基于蒙特卡罗方法的试卷难度分布研究

姚　进　段会川　寿思诚　刘　弘
(山东师范大学信息管理学院　济南 250014)

摘　要　本文根据教育测量学理论对试题难度和试卷难度分布问题进行了概率学研究 ,提出了基于蒙特卡罗方法的试卷难度分布求解思路并进行了实现。文中针对常模参照性考试详细讨论了正态分布的适用范围 ,并讨论了蒙特卡罗方法的收敛性 ,同时给出了该方法的计算结果。研究表明 ,蒙特卡罗方法在解决试卷难度分布问题方面具有较好的优势。

关键词　蒙特卡罗方法　试卷难度分布　常模参照性考试　正态分布

RESEARCH ON THE DIFFICULTY DISTRIBUTION OF TEST PAPERS BASED ON THE MONTE-CARLO METHOD

Yao Jin　Duan Huichuan　Shou Sicheng　Liu Hong
(School of Information Management , Shandong Normal University , Jinan 250014)

Abstract　This paper discusses particularly the degree of difficulty of test questions and the difficulty distribution of test papers in the light of the educational measurement methodology. The thought of Monte-Carlo method is proposed and realized to solve the difficulty distribution of test papers. The applicable scope ,especially for the criterion-related norm-referenced test ,of normal distribution is analyzed in detail. The astringency of the Monte-Carlo method is demonstrated ,and the calculation result is listed. The paper shows that the Monte-Carlo method has obvious advantage in solving the difficulty distribution of test papers.

Keywords　Monte-Carlo method　Difficulty distribution of test papers　Criterion-related norm-referenced test　Normal distribution

1　引　言

组卷智能化是考试系统中的一个重点和难点问题 ,确定试卷中的试题难度分布是其中的一项重要任务。

试卷难度分布是由试卷平均难度的期望值及分散度(即标准差)所决定的 ,考试系统应选择适宜的数学模型计算试题难度分布。随着 PC 计算机性能和功能的不断增强 ,蒙特卡罗这种古老但是有效的方法在实际应用中越来越能够扬其长处 ,本文根据教育测量学理论对试题难度及其在试卷中的分布进行了详细的概率学研究 ,古为今用地提出了用蒙特卡罗方法确定试题难度分布的思路 ,并以正态分布函数作为难度的分布模型进行了实现。文章对正态分布模型在常模参照性考试中的适用性进行了详尽的分析和讨论 ,较为充分地说明了正态分布的有效范围是难度值区间[0.3 ,0.7]。分析和研究表明蒙特卡罗方法特别适宜于解决试卷难度分布问题。

2　试题难度及试卷的难度分布

2.1 教育测量学中的难度理论

试题难度是描述试题对学生知识和能力水平的适合程度的一个重要指标 ,它具有相对意义 ,难度的高低与被试群体的整体水平有关。

在教育测量学中 ,试题难度有基本计算公式和极端分组计算公式两种计算方法。用 P 表示难度指标 ,则客观题的难度基本计算公式为:

$$P = \frac{R}{N} \tag{1}$$

其中 ,N 表示参加考试的总人数 ,R 表示答对该题的人数。主观题的难度基本计算公式为:

$$P = \frac{\overline{X}}{K} \tag{2}$$

其中 ,$\overline{X}$ 表示所有考生在该题上的平均得分 ,K 表示该题的满分值。

当考试者数目很大时 ,通常采用极端分组法计算试题难度。即将考试成绩由高到低排序 ,并按照相同比例确定高分组和低分组 ,分别用基本公式计算高分组难度 P_H 和低分组难度 P_L ,则客观题的难度计算公式如下:

$$P = \frac{P_H + P_L}{2} \tag{3}$$

对于主观题 ,可设相同的高分组和低分组人数 n ,并分别计算高分组和低分组的得分总和 X_H、X_L ,以 H、L 分别表示该题目的最高和最低得分 ,则难度计算公式可表示为:

$$P = \frac{X_H + X_L - 2nL}{2n(H - L)} \tag{4}$$

可以看出 ,上面讨论的试题难度实际上是试题的容易度 ,即难度值越大试题越容易。为了与习惯保持一致 ,可对试题难度公式进行变换 ,如公式(1)可变换为:

收稿日期:2004 - 02 - 24。姚进 ,硕士生 ,主研领域:计算机应用技术。

$$P = 1 - \frac{R}{N} \tag{5}$$

后文所述的试题难度都将是公式(5)意义上的试题难度,即难度值高的试题难度大,难度值低的试题难度小。

一般地,试题难度值为[0,1]区间上的实数,但习惯上常将试题难度划分为表1所示的5个等级。

表1　试题难度等级

级别	A(易)	B(较易)	C(中等)	D(较难)	E(难)
区间	[0,0.2]	[0.2,0.4]	[0.4,0.6]	[0.6,0.8]	[0.8,1]

2.2 试卷中的难度分布

一份试卷是由难度不同的各种试题组成的,试卷的合理性决定于各种不同难度试题所占比例的合理性,即难度分布的合理性。

试卷的平均难度无疑是衡量试卷难度的最重要指标之一,它与考生成绩密切相关。用 $\overline{D}$ 表示试卷的平均难度值,$\overline{S}$ 表示全体考生成绩的平均值,则它们之间的关系可表示为:

$$\overline{D} = 1 - \frac{\overline{S}}{S} \tag{6}$$

其中,S 表示试卷的满分值。

可见,(6)式定义的试卷难度的意义与(5)式定义的试题难度的意义相同,都在区间[0,1]上取值,且取值越大,难度越高。

试卷中试题难度的分散程度是描述试卷难度的另一个重要指标,它与考试成绩的分散程度密切相关。根据大量样本试卷的分析,对于颇具代表性的常模参照性考试,考试成绩应该呈正态分布,这类考试试卷中大多数题目的难度在0.3～0.7之间。对于竞赛类型的考试,高难度试题比重应比较大一些,而过关类型的考试,低难度的试题应占较大的比重。后两种考试的考试成绩分布将具有一定的不对称性,即对正态分布有一定的偏离,常称为偏态分布。

本文选择最具普遍性的常模参照性考试试卷难度分布问题进行讨论。根据上文的叙述,此类考试的成绩分布应该服从正态分布,其分布密度函数描述为:

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in (-\infty, +\infty) \tag{7}$$

其中,随机变量 x 代表考生的成绩,期望 μ 即是考生成绩的平均分 $\overline{S}$,它决定了分布曲线的峰值位置,标准差 σ 表示考生成绩的分散程度,它决定了分布曲线的形态。

从公式(6)可知,试卷平均难度 $\overline{D}$ 是考生成绩平均值 $\overline{S}$ 的线性函数。根据概率论中“随机变量函数的分布”定理,试题难度 D 也应是一个服从正态分布的随机变量,难度均值 $\overline{D}$ 即是该分布的数学期望。

正态分布函数的定义域是[$-\infty, +\infty$],而难度取值范围是在有界的区间[0,1]上,因此,要使正态分布有效地描述试题难度问题,必须使其分布曲线在区间[0,1]上“有效”。

确定“有效”需要有科学的原则,这种原则应该根据试题组卷问题自身的特点来确定。一般来说,试题的最小分值是1分,我们可以认为,当正态分布曲线在[0,1]区间外的面积小于0.5%(4舍5入后为零),而区间[0,1]内的面积大于99.5%(4舍5入后为100%)时,用它描述试题难度分布是“有效”的。

正态分布的曲线形态是由标准差 σ 决定的。σ 值的大小体现了随机变量在其期望 μ 附近的聚散程度,σ 越小,随机变量取值越集中,σ 越大,随机变量取值越分散。上面确定的99.5%的“有效”性原则,实际就是对 σ 最大值 σ_{max} 的限制。我们对[0,1]区间上各期望值下的 σ_{max} 进行了计算,结果如图1所示。

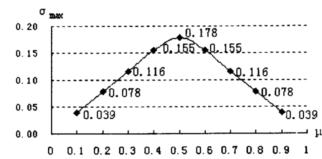

图1　常模参照性考试下难度正态分布的最大标准差

一份理想试卷中的试题应该在满足正态分布的前提下,具有尽量宽的覆盖面。我们用图1中的数据绘制了如图2所示的不同 μ 下的正态分布曲线 $N(\mu, \sigma_{max}^2)$(从对称性方面考虑,我们只给出了期望值高于0.5的曲线)。从图2可以看出,当期望值在[0.3,0.7]之外时,试卷中试题的覆盖面将不足题库试题的50%,我们认为这时的正态分布已经很难给出合理的试卷难度分布了。而[0.3,0.7]的期望值区间正是前面所提到的常模参照性考试的平均难度值区间,这也从另一个方面说明我们分析的合理性。

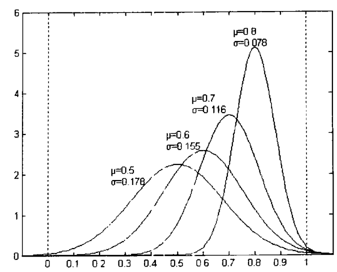

图2　不同期望与标准差下的正态分布曲线

2.1节最后曾经指出,试题难度通常分为5个等级,为不失一般性,我们接下来讨论的试卷难度分布问题即是指在此5个等级上的难度分布问题。

3　基于蒙特卡罗方法的试卷难度分布计算

3.1 蒙特卡罗方法

蒙特卡罗方法,又称为随机抽样方法或统计试验方法,是一种基于“随机数”的计算方法。其基本思想是当所要求解的问题是某种事件出现的概率,或者是某个随机变量的数学期望时,可以通过“抽样试验”的方法,得到这种事件出现的频率或者这个随机变量的平均值,并近似地将其作为问题的解。

用蒙特卡罗方法求解问题的过程包括以下三个主要步骤:

(1)构造或描述一个与求解问题有关的概率数学模型;

(2)依据该模型进行抽样,即产生随机变量的模拟值;

(3)建立各种估计量或求解事件发生的频率并将其作为问题的近似解。

通过前面的分析,我们已经确定难度随机变量服从正态分布,即蒙特卡罗方法的第1步已经完成。

3.2 正态分布抽样算法

蒙特卡罗方法的关键是其第2步,即建立基于第1步中得到的概率数学模型的抽样算法。对于本文所讨论的问题,即是建立正态分布 $N(\mu, \sigma^2)$ 的随机数发生器。

蒙特卡罗模拟通常以均匀分布的随机数发生器为基础,通过某种变换得到需要的随机数分布。对于正态分布,采用比较经典的博克斯-马勒(Box-Muller)变换,其算法如下:

1. 用均匀分布 $U(0,1)$ 的随机数发生器产生随机数 $U1$、$U2$。

2. 对 $U1$、$U2$ 施行博克斯-马勒变换:

$$N_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2)$$

$$N_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

则 $N1$、$N2$ 便是服从标准正态分布 $N(0,1)$ 的随机数。

3. 计算 $X_i = \mu + \sigma N_i$,则 X_i 即是满足正态分布$N(\mu,\sigma^2)$的随机数。

3.3 试卷难度分布的计算

基于蒙特卡罗方法求解该问题的依据是贝努里大数定律[1]的拓展应用,即对于有 m 种事件的随机变量 X 进行 n 次抽样(n 称为样本量),对其第 $i(i=1,\cdots,m)$个事件 A_i 的出现次数进行计数,假设频数值为 d_i,则由 $P^*(A_i)=\frac{d_i}{n}$式可计算第 i 事件出现的频率。对于足够大的 n,$P^*(A_i)$将依概率收敛于第 i 个事件的理论概率 $P(A_i)$。

我们将在难度等级 i 上选择试题看作一个随机事件 A_i($i=1,2,3,4,5$),将难度值 j(j 在$[0,1]$上取值)在某个难度等级区间 i 范围内的试题被组入试卷看作是随机事件 $A_{i,j}$,则 A_i 事件发生的概率可认为是 i 难度等级上应该分配的试题比例,它应该是该范围内所有 $A_{i,j}$ 事件发生的概率的和。这样,便将试题难度分布求解问题转化成为各难度等级上抽题事件发生的概率的计算问题。进行了正态分布抽样之后,第三步即求解 A_i 事件发生的频率并将之作为概率 $P(A_i)$的近似值。

4　算法收敛性及结果与分析

4.1 算法收敛性

蒙特卡罗方法是一种基于概率的方法,该方法计算结果的收敛性与效率是人们非常关心的。

为了分析该算法的收敛性,我们进行了大量的计算。我们选取难度期望值 0.5,标准差 0.178 为典型的组卷要求,分别以 128 到 4096 之间 6 种不同规模的样本量抽取随机数,对每一种规模样本量进行 2000 次抽样,计算所得 2000 个平均难度值的标准差。我们发现,在样本量达到 512 时,标准差已经小于 0.01;样本量超过 1000 时,平均难度即已达到较理想的收敛结果。图 3 给出了难度期望值 0.5、标准差 0.178 时计算各样本量下平均难度的标准差的变化情况。其他难度期望值下的结果与此相似。

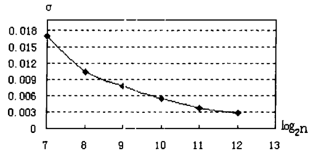

图 3　正态分布抽样算法的收敛情况

从图 3 可以看出,当样本量为 2048 时,可得到相当可信的计算结果,这也说明该方法的效率是比较高的。

4.2 计算结果及其分析

根据上面的收敛性分析,我们取样本量为 2000,对难度期望和标准差分别取图 1 所示的数据即(0.5,0.178)、(0.6,0.155)、(0.7,0.116)、(0.8,0.078)作为组卷参数进行了试卷难度分布计算,所得结果示于表 2(考虑对称性,期望值低于 0.5 的情况与此类似)。

从表 2 可以看出,取样本量为 2000 时,所得平均难度与期望难度的偏差不超过 2 ‰,这足以说明该方法的可用性。

表 2　使用蒙特卡罗方法计算的难度分布

组卷参数		计算结果(%)					
难度期望	标准差	A(易)	B(较易)	C(中等)	D(较难)	E(难)	平均难度
0.5	0.178	4	25	43	22	6	0.499
0.6	0.155	0	10	40	40	10	0.599
0.7	0.116	0	1	20	60	19	0.698
0.8	0.078	0	0	1	50	49	0.798

表 2 以数字形式表明了随着难度期望与 0.5 距离的增大,试题向少数难度等级集中的趋势。当难度期望为 0.7 时,难度分布集中在 3 个难度等级上,而当难度期望为 0.8 时,难度分布仅集中在 2 个难度等级上。这再一次说明,正态分布对于[0.3,0.7]范围内的难度期望是合理的,而在该范围之外,试题过分地集中于狭窄的难度等级上,因而其适用性值得怀疑。

采用蒙特卡罗方法计算得到的表 2 与使用积分法对正态分布函数的计算得到的结果基本一致,而且我们可以期望对于足够大的样本量,蒙特卡罗方法的结果应该收敛于积分结果。但是,作者认为,使用蒙特卡罗方法解决试卷难度分布问题较积分法更为合理而且灵活。

首先,试卷难度分布与组卷是一个概率学问题,因为试题库各个试题的难度具有一定的不确定性,它与命题专家考虑问题的角度和确定难度所针对的被试群体有很大的关系。其次,不同次考试的被考对象的水平存在不同程度上的不确定性。再次。目前越来越普遍的基于计算机网络的考试,对组卷的随机性和被考对象试卷的区别度提出了较高的要求。

蒙特卡罗方法是一个概率学方法,用它解决试卷难度分布这一概率学问题是非常合适的。通过调整蒙特卡罗方法中的参数,可以容易地实现随机性好和区分度高的组卷。例如,在基于网络的考试中,可通过设置σ值具有一定的扰动区间,或通过设置低于 2000(因而提高难度分散度)的抽样样本量,均可增大不同考生试卷的差异性。

5　结束语

本文就考试系统组卷智能化的一个重要方面 ——试卷难度分布问题进行了详细的概率学讨论,并且运用蒙特卡罗方法实现了试卷难度分布的计算。论文从理论和计算两个方面说明了正态分布对常模参照性考试的适用性。分析和研究表明,基于概率学的蒙特卡罗方法对解决试题难度分布这一概率相关的问题具有明显的优势。本文所述的方法已经在山东省中小学信息技术考试系统中得到了应用,并取得了理想的效果。

参　考　文　献

[1] 魏宗舒,概率论与数理统计教程[M],北京:高等教育出版社,1983.

[2] 《现代数学手册》编纂委员会,随机数学卷[M],武汉:华中科技大学出版社,1999.

[3] 胡中锋、李方,教育测量与评价[M],广东:广东高等教育出版社,1999.

[4] 毛秉毅,"一种计算试卷中试题难度分布的有效方法"[J],《计算机工程》,2002,28(6):280～281.