CHOUYANG DIAOCHA LILUN YU JISHU YINGYONG

抽样调查理论与技术应用

李晓风　马江洪　编著

人民交通出版社

Chouyang Diaocha Lilun yu Jishu Yingyong

抽样调查理论与技术应用

李纪治　李晓风　马江洪　编著

人民交通出版社

内 容 提 要

本书介绍抽样调查理论与技术应用方法，主要内容包括：抽样调查概论，简单随机抽样，分层随机抽样，比率估计量(法)，回归估计量(法)，等距抽样，不等概抽样，整群抽样，阶段抽样，汽车运输抽样调查方法等。

本书主要内容分别于1991年和1996年获交通部科技进步三等奖。

本书可供从事公路运输管理的技术人员、大专院校有关专业师生使用参考。

图书在版编目(CIP)数据

抽样调查理论与技术应用/李纪治等编著. —北京：
人民交通出版社，1997.9
ISBN 7-114-02817-2
Ⅰ.抽… Ⅱ.李… Ⅲ.抽样理论 Ⅳ.0212.2
中国版本图书馆CIP数据核字(97)第22660号

抽样调查理论与技术应用

李纪治 李晓风 马江洪 编著
版式设计：刘晓方 责任校对：梁秀青 责任印制：张 凯
人民交通出版社出版发行
(100013 北京和平里东街10号)
各地新华书店经销
北京牛山世兴印刷厂印刷
开本：787×1092 1/16 印张：9.25 字数：236千
1997年12月 第1版
1997年12月 第1版 第1次印刷
印数：0001—3 000册 定价：15.00元
ISBN 7-114-02817-2
U·02008

序

改革开放以来，我国公路运输业得到飞速发展。到1996年底，全国公路总里程已达118万公里，其中高速公路3 422公里，民用汽车保有量超过1 100万辆，是建国以来公路交通运输发展最快、最好的时期。公路运输的快速发展，较大程度地缓和了我国交通运输的“瓶颈”状态，强有力地支持了国民经济健康、快速、持续发展，得到全国各界的关注和好评。这些成绩的取得，是与科学合理地进行公路交通运输网规划和设计，制定并实施公路运输发展宏观政策分不开的。

发达国家的发展经验告诉我们，社会经济发展规划的制定，各项政策的出台和实施，都离不开真实、全面、系统、及时、准确的依据。因此，系统而科学地研究反映宏观经济状况的信息资料的采集和处理方法，具有很重要的应用价值和现实意义。

抽样调查作为应用统计的一个重要分支，在各个领域，特别是社会经济领域中有极其重要的应用价值。随着我国改革开放的不断深入和社会主义市场经济体制的逐步建立，抽样调查以其花费少和适时性强的特点，在调查方法中将逐渐占据主导地位；而随着它愈来愈广泛的应用，对方法与理论的需求也愈加迫切。根据实际工作需要和交通部有关职能部门的安排，西安公路交通大学会同辽宁省交通厅、陕西省交通厅，较早着手研究公路运输行业信息资料的采集和处理方法，并分别于“七五”和“八五”期间承担了交通部重点科研项目《公路运输统计体系与方法研究》和《汽车运输轴样调查方法研究》。这两个课题分别于1990年和1995年通过了交通部科学技术司主持的课题评审鉴定会，其中《汽车运输抽样调查方法》课题的核心内容被交通部采纳，在向全国颁布实施的“汽车运输抽样调查方案”中得到应用。该方案从1993年开始在全国实施以来，得到各地交通运输部门的好评，所获取的公路运输统计资料，对反映公路运输行业实际，正确进行宏观决策有参考价值。这两项课题分别于1991年和1996年被交通部评为科技进步三等奖。

为了加强公路运输行业统计队伍建设，“交通部公路运输行业统计培训中心”于1993年9月正式成立，机构设在西安公路交通大学。为满足人员培训及部属高等院校教学的需要，西安公路交通大学组织有经验的直接参加课题研究的教师，在课题报告的基础上编写了这本教材，这是一件好事。其意义表现在以下几个方面：

首先，出版这本教材，对总结“七五”、“八五”期间汽车运输抽样调查研究工作的成果，推广、普及抽样调查方法在整个交通运输统计中的应用，将起到积极的作

用。将科研成果尽快转化为现实生产力，推动社会经济的发展，是世界科技发展的趋势。邹家华副总理在1996年12月召开的全国统计工作会议上指出："要适应转变经济增长方式和实施科教兴国战略的要求，通过广泛运用现代科学技术，大力推动统计科技进步，进一步提高统计工作的科技含量，把统计事业快速发展和统计整体功能的有效发挥建立在依靠科技进步和提高统计工作者素质的基础之上。充分发挥好现代科学技术在准确快捷地反馈纷繁复杂、瞬息万变的社会经济生活，预防和排除对统计信息的各种干扰等方面的作用。"将先进的抽样技术编写成教材，供院校学生和工作人员进行学习和运用，这是提高统计工作科技含量的一种有效形式，是发挥社会效益的一种体现。其次，该教材较系统地介绍了抽样技术理论和方法，便于读者较全面地掌握抽样技术的理论、方法及其适用范围和使用条件，为创造性地实施交通部"汽车运输抽样方案"奠定了理论基础。再次，通过对各种常用的抽样方法的学习和研究，便于从科学的角度去把握交通部"汽车运输抽样方案"中规定的"分层随机抽样"的根据和缘由，提高抽样精度，降低抽样误差。最后，在公路运输业中，科学地进行调查并取得成功很不容易，主要原因是公路运输具有点多、面广、流动、分散的特点。抽样调查以其投入少、见效快、精度高、易实施的优势，在公路运输业信息资料采集工作中具有广阔的应用前景。这项工作的开展并取得成效，在一定程度上也丰富了我国的抽样调查理论和方法体系。

从事公路运输管理、教学与抽样调查工作的人员，可结合实际情况，参考这本教材，以进一步改进公路运输行业信息采集方法，为科学决策提供更为有效的依据，推进公路运输业早日实现两个根本性转变。

交通部副部长

1997年6月20日

前　言

《抽样调查理论与技术应用》一书是在交通部“八五”重点科研项目《公路运输统计体系——汽车运输抽样调查方法研究》课题研究报告基础上形成的一本教科书。交通部副部长胡希捷同志专为本书作序。国家和交通部已将公路、水路运输行业抽样调查工作列为“九五”期间的推广项目，为配合该项工作的顺利进行以及交通部部属高校经济管理、计划统计专业的教学需要，西安公路交通大学组织有关教师编写了这本书。本书分为三个部分：第一部分属于抽样调查理论与技术的基础知识：统计初步，简要地介绍了有关的概率论和数理统计知识理论，如概率及其性质，随机变量的概率分布及数字特征，大数定律和中心极限定理，参数估计初步等内容。其目的是为了使读者在系统学习和掌握抽样理论与技术时奠定必要的基础知识。第二部分是本书的核心内容，基本包括了目前已经成熟的所有抽样调查方法，如简单随机抽样、分层随机抽样、比率估计法、回归估计法、等距抽样、不等概抽样、整群抽样、阶段抽样等内容。读者在了解和掌握这几种常用的抽样方法后，基本上可以解决社会经济、工程技术、科学研究中的具体问题。第三部分属于抽样调查理论和技术在公路运输行业中的应用，以课题报告为基础，系统介绍了交通部颁布的“汽车运输抽样方案”中的具体内容，其中涉及到公路运输行业状况，汽车运输特点，分层随机抽样以及比率估计和回归估计结合的技术处理等，使读者能系统地理解交通部方案的产生过程及原因，并能科学合理的实施。

在课题研究中，本书的作者作为课题组主要成员为了高质量地提出适合我国公路运输实际状况，便于具体操作的抽样方案，系统地研究了美国、日本、印度、英国对汽车运输组织的调查方法，以及各自提出的抽样理论依据。利用较长的时间，投入相当大的精力走访了交通部综合计划司、科学技术司、公路管理司、国家计委、国家统计局、国家计委综合运输研究所、交通部科学研究院、交通部公路科学研究所、山东省交通厅、广东省交通厅、山西省交通厅、江苏省交通厅、四川省交通厅、河南省交通厅、黑龙江省交通厅、吉林省交通厅、湖北省交通厅、湖南省交通厅、河北省交通厅、新疆维吾尔自治区交通厅、甘肃省交通厅、宁夏回族自治区交通厅、青海省交通厅、内蒙古自治区交通厅、贵州省交通厅、云南省交通厅、广西壮族自治区交通厅、福建省交通厅、浙江省交通厅、安徽省交通厅、北京市交通局、天津市交通局、上海市交通局、海南省交通厅和西藏自治区交通厅。这些地方和单位给予课题组大力支持与协助，使课题组获取了大量的第一手资料。为了了解中国的公路运输行业状况，各地区间发展不平衡的现实奠定了基础。同时作为研究单位的陕西省交通厅和辽宁省交通厅尽其所能为课题研究提供了方便，课题组在陕西省的西安市、汉中市、渭南市以及辽宁省的大连市、抚顺市、营口市进行了数月的反复测试，西安公路交通大学左庆乐、邵荣昭、丁岳维等老师付出了许多心血和汗水。经课题组对几种方案比较使用，最终向交通部提交了“分层随机抽样”方案。西安公路交通大学的王健伟老师还编制了汽车运输抽样调查系统的计算机运行软件，已在全国各地交通部门使用。在一系列关键技术细节的处理方面，中国科学院系统工程研究所的冯士雍研究员给课题组极大的帮助，使

方案中的事后分层技术处理，大小月调查技术，抽样精度的控制等都得到了较理想的解决，取得良好的效果。课题组的主要成员有李纪治、张水仙、刘进敏、王健伟、李晓风，以及已故的张中桂老师。

在编著过程中，作者始终贯彻系统介绍和理论探讨相结合的原则。在介绍有关定义和定理及推论的同时，尽可能用简单的方法进行必要的论证，以适应不用层次的读者在掌握和了解抽样理论中的不同需求。各章都配有相应的练习题，便于读者判断是否已掌握了相应的内容。一般地讲，对于已经学习过概率论与数理统计内容的读者，可以直接从第二部分开始，顺序地学习具体的抽样方法。若仅对一两种抽样方法感兴趣的读者，建议学习顺序应从第二、三章开始。比如，某读者对整群抽样有兴趣，且又学习过基础知识，其阅读顺序应该是第二、三章和第九章。可以认为第二、三章是第四章到第十章的基础。对于一般的读者，不必苛求从理论角度来理解各种抽样方法，但要特别注意各种抽样方法使用的条件，尤其是一些量化的限制条件。几乎所有的结论都是有条件的，超越或达不到这些条件该结论也就无法成立。若仅摘取个别结论而忽略该结论形成的依据，肯定会导致工作失误，这本身就不是科学地研究和解决问题的方法。

在本书的章节顺序，各章节讨论的重点也都得到冯士雍老师的指点。作者受益很多很多。李纪治副教授负责第四、五、六、十一章的编著工作，李晓风副教授负责第一、二、三章的编著工作，马江洪副教授负责第七、八、九、十章的编著工作。

借这本集许多人心血的教科书出版之际，请允许作者向交通部副部长胡希捷同志，交通部综合计划司以及所有支持和帮助过我们的单位和个人再次深表谢意，并期望继续支持我们在交通运输经济管理领域中的研究工作。

编著者

1997 年 5 月于西安公路交通大学

目　　录

第一章　统计初步

抽样调查是应用最为广泛的数理统计方法之一，它的原理是数理统计理论的重要组成部分，它的理论依据是概率论和数理统计中的一些基本理论。本章简要介绍这些基本理论，包括：概率及其性质，随机变量及其分布，数字特征，大数定律和中心极限定理，参数估计等内容。

第一节　概率及其性质

自然现象和社会经济现象大体上可以分为两类，一类称其为确定性现象或必然现象，即在一定条件下，其结果必然出现或必然不出现，具有确定性。例如在标准大气压下，水加热至100℃时，必然会沸腾；在大气层中，物体在重力作用下必然自由下落等等。另一类称其为偶然性现象或随机现象，即在一定条件下，其结果可能出现，也可能不出现，具有偶然性。例如在相同条件下抛一枚硬币，其结果可能是正面（有花的一面）向上，也可能是反面（有数字的一面）向上，无论如何控制抛掷条件，在抛掷前都无法确定出现的结果是什么？某两地间的客运公共车辆，每辆车每天的乘客数和营运收入就带有一定的偶然性，无论怎样改进服务条件和其它条件，都不可能准确地预测未来某一天的乘客数和营运收入；这类现象从表面上看，结果的出现与否都带有偶然性，似乎没有规律性，但随着长时间的观察和研究，人们发现：只要将这类现象反复多次观察和试验，它的结果就会呈现出某种规律性。例如反复抛掷一枚硬币，会发现正面出现的次数与反面出现的次数大体相等；反复观察一辆车每天的乘客数和营运收入，会发现乘客数和营运收入各呈一定规律分布。像这种规律，我们称其为随机现象的统计规律性，概率论和数理统计就是研究随机现象的统计规律性的数学学科。

一、随机事件的概率

我们把对随机现象的观察或实验统称为随机试验，把随机试验中可能出现，也可能不出现的结果称为随机事件，简称为事件，并用大写字母 A、B、C…等表示。另外，把那些最简单的，不能再分的事件称为基本事件，而由基本事件构成的事件称为复合事件。由所有基本事件构成的集合称为基本空间或样本空间，记为 Ω。

例 1　在 0，1，2，…，9 十个数中任意选取一个，可能有十种不同的结果："取得数字 0"，"取得数字 1"，…，"取得数字 9"。但还有其它可能结果："取得大于 4 的数字"，"取得数字是 3 的倍数"等等。其中"取得数字 0"，"取得数字 1"，…，"取得数字 9"都是基本事件，而"取得数字是 3 的倍数"是一个复合事件，它由"取得数字 3"，"取得数字 6"，"取得数字 9"三个基本事件组合而成，如果用 i 表示"取得数字 i"这一基本事件，其中 $i=0,1,2,\cdots,9$，则样本空间

$$\Omega=\{0,1,2,\cdots,9\}$$

显然，基本事件、复合事件都是样本空间 Ω 的子集。

在每次试验中一定出现的事件称为必然事件。我们把样本空间 Ω 也看作一个事件。由于在每次试验中必然出现 Ω 中的一个基本事件，也即 Ω 必然出现，因此 Ω 是一个必然事件。这样

我们就把必然事件记作 Ω。而在每次试验中一定不出现的事件称为不可能事件，记作 ϕ，例如在上述例 1 中，“取得的数字小于 18”这一事件就是必然事件，“取得的数字大于 19”这一事件则是不可能事件。

随机事件在一次试验中可能出现，也可能不出现，但反复进行大量试验，可以发现随机事件的出现与否会呈现一定的规律性，即统计规律性。在对随机现象进行研究的时候，人们经常希望能够对有关的随机事件所遵循的统计规律性给出定量的客观描述。为此，人们引入概率这一数量指标来描述随机事件在一次试验中出现的可能性大小。

定义 1.1 如果在一次试验中，共有 n 个等可能出现的基本事件，其中只有 k 个能使事件 A 出现（即事件 A 由 k 个基本事件组成），则定义事件 A 的概率 $P(A)$ 为

$$P(A)=\frac{k}{n} \tag{1.1}$$

这就是概率的古典定义，由于极端的情形是每个基本事件都能使事件 A 出现或没有一个基本事件能使事件 A 出现，所以由式(1.1)有

$$0\leqslant P(A)\leqslant 1$$

例 2 抛掷硬币一次，可能出现的基本事件只有两个，$A=\{$正面向上$\}$；$B=\{$反面向上$\}$由于 A 和 B 都有可能出现。而没有理由认为一个会比另一个出现的可能性大。所以，A 和 B 是等可能出现的，根据(1.1)式得

$$P(A)=P(B)=\frac{1}{2}$$

例 3 一口袋中装有标号分别为 $1,2,\cdots,10$ 的十只同样的球，从中任取一只，问取到号数不小于 8 的球的概率是多少？

〔解〕记 $A=\{$取到号数不小于 8 的球$\}$，从口袋中任取一只球，每只球被取得的可能性相同。那么，共有十个基本事件，其中能使 A 出现的是“取得 8 号球”、“取得 9 号球”和“取得 10 号球”等 3 个基本事件，故按(1.1)式得

$$P(A)=\frac{3}{10}$$

由上面可以看到按古典定义计算事件的概率时，试验必然具备以下特征：

(1)在一次试验中，只有有限个基本事件；

(2)试验中任一基本事件出现的可能性相同。

但是，实际情况中，有许多试验不具备这两点，如在某无线电传呼台观察(O.T)时间段内收到的呼叫次数 m，这里 m 有可能趋于无穷大，那么，如何确定诸如“收到 10 次呼叫”这一类事件的概率呢？为此，人们又引入以下定义。

定义 1.2 假设在相同条件下重复进行 n 次试验，在 n 次试验中事件 A 出现的次数为 k 次，那么称事件 A 在 n 次试验中出现的频率为 $\frac{k}{n}$，如果试验次数增大时，频率 $\frac{k}{n}$ 稳定地在某一常数 p 附近摆动，而且摆动的幅度随试验次数的增大而变小，则定义事件 A 的概率为 p，记为

$$P(A)=p \tag{1.2}$$

这就是概率的统计定义，从统计定义可以看到：事件 A 出现的频率与概率有密切的关系，只要试验次数相当多，我们就可以取频率 $\frac{k}{n}$ 作为事件 A 的概率 p 的近似值。

概率的统计定义和古典定义是一致的。以抛掷一枚硬币的试验为例，由古典定义计算出事

件 $A=\{$正面向上$\}$的概率为$\frac{1}{2}$。而历史上许多学者做过抛掷硬币的试验，下表所列的是其中几次比较著名的试验记录。

实验者	抛掷次数	"正面向上"的次数	频率	实验者	抛掷次数	"正面向上"的次数	频率
隶莫根	2 048	1 061	0.518	皮尔逊	12 000	6 019	0.501 6
蒲丰	4 040	2 048	0.506 9	皮尔逊	24 000	12 012	0.500 5

由表可见，"正面向上"的频率在$\frac{1}{2}$附近摆动，而且抛掷次数越多，频率越接近 0.5，与古典定义计算的概率$\frac{1}{2}$相符。

二、概率的简单性质

在概率论中，经常遇到讨论由多个事件组成的复合事件的概率，为此引入如下概念：

和事件："事件 A 和事件 B 中至少有一个出现"是一个复合事件，称为 A 与 B 的和事件，记作 $A\cup B$。

积事件："事件 A 和事件 B 同时出现"是一个复合事件，称为 A 与 B 的积事件，记作 AB。

和事件和积事件还可推广到更多个事件的情形。

另外，如果 $AB=\phi$，则称事件 A 与事件 B 互斥，此时，$A\cup B$ 特别记作 $A+B$。如果$A\cup B=\Omega$，$AB=\phi$，则称事件 A 与事件 B 互逆，即 A 的逆事件是 B，记作 $\overline{A}=B$，同样，B 的逆事件是 A，记作 $\overline{B}=A$。

如果事件 A 出现与事件B 出现互不影响，则称 A 与 B 相互独立，对多个事件的情形类似。

例 4 设有两个口袋，甲袋中装有 3 个红球和 3 个白球，乙袋中装有 5 个红球和 2 个黄球。现从甲、乙两袋中任意各取一球，若用 A 表示从甲袋中取一红球，B 表示从乙袋中取一黄球，那么，和事件 $A\cup B$ 表示从甲袋中取一红球或者从乙袋中取一黄球；积事件 AB 表示从甲袋中取一红球，同时从乙袋中取一黄球，事件 A 的逆事件 $\overline{A}$ 表示从甲袋中取一白球，事件 B 的逆事件 $\overline{B}$ 表示从乙袋中一红球，由于从甲袋中任取一球与从乙袋中任取一球，其结果互不影响，故事件 A 与事件 B 相互独立。

根据概率的定义，可以推出概率具有以下基本性质：

(1)$0\leqslant P(A)\leqslant 1$

(2)$P(\phi)=0$；$P(\Omega)=1$

(3)$P(A+B)=P(A)+P(B)$

(4)$P(A)=1-P(\overline{A})$

(5)$P(A\cup B)=P(A)+P(B)-P(AB)$

(6)若 A 与 B 相互独立，则

$$P(AB)=P(A)P(B)$$

特别是性质(3)与(6)的推广情形，我们以后经常用到。

对事件 $A_1,A_2,\cdots,A_n$

若 $A_1,A_2,\cdots,A_n$ 两两互斥，则

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n) \tag{1.3}$$

若 $A_1,A_2,\cdots,A_n$ 相互独立，则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n) \tag{1.4}$$

例 5 计算例 4 中，事件 $AB,\overline{A},\overline{B}$ 及 $A\cup B$ 的概率。

〔解〕由概率的性质可得

$$P(AB)=P(A)P(B)=\frac{3}{6}\times\frac{2}{7}=\frac{1}{7}$$

$$P(\overline{A})=1-P(A)=1-\frac{3}{6}=\frac{1}{2}$$

$$P(\overline{B})=1-P(B)=1-\frac{2}{7}=\frac{5}{7}$$

$$P(A\cup B)=P(A)+P(B)-P(AB)=$$
$$\frac{3}{6}+\frac{2}{7}-\frac{1}{7}=\frac{9}{14}$$

三、条件概率及有关公式

在实际问题中，一般除了要考虑事件 A 的概率 $P(A)$，还须考虑在“事件 B 已经出现”这一条件下，事件 A 出现的概率。一般地说，后者的概率与前者的概率未必相同。为了区别起见，我们把后者称为条件概率，记为 $P(A|B)$，读作在条件 B 下，事件 A 的概率。

定义 1.3 设事件 B 的概率 $P(B)>0$，则在事件 B 已经出现的条件下，事件 A 的条件概率 $P(A|B)$ 定义为

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{1.5}$$

例 6 某客运汽车公司有甲、乙两个车队，甲队有 40 座客车 4 辆，50 座客车 7 辆，60 座客车 4 辆；乙队有 40 座客车 5 辆，50 座客车 8 辆，60 座客车 2 辆。现从该公司的所有客车中任意抽取一辆，记 A 为“抽到甲队车辆”，B 为“抽到 40 座客车”，计算 $P(A|B)$ 及 $P(B|A)$。

〔解〕由于

$$P(A)=\frac{15}{30}=\frac{1}{2}$$

$$P(B)=\frac{9}{30}=\frac{3}{10}$$

$$P(AB)=\frac{4}{30}$$

所以

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{\frac{4}{30}}{\frac{3}{10}}=\frac{4}{9}$$

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{\frac{4}{30}}{\frac{1}{2}}=\frac{4}{15}$$

可见

$$P(A)\neq P(A|B)\text{ 及 }P(B)\neq P(B|A)$$

由条件概率的定义，可以导出几个常用的公式。

(1)乘法公式

若 $P(A)>0$，则由定义 1.3 知

$$P(B|A)=\frac{P(AB)}{P(A)}$$

两边乘以 $P(A)$ 得

$$P(AB)=P(A)P(B|A) \tag{1.6}$$

称为乘法公式，推广到多个事件积的情形有

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{1.7}$$

(2)全概率公式

若 A_1 与 A_2 互逆，且 $P(A_i)>0,i=1,2$，则由积事件的定义及概率的性质得

$$P(B)=P(\Omega B)=P\{(A_1+A_2)B\}=$$
$$P(A_1B+A_2B)=$$
$$P(A_1B)+P(A_2B)$$

再根据乘法公式得

$$P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2) \tag{1.8}$$

称为全概率公式，其一般形式为

若 $A_1,A_2,\cdots,A_n$ 两两互斥，$A_1+A_2+\cdots+A_n=\Omega$ 且 $P(A_i)>0,i=1,2\cdots,n$ 则

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i) \tag{1.9}$$

这里，$A_i,i=1,2,\cdots,n$，可以看作导致 B 出现的诸“原因”。在实际问题中，“原因”A_i 出现的概率及在“原因”A_i 出现的条件下，B 出现的概率一般容易确定，这样便可以由全概率公式计算 B 的概率 $P(B)$。

例 7 设一个仓库中有 10 箱同样规格的产品，其中有甲厂生产的 5 箱，乙厂生产的 3 箱，丙厂生产的 2 箱，已知甲、乙、丙三厂生产中次品率分别为 $\frac{1}{10},\frac{1}{15},\frac{1}{20}$，从这 10 箱产品中任取一箱并任取其中 1 件产品，求取出这个产品是正品的概率。

〔解〕以 A_1,A_2,A_3 分别表示“取到的产品是甲、乙、丙厂生产的”，以 B 表示“取到的产品为正品”。依题意，A_1,A_2,A_3 两两互斥，且以 $A_1+A_2+A_3=\Omega$

由于

$$P(A_1)=\frac{5}{10}=\frac{1}{2}$$
$$P(A_2)=\frac{3}{10}$$
$$P(A_3)=\frac{2}{10}=\frac{1}{5}$$

而

$$P(B|A_1)=\frac{9}{10},\quad P(B|A_2)=\frac{14}{15}$$
$$P(B|A_3)=\frac{19}{20}$$

所以，由全概率公式得

$$P(B)=\sum_{i=1}^{3}P(A_i)P(B|A_i)=$$
$$\frac{1}{2}\times\frac{9}{10}+\frac{3}{10}\times\frac{14}{15}+\frac{1}{5}\times\frac{19}{20}=0.92$$

第二节　随机变量的概率分布及数字特征

随机事件的概率仅揭示了事件出现的统计规律性，而要揭示随机现象的统计规律性，就必

须了解随机现象的所有可能的结果(即事件)出现的概率。为此,我们引入随机变量的概念,以便更加全面地研究随机现象有多少种可能的结果,以及每一可能结果会以多大的概率出现(即所谓概率分布)。此外,我们还将介绍随机变量的数字特征的有关概念,它描述了随机现象的某些基本特征。

一、随机变量及其概率分布

一般说来,随机事件都可以用数量来描述。一些随机事件本身就带有数量指标,如一个汽车运输公司在某一时间段内投入运营的车辆数,某个电话总机在一定的时间段(O.T)内收到呼叫次数,投掷一枚骰子可能出现的点数等等。另外,一些随机事件虽然本身并不带有数量指标,如投掷一枚硬币,结果可能出现正面或反面,检验一个(或一批)产品质量的结果可能是合格或不合格等等。但是,对这类随机事件,我们可以采取适当方式给它们记以数量指标,如在投掷硬币试验中,可以把正面向上记为1,反面向上记为0,在产品质量抽检中,可以把产品质量合格记为1,不合格记为0等等。总之,我们可以建立随机事件和数量之间的一个对应关系,并且,一般说来,不同随机事件所对应的数量也会是不相同的。从而,引入以下定义。

定义 1.4 若随机试验中所出现的每种结果(即随机事件)e 都可以唯一地对应着一个数值 $X(e)$,则变量 $X(e)$ 称为随机变量。

根据上述定义,随机变量实际上是事件和数之间的一种对应关系,通过这种关系,样本空间的任一部分基本事件的集合(即事件)都可以用数轴上一部分点或一个区间来表示。反过来,根据随机变量的取值情况可以将随机试验的所有结果都表示出来。

例 8 (1)投掷一枚硬币,结果有"正面向上"和"反面向上"两种。令

$$X=\begin{cases}0,\text{当正面向上时}\\1,\text{当反面向上时}\end{cases}$$

则 X 是一个随机变量,这里"$X=0$"表示事件"正面向上",而"$X=1$"表示事件"反面向上"。

(2)一射手一次射击命中目标的概率为0.8,现在连续射击20次,则命中目标的次数 X 是一个随机变量,这里,X 可能取0,1,2,…,20等21个整数值,显然,"$X=5$"表示事件在"20次射击中,命中目标5次","$X\leqslant 10$"表示事件"在20次射击中,命中目标的次数不超过10次"。

(3)在公路上某一位置观察每天过往的车辆数 X,则 X 是一个随机变量,这里 X 可能取0,1,2,…,等无穷多个整数值,而"$X=0$""$X=1$"…,"$X=k$"…等都是随机事件。

(4)某公共汽车站每隔15min有一辆公共汽车通过,对任一时刻到来的乘客而言,他候车时间 X 是一个随机变量,显然 $0\leqslant X\leqslant 15$,如"$X=2$","$X=11.2$"等都是随机事件。

一般,根据随机变量的取值特征将其分为两类,一类称为离散型随机变量,这类随机变量可能取的值可以一一列举出来,像例8中的(1)、(2)、(3),都是离散型随机变量的例子,它们要么取有限个值,要么虽然取无穷多个值,但是可以一一列举出来;另一类称为非离散型随机变量,这类随机变量可能取的值不能一一列举出来,所包含的范围很广,其中最主要也是实际中经常遇到的是所谓连续型随机变量,这种随机变量可以取某一区间内的任一值,像例8中的(4)就是连续型随机变量的例子。

由于随机变量的取值是由随机试验的结果确定的,而随机试验的任一结果在一次试验中是否出现带有偶然性,所以随机变量取值时就带有偶然性。因此,在讨论随机变量时,不仅要了解随机变量都可能取哪些数值,更重要的是还要了解它以怎样的概率取这些数值,这就是随机

变量的概率分布问题，对离散型随机变量和连续型随机变量分别引入如下定义。

定义 1.5 若随机变量 X 所有可能取的值为 $x_1, x_2, \cdots, x_i, \cdots$，且

$$P\{X=x_i\}=p_i, i=1,2,\cdots \tag{1.10}$$

（其中 i 可以取到某个自然数 n 为止，即 $i=1,2,\cdots,n$）则称随机变量 X 为离散型随机变量，称式(1.10)为 X 的概率分布列或分布律。

式(1.10)通常也写成下列形式

X	x_1,	x_2,	…	x_i	…,
P	P_1,	P_2,	…	P_i	…,

(1.11)

显然，分布律中的 $p_i(i=1,2,\cdots)$ 应满足

(1) $p_i \geqslant 0$

(2) $\sum\limits_i p_i=1$

例 9 设一口袋中装有 6 个同样的球，其中 3 个标有号码 1，2 个标有号码 3，一个标有号码 5，现从口袋中任取一球，观察其号码 X，求 X 的分布律。

〔解〕由题意 X 的分布律如下

$$P\{X=1\}=\frac{3}{6}=\frac{1}{2}$$

$$P\{X=2\}=\frac{2}{6}=\frac{1}{3}$$

$$P\{X=5\}=\frac{1}{6}=\frac{1}{6}$$

或写成

X	1	3	5
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

定义 1.6 对随机变量 X，如果存在一个非负的可积函数 $f(x)(-\infty<x<+\infty)$，使对任意的 $a, b(a<b)$，都有

$$P\{a<X<b\}=\int_b^a f(x)\mathrm{d}x \tag{1.12}$$

则称随机变量 X 为连续型随机变量，并称 $f(x)$ 为 X 的概率密度函数，简称为密度。

同样，X 的密度 $f(x)$ 满足

(1) $f(x)\geqslant 0 \quad (-\infty<x<+\infty)$

(2) $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$

一般说来，$f(x)$ 的大小反映了 X 在点 x“附近”取值的概率的大小，所以，我们用 X 的密度 $f(x)$ 来描述连续型随机变量 X 取值的概率分布。

二、数字特征

随机变量的概率分布完全揭示了随机变量所表示的随机现象的统计规律性。但是，在实际问题中，要确定一个随机变量的概率分布，往往是做不到的。另外，在许多情况下，只要知道了反映随机变量一些基本特征的指标就能满足需要，对此引入一些能在一定程度上刻划随机变量的某些特征的指标，即随机变量的数字特征，其中最重要也是最常用的是均值和方差。

(一)均值(数学期望)

随机变量的均值又称为数学期望,它是随机变量的最重要的数字特征之一,定义如下

定义 1.7 对离散型随机变量 X,设其分布律为:$P\{X=x_i\}=p_i, i=1,2,\cdots$,则称 $\sum_i x_i p_i$(在 X 取无穷多个值时,要求此级数绝对收敛)为 X 的均值或数学期望,记为$E(X)$,即

$$E(X)=\sum_i x_i p_i \tag{1.13}$$

对连续型随机变量 X,设其密度函数为 $f(x)$,$-\infty<x<+\infty$,则称$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$(这里要求此积分绝对收敛)为 X 的均值或数学期望,记为 $E(X)$

即

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x \tag{1.14}$$

例 10 设一盒子中装有分别标有不同数字的同种卡片 10 张,其中 3 张上标有数字 0,2 张上标有数字 5,4 张上标有数字 10,1 张上标有数字 15,现从盒中任取一张卡片观察其数字 X,求 X 的均值。

〔解〕由题意知,X 是一个离散型随机变量,其分布律如下

X	0	5	10	15
P	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{10}$

从而,X 的均值 $E(X)$为

$$E(X)=0\times\frac{3}{10}+5\times\frac{1}{5}+10\times\frac{2}{5}+15\times\frac{1}{10}=6.5$$

由上述计算可以看到离散型随机变量的均值 $E(X)$实际上是 X 所取值的"加权"平均值,其中 x_i 在均值中所占的"权数"就是事件$\{X=x_i\}$出现的概率 p_i,对连续型随机变量的情形也类似。所以,均值 $E(X)$反映了随机变量 X 取值的"平均"大小。

均值具有以下性质

(1) $E(C)=C$ (C 为常数)

(2) $E(aX+b)=aE(X)+b$ (a,b 为常数)

(3) $E(X_1+X_2)=E(X_1)+E(X_2)$

(4) 若随机变量 X_1 与 X_2 相互独立(即对任意实数 a_1,a_2,b_1,b_2,事件$\{a_1<X_1<b_2\}$与事件$\{a_2<Z_2<b_2\}$相互独立),则 $E(X_1X_2)=E(X_1)E(X_2)$

其中性质(3)和(4)可以推广到任意有限个随机变量的情形。

(二)方差

随机变量 X 的均值 $E(X)$虽然反映了 X 取值的"平均"大小,但是,它不能反映 X 取值的分散程度。为了更加全面地了解 X 取值的变化情况,我们除了要知道 X 的均值,还希望知道 X 取值关于均值的离散程度。随机变量的方差就是为此目的而引入的一个重要指标,定义如下。

定义 1.8 设 X 是一个随机变量,且 X 的均值 $E(X)$存在,称

$$E\{[X-E(X)]^2\}$$

为 X 的方差,记为 $V(X)$,而 $\sqrt{V(X)}$称为标准差。

即

$$V(X)=E\{[X-E(X)]\}^2 \tag{1.15}$$

从而,离散型随机变量 X 的方差为

$$V(X)=\sum_i [x_i-E(X)]^2 p_i \tag{1.16}$$

其中,$p_i=P\{X=x_i\}, i=1,2,\cdots$。