研究生教材

YANJIUSHENGJIAOCAI

应用数理统计

YINGYONGSHULITONGJI

主编　唐湘晋　陈家清　毛树华

武汉理工大学出版社
WUTP Wuhan University of Technology Press

研究生教材

应用数理统计

主　编　唐湘晋　陈家清
　　　　毛树华
参　编　李　丹

武汉理工大学出版社
·武　汉·

内 容 提 要

本书是为高等院校非数学专业高年级学生和研究生编写的教材。内容包括概率论基础知识、统计量与抽样分布、参数估计、假设检验、贝叶斯统计及决策理论、试验设计与方差分析等。

本书可作为高等院校工科类以及管理、经济与金融类本科生、研究生的教材，也可供从事相关工作的技术人员参考。

图书在版编目(CIP)数据

应用数理统计/唐湘晋，陈家清，毛树华主编．—武汉：武汉理工大学出版社，2013.10

ISBN 978-7-5629-4165-1

Ⅰ.①应…

Ⅱ.①唐…　②陈…　③毛…

Ⅲ.①数理统计-高等学校-教材

Ⅳ.①O212

中国版本图书馆 CIP 数据核字(2013)第 219247 号

项目负责人：陈军东　彭佳佳　　**责任编辑**：彭佳佳

责 任 校 对：段　智　　**装帧设计**：芳华时代

出 版 发 行：武汉理工大学出版社

社　　　址：武汉市洪山区珞狮路 122 号

邮　　　编：430070

网　　　址：http://www.techbook.com.cn

经　　　销：各地新华书店

印　　　刷：荆州市鸿盛印务有限公司

开　　　本：787×960　1/16

印　　　张：15

字　　　数：310 千字

版　　　次：2013 年 10 月第 1 版

印　　　次：2013 年 10 月第 1 次印刷

印　　　数：1—2000 册

定　　　价：23.00 元

凡购本书，如有缺页、倒页、脱页等印装质量问题，请向出版社发行部调换。

本社购书热线电话：027-87523148　87664138　87515798　87165708(传真)

前　　言

数理统计学是一门应用性很强的学科，其理论和方法已被广泛应用于自然科学、工程技术、社会科学、经济与金融以及人文科学等各个领域。而计算机的不断普及和信息技术的飞速发展又为数理统计的应用注入了新的活力，同时也为数理统计的理论与方法提供了更加广阔的应用空间；数理统计学已成为数据处理、科学决策的重要理论和方法。因此，要想更好地处理大量的数据并从中得出有助于科学决策的定量化结论，就必须学习和运用数理统计的理论与方法。

本书是根据全国工科院校硕士研究生"数理统计"课程的教学基本要求而编写的，在编写过程中充分汲取了在研究生数理统计教学实践中积累的宝贵经验。在教材的内容选取方面，既涵盖了数理统计基本内容、基本思想和基本方法，又包括那些既有深刻的理论意义、又有重要实用价值的数理统计的概念和方法。本书在编写过程中尽量做到从实际出发，注重概念与定理的直观描述和实际背景，强调数理统计方法的具体应用，通过典型实例的分析来介绍方法，培养学生应用数理统计的理论与方法解决实际问题的能力。

参加本书编写的人员有唐湘晋、陈家清、毛树华和李丹，他们都是多年来从事研究生、本科生概率统计教学工作，具有丰富教学实践经验的教师。

这里特别感谢武汉理工大学研究生院培养处的各位同仁，他们资助和鼓励编者完成了这本书的编写；王卫华教授认真审阅了书稿，并提出了许多宝贵意见，在此表示衷心感谢。同时还要感谢武汉理工大学出版社的支持和帮助！

书中难免会有不妥之处，欢迎广大读者批评指正。

编　　者

2013 年 8 月于马房山

目　　录

第 1 章　概率论基础知识 ………………………………………… (1)

1.1　随机事件与概率 ………………………………………… (1)

1.2　计数技术和概率的计算 ………………………………… (2)

1.3　条件概率、贝叶斯公式及事件的独立性 ……………… (5)

1.4　随机变量和概率分布 …………………………………… (6)

1.5　矩和矩母函数 …………………………………………… (11)

1.6　常见的分布 ……………………………………………… (16)

1.7　协方差与相关系数 ……………………………………… (25)

1.8　随机变量的函数 ………………………………………… (27)

习题 1 ………………………………………………………… (33)

第 2 章　统计量与抽样分布 ………………………………… (35)

2.1　总体和样本 ……………………………………………… (35)

2.2　统计量 …………………………………………………… (37)

2.3　样本数字特征 …………………………………………… (41)

2.4　三类重要分布 …………………………………………… (45)

2.5　抽样分布 ………………………………………………… (49)

习题 2 ………………………………………………………… (57)

第 3 章　参数估计 …………………………………………… (60)

3.1　矩法估计和极大似然估计 ……………………………… (61)

3.2　点估计量的优良性 ……………………………………… (69)

3.3　有效估计与一致最小方差无偏估计 …………………… (75)

3.4　区间估计 ………………………………………………… (83)

习题 3 ………………………………………………………… (95)

第 4 章　假设检验 …………………………………………… (99)

4.1　假设检验的概念与步骤 ………………………………… (99)

4.2　正态总体参数的假设检验 ……………………………… (105)

4.3　计数的卡方检验 ………………………………………… (120)

4.4　似然比检验 ……………………………………………… (129)

习题 4 …… (132)
第 5 章 贝叶斯统计及决策理论 …… (136)
5.1 引言 …… (136)
5.2 贝叶斯点估计 …… (138)
5.3 贝叶斯置信区间或可信区间 …… (148)
5.4 贝叶斯假设检验 …… (151)
5.5 贝叶斯决策理论 …… (153)
习题 5 …… (174)
第 6 章 试验设计与方差分析 …… (177)
6.1 试验设计的基本概念 …… (177)
6.2 完全随机设计 …… (180)
6.3 随机单位组设计 …… (182)
6.4 拉丁方设计 …… (185)
6.5 因子设计 …… (188)
6.6 优化设计 …… (190)
6.7 方差分析的基本原理 …… (191)
6.8 单因素方差分析 …… (191)
6.9 两因素方差分析 …… (204)
习题 6 …… (214)
附录 1 标准正态分布表 …… (217)
附录 2 χ^2 分布表 …… (219)
附录 3 F 分布临界值表 …… (222)
附录 4 t 分布表 …… (230)

第 1 章　概率论基础知识

概率论是数理统计的理论基础，为了使它们能更好地衔接起来，本章扼要地阐述了概率论的基本概念、定理与公式，并补充了特征函数等工程数学中选修的内容.

1.1　随机事件与概率

1.1.1　随机现象

在一定条件下时而出现这样的结果，时而又出现那样的结果，而且事先无法断言出现的究竟是哪一种结果，这类现象就称为随机现象.

一般来说，随机现象具有两重性：表面上的偶然性与内部蕴含着的必然性. 随机现象的偶然性又称为它的随机性. 在一次试验或观察中，结果的不确定性就是随机现象的一面；在相同的条件下进行大量重复试验或观察时呈现出来的规律性是随机现象的必然性的一面，称随机现象的必然性为统计规律性.

1.1.2　随机试验

若试验具有下列共同特征：

(1) 在相同的条件下，试验可重复进行；

(2) 试验的一切可能结果是预先可以明确的，但每次试验前无法预先断言究竟会出现哪个结果.

则称之为随机试验，简称试验，记作 E 或 E_1、E_2 等.

1.1.3　样本空间

对于随机试验 E，以 ω 表示它的一个可能出现的试验结果，称 ω 为 E 的一个样本点. 样本点的全体称为样本空间，用 Ω 表示，即 $\Omega = \{\omega\}$.

从集合论的观点看,样本空间 Ω 是由一切可能的结果所构成的集合,而每个样本点 ω 是集合 Ω 中的元素.

1.1.4 随机事件

为了便于直观理解,不妨在这一小节先设样本空间 $\Omega = \{\omega\}$ 为可列集. 通常,对于某个随机试验来说,在一次试验中可能出现也可能不出现的事件,就称为随机事件,用大写英文字母 A、B、C、A_i 等来表示. 在引入了样本空间的定义后,可从集合论的观点看:粗略地说,样本空间 Ω 的子集就是随机事件.

1.1.5 随机事件的概率

随机事件在一次试验中,可能发生也可能不发生,具有偶然性. 但是,人们从实践中认识到,在相同的条件下,进行的大量的重复试验中,试验的结果具有某种内在的规律性,即随机事件发生的可能性大小是可以比较的,是可以用一个数字进行度量的. 例如,在投掷一枚均匀骰子的试验中,对于事件 A:"掷出偶数点",B:"掷出2点",显然事件 A 比事件 B 发生的可能性要大.

对于一个随机试验,我们不仅要知道它可能出现哪些结果,更重要的是还要研究各种结果发生的可能性的大小,从而揭示其内在的规律性.

概率就是随机事件发生的可能性大小的数量表征. 对于事件 A,通常用 $P(A)$ 来表示事件 A 发生的可能性大小,即 A 发生的概率. 但是,如何对事件的概率进行定义呢?在下一节将给出具体表述.

1.2 计数技术和概率的计算

1.2.1 频率及其性质

1. 频率的定义

定义 1.1 在相同的条件下,重复进行了 N 次试验,若事件 A 发生了 μ 次,则称比值 $\frac{\mu}{N}$ 为事件 A 在 N 次试验中出现的频率,记为 $f_N(A) = \frac{\mu}{N}$.

2. 频率的性质

① 非负性:对任意 A,有 $f_N(A) \geqslant 0$;

② 规范性：$f_N(\Omega)=1$；

③ 可加性：若 A、B 互斥，则 $f_N(A\cup B)=f_N(A)+f_N(B)$.

3. 频率的稳定性

大量的重复试验中，频率常常稳定于某个常数，称为频率的稳定性.

通过大量的实践，易知，若随机事件 A 出现的可能性越大，一般来讲，其频率 $f_N(A)$ 也越大. 由于事件 A 发生的可能性大小与其频率大小有如此密切的关系，且频率又具有稳定性，故而可通过频率来定义概率.

1.2.2 概率的统计定义（基于计数技术的概率定义）

定义 1.2 在相同的条件下，独立重复地做 N 次试验，当试验次数 N 很大时，如果某事件 A 发生的频率 $f_N(A)$ 稳定地在 $[0,1]$ 上的某一数值 p 附近摆动，而且一般来说随着试验次数的增多，这种摆动的幅度会越来越小，则称数值 p 为事件 A 发生的概率，记为 $P(A)=p$.

概率的统计定义一方面肯定了任一事件的概率是存在的；另一方面又给出了一个近似计算概率的方法，但其不足之处是要进行大量的重复试验.

注：$\lim\limits_{N\to\infty} f_N(A)\neq p$.

1.2.3 古典概型

若随机试验 E 及其样本空间 Ω 有下列特性：

(1) 样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$ 只有有限个样本点；

(2) 试验中每个样本点出现的可能性相同（等可能性）；

则称定义在该样本空间 Ω 上的概率模型为古典概型.

定义 1.3 设随机试验 E 为古典概型，其样本空间为 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$；对于任一事件 A，其概率定义为：

$$P(A)=\frac{A\text{ 中包含的样本点数}}{\Omega\text{ 中包含的样本点数}}$$

1.2.4 几何概型

若随机试验具有特性：

① 试验的结果是无限且不可列的；

② 每个结果出现的可能性相同（等可能性）；

则称其为几何概型.在几何概型中,我们通过几何度量(长度、面积、体积等)来计算事件出现的可能性.

定义 1.4 设随机试验 E 的样本空间 Ω 是 R^n 中的可测子集,具有有限的测度 $\mu(\Omega)>0$(记号 $\mu(\Omega)$ 表示集合 Ω 的测度).若 $A\subset\Omega$,记 $P(A)=\dfrac{\mu(A)}{\mu(\Omega)}$,称 A 为事件,$P(A)$ 为事件 A 的概率.

1.2.5 概率的公理化定义

定义 1.5 设 E 为一个随机试验,Ω 为它的样本空间,以 E 中所有的事件组成的集合为定义域,定义一个函数 $P(A)$(其中 A 为任意一个随机事件),且 $P(A)$ 满足:

① 非负性:对任一事件 A,$P(A)\geqslant 0$;

② 规范性:$P(\Omega)=1$;

③ 可列可加性:若事件 $A_i(i=1,2,\cdots)$ 两两互斥,即满足 $A_iA_j=\varnothing(i\neq j)$,则

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$$

则称 $P(A)$ 为随机事件 A 的概率.

1.2.6 概率的性质

从概率的公理化定义,我们可以推出概率的性质:

① $P(\varnothing)=0$;

② $P(\overline{A})=1-P(A)$;

③ 有限可加性:若事件 $A_1,A_2,\cdots,A_n$ 互不相容,则

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)$$

④ 设 A,B 为任意两个事件,则

i. $P(A\backslash B)=P(A)-P(AB)$;

ii. $P(A\cap B)=P(A)+P(B)-P(AB)$;

iii. 若 $A\subset B$,则 $P(A)\leqslant P(B)$;

⑤ 设 $A_1,A_2,\cdots,A_n$ 为任意 n 个事件,则

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

推论 ① $P(\bigcup_{k=1}^{n} A_k) \leqslant \sum_{k=1}^{n} P(A_k)$（有限次可加性）

② $P(\bigcup_{k=1}^{n} A_k) \geqslant \sum_{k=1}^{n} P(A_k) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j)$

1.3 条件概率、贝叶斯公式及事件的独立性

1.3.1 条件概率

定义 1.6 设 A, B 是两个随机事件，且 $P(B) > 0$，称 $P(A \mid B) = P(AB)/P(B)$ 为在事件 B 发生条件下事件 A 发生的条件概率.

易验证上面定义的条件概率满足概率公理化定义，因此它具有概率的相应性质：

① $P(\varnothing \mid B) = 0$

② $P(\overline{A} \mid B) = 1 - P(A \mid B)$

③ $P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 A_2 \mid B)$

1. 乘法公式

由条件概率的定义：

$$P(A \mid B) = P(AB)/P(B) \Rightarrow P(AB) = P(B)P(A \mid B) \quad (P(B) > 0)$$

$$P(B \mid A) = P(AB)/P(A) \Rightarrow P(AB) = P(A)P(B \mid A) \quad (P(A) > 0)$$

定理 1.1 一般地，对任意 n 个事件 $A_1, A_2, \cdots, A_n$，若 $P(A_1 A_2 \cdots A_n) > 0$，则

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 A_2) \cdots P(A_n \mid A_1 A_2 \cdots A_{n-1})$$

2. 全概率公式

定理 1.2 若 $B_1, B_2, \cdots, B_n$ 为样本空间 Ω 的一个分解，即：① $B_1, B_2, \cdots, B_n$ 两两互斥；② $\bigcup_{i=1}^{n} B_i = \Omega$. 设 A 为任意一个事件，则

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i)$$

上述公式称为全概率公式.

3. 贝叶斯公式(Bayes 公式)

定理 1.3　若 $B_1, B_2, \cdots, B_n$ 为样本空间 Ω 的一个分解，A 为任意一个事件，则

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}$$

上述公式称为贝叶斯公式.

1.3.2　事件的独立性

定义 1.7　若事件 A、B 满足：$P(AB) = P(A)P(B)$，则称 A、B 相互独立.

显然，下列命题等价：

① A 与 B 相互独立；

② A 与 $\bar{B}$ 独立；

③ $\bar{A}$ 与 B 独立；

④ $\bar{A}$ 与 $\bar{B}$ 独立；

⑤ $P(A \mid B) = P(A), (P(B) > 0)$；

⑥ $P(A \mid \bar{B}) = P(A), (P(\bar{B}) > 0)$；

⑦ $P(\bar{A} \mid \bar{B}) = P(\bar{A}), (P(\bar{B}) > 0)$；

⑧ $P(\bar{A} \mid B) = P(\bar{A}), (P(B) > 0)$.

定义 1.8　对任意 n 个事件 $A_1, A_2, \cdots, A_n$，若：

$$P(A_iA_j) = P(A_i)P(A_j), 1 \leqslant i < j \leqslant n$$

$$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k), 1 \leqslant i < j < k \leqslant n$$

$$\cdots\cdots$$

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)（共 2^n - n - 1 个式子）$$

均成立，则称 $A_1, A_2, \cdots, A_n$ 相互独立.

1.4　随机变量和概率分布

1.4.1　一维随机变量及其分布

1. 随机变量的概念及分类

定义 1.9 定义在样本空间 Ω 上的一个实值函数 $X = X(\omega)$，使随机试验的每一个结果 ω 都可用一个实数 $X(\omega)$ 来表示，且实数 X 满足：

① X 是由 ω 唯一确定；

② 对于任意给定的实数 x，事件 $\{X \leqslant x\}$ 都是有概率的.

则称 X 为一随机变量. 随机变量一般用大写字母 X,Y,Z 等表示.

2. 随机变量的分布函数及其性质

定义 1.10 设 X 为一随机变量，x 是任意实数，称函数

$$F(x) = P(X \leqslant x) \quad (-\infty < x < +\infty)$$

为随机变量 X 的分布函数.

分布函数是一个以全体实数为其定义域，以事件 $\{\omega \mid X(\omega) \leqslant x\}$ 的概率为函数值的一个实值函数. 分布函数具有以下基本性质：

① $0 \leqslant F(x) \leqslant 1$；

② $F(x)$ 是非减函数；

③ $F(x)$ 是右连续的；

④ $\lim\limits_{x\to-\infty} F(x) = 0$，$\lim\limits_{x\to+\infty} F(x) = 1$.

设随机变量 X 的分布函数为 $F(x)$，则可用 $F(x)$ 来表示下列概率：

① $P(X \leqslant a) = F(a)$；

② $P(X < a) = F(a-0)$；

③ $P(X > a) = 1 - P(X \leqslant a) = 1 - F(a)$；

④ $P(X \geqslant a) = 1 - P(X < a) = 1 - F(a-0)$；

⑤ $P(X = a) = P(X \leqslant a) - P(X < a) = F(a) - F(a-0)$；

⑥ $P(|X| < a) = P(-a < X < a) = P(X < a) - P(X \leqslant -a)$
$$= F(a-0) - F(-a).$$

3. 离散型随机变量

定义 1.11 如果随机变量 $X(\omega)$ 所有可能取值是有限个或可列多个，则称 $X(\omega)$ 为离散型随机变量.

定义 1.12 设离散型随机变量 $X(\omega)$ 所有可能取的值为 $x_k, k \in N$，$X(\omega)$ 取各个值的概率为：

$$P\{X = x_k\} = p_k, k \in N$$

称 $P\{X = x_k\} = p_k, k \in N$ 为 $X(\omega)$ 的概率分布或分布律.

易知 p_k 满足如下两个条件：

① $p_k \geqslant 0, k \in N$ （非负性）

② $\sum_{k=1}^{\infty} p_k = 1$ （规一性）

分布律也可以用表格形式来表示(表1.1)：

表1.1 分布律表

X	x_1	x_2	x_3	$\cdots$	x_k $\cdots$
p_i	p_1	p_2	p_3	$\cdots$	p_k $\cdots$

4. 连续型随机变量

定义1.13 若对于随机变量X，存在一定义在R上的非负函数$f(x)$，使对$\forall a \in R$，满足：

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

则称X为连续型随机变量；其中函数$f(x)$称为X的概率密度函数，简称概率密度.

由定义1.13易知，概率密度具有下列性质：

① $f(x) \geqslant 0$ （非负性）

② $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$ （规一性）

③ $P(a < X \leqslant b) = \int_{a}^{b} f(x)\mathrm{d}x \quad (a \leqslant b)$

④ 如果随机变量$X(\omega)$有概率密度$f(x)$，则$P\{\omega: X(\omega) = a\} = 0 (\forall a \in R)$

⑤ 若$f(x)$在x处连续，则$\lim\limits_{h \to 0} \dfrac{P(x < X \leqslant x + h)}{h} = f(x)$.

1.4.2 多维随机变量及其概率分布

在许多随机现象中，对试验的每个结果ω只用一个随机变量$X(\omega)$去描述是不够的，而是需要同时用多个随机变量去描述. 因此，我们引入多维随机变量的概念.

定义1.14 设$X_1, X_2, \cdots, X_n$都为试验E的样本空间Ω上的随机变量，则称$(X_1, X_2, \cdots, X_n)$为Ω上的n维随机变量或n维随机向量. 称

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}, (x_1, x_2, \cdots, x_n) \in R^n$$

为n维随机变量$(X_1, X_2, \cdots, X_n)$的联合分布函数. 称$F_{X_i}(x) = F(+\infty, \cdots, +\infty, x, +\infty, \cdots, +\infty)$为随机变量$X_i$的(一维)边缘分布函数.

重点了解和掌握二维随机变量(X, Y). 对于二维随机变量(X, Y)，只讨论离散型和连续型两大类.

1. 二维离散型随机变量及其分布

(1) 联合分布律

定义 1.15 若二维随机变量(X,Y)可能取的值(向量)是有限多个或可列无穷多个,则称(X,Y)为二维离散型随机变量.

设二维离散型随机变量(X,Y)可能取的值为$(x_i,y_j)(i=1,2,\cdots;j=1,2,\cdots)$. 则取这些值的概率为

$$p_{ij}=P\{X=x_i,Y=y_j\}\quad(i,j=1,2,\cdots)$$

称上式为(X,Y)的联合分布律.

(X,Y)的联合分布律可以用表格的形式表示(表 1.2).

表 1.2 联合分布律表

$X\backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$

这里 p_{ij} 具有下面两个性质:

① $p_{ij}\geqslant 0(i,j=1,2,\cdots)$;

② $\sum\limits_i\sum\limits_j p_{ij}=1$.

知道了(X,Y)的联合分布律以后,可以求其联合分布函数

$$F(x,y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}p_{ij}$$

(2) 边缘分布律

设二维离散型随机变量(X,Y)的联合分布律为

$$p_{ij}=P\{X=x_i,Y=y_j\}\quad(i,j=1,2,\cdots)$$

则称

$$p_{i\cdot}=P\{X=x_i\}=\sum_j P\{X=x_i,Y=y_j\}=\sum_j p_{ij}\quad(i=1,2,\cdots)$$

和

$$p_{\cdot j}=P\{Y=y_j\}=\sum_i P\{X=x_i,Y=y_j\}=\sum_i p_{ij}\quad(j=1,2,\cdots)$$

分别为(X,Y)关于 X 和 Y 的边缘分布律,简称为(X,Y)的边缘分布律.

X 和 Y 的边缘分布律满足:

① $p_{i\cdot} \geqslant 0, p_{\cdot j} \geqslant 0$；

② $\sum_i p_{i\cdot} = 1, \sum_j p_{\cdot j} = 1$.

2. 二维连续型随机变量及其分布

(1) 联合概率密度

定义 1.16 设 $F(x,y)$ 为二维随机变量 (X,Y) 的分布函数，若非负函数 $f(x,y)$ 使得

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\mathrm{d}x\mathrm{d}y, (x,y) \in R^2$$

则称 (X,Y) 为二维连续型随机变量，且称 $f(x,y)$ 为二维连续型随机变量 (X,Y) 的联合概率密度函数.

分布密度函数满足：

① $f(x,y) \geqslant 0$；

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

③ 在 $f(x,y)$ 的连续点 (x,y) 处有 $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$；

④ 对于坐标平面上的区域 D，$P\{(x,y) \in D\} = \iint\limits_D f(x,y)\mathrm{d}x\mathrm{d}y$.

(2) 边缘概率密度

设二维连续型随机变量 (X,Y) 的联合分布函数、联合概率密度依次为 $F(x,y)$、$f(x,y)$，分量 X、Y 的边缘分布函数分别为 $F_X(x)$、$F_Y(y)$. 利用边缘分布函数与联合分布函数的关系，可得

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)\mathrm{d}y\right]\mathrm{d}u$$

$$F_Y(y) = F(+\infty, y) = \int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(x,v)\mathrm{d}x\right]\mathrm{d}v$$

则称 $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$ 为二维随机变量 (X,Y) 关于 X 的边缘概率密度；

$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x$ 为二维随机变量 (X,Y) 关于 Y 的边缘概率密度.

1.5 矩和矩母函数

1.5.1 随机变量的数学期望

1. 离散型随机变量的数学期望

定义 1.17 设 X 为离散型随机变量，其分布列为：$\begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$，若 $\sum_{i=1}^{\infty}|x_i|p_i<+\infty$，则称 $EX=\sum_{i=1}^{\infty}x_ip_i$ 为 X 的数学期望或均值.

注：为使 EX 与级数各项的次序无关，必须要求 $\sum_{i=1}^{\infty}|x_i|p_i$ 收敛；否则，EX 不存在.

【例 1.1】 设用一个匀称的骰子来玩游戏.在这样的游戏中，将此骰子抛掷一次，若骰子向上为 2，则玩游戏的人赢 20 元，若向上为 4 则赢 40 元，若向上为 6 则输 30 元，若其他的面向上，则玩游戏的人既不赢也不输，求玩游戏的人赢得钱数的期望.

解 令 X 为任何一次抛掷中赢得钱数，则 $X\sim\begin{pmatrix} 0 & 20 & 40 & -30 \\ 1/2 & 1/6 & 1/6 & 1/6 \end{pmatrix}$，则由离散型随机变量数学期望的定义可知：

$$EX=0\times 1/2+20\times 1/6+40\times 1/6-30\times 1/6=5$$

从而玩游戏的人可期望赢 5 元.因此，在一个公正的游戏中，玩游戏的人为了参加游戏应当付 5 元底金.

2. 连续型随机变量的数学期望

由离散型随机变量数学期望的定义，我们自然可以设想取很密的分点 $x_0<x_1<\cdots<x_n$，将$(-\infty,+\infty)$分割成 n 个小区间，则 X 落在$(x_i,x_{i+1}]$内的概率等于

$$P\{x_i<X\leqslant x_{i+1}\}=\int_{x_i}^{x_{i+1}}f(x)\mathrm{d}x=f(\xi)(x_{i+1}-x_i)\text{（其中 }\xi\text{ 落在 }x_i\text{ 与 }x_{i+1}\text{ 之间）}$$

$$\approx f(x_{i+1})(x_{i+1}-x_i)\text{（积分中值定理）}$$

因此当 $n\to\infty$ 时，X 与以概率 $f(x_{i+1})(x_{i+1}-x_i)$ 取值 x_{i+1} 的离散型随机变量相似，而后者的数学期望为 $\sum_i x_{i+1}f(x_{i+1})(x_{i+1}-x_i)$，并且这个和式的极限就是 $\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$.