范围,使我们有可能从仪器分析的测量信号中获得更多、更准确的信息。同时,数学方法的广泛使用也使仪器分析的计算公式呈现愈来愈复杂的发展趋势。如有关文献^[7]探讨了激光显微发射光谱定量分析的基本公式:

$$I_i = \frac{W_i^b}{u_i^b} a_i C_i^b$$

这些复杂的算式,很少有各参量之间相互独立的情形,有时甚至根本无法表达为待求量的显函数。此时,用蒙特卡罗模拟处理随机误差的传递与合成是一种成功的方法。

4 参考文献

- 1 宋清. 定量分析中的误差和数据评价. 北京:高等教育出版社,1982. 58.
- 2 邓勃. 数理统计方法在分析测试中的应用. 北京:化学工业出版社,1984. 269.
- 3 赵新那等. 数值方法在分析化学中的应用. 长沙:中南工业大学出版社,1987. 9.
- 4 徐钟济. 蒙特卡罗方法. 上海:上海科学技术出版社,1985.
- 5 李利民,赵新那. ISE 多次标准加入法误差传播的 Monte Carlo 模拟. 计算机与应用化学. 1984,1(3): 217.
- 6 建筑材料科学研究院. 水泥化学分析. 北京:中国建筑工业出版社,1982. 515.
- 7 王昭宏,朱祝彪. 激光显微发射光谱定量分析基本公式的探讨. 岩矿测试. 1990,9(2):149.

(收稿日期:1993-02-25,修回日期:1993-9-11)

(英文下转第 160 页)

~~~~~  
(上接第 158 页)

## Propagation and Combination of Random Errors and Monte Carlo Simulation

*Zhang Dakang*

(Jidong Cement Plant, Tangshan, Hebei, 063031)

A BASIC programme has been compiled for the Monte Carlo simulation as applied to the propagation and combination of random errors occurred in the testing. Using this method, one is able to overcome the difficulties involved in the application of Taylor series in case where the combined errors are not mutually independent. When the combined errors are independent the results are consistent with the Taylor series.

**Key words:** Monte Carlo simulation, random errors, BASIC programme