全国工程硕士专业学位教育指导委员会推荐教材

杨虎 钟波 刘琼荪 编著

应用数理统计

http://www.tup.com.cn

清华大学出版社

全国工程硕士专业学位教育指导委员会推荐教材

杨虎 钟波 刘琼荪 编著

应用数理统计

清华大学出版社
北京

内容简介

本书根据全国工程硕士专业学位教育指导委员会数学公共课改革协调小组制定的工程硕士数理统计课程教学基本要求，着重介绍统计思想和应用方法. 内容包括概率知识、统计概念、参数估计、假设检验、回归分析、方差分析、试验设计等，在应用上增加了许多新的内容，如：非参数方法、回归诊断、因子分析等. 为了方便实际分析和数据处理，本书采用常用的 Excel 软件设计了各类统计算法和应用案例. 全书论述深入浅出，通俗易懂，富有启发性.

本书读者对象为各类工程硕士研究生，也可作为理工科本科生、教师、科技工作者和工程技术人员的参考书.

图书在版编目（CIP）数据

应用数理统计/杨虎，钟波，刘琼荪编著. —北京：清华大学出版社，2006.12

（全国工程硕士专业学位教育指导委员会推荐教材）

ISBN 7-302-13787-0

Ⅰ. 应…　Ⅱ. ①杨…　②钟…　③刘…　Ⅲ. 数理统计－研究生－教材　Ⅳ. O212

中国版本图书馆 CIP 数据核字(2006)第 108401 号

责任编辑：刘　颖　赵从棉

责任校对：刘玉霞

责任印制：何　芊

出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座

http://www.tup.com.cn　　**邮　　编**：100084

c-service@tup.tsinghua.edu.cn

社 总 机：010-62770175　　**邮购热线**：010-62786544

投稿咨询：010-62772015　　**客户服务**：010-62776969

印 刷 者：北京鑫丰华彩印有限公司

装 订 者：三河市李旗庄少明装订厂

经　　销：全国新华书店

开　　本：185×230　**印　张**：13.5　**字　数**：273 千字

版　　次：2006 年 12 月第 1 版　**印　次**：2006 年 12 月第 1 次印刷

印　　数：1～4000

定　　价：29.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：010-62770177 转 3103　　产品编号：021799-01

前　言

数理统计学是一门应用性很强的学科，其方法被广泛应用于现实社会的信息、经济、工程等各个领域，学习和运用数理统计方法已成为当今技术领域里的一种时尚．面对信息时代，要想处理大量的数据并从中得出有助于决策的量化结论，就必须掌握不断更新的数理统计知识．

工程硕士研究生究竟需要什么样的数理统计教材？重庆大学从1997年开始招收工程硕士研究生以来就一直在这方面进行探索，先是和工科研究生共用同样的教材，后来又在研究生院的资助下专门出版了一本工程硕士研究生教材．使用几年后感觉有不少问题，虽然减少了篇幅，但由于解释不够充分，证明很多，更像一本数学教材，工程硕士研究生学起来很费力，更谈不上对统计思想和方法的全面了解．2002年在重庆大学研究生院的支持下，我们开始全面分析工程硕士研究生的数理统计教材内容．感觉工程硕士研究生虽然没有全日制研究生那样多的时间用于消化课堂内容和课后训练，但大多具有丰富的实践经验，应用方法很容易上手并乐于在具体项目中采用．因此，我们认为训练思维和强化数学基础不是工程硕士研究生数理统计教学的目的，而应该更多地着眼于灌输统计思想和介绍应用统计方法．当然数理统计传统内容的介绍颇费精力，最后我们的思路是重点讲清楚原理，不过分强调数学的严密和论证的充分，毕竟工程硕士研究生基本上已经有稳定的工作，如何在实际中应用所学更为迫切．

本书对参数估计和假设检验作了局部的改动，主要是舍弃了理论上的严格阐述和论证，符号也尽量简化；应用上增加了很多新的内容，如：假设检验强调了非参数方法的应用，回归分析增加了回归诊断，第8章介绍了因子分析，这是心理统计分析的基础，在各个领域有广泛的应用前景，尤其适合工程硕士研究生在工作中应用．

对于如此丰富的内容，我们在编排上尽量用浅显的描述和说明文字，避免过多的符号演算，因此，阅读本书不需要太多的数学知识，学生必备的知识仅需要高等数学和线性代数，加 * 号的内容供教学取舍和学有余力的学生自学．当然，为了弥补本书在数学理论上的不足，书末附有部分参考书籍，供研究生进一步学习和科研查阅之需．因此本书更多的角色是充当传播统计思想、学习统计方法和研究统计应用的工具书和入门读物，以适

应工程硕士各行业学生的需要.

本书需要 48 学时左右的课堂讲授,为了方便实际分析和数据处理,我们针对 Excel 设计了算法和实例,而不是选用流行的大型统计软件(工程硕士研究生比较分散,无法使用高校的实验设施,为了学习购置这些软件意义不大).这样做对于学生理解应用统计方法的整个计算过程很有必要,比如因子分析的 Excel 求解,是本书的亮点之一,学生过度依赖软件而不是自己编程,对计算过程不求甚解,将无法真正了解统计方法的设计思想.全书各章节的内容均被编者多次在重庆大学工程硕士研究生课程中讲授过,很多授课老师提出的宝贵意见对全书的完善功不可没,恕不一一致谢.本书的出版得到清华大学出版社全国工程硕士核心教材基金资助和重庆大学研究生院的专项资助,特在此表示衷心的感谢!

本书属于全新的尝试,效果如何尚待检验,书中使用的数据和 Excel 模板可以在重庆大学数理学院 Sci 论坛公共课辅导栏目里下载,必要时将出版相应的课件和应用光盘以辅助教学.限于编者水平,全书错谬之处一定不少,欢迎读者批评指正!

编 者

2006 年 5 月于重庆大学数理学院统计与精算科学系

符 号 说 明

样本空间：Ω

参数空间：Θ

集合、随机事件：采用大括号$\{\}$

概率：$P(\cdot)$或 p

随机变量：$X,Y,\cdots$

分布函数：$F(x)$，标准正态分布函数 $\Phi(x)$

密度函数：$f(x)$或 $f(x,\theta)$，其中 θ 为参数

条件密度函数：$f(x|y)$，表示随机变量 $Y=y$ 时，X 的密度函数

常数：a,b,c,d,l,m,n,k

变量：x,y,z,t,u,v,w

样本：$X_1,X_2,\cdots,X_n$

样本观测值：$x_1,x_2,\cdots,x_n$

样本均值：$\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$

样本中位数：$\widetilde{X}$

样本方差：$S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$

样本标准差：$S=\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$

样本 k 阶原点矩：$M_k=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k, k=1,2,\cdots$

样本 k 阶中心矩：$M_k^*=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k, k=1,2,\cdots$

总体均值：EX

总体方差：DX, $\mathrm{var}\, X$

协方差：$\mathrm{cov}(X,Y)$

当 $\boldsymbol{X},\boldsymbol{Y}$ 分别表示随机向量时，$E\boldsymbol{X}$，$D\boldsymbol{X}$，$\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})$分别表示数学期望、方差向量和协方差矩阵，$\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})$可简单表示为 $\mathrm{cov}(\boldsymbol{X})$.

相关系数矩阵：$\boldsymbol{R}$

相关系数：r

n 阶单位矩阵：$\boldsymbol{I}_n$（常略去下标 n，可根据前后文判断其阶数）

目 录

第 1 章　概率知识 …… 1

1.1　概率的计算 …… 1
1.2　一维随机变量 …… 5
1.3　多维随机变量 …… 11
1.4　数字特征 …… 13
*1.5　大数定律和中心极限定理 …… 18
习题 1 …… 20

第 2 章　统计概念 …… 22

2.1　数理统计的涵义 …… 22
2.2　总体、样本与统计量 …… 24
2.3　顺序统计量、经验分布函数和直方图 …… 27
2.4　抽样分布 …… 31
2.5　应用案例 …… 37
习题 2 …… 38

第 3 章　参数估计 …… 41

3.1　点估计和区间估计的概念 …… 41
3.2　矩估计和极大似然估计 …… 41
*3.3　点估计的优良性准则 …… 46
3.4　区间估计 …… 52
3.5　应用案例 …… 59
习题 3 …… 60

第 4 章　假设检验 …… 63

4.1　假设检验的基本概念 …… 63
4.2　参数假设检验 …… 67
4.3　非参数假设检验 …… 75
4.4　应用案例 …… 83
习题 4 …… 88

第 5 章　回归分析 …… 91

5.1　回归分析的基本概念 …… 91
5.2　一元线性回归 …… 92
*5.3　多元线性回归 …… 106
5.4　应用案例 …… 113
习题 5 …… 119

第 6 章　方差分析 …… 123

6.1　方差分析的基本原理 …… 123
6.2　单因素方差分析 …… 124
*6.3　双因素方差分析 …… 130
6.4　应用案例 …… 135
习题 6 …… 138

第 7 章　试验设计 …… 141

7.1　正交设计的基本概念 …… 141
7.2　无交互效应的正交设计与数据分析 …… 145
7.3　有交互效应的正交设计与数据分析 …… 150
7.4　应用案例 …… 154
习题 7 …… 158

第 8 章　因子分析 …… 161

8.1　因子分析的基本原理 …… 161
8.2　因子分析模型 …… 162
8.3　因子分析模型中参数的估计方法 …… 165
8.4　因子旋转 …… 167

8.5 因子得分 …… 168
8.6 应用案例 …… 169
习题 8 …… 173

附录 常用数理统计表 …… 176

习题提示与解答 …… 194

参考文献 …… 203

第1章

概率知识

概率论是应用数理统计的重要工具，是研究随机现象统计规律性的数学学科，要用一章的篇幅介绍整个概率论的内容是不现实的，因此本书假定读者已经具备必要的概率论知识，仅仅为了检索或便于复习的考虑，对必备的概率论知识进行简要的归纳和总结，完全没有基础的读者应该参阅相关的概率论教材.

随机现象、随机试验、随机事件是我们进入概率论世界的三把钥匙. 自然界中存在着大量的随机现象，称发生与否具有偶然性的事件为随机事件，通过随机试验可以观察到随机事件. 第一个问题是既然是随机事件，规律性作何解释？这里指的是统计规律，即在大量重复试验时，这些随机事件所显示的必然的本质特征. 第二个问题是为什么说它是一门数学学科？不是研究具体的试验和具体的事件，而是通过数学抽象和概括，用定量化的数学语言去研究一般的规律性，因此是数学的一个分支.

关于概率论的起源简单说明如下：作为机会游戏在纪元前就开始了，毋庸讳言，概率论萌芽于早期的赌博这一今天比较忌讳但又普遍存在的社会现象. 概率论起源于赌博，还可以用 1494 年 Pacciolo 的著名分点问题①(the problem of points)来说明. 围绕着这一问题，Cardano 及随后的很多人进行了更深入的研究，特别是 Pascal 和 Fermat 一生致力于赌博问题的研究并把概率论引入正途. 到 20 世纪初，概率论作为一门数学分支已经牢牢的站住了脚，1933 年 Kolmogorov 提出的概率空间及公理化定义使这一学科进一步完善.

---

① 分点问题：1494 年，在意大利出版的一本计算数学的教科书中，Pacciolo 写道：若一个比赛赢 6 次才算赢，两个赌徒，1 胜 5 次，另一胜 2 次，因故赌博中断，赌金应按 5∶2 分给他们. Cardano 认为第一个赌徒在以后的比赛中只有 5 种可能的结果，胜第 1 次，胜第 2 次，胜第 3 次，胜第 4 次或者全部输掉，因此应按(1+2+3+4)∶1=10∶1 来分. 实际上正确的结果为 15∶1.

1.1 概率的计算

将复杂的事件化为简单事件的关系和运算，有利于对复杂事件统计规律性的描述和研究.

1.1.1 事件的关系与运算

如果用英文字母 A,B,C 等来表示事件，通过集合论知识，可以方便地表示子事件 $A\subset B$，两事件相等 $A=B$，和事件 $A\cup B$，积事件 AB(或 $A\cap B$)，差事件 $C=A-B$，互斥事件 $AB=\varnothing$，逆事件 $\overline{A}$，完备事件组 $A_1,A_2,\cdots,A_n$(两两互斥且 $\bigcup\limits_{i=1}^{n}A_i=\Omega$)等，并可以建立下面的运算规律.

(1) 交换律：$A\cup B=B\cup A,AB=BA$；

(2) 结合律：$(A\cup B)\cup C=A\cup(B\cup C),(AB)C=A(BC)$；

(3) 分配律：$(A\cup B)\cap C=AC\cup BC,(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$；

(4) 德摩根(De Morgan)定理(对偶律)：$\overline{\bigcup\limits_{i=1}^{n}A_i}=\bigcap\limits_{i=1}^{n}\overline{A}_i,\overline{\bigcap\limits_{i=1}^{n}A_i}=\bigcup\limits_{i=1}^{n}\overline{A}_i$；

(5) 差化积：$A-B=A\overline{B}$；

(6) 吸收律：如果 $A\subset B$，则 $A\cup B=B,AB=A$.

注意：当 $AB=\varnothing$ 时，$A\cup B$ 常表示为 $A+B$.

1.1.2 概率的统计定义与公理化定义

在概率论早期的实践中，往往将随机事件的频率 $f_n(A)=\dfrac{r}{n}$ 当作概率. 人们发现，随着试验次数的增加，频率具有一定的稳定性，并且会接近事件的概率. 直到 20 世纪上半叶，一系列极限定理的建立，从理论上支持了这种观点. 建立在这种观点之上的概率定义就称为概率的统计定义.

众所周知，数学在 20 世纪初开始了一场公理化革命，作为数学分支的概率论也不例外，在 1933 年苏联著名数学家 Kolmogorov 提出了概率的公理化定义.

定义 1.1.1 设 Ω 是试验 E 的样本空间，对于试验 E 的每个随机事件 $A(A\subset\Omega)$，有一个实数 $P(A)$ 与之对应，且 $P(A)$ 满足

公理 1(非负性)：$P(A)\geqslant 0$；

公理 2(规范性)：$P(\Omega)=1$；

公理 3(可列可加性)：$P\left(\sum\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}P(A_i)$；

则称 $P(A)$为事件 A 的概率.

利用这三条公理，可以推导出概率的许多性质，下面是其中的一些，证明请读者自行补上或参阅相关教材.

(1) 不可能事件的概率为零 $P(\varnothing)=0$；

(2) (有限可加性) $P\left(\sum\limits_{i=1}^{m} A_i\right)=\sum\limits_{i=1}^{m} P(A_i)$；

(3) 设 A 为任一随机事件，则 $P(A)=1-P(\bar{A})$；

(4) 设 A,B 为任两随机事件，则 $P(B-A)=P(B)-P(AB)$；

(5) (单调性)若 $A\subset B$，则 $P(A)\leqslant P(B)$；

(6) 设 A,B 为任两事件，则 $P(A\cup B)=P(A)+P(B)-P(AB)$.

1.1.3 古典概型与古典概率计算

古典概型是这样的随机试验：①基本事件总数是有限的；②基本事件发生的可能性相等. 在古典概型下，概率的计算很简单，为

$$P(A)=\frac{r}{n}=\frac{A\text{ 包含的基本事件数}}{\text{基本事件总数}} \tag{1.1.1}$$

某些时候或者对于某些特定的问题，n 和 r 的求取会很困难，需要用到排列组合的知识，对这方面不熟悉的读者请参考相关教材. 对于古典概型中等可能的判断可以通过对称性、均衡性、平凡性等作出. 古典概率具有如下性质：

(1) 非负有界性 $0\leqslant P(A)\leqslant 1$；

(2) 规范性 $P(\Omega)=1$；

(3) 有限可加性 $P\left(\sum\limits_{i=1}^{m} A_i\right)=\sum\limits_{i=1}^{m} P(A_i)$.

例 1.1.1 设有 40 件产品，其中有 10 件为次品其余为正品，现从中任取 5 件，求取出的 5 件产品中至少有 4 件次品的概率.

解 “至少有 4 件次品”这一事件可表示成“恰有 4 件次品”与“5 件全是次品”这两个互斥事件的和事件.

设 A=“恰有 4 件次品”，B=“5 件全是次品”，C=“至少有 4 件次品”，则

$$C=A\cup B,\quad \text{且}\quad AB=\varnothing$$

利用概率的有限可加性，有

$$P(C)=P(A\cup B)=P(A)+P(B)$$

而基本事件总数 $n=\mathrm{C}_{40}^5$(这里 C_n^m 表示从 n 个元素中抽取 m 个元素的全部组合数)，A 所含基本事件数为 $\mathrm{C}_{10}^4\mathrm{C}_{30}^1$，$B$ 所含基本事件数为 C_{10}^5，故

$$P(A)=\frac{\mathrm{C}_{10}^4\mathrm{C}_{30}^1}{\mathrm{C}_{40}^5}=\frac{175}{18\,278}\approx 0.0096$$

$$P(B)=\frac{C_{10}^{5}}{C_{40}^{5}}=\frac{7}{18\,278}\approx 0.0004$$

因此

$$P(C)=P(A)+P(B)\approx 0.01$$

1.1.4 几何概型与几何概率计算

与古典概型相比，几何概型的基本试验结果可以是无穷多个，但它要求试验结果在空间区域中是均匀的，同时，该空间区域是可测的，整个概率空间具有有限测度. 在几何概型下，几何概率的计算公式如下：

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{A\text{ 的几何测度}}{\text{概率空间的几何测度}} \tag{1.1.2}$$

几何概率也有如下性质：

(1) 非负有界性 $0\leqslant P(A)\leqslant 1$；

(2) 规范性 $P(\Omega)=1$；

(3) 可列可加性 $P\left(\sum\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}P(A_i)$.

例 1.1.2(蒲丰投针问题) 在平面上画一些距离都等于 a 的平行线，向此平面上任投一长为 $l(l<a)$的针，试求此针与任一平行线相交的概率.

解 以 x 表示针的中点 M 到最近一条平行线的距离，φ 表示针与平行线的交角. 据针与平行线的位置关系(如图 1.1.1)，显然有

$$0\leqslant x\leqslant\frac{a}{2},\quad 0\leqslant\varphi\leqslant\pi$$

满足上面两个不等式的点(φ,x)在 φ-x 平面上构成一个矩形 G，如图 1.1.2.

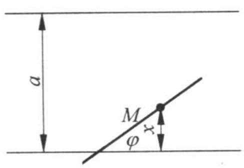

图 1.1.1

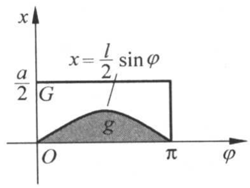

图 1.1.2

为使针与平行线相交，充要条件是

$$x\leqslant\frac{l}{2}\sin\varphi$$

此不等式构成图 1.1.2 中 G 的一个子区域 g. 于是，投针问题就等价于向 G 内投掷一点，求点落入 g 内的概率问题. 这是一个几何概率问题，于是得

$$P(\text{针与任一平行线相交}) = \frac{g\text{ 的面积}}{G\text{ 的面积}} = \frac{\frac{1}{2}\int_0^{\pi} l\sin\varphi\mathrm{d}\varphi}{\frac{1}{2}\pi a} = \frac{2l}{\pi a}$$

通过这个公式，Fox 在 1884 年取 $a=1, l=0.75$，投针 1030 次试验，实际相交次数 489 次，先计算出“针与任一平行线的相交”频率作为其概率的近似值，代入上面的公式，求出 π 的近似值为 3.1595.

1.2 一维随机变量

描述随机试验结果的变量 X 取什么值，在每次试验之前是不能确定的（试验前只知道它所有可能的取值或取值范围），它们的取值依赖于试验的结果，由于试验的结果是随机的，所以它们的取值具有随机性. 像这种取值由随机试验的结果所确定的变量称为随机变量. 随机变量主要有离散型随机变量与连续型随机变量两大类.

1.2.1 一维离散型随机变量及其分布列

设 X 是一个随机变量，如果 X 可能取的值只有有限个或可数个，则称 X 是一维离散型随机变量，简称离散型随机变量.

设离散型随机变量 X 可能取值为 $x_1, x_2, \cdots$（其中任何两个都不相同），X 取各个值的概率分别是 $p_1, p_2, \cdots$，记成

$$P(X = x_k) = p_k, \quad k = 1, 2, \cdots \tag{1.2.1}$$

称 p_k 为离散型随机变量 X 的分布列，也称为概率分布. 显然分布列满足如下条件：

$$p_k \geqslant 0 (k = 1, 2, \cdots) \quad \text{且} \quad \sum_{k=1}^{\infty} p_k = 1$$

1.2.2 一维连续型随机变量及其密度函数

对于随机变量 X，如果存在一个非负可积函数 $f(x)$，对任意实数 x，有

$$P(X \leqslant x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \tag{1.2.2}$$

则称 X 为连续型随机变量，称 $f(x)$ 为 X 的密度函数或分布密度. 密度函数具有如下两个性质：

(1) $f(x) \geqslant 0, x \in (-\infty, +\infty)$；

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

只要得到了连续型随机变量 X 的密度函数 $f(x)$，就可以解决任何事件的概率计算问题，比如

$$P(a < X < b) = \int_a^b f(x)\mathrm{d}x, \quad \text{对 } a, b \text{ 取任意实数} \tag{1.2.3}$$

1.2.3　一维随机变量的分布函数

设 X 为随机变量，x 为任意实数，称 $F(x)=P(X\leqslant x)$ 为随机变量 X 的分布函数，简称分布. 分布函数 $F(x)$ 与离散型随机变量 X 的分布列以及连续型随机变量 X 的密度函数 $f(x)$ 有如下关系：

$$F(x) = P(X \leqslant x) = \begin{cases} \sum\limits_{x_k \leqslant x} P(X = x_k), & X \text{ 是离散随机变量} \\ \int_{-\infty}^{x} f(x)\mathrm{d}x, & X \text{ 是连续随机变量} \end{cases} \tag{1.2.4}$$

分布函数 $F(x)$ 有如下性质：

(1) $0\leqslant F(x)\leqslant 1$；

(2) $F(-\infty)=0, F(+\infty)=1$；

(3) $F(x)$ 是单调递增函数；

(4) $F(x)$ 是右连续函数.

特别地，连续型随机变量 X 的分布有如下性质：

(1) 若 $f(x)$ 在 x 点连续，则 $F(x)$ 在 x 点可导，且 $F'(x)=f(x)$；

(2) $F(x)$ 是连续函数.

例 1.2.1　现有 7 件产品，其中一等品 4 件，二等品 3 件，从中任取 3 件(不放回抽取). 求：

(1) 抽取的 3 件产品中含一等品件数 X 的分布列；

(2) X 的分布函数；

(3) 抽取的 3 件产品中至少含一件一等品的概率.

解　(1) X 的可能取值为：0,1,2,3，利用古典概率及性质计算：

$$P(X=0) = \frac{C_4^0 C_3^3}{C_7^3} = \frac{1}{35}, \quad P(X=1) = \frac{C_4^1 C_3^2}{C_7^3} = \frac{12}{35}$$

$$P(X=2) = \frac{C_4^2 C_3^1}{C_7^3} = \frac{18}{35}, \quad P(X=3) = \frac{C_4^3 C_3^0}{C_7^3} = \frac{4}{35}$$

X 的分布列为

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{35} & \frac{12}{35} & \frac{18}{35} & \frac{4}{35} \end{bmatrix}$$

显然，该分布列满足性质：$p_k \geqslant 0(k=1,2,\cdots)$ 和 $\sum\limits_{k=1}^{\infty} p_k = 1$.

(2) 根据分布函数定义,即参照公式(1.2.4),计算 X 的分布函数如下:

$$F(x)=P(X\leqslant x)=\begin{cases}0, & x<0\\ \dfrac{1}{35}, & 0\leqslant x<1\\ \dfrac{13}{35}, & 1\leqslant x<2\\ \dfrac{31}{35}, & 2\leqslant x<3\\ 1, & x\geqslant 3\end{cases}$$

(3) 所求概率可以表示为$\{X\geqslant 1\}$,即

$$P(X\geqslant 1)=1-P(X<1)=1-P(X=0)=\frac{34}{35}$$

例 1.2.2 设连续型随机变量 X 的分布函数为

$$F(x)=\begin{cases}Ae^{x}, & x<0\\ B, & 0\leqslant x<1\\ 1-Ae^{-(x-1)}, & x\geqslant 1\end{cases}$$

求:(1)常数 A,B;(2)X 的密度函数;(3)概率 $P\left(X>\dfrac{1}{3}\right)$.

解 (1) 由于 X 是连续型随机变量,故 $F(x)$是连续函数,所以 $F(x)$在 $x=0$ 和 $x=1$ 两点连续,由

$$\lim_{x\to 0^-}F(x)=\lim_{x\to 0^-}Ae^{x}=A,\quad \lim_{x\to 0^+}F(x)=\lim_{x\to 0^+}B=B$$

可知 $A=B$. 又由于

$$\lim_{x\to 1^-}F(x)=\lim_{x\to 1^-}B=B$$

$$\lim_{x\to 1^+}F(x)=\lim_{x\to 1^+}(1-Ae^{-(x-1)})=1-A$$

可知 $B=1-A$,故有 $A=B=\dfrac{1}{2}$,于是

$$F(x)=\begin{cases}\dfrac{1}{2}e^{x}, & x<0\\ \dfrac{1}{2}, & 0\leqslant x<1\\ 1-\dfrac{1}{2}e^{-(x-1)}, & x\geqslant 1\end{cases}$$

(2) 因为 X 的密度函数在连续点处与分布函数有关系 $f(x)=F'(x)$,所以 X 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2}e^{x}, & x<0\\ 0, & 0\leqslant x<1\\ \dfrac{1}{2}e^{-(x-1)}, & x\geqslant 1\end{cases}$$

(3) $P\left(X>\dfrac{1}{3}\right)=1-P\left(X\leqslant\dfrac{1}{3}\right)=1-F\left(\dfrac{1}{3}\right)=1-\dfrac{1}{2}=\dfrac{1}{2}$

或

$$P\left(X>\frac{1}{3}\right)=\int_{\frac{1}{3}}^{+\infty}f(x)\mathrm{d}x=\int_{\frac{1}{3}}^{1}0\mathrm{d}x+\int_{1}^{+\infty}\frac{1}{2}e^{-(x-1)}\mathrm{d}x=\frac{1}{2}$$

1.2.4　常见的分布列或密度函数

1. 二项分布

若随机变量 X 满足 $P(X=k)=C_n^k p^k q^{n-k}$，$k=0,1,\cdots,n$，其中 $0<p<1$，$q=1-p$，则称 X 服从参数为 (n,p) 的二项分布，记作 $X\sim B(n,p)$. 当 $n=1$ 时，称 $B(1,p)$ 为两点分布（或 0-1 分布）. 图 1.2.1 给出了 $B(10,0.2)$ 的概率分布图形.

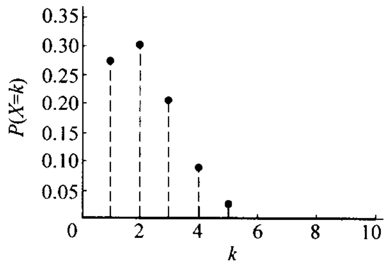

图 1.2.1　$B(10,0.2)$ 概率分布

二项分布的实际背景：在 n 重 Bernoulli 试验（n 次重复独立试验，每次试验只有两种可能结果）中，若每次试验事件 A 出现的概率为 p，则事件 A 出现的次数 X 就服从参数为 (n,p) 的二项分布. 例如在产品检验中，观察次品数 X 的分布状况；又如观察某幢大楼电梯的运行状况，需要了解每年电梯出故障次数 X 的分布情况等.

二项分布具有线性可加性：若 $X_1\sim B(n_1,p)$，$X_2\sim B(n_2,p)$，并且 X_1 与 X_2 独立，则 $X_1+X_2\sim B(n_1+n_2,p)$；特别地，若 $X_1,X_2,\cdots,X_n$ 独立同分布，均服从 0-1 分布，则 $\sum\limits_{i=1}^{n}X_i\sim B(n,p)$.

2. Poisson 分布

若随机变量 X 满足 $P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,2,\cdots$，$\lambda>0$，则称 X 服从参数为 λ 的 Poisson 分布，记作 $X\sim P(\lambda)$. 图 1.2.2 给出了 $\lambda=1$ 的 Poisson 分布图形.

常见的例子有：一块钢板上出现的气泡数 X，一本书上的印刷错误数 X，某超市排队等候的人数 X 等都服从 Poisson 分布.