

# 在高考数据处理中使用蒙特卡洛方法

广西计算中心 罗海鹏 符华儿

## 提 要

本文叙述蒙特卡洛方法在一次高考数据处理中的应用情况。

## 一、问题的提出

1981年5月,我们用一台微型计算机对南宁市高考预选成绩进行了统计处理。高考数据处理是一项细致的工作,一个数据错误都可能影响某个考生的录取,甚至影响整个统计工作的进行。此外,这次高考统计的数据量较大,共一万多考生,近十万个数据;而所允许的统计时间较短,总共四天左右。整个工作,时间仓促,回旋余地少;而关系却很重大,一旦统计失败,影响千家万户,后果不堪设想。当时,在国内还未见有用微型计算机统计高考分数的报导,没有什么现成的经验可参考。而在前一两年,国内有几个比较先进的地区,在用大、中型计算机统计高考分数时,出了一些问题。我们也是第一次做这项工作,没有什么实践经验,从技术上说,有许多东西还无把握。比如,一个磁盘至多能建立多少个文件?在建立了许多文件档的情况下,一个磁盘能装多少数据?由于同时要做许多项目的计算和统计,而微型计算机内存较小,实际运算时内存会不会溢出?大规模地排序,在具体实现时,在存储容量上、运行时间上,会遇到什么问题?鉴于以上种种,为了这次统计工作的万无一失,非得做大规模的模拟试验才行。但当时统计的时间已很紧迫,需要试验的数据量要等于或大于实际高考数据量,靠手工输入数据做这项模拟试验已不可能。因此,我们使用了蒙特卡洛方法,用计算机产生的随机数来模拟高考数据,进行各项试验,取得了各科第一手材料,顺利地解决了上述问题。

## 二、试验设计

蒙特卡洛(Monte Carlo)方法也叫统计试验方法,它是随着电子计算机的产生而发展起来的一种数值方法。蒙特卡洛方法的基本思想在于,根据问题的特征和数值的规律性,利用计算机作大量的随机试验,将试验的结果进行加工,从而得到问题的解决。(1)

试验规模以一万考生为例,每个考生有一个准考证号码,七个单科成绩。随机数发生器产生的数是在0—1区间均匀分布的,把产生出来的数乘以一万,取整,即得到一个准考证号码(0—10000之间),乘以一百,取整,即得到一项单科成绩(0—100之间)。但对于一个考生来说,如果各科成绩单独产生,常常会出现单科成绩分数相差太大,比如可能有数学95分,政治3分,这往往不符合实际情况。故我们对于每一个考生,首先产生一个基本分

数，再产生一个浮动分数，把这个浮动分数，加在基本分数上或从基本分数上减去它，以这个值作为这个考生的某项单科的成绩。这样，产生一个考生的所有数据，使用随机数发生器共十六次。下面是使用随机数发生器的程序框图。

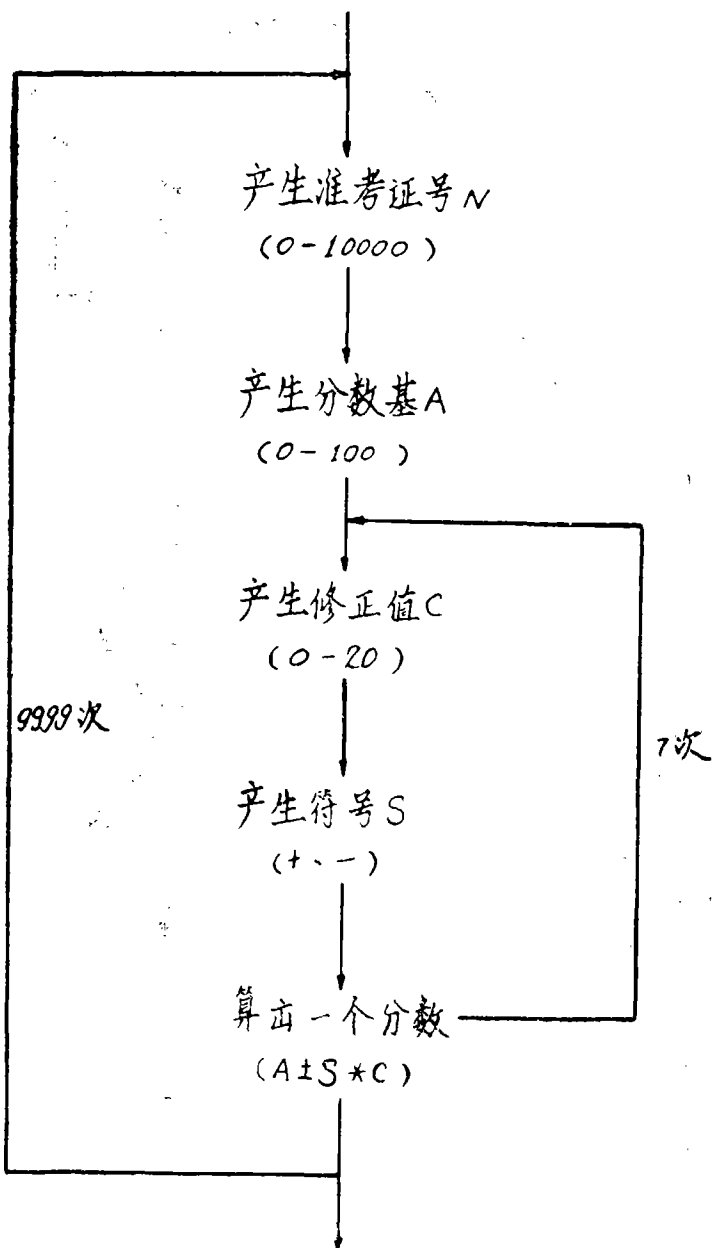
在半天左右的时间里，计算机即可自动产生这一万组数据。下一步就是做各项检验了。我们用随机数发生器产生的数据进行了58个文件档的磁盘容量检验，内存容量检验，打印成绩单检验，排序检验和打印机调试等。由于使用了蒙特卡洛方法，争取了时间，及早发现了一些问题，使我们能够在程序设计上搞得更细致，堵塞了许多漏洞。

### 三、事后的考虑

1. 由于随机数发生器产生的0—1之间的值不会绝对均匀，当我们取小数点后4位的值乘以10000时，所产生的一万个准考证号码可能会有重复。实际上这一项数据本来可以不用随机数发生器产生，准考证号按照自然顺序，用一个计数器1、2、3、...地编排下去就可以了。

2. 试验中所产生的各科成绩和实际情况相比较，也还是有许多不同的。比如，较多考生的成绩集中在中间偏低的水平，符合一定的偏态分布规律，而我们试验产生的数据在10—90之间是均匀分布的，又各科总平均分数可能是有一定的差距的，而试验

产生的数据在总平均分数上是一样的；等等。所有这些，我们都还可以根据随机数发生器的特点来做更仔细的安排，使试验产生的数据与实际情况更接近。但对于高考分数处理的试算，也许已不必要做更细致的模拟了。



### 参 考 书

(1) 雷光耀，蒙特卡洛方法简介，全国第三届计算物理会议讲义，1982.11。