复习&答疑

王笑尘

北京邮电大学网络空间安全学院

wxiaochen@bupt.edu.cn

课程考核

- **期末考试：** 统一执行闭卷书面考试。
  - 涵盖课程所有内容；
  - 考试时间：**2024年12月31日 8:00-10:00（不要错过！）**
  - 考试地点：沙河校区N318
  - 一般包括**填空题**10道（共计40分）；**计算题**5-6道，每道计算题包含1-3个小题（共计60分）。

课程内容—概率论

教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—数理统计

教学内容	知识点
第六章 数理统计部分（样本及抽样分布）	总体、样本统计量，χ^2分布、t分布、F分布，抽样分布，顺序统计量
第七章 参数估计	矩估计及极大似然估计，估计量的评选标准，区间估计
第八章 假设检验	假设检验问题与基本概念，正态总体期望的假设检验，正态总体方差的假设检验
第九章 方差分析及回归分析	单因素方差分析，双因素方差分析，一元线性回归，多元线性回归

课程内容—第一章

教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—第一章

研究随机现象

随机试验所有可能结果组成的集合

样本空间的子集

基本概念与运算

（随机试验，样本空间，事件）

1、可重复
2、事先明确所有可能结果
3、每次试验结果不确定

a. 事件关系（包含、并、交、差、互斥、逆）
b. 事件运算（交换、结合、分配、德摩根律）

频率与概率

统计定义

公理化定义

古典定义

课程内容—第一章

★ 事件运算所满足的下述定律：

1．交换律：$A\cup B=B\cup A, A\cap B=B\cap A$

2．结合律：

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

$$A\cap(B\cap C)=(A\cap B)\cap C$$

3．分配律：

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$$

4．德摩根律：

$$\overline{A\cup B}=\overline{A}\cap\overline{B}$$

$$\overline{A\cap B}=\overline{A}\cup\overline{B}$$

课程内容—第一章

基本概念与运算

（随机试验，样本空间，事件）

频率与概率

- 统计定义
  - 频率的稳定性
- 公理化定义
  - 1、非负性
    2、规范性
    3、可列可加性（两两互不相容）
  - 概率性质、多个事件加法定理
- 古典定义
  - 1、元素有限性
    2、等可能性
  - 排列组合：所有/某条件下基本事件数

课程内容—第一章

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

性质、
计算（定义、缩减样本空间）、
乘法定理（多个事件积事件）

古典定义

条件概率

独立性

全概率公式与贝叶斯公式

用条件概率求非条件概率：
1、样本空间划分：互斥完备组
2、基于划分，综合加法公式和乘法公式，

$$P(A) = P(A|B_1)\cdot P(B_1) + P(A|B_2)\cdot P(B_2) + \cdots$$
$$= P(A|B_n)\cdot P(B_n) = \sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)$$

已知结果求原因：用条件概率求条件概率

贝叶斯公式 $P(B_i|A) = \dfrac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)},$

课程内容—第一章

- 古典定义
  - 条件概率
    - **独立性**
  - 全概率公式与贝叶斯公式

1、P(AB)=P(A)P(B)
2、P(A)>0,P(B)>0,独立和互不相容：不同时成立
3、区分两两独立和相互独立
4、用独立的乘积特性、德摩根律求独立事件和的概率

课程内容—第一章

▲ 相互独立与两两独立的关系：

两两独立——n 个事件任何两个彼此独立

相互独立—— n个事件任意 k个 $(k \le n)$都是独立的

故相互独立 $\Rightarrow$ 两两独立, 反之则不真

课程内容—第一章

▲ n 个独立事件**和** 的概率公式:

设事件 $A_1, A_2, \cdots, A_n$ 相互独立, 则

$$P(A_1+\ldots+A_n) = 1 - P(\overline{A_1 + A_2 + \cdots + A_n})$$

$$= 1 - P(\overline{A}_1\overline{A}_2 \cdots \overline{A}_n)$$

$$= 1 - P(\overline{A}_1)P(\overline{A}_2)\cdots P(\overline{A}_n)$$

由德摩根律

$\overline{A}_1, \overline{A}_2, \cdots, \overline{A}_n$ 也相互独立

▲ 类似地, $A_1, A_2, \cdots, A_n$ 至少有一个不发生的概率（1-都发生的概率）：

$$P(\overline{A}_1 + \overline{A}_2 + \cdots + \overline{A}_n) = 1 - P(A_1)P(A_2)\cdots P(A_n)$$

课程内容—第二章

教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—第二章

- 随机变量
  - 离散型随机变量
    - 分布律
      - 定义
      - 常用分布
    - 分布函数
    - 函数的分布
  - 连续型随机变量
    - 概率密度
      - 定义
      - 常用分布
    - 分布函数
    - 函数的分布

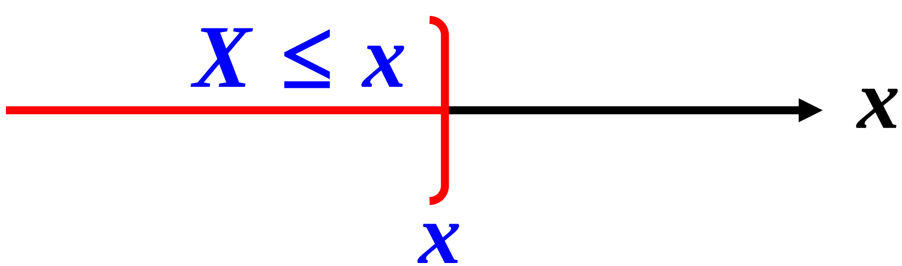
课程内容—第二章

1. 定义：设 X 是一个随机变量，x是任意实数，称函数：

$$\boxed{F(x)=P(X\le x)} \quad (-\infty<x<+\infty)$$

为 X 的分布函数。 记作：$X \sim F(x)$ 或 $F_X(x)$.

注 如果将 X 看作数轴上随机点的坐标，则分布函数 $F(x)$ 的值就表示 X落在区间$(-\infty, x]$上的概率



课程内容—第二章

2. 性质

性质1 $F(x)$是一个不减函数，即若 $x_1 \le x_2$，

则：$F(x_1) - F(x_2) \le 0$

证：$F(x_2) - F(x_1) = P(x_1 < X \le x_2) \ge 0$,

即 $F(x_1) \le F(x_2)$

课程内容—第二章

性质2 $0 \le F(x) \le 1$ 且: $\begin{cases} F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0 \\ F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1 \end{cases}$

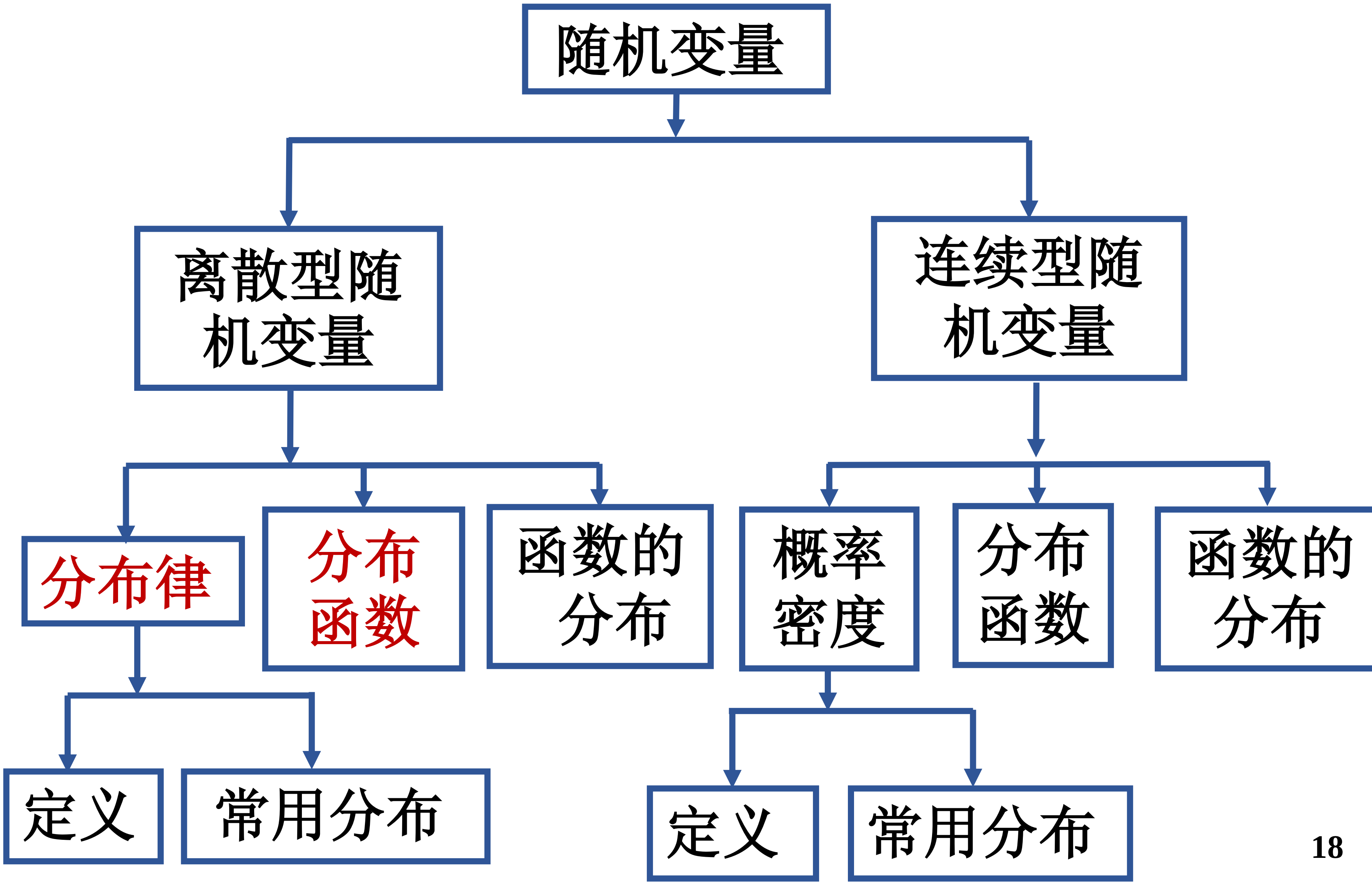
特别: 若 X 仅在 $(a,b]$ 内取值，则有:

$$F(a) = P(X \le a) = 0, \quad F(b) = P(X \le b) = 1$$

性质3 $F(x)$是右连续的函数，即 $\lim\limits_{x \to x_0^+} F(x) = F(x_0)$

反之，具备性质1、2、3的函数$F(X)$必是某个随机变量的分布函数

课程内容—第二章


随机变量
离散型随机变量
连续型随机变量
分布律
分布函数
函数的分布
概率密度
分布函数
函数的分布
定义
常用分布
定义
常用分布

课程内容—第二章

1.(0-1)分布

若随机变量X只能取 0 与 1 两个值，它的分布律为:

$$P(X=k)=p^k(1-p)^{(1-k)} \qquad k=0,1. \quad 0<p<1$$

则称 X 服从 (0-1)分布，记为：$X\sim(0,1)$

列表:

X	0	1
P_k	$1-p$	p

(2). 二项分布

若用X表示 n 重伯努利概型中事件A 发生的次数，它的分布律为：

$$P_n(k) = C_n^k p^k (1-p)^{n-k} \quad k = 0, 1, 2\cdots n$$

则称 X 服从参数为 n, p (0<p<1) 的二项分布，

记为: $X \sim b(n, p)$

分布律列表

X	0	1	$2 \cdots\cdots n$
$P_n(k)$	$P_n(0)$	$P_n(1)$	$P_n(2) \cdots\cdots P_n(n)$

3. 泊松分布

定理： 若随机变量X 的所有可能取值为: $0, 1, 2, \cdots\cdots$

而它的分布律(它所取值的各个概率)为:

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!} \quad k=0,1,2,\cdots$$

其中 $\lambda>0$ 是常数.

则称 X 服从参数为 λ 的 泊松分布，记为$X \sim P(\lambda)$

课程内容—第二章

- 随机变量
  - 离散型随机变量
    - 分布律
      - 定义
      - 常用分布
    - 分布函数
    - 函数的分布
  - 连续型随机变量
    - 概率密度
      - 定义
      - 常用分布
    - 分布函数
    - 函数的分布

$$F(x)=\int_{-\infty}^{x} f(t)dt \quad (=P(X\le x))$$

课程内容—第二章

1. 均匀分布

若连续型随机变量 X 具有概率密度 f (x)为:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其它}\end{cases}$$

则称 X 在区间 (a, b)上服从均匀分布 (或等概率分布)

注: ▲ 易证 $f(x)$ 满足:

$$1^0.\ f(x)\geq 0\,,\qquad 2^0.\int_{-\infty}^{+\infty} f(x)dx=1$$

2. 指数分布

若连续型随机变量 X 具有概率密度 f (x)为：

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & 其它\end{cases}$$

其中 $\theta>0$ 为常数

则称 X 为服从参数 θ 的指数分布

注：易证 $f(x)$ 满足：

$$1^0.\ f(x)\geq 0,\quad 2^0.\int_{-\infty}^{+\infty}f(x)dx=1$$

(1). 正态分布的定义

若随机变量 X 的概率密度为：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty$$

其中：μ和σ^2都是常数，μ任意，$\sigma>0$，
则称X服从参数为μ和σ^2的正态分布。记作：

$$X\sim N(\mu,\sigma^2)$$

$f(x)$ 所确定的曲线叫作正态曲线。

引理：若 $X\sim N(\mu,\sigma^2)$，则：$Z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$

课程内容—第二章

离散型随机变量

连续型随机变量

函数的分布

求Y=g(X)概率分布/概率密度的一般思路：

1、确定Y的取值/取值范围；

2、确定X对应的取值/范围；

3、求Y在某点/某范围取值的概率，即为求X在对应点/对应范围的取值概率(本质、理解)；

4、根据步骤3，即可得到离散型随机变量Y的分布律，或连续型随机变量Y的分布函数。

课程内容—第二章

北京邮电大学 Beijing University of Posts and Telecommunications

离散型随机变量

连续型随机变量

函数的分布

其中，对于连续型随机变量：
对步骤3得到的分布函数进行求导，即得到其概率密度函数。

定理： 设随机变量 X 具有概率密度 $f_X(x)$ $(-\infty < x < +\infty)$,又设函数 g(x) 处处可导，且恒有 $g'(x) > 0$ (或 $g'(x) < 0$)

则 Y= g(X) 是连续型随机变量，

其概率密度为:

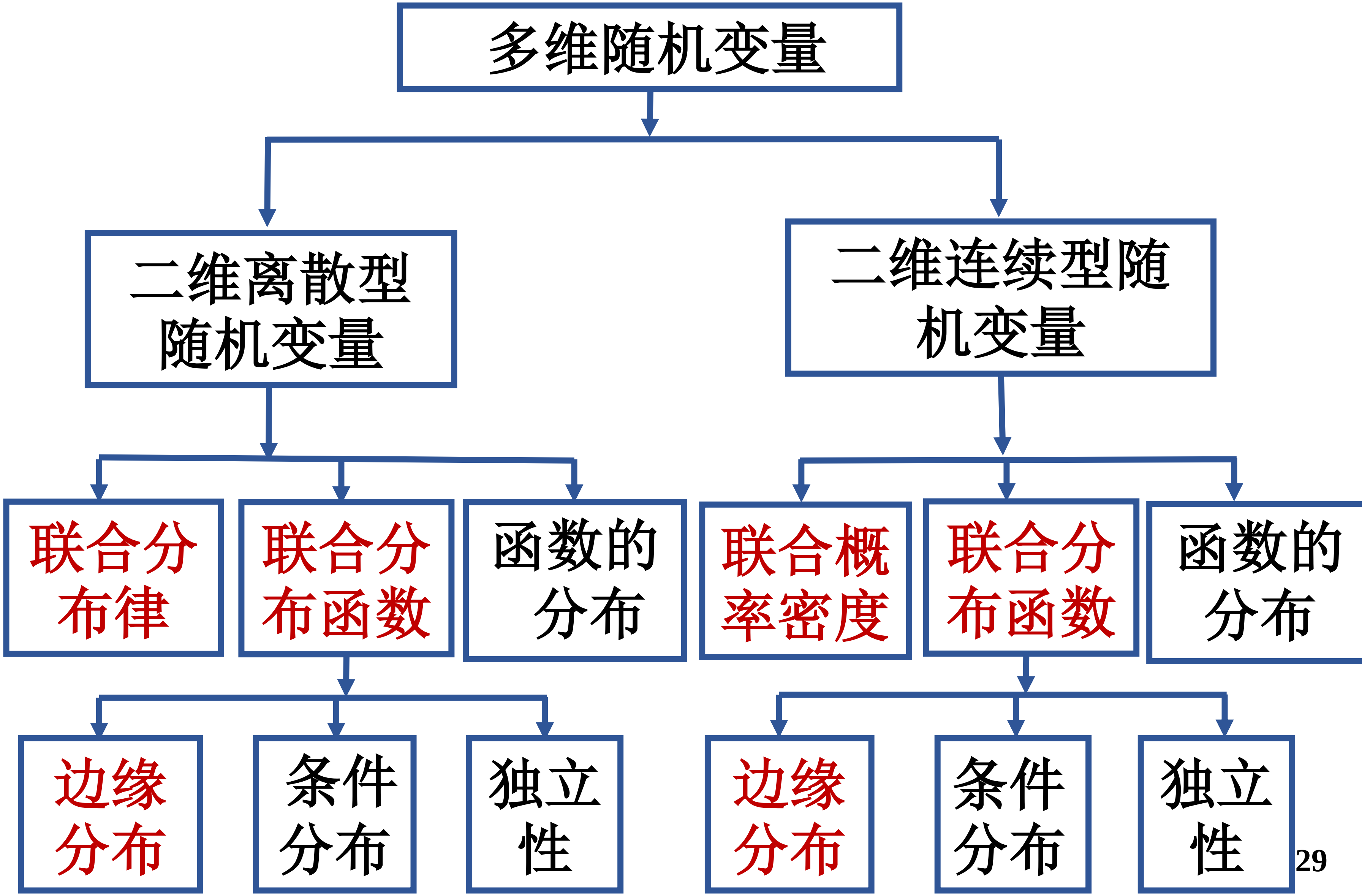
严格单调可微的函数

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & \alpha < y < \beta \\ 0 & \text{其它} \end{cases}$$

课程内容—第三章

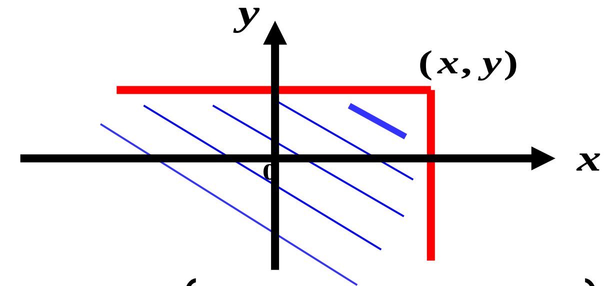
教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—第三章



多维随机变量
二维离散型随机变量
二维连续型随机变量
联合分布律
联合分布函数
函数的分布
联合概率密度
联合分布函数
函数的分布
边缘分布
条件分布
独立性
边缘分布
条件分布
独立性

课程内容—第三章



定义在同一样本空间的（X,Y）

$$F(x,y)=P\{(X\le x)\cap(Y\le y)\}$$

分布函数四个性质

（1）非减
（2）右连续
（3）[0,1]
（4）矩形区域内概率非负

多维随机变量

二维离散型随机变量

取值(x,y)只能是有限对或可列无限多对

$$P(X=x_i,Y=y_j)$$

$$=p_{ij}$$

$$i,j=1,2,\cdots\cdots$$

(1) $p_{ij}\ge 0$

(2) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1$

联合分布律

联合分布函数

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

课程内容—第三章

多维随机变量

二维连续型随机变量

联合概率密度

联合分布函数

4个性质：
（a）非负
（b）规范性
（c）分布函数与概率密度的关系（二阶偏导）
（d）某区域内概率的求解

$$\boxed{F(x,y)} = \int_{-\infty}^{y}\int_{-\infty}^{x} \boxed{f(u,v)}\,dudv$$

课程内容—第三章

已知 $P(X=x_i, Y=y_j)=p_{ij}$ 为(X,Y)的联合分布律

则 X 边缘分布函数 $\boxed{F_X(x)=F(x,+\infty)}=\sum\limits_{x_i\le x}\boxed{\sum\limits_{j=1}^{\infty}p_{ij}}$

边缘分布律 $p_{i\cdot}=P(X=x_i)=\sum\limits_{j=1}^{\infty}p_{ij} \quad i=1,2\cdots$

则 Y 边缘分布函数 $\boxed{F_Y(y)=F(+\infty,y)}=\sum\limits_{y_j\le y}\boxed{\sum\limits_{i=1}^{\infty}p_{ij}}$

边缘分布律 $p_{\cdot j}=P(Y=y_j)=\sum\limits_{i=1}^{\infty}p_{ij} \quad j=1,2\cdots$

$p_{i\cdot}$ 是由 p_{ij}关于 j 求和得到；$p_{\cdot j}$ 是由 p_{ij} 关于 i 求和得到。

课程内容—第三章

已知连续型随机变量(X,Y)的联合概率密度 $f(x,y)$

及联合分布函数 $F(x,y)$

则 X 的

边缘分布函数:

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\boxed{\left[\int_{-\infty}^{+\infty} f(x,y)dy\right]}dx$$

边缘概率密度:

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy$$

则 Y 的

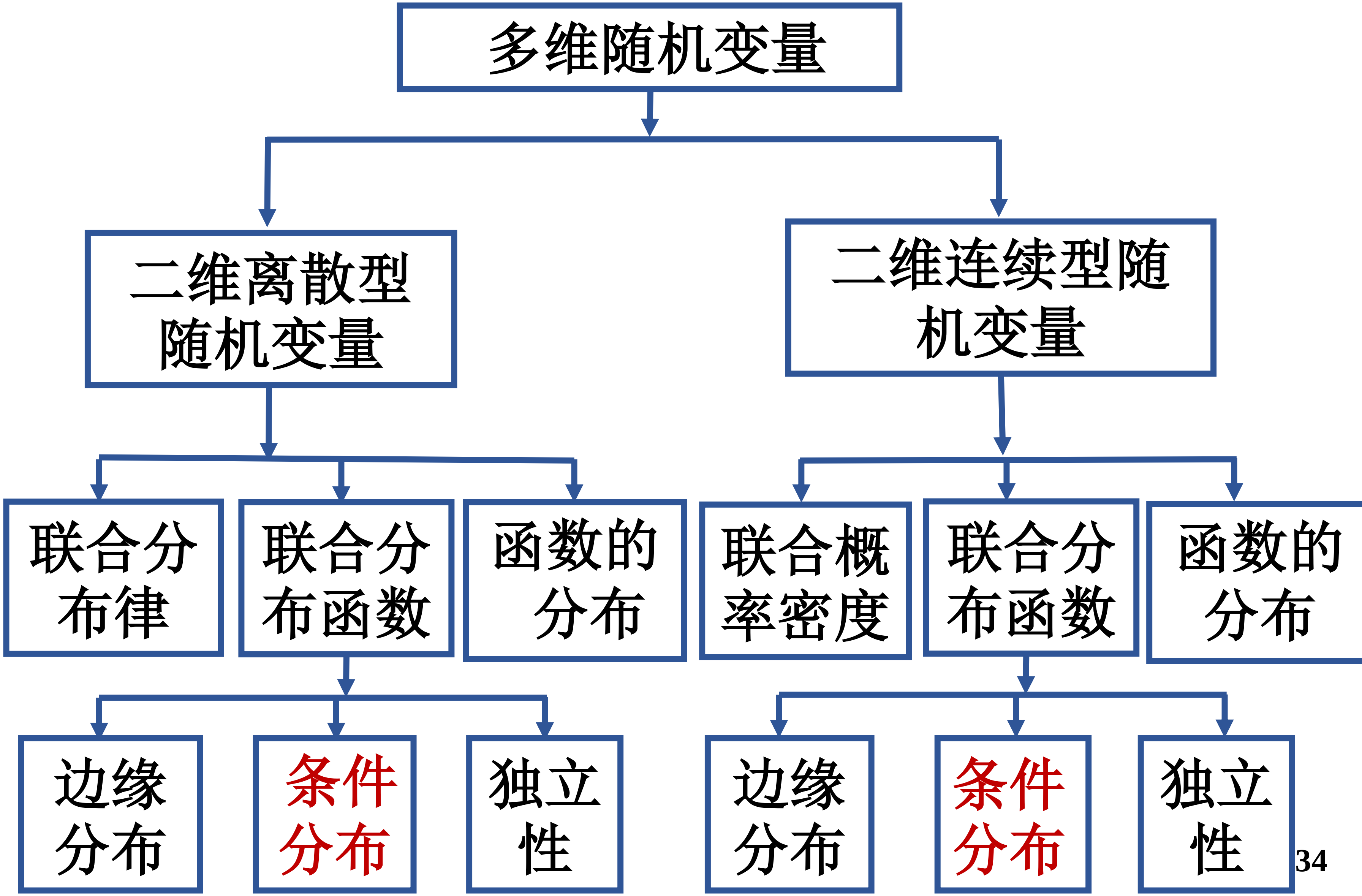
边缘分布函数:

$$F_Y(y)=F(+\infty,y)=\int_{-\infty}^{y}\boxed{\left[\int_{-\infty}^{+\infty} f(x,y)dx\right]}dy$$

边缘概率密度:

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx$$

课程内容—第三章



多维随机变量
二维离散型随机变量
二维连续型随机变量
联合分布律
联合分布函数
函数的分布
联合概率密度
联合分布函数
函数的分布
边缘分布
条件分布
独立性
边缘分布
条件分布
独立性

课程内容—第三章

二维离散型随机变量

联合分布函数

条件分布

在 $\{Y=y_j\}$ 条件下，随机变量X的条件分布律为（=联合分布/边缘分布）：

$$P(X=x_i \mid Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}}$$

2. 性质:

1^0. $P(X=x_i \mid Y=y_j)\geq 0$

2^0. $\sum_{i=1}^{\infty}P(X=x_i \mid Y=y_j)=\sum_{i=1}^{\infty}\frac{p_{ij}}{p_{\cdot j}}=\frac{1}{p_{\cdot j}}\sum_{i=1}^{\infty}p_{ij}=\frac{1}{p_{\cdot j}}\cdot p_{\cdot j}=1$

3^0. $\sum_{j=1}^{\infty}P(Y=y_j \mid X=x_i)=1$

课程内容—第三章

二维连续型随机变量

联合分布函数

条件分布

在条件 $Y=y$ 下，X 的条件分布函数为：

$$F_{X|Y}(x|y)=P(X\le x|Y=y)$$

对于 $f_Y(y)>0$，在条件 $Y=y$ 下的X的条件概率密度：

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

$$F_{X|Y}(x|y)=\int_{-\infty}^{x}f_{X|Y}(x|y)\,dx$$

课程内容—第三章

- 多维随机变量
  - 二维离散型随机变量
    - 联合分布律
    - 联合分布函数
      - 边缘分布
      - 条件分布
      - 独立性
    - 函数的分布
  - 二维连续型随机变量
    - 联合概率密度
    - 联合分布函数
      - 边缘分布
      - 条件分布
      - 独立性
    - 函数的分布

一. 随机变量相互独立的定义

设 (X,Y) 的联合分布函数及边缘分布函数为$F(x,y)$及 $F_X(x)$，$F_Y(y)$。若对任意的 x, y 都有:

$$P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y)$$

即 $$F(x,y) = F_X(x) \cdot F_Y(y)$$

则 称 随机变量 X和 Y是相互独立的。

二. 当 (X,Y) 为离散型随机变量

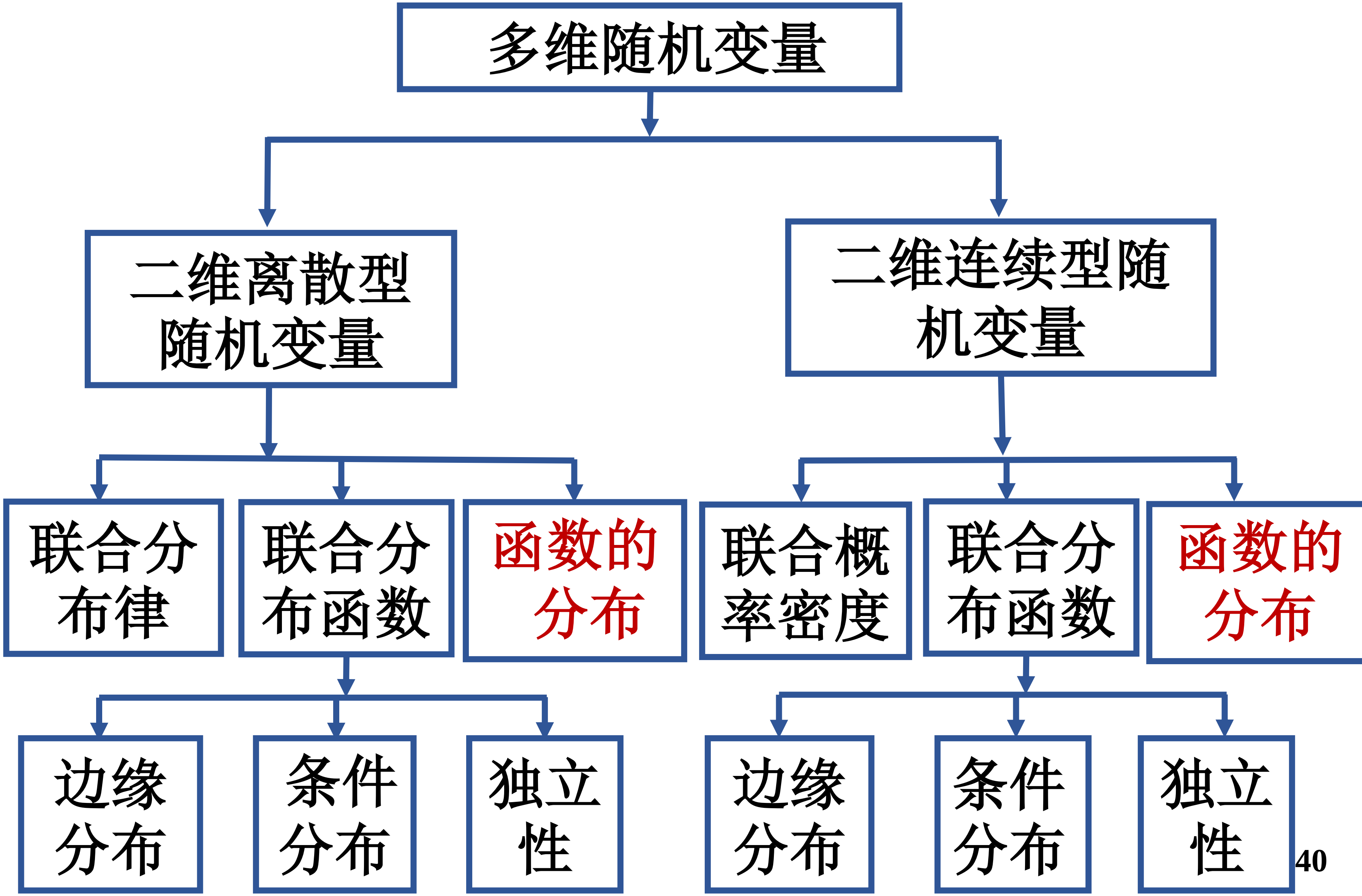
X和Y相互独立 $\Longleftrightarrow$ 对于 (X,Y) 的所有可能的取值 (x_i,y_j) 有

$$P(X=x_i,Y=y_j)=P(X=x_i)\cdot P(Y=y_j)$$

三. 当 (X,Y) 为连续型随机变量

X 和 Y 相互独立 $\Longleftrightarrow$ $f(x,y)=f_X(x)\cdot f_Y(y)$

课程内容—第三章


多维随机变量
二维离散型随机变量
二维连续型随机变量
联合分布律
联合分布函数
函数的分布
联合概率密度
联合分布函数
函数的分布
边缘分布
条件分布
独立性
边缘分布
条件分布
独立性

课程内容—第三章

二维连续型随机变量

函数的分布

一、和的分布(Z=X+Y)

$$f_Z(z)=\int_{-\infty}^{+\infty} f(x, z-x)\,dx$$

X和Y相互独立时，和的分布(Z=X+Y)

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)\,dx$$

结论（可加性）：对于相互独立的服从正态分布，Γ分布或泊松分布的随机变量，它们的和仍服从正态分布，Γ分布或泊松分布，并且参数是单个参数的相加。

课程内容—第三章

二维连续型随机变量

函数的分布

二、商的分布(Z=X/Y)

$$f_Z(z) = \int_{-\infty}^{+\infty} |y|\, f(yz, y)\, dy$$

X和Y相互独立时，商的分布(Z=X/Y)

$$f_Z(z) = \int_{-\infty}^{+\infty} |y| \cdot f_X(yz) \cdot f_Y(y)\, dy$$

类似地，乘积的分布（Z=XY）

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx$$

课程内容—第三章

三、最值的分布

二维连续型随机变量

函数的分布

1. $M = \max(X, Y)$ 的分布

$$F_{\max}(m) = F_X(m) \cdot F_Y(m)$$

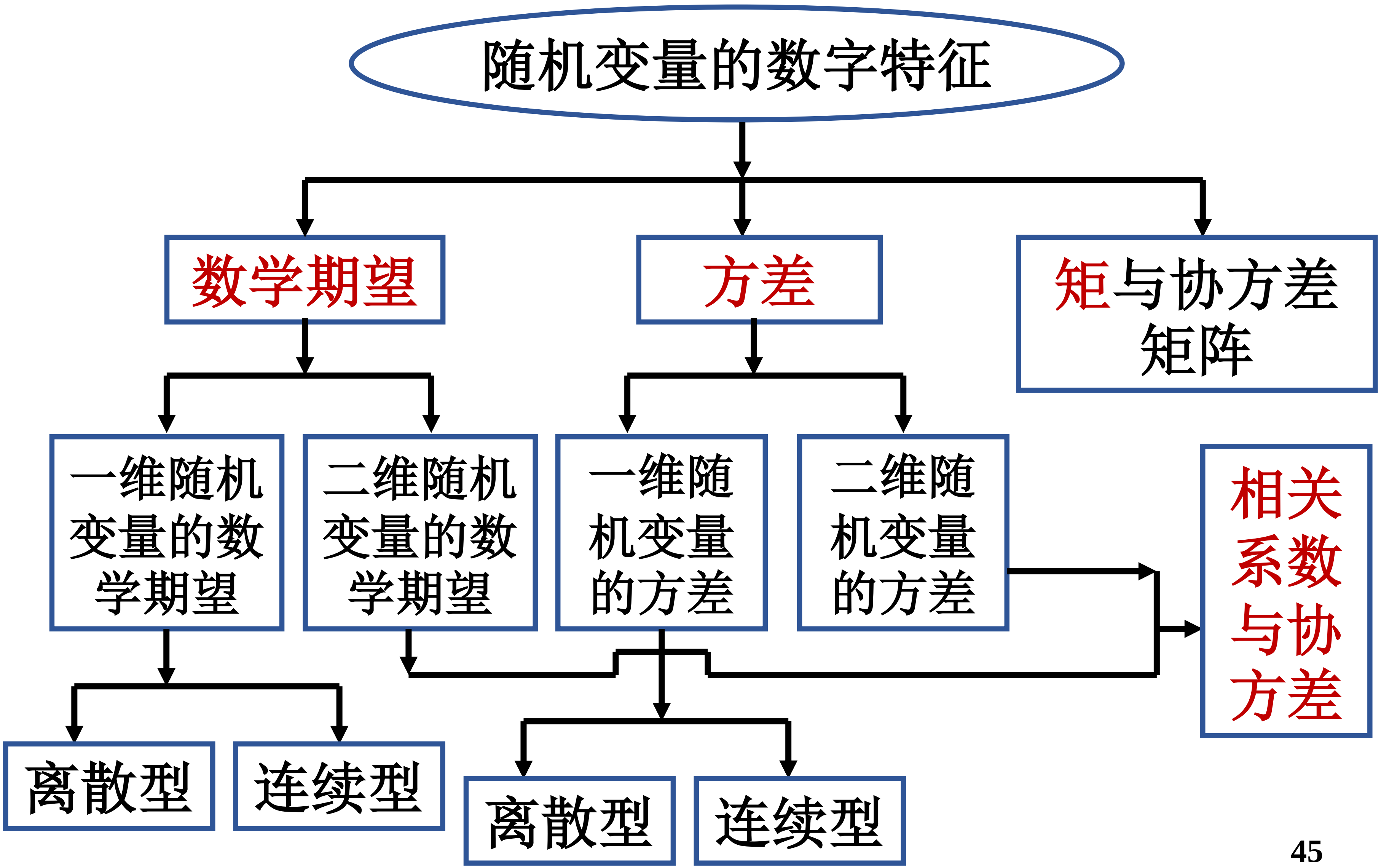
2. $N = \min(X, Y)$ 分布

$$F_{\min}(n) = 1 - [1 - F_X(n)] \cdot [1 - F_Y(n)]$$

课程内容—第四章

教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—第四章


随机变量的数字特征
数学期望
方差
矩与协方差矩阵
一维随机变量的数学期望
二维随机变量的数学期望
一维随机变量的方差
二维随机变量的方差
相关系数与协方差
离散型
连续型
离散型
连续型

课程内容—第四章

数学期望

一维随机变量的数学期望

二维随机变量的数学期望

随机变量**函数**的数学期望

$$E(X)=\sum_{k=1}^{\infty} x_k\ p_k$$

$$E(X)=\int_{-\infty}^{\infty} x\,f(x)dx$$

$$E(X)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i\ p_{ij}$$

$$E(Y)=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} y_j\ p_{ij}$$

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f_X(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\cdot f(x,y)dydx$$

$$E(Y)=\int_{-\infty}^{+\infty} y\cdot f_Y(y)dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\cdot f(x,y)dxdy$$

$$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k$$

$$E(Y)=E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)\,dx$$

数学期望的性质

$$E(X+Y)=E(X)+E(Y)$$

相互独立，则

$$E(XY)=E(X)E(Y)$$

课程内容—第四章

方差

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

性质

$D(X)=E(X^2)-[E(X)]^2$（常用计算公式）

$D(cX)=c^2D(X)$

$D(X+C)=D(X)$

$D(X+Y)=D(X)+D(Y)$（相互独立）

课程内容—第四章

方差 $D(X)=E\{[X-E(X)]^2\}$

性质

离散型

连续型

(1)均匀分布：$E(X)=\frac{a+b}{2}$, $D(X)=\frac{(b-a)^2}{12}$

(2)指数分布：$E(X)=\theta$, $D(X)=\theta^2$

(3)正态分布：$E(X)=\mu$, $D(X)=\sigma^2$

(1)(0-1)分布：$E(X)=p$, $D(X)=pq$

(2)二项分布：$E(X)=np$, $D(X)=npq$

(3)泊松分布：$E(X)=D(X)=\lambda$

课程内容—第四章

协方差（刻画两个随机变量之间的关系）

$$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$$

(1). $Cov(X,Y)=Cov(Y,X)$

(2). $Cov(aX,bY)=ab\,Cov(X,Y)$

(3). $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

$$D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$$

若 X 与 Y 相互独立则： $Cov(X, Y)= 0$

课程内容—第四章

相关系数

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

(1). $|\rho_{XY}|\le 1$

(2). $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数 a,b，使得：

$$P(Y=aX+b)=1$$

X 和 Y 以概率 1 线性相关

注：X 和 Y 独立时，$\rho_{XY}=0$ 但其逆不真。

课程内容—第四章

一. 矩

矩是随机变量的更为广泛的一种数字特征，前面介绍的数学期望及方差都是某种矩。

定义: 设 X 和 Y 是随机变量

(1). 若 $E(X^k)$ 存在，则称它为 X 的 k 阶原点矩，$k=1,2,\cdots$ 简称 k 阶矩。

(2). 若 $E\{[X-E(X)]^k\}$ 存在，$k=2,3,\cdots$ 则称它为 X 的 k 阶中心矩。

(3). 若 $E(X^kY^l)$存在， $k,l=1,2,\cdots$ 则称它为 X 和 Y的 $k+l$ 阶混合矩。

(4). 若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$存在，$k,l=1,2,\cdots$ 则称它为 X 和 Y的 $k+l$ 阶混合中心矩。

注：

数学期望$E(X)$ 是随机变量X 的一阶原点矩；
方差$D(X)$是随机变量X的二阶中心矩；
协方差$Cov(X,Y)$是随机变量X和Y的二阶混合中心矩。

课程内容—第五章

教学内容	知识点
第一章 概率论的基本概念	随机事件及其运算，事件的概率及其性质，条件概率，事件的独立性
第二章 随机变量及其分布函数	离散型随机变量及其分布律，连续型随机变量及其概率密度，随机变量函数及其分布
第三章 多维随机变量及其分布	多维随机变量及其联合分布，边缘分布，条件分布，随机变量的独立性，多维随机变量函数及其分布，n维随机变量简介
第四章 随机变量的数字特征	数学期望，方差，协方差与相关系数，矩
第五章 大数定律及中心极限定理	大数定律，中心极限定理

课程内容—第五章

概率论的理论结果

- 大数定律
  - 切比雪夫大数定理
  - 伯努利大数定理
  - 辛钦(大数)定理
- 中心极限定理
  - 独立同分布的中心极限定理
  - 李雅普诺夫中心极限定理
  - 棣莫弗—拉普拉斯中心极限定理

课程内容—第五章

➢在一定条件下,一列随机变量的算术平均值（按某种意义）收敛于这些项的均值的定理。（为用平均值估计期望提供了理论依据）

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} = 1$$

大数定律

切比雪夫大数定理

- 相互独立
- 方差有共同上界

伯努利大数定理

- 相互独立
- 各变量同时服从以p为参数的（0-1）分布

辛钦大数定理

- 相互独立
- 各变量服从同一分布，且具有数学期望

课程内容—第五章

- 背景：如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大， 则这种量一般都近似服从正态分布。

- 在一定条件下，大量相互独立的随机变量之和的概率分布近似于正态分布的定理—中心极限定理。

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\boxed{\sum_{k=1}^{n} X_k} - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} \le x \right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

课程内容—第五章

中心极限定理

独立同分布的中心极限定理

棣莫弗—拉普拉斯中心极限定理

李雅普诺夫中心极限定理

- 相互独立
- 同分布
- 有限数学期望和方差

特例

- 相互独立
- 服从二项分布

- “由大量微小的、独立的随机因素”（不要求同分布）累积成的变量（即，满足具体条件）
- 随机因素个数趋于无穷时，以正态分布为极限。

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

课程内容—第六章

教学内容	知识点
第六章 数理统计部分（样本及抽样分布）	总体、样本统计量，χ^2分布、t分布、F分布，抽样分布，顺序统计量
第七章 参数估计	矩估计及极大似然估计，估计量的评选标准，区间估计
第八章 假设检验	假设检验问题与基本概念，正态总体期望的假设检验，正态总体方差的假设检验
第九章 方差分析及回归分析	单因素方差分析，双因素方差分析，一元线性回归，多元线性回归

几个常用的统计量

(1). 样本均值： $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

它反映了总体均值的信息

(2). 样本方差： $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)$

它反映了总体方差的信息，是总体方差的无偏估计(P145)

(3). 样本标准差： $\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$

(4). 样本 k 阶原点矩： $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ $k=1, 2,\ldots$

它反映了总体 k 阶矩的信息

(5). 样本 k 阶中心矩： $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$

它反映了总体 k 阶中心矩的信息

课程内容—第六章

▲ 若总体 X 的 k 阶原点矩 $E(X^k)=\mu_k$ 存在，

则当 $n\to\infty$ 时有：$A_k \xrightarrow{p} \mu_k,\quad k=1,2,\cdots$

▲ 设总体X的均值为μ，方差为σ^2，$X_1, X_2, \ldots, X_n$ 为取自总体X的样本，则 (教材第5版P145)

$$E(\bar{X})=\mu,\quad D(\bar{X})=\frac{\sigma^2}{n},\quad E(S^2)=\sigma^2$$

课程内容—第六章

数理统计

抽样分布

常用的统计量

四个重要分布

1. χ^2 分布

定义. 设 $X_1, X_2, \cdots, X_n$ 是来自正态分布 $N(0, 1)$ 的样本，则称统计量：

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

为服从自由度为 n 的 χ^2 分布。记为：$\chi^2 \sim \chi^2(n)$

▲ 若 $\chi^2 \sim \chi^2(n)$，则 $E(\chi^2) = n$, $D(\chi^2) = 2n$

▲ χ^2 分布的可加性：

若 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$ 且

χ_1^2, χ_2^2 相互独立，则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$

课程内容—第六章

数理统计

抽样分布

常用的统计量

四个重要分布

2. ***t*** **分布**

定义. 设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X 与 Y 相互独立，则称随机变量:

$$T = \frac{X}{\sqrt{Y/n}}$$

为服从自由度为 n 的 t 分布。记为 $T \sim t(n)$

▲ 当 n 充分大时, $\lim\limits_{n\to\infty} h(x;n) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

▲ 由上 α 分位点定义及 $h(t)$ 对称性得:

$$t_{1-\alpha}(n) = -t_{\alpha}(n)$$

课程内容—第六章

数理统计

抽样分布

常用的统计量

四个重要分布

3. F分布

定义. 设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, X 与 Y 相互独立，则称统计量：

$$F = \frac{X/n_1}{Y/n_2}$$

为服从自由度 n_1 及 n_2 的 F 分布，记作：$F \sim F(n_1, n_2)$

▲ 若 $F \sim F(n_1, n_2)$ 则 $\frac{1}{F} \sim F(n_2, n_1)$

▲ 若 $X \sim t(n)$ 则：$X^2 \sim F(1, n)$

▲ F 分布的上 α 分位点的性质(推导见P145)：

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$$

课程内容—第六章

数理统计

抽样分布

常用的统计量

四个重要分布

三. 正态分布的样本均值与样本方差的分布

定理 1 (样本均值和样本方差的分布)

设 $X_1, X_2, \ldots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 是其样本均值和样本方差

则 (1) $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$

(2) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

(3) $\bar{X}$ 和 S^2 相互独立

(4) $E(S^2) = \sigma^2$

课程内容—第六章

数理统计

抽样分布

常用的统计量

四个重要分布

定理 2. 设 $X_1, X_2, \ldots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，$\bar{X}$ 和 S^2 分别为样本均值和样本方差，

则有：$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

(1) $$\frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1-1, n_2-1);$$

(2) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时，

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_W \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

其中：$$s_W^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

课程内容—第七章

教学内容	知识点
第六章 数理统计部分（样本及抽样分布）	总体、样本统计量，χ^2分布、t分布、F分布，抽样分布，顺序统计量
第七章 参数估计	矩估计及极大似然估计，估计量的评选标准，区间估计
第八章 假设检验	假设检验问题与基本概念，正态总体期望的假设检验，正态总体方差的假设检验
第九章 方差分析及回归分析	单因素方差分析，双因素方差分析，一元线性回归，多元线性回归

课程内容—第七章

数理统计—统计推断

参数估计

点估计

矩估计法

极大似然估计法

矩估计法的依据：
用样本的各阶矩来代替总体的各阶矩

1、通过计算总体各阶矩，得到未知参数与总体各阶矩之间的关系；
2、样本矩代替总体矩，得到未知参数的估计量。

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(1) $\mu_i = E(X^i)$ 是参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的函数，

记为：

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k) \\ \mu_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k) \end{cases} \qquad (*)$$

(2) 解（*）式解出 $\theta_1,\theta_2,\cdots,\theta_k$ 得到：

$$\begin{cases} \theta_1 = \theta_1(\mu_1,\mu_2,\cdots,\mu_k) \\ \theta_2 = \theta_2(\mu_1,\mu_2,\cdots,\mu_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \theta_k = \theta_2(\mu_1,\mu_2,\cdots,\mu_k) \end{cases} \qquad (**)$$

(3) 用 μ_i 的估计量 $A_i = \frac{1}{n}\sum_{j=1}^{n} X_j^i$ 分别代替（**）中的 μ_i，则得 θ_i 的矩估计量 $\hat{\theta}_i$：

$$\hat{\theta}_i = \hat{\theta}_i(A_1,A_2,\cdots,A_k),\quad i = 1,2,\cdots,k$$

课程内容—第七章

数理统计—统计推断

参数估计

点估计

矩估计法

极大似然估计法

极大似然估计法的基本思想：当试验得到一个结果时，应选择**使得这个试验结果出现概率达到最大**的这个参数值作为参数的估计值。

1、构建样本对应的似然函数（即，样本结果出现的概率）；

2、若似然函数可微，对未知参数求导，求使得$\frac{\partial L(\theta)}{\partial \theta}=0$或$\frac{\partial \ln L(\theta)}{\partial \theta}=0$的解。

$$L=\prod_{k=1}^{n} f(x_k,\theta_1,\theta_2,\cdots\theta_l)$$

或 $$=\prod_{k=1}^{n} P(x_k,\theta_1,\theta_2,\cdots\theta_l)$$

课程内容—第七章

- 参数估计
  - 点估计
    - 矩估计法
    - 极大似然估计法
  - 区间估计
    - 均值的区间估计
    - 方差的区间估计

定义: 设总体 X 的分布函数 $F(x,\theta)$ 含有一个未知参数 θ，对于给定的值 α $(0<\alpha<1)$，若由样本 $X_1,X_2\cdots X_n$ 确定的两个统计量：

$$\bar{\theta}=\bar{\theta}(X_1,X_2,\cdots,X_n) \text{和} \underline{\theta}=\underline{\theta}(X_1,X_2\cdots,X_n)$$

满足：$P(\underline{\theta}<\theta<\bar{\theta})=1-\alpha$ ，则称随机区间 $(\underline{\theta},\bar{\theta})$ 是 θ 的置信度为 $1-\alpha$ 的置信区间。

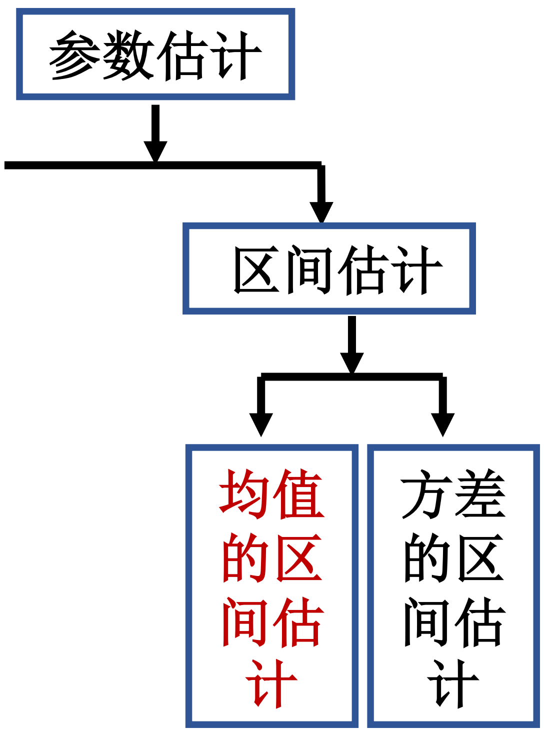
课程内容—第七章

1. **构建枢轴量$W(\theta,\hat{\theta}(X_1,X_2,\dots,X_n))$：** 可从待估参数$\theta$的一个点估计量$\hat{\theta}$出发，构建关于$\theta$和$\hat{\theta}$的函数$W(\theta,\hat{\theta})$，且$W(\theta,\hat{\theta})\sim$F(x)为已知不含未知参数的分布（此处综合考虑点估计量和第六章各抽样分布）。

2. **构造区间：** 在F(x)分布中，构造区间$P\left(a < W(\theta,\hat{\theta}) < b\right) = 1-\alpha$（置信水平）。

3. **求解区间：** 根据a< $W(\theta,\hat{\theta}) < b$，即得出取值范围$\underline{\theta} < \theta < \overline{\theta}$，即为其置信水平为$1-\alpha$的置信区间。

课程内容—第七章



1、单个正态总体，方差已知

➢ 枢轴量为：

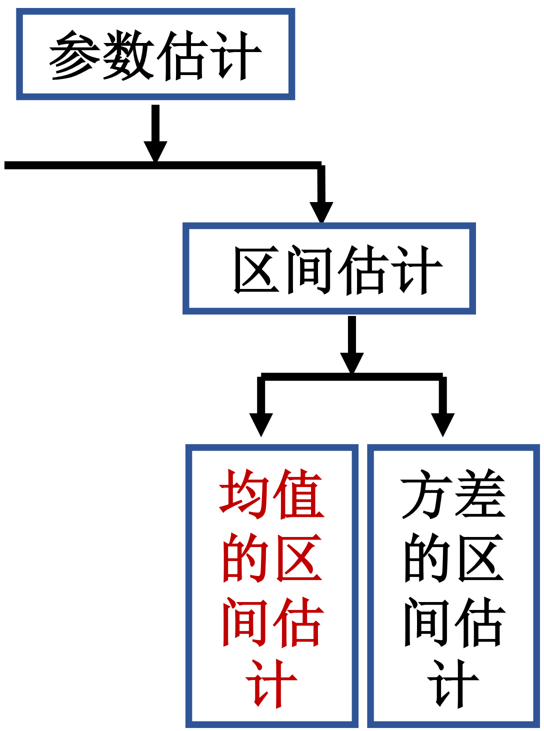
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

2、单个正态总体，方差未知

➢ 枢轴量为：

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

课程内容—第七章



3、两个正态总体，方差已知

➢ 枢轴量为：

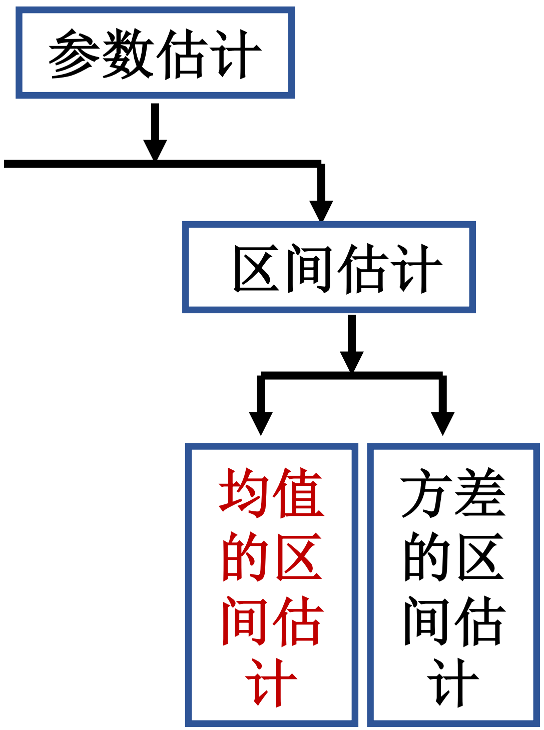
$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$$

4、两个正态总体，方差未知

➢ 枢轴量为：

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}}\sim N(0,1)$$

课程内容—第七章



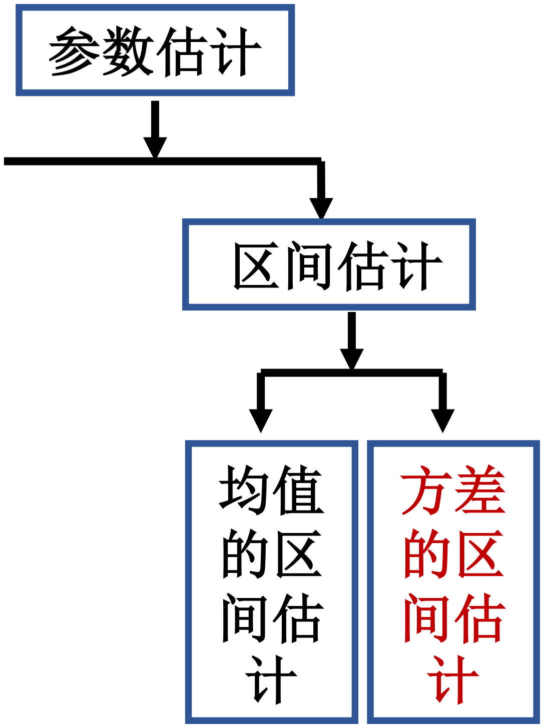
5、两个正态总体，方差未知但相等

➢ 枢轴量为：

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_w\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

$$s_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2},$$

课程内容—第七章



1、单个正态总体，均值未知

➢ 枢轴量为：

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

2、两个正态总体，均值未知

➢ 枢轴量

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

课程内容—第八章

教学内容	知识点
第六章 数理统计部分（样本及抽样分布）	总体、样本统计量，χ^2分布、t分布、F分布，抽样分布，顺序统计量
第七章 参数估计	矩估计及极大似然估计，估计量的评选标准，区间估计
第八章 假设检验	假设检验问题与基本概念，正态总体期望的假设检验，正态总体方差的假设检验
第九章 方差分析及回归分析	单因素方差分析，双因素方差分析，一元线性回归，多元线性回归

课程内容——第八章

统计推断

参数估计　　假设检验

一、假设检验基本思想

设总体 X 含有未知参数θ (或总体分布函数$F(x)$未知)

检验下述假设:

假设 $H_0:\ \theta=\theta_0$ 或 $F(x)=F_0(x)$

二、检验依据

小概率事件原理

小概率事件在一次试验中是几乎不可能发生的。

如果在假设 H_0成立的条件下某事件是小概率事件,但在一次试验中却发生了,于是就可怀疑假设 H_0 的正确性从而拒绝 H_0 。

在假设检验中，常称这个小概率为显著性水平，用 α 表示。 $P\{\text{拒绝}H_0\,|\,H_0\text{为真}\}=\alpha$

课程内容——第八章

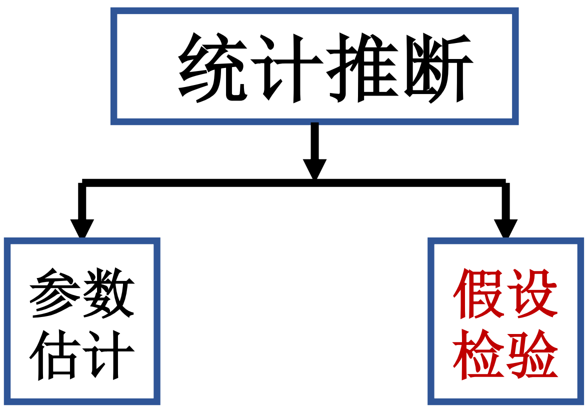
三、本质（理解）

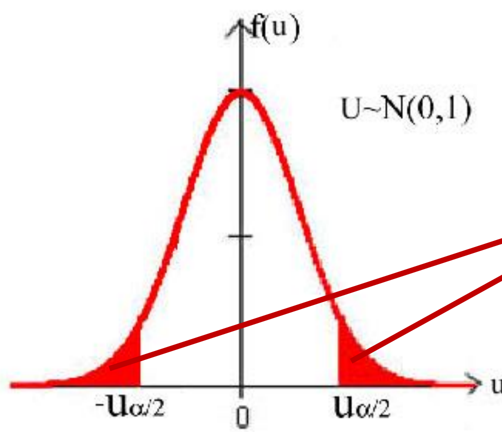
$$H_0:\ \mu=\mu_0\ (\mu_0=355)$$

构造一个分布已知且表征样本结果与假设结果差异的统计量，例如

偏差

$$U=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$$





如果偏差大（即落在一定显著性水平的拒绝域中），则拒绝原假设（即差异显著）。

课程内容—第八章

假设检验问题的步骤：

1. 根据实际问题要求，提出原假设H_0及备择假设H_1

2. 确定检验统计量

3. 按 $P(\text{拒绝}H_0 \mid H_0\text{为真}) = \alpha$，求出拒绝域

4. 取样本，将样本观察值代入统计量，观察统计量是否在拒绝域以确定接受 H_0 还是拒绝 H_0

课程内容—第八章

一. 单个正态总体 $N(\mu,\sigma^2)$ 均值 μ 的检验

1. σ^2 已知，关于 μ 的检验(Z 检验)

都取检验统计量: $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

拒绝域:

双边假设检验

$$\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{\frac{\alpha}{2}}$$

右边假设检验

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq z_{\alpha}$$

左边假设检验

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq -z_{\alpha}$$

课程内容—第八章

一. 单个正态总体 $N(\mu,\sigma^2)$ 均值 μ 的检验

1. σ^2 未知，关于 μ 的检验(t 检验)

都取检验统计量: $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

拒绝域:

双边假设检验	右边假设检验	左边假设检验
$\left\lvert \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \right\rvert \geq t_{\frac{\alpha}{2}}(n-1)$	$\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \geq t_{\alpha}(n-1)$	$\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq -t_{\alpha}(n-1)$

注:　当总体服从正态分布，μ 未知

双边假设检验: $H_0: \sigma^2 = \sigma_0^2$,　$H_1: \sigma^2 \neq \sigma_0^2$

右边假设检验: $H_0: \sigma^2 \leq \sigma_0^2$,　$H_1: \sigma^2 > \sigma_0^2$

左边假设检验: $H_0: \sigma^2 \geq \sigma_0^2$,　$H_1: \sigma^2 < \sigma_0^2$

在显著性水平 α 下, H_0 的拒绝域分别为:

双边假设检验: $$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{1-\frac{\alpha}{2}}(n-1) \cup \frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{\frac{\alpha}{2}}(n-1)$$

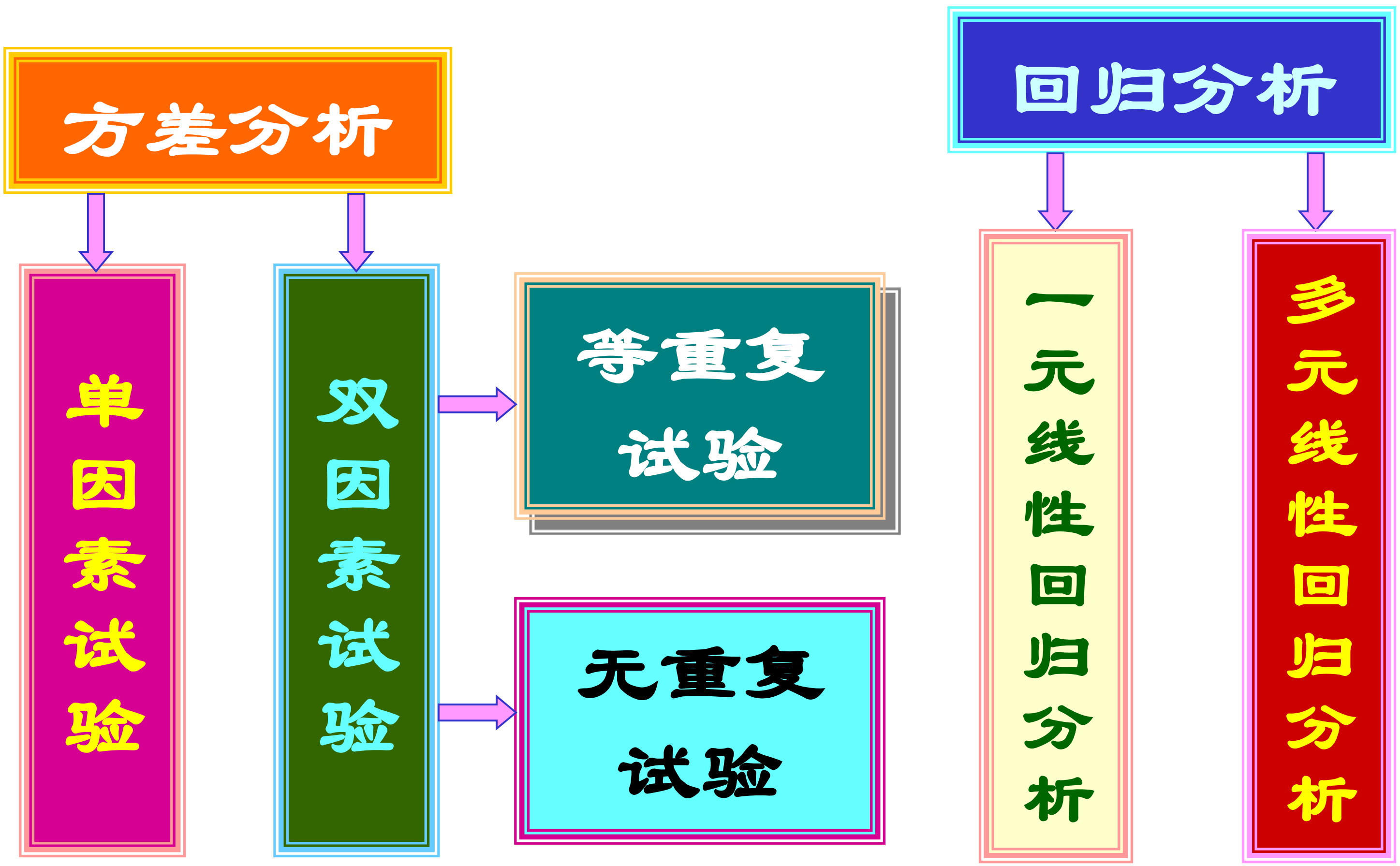
右边假设检验: $$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{\alpha}(n-1)$$

左边假设检验: $$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{1-\alpha}(n-1)$$

课程内容—第九章

教学内容	知识点
第六章 数理统计部分（样本及抽样分布）	总体、样本统计量，χ^2分布、t分布、F分布，抽样分布，顺序统计量
第七章 参数估计	矩估计及极大似然估计，估计量的评选标准，区间估计
第八章 假设检验	假设检验问题与基本概念，正态总体期望的假设检验，正态总体方差的假设检验
第九章 方差分析及回归分析	单因素方差分析，双因素方差分析，一元线性回归，多元线性回归

课程内容—第九章



课程内容—第九章

单/双因素试验的方差分析步骤

(1) 建立数学模型;
(2) 分解平方和;
(3) 研究统计特性;
(4) 确定拒绝域.

单因素方差分析的问题是检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_s,$$
$$H_1: \mu_1, \mu_2, \cdots, \mu_s \text{不全相等.}$$

或等价于检验假设 $H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0,$

$$H_1: \delta_1, \delta_2, \cdots, \delta_s \text{不全为零.}$$

单因素试验方差分析表

方差来源	平方和	自由度	均　方	F　比
因　素A	S_A	$s-1$	$\bar{S}_A = \dfrac{S_A}{s-1}$	$F = \bar{S}_A / \bar{S}_E$
误　差	S_E	$n-s$	$\bar{S}_E = \dfrac{S_E}{n-s}$	
总　和	S_T	$n-1$		

单因素方差分析问题的检验假设

当H_0成立时，$S_E/\sigma^2 \sim \chi^2(n-s)$, $S_A/\sigma^2 \sim \chi^2(s-1)$,

且S_E和S_A相互独立,从而

$$F=\frac{S_A/(s-1)}{S_E/(n-s)} \sim F(s-1,n-s).$$

对给定的检验水平α,若$F \ge F_\alpha(s-1,n-s)$,则拒绝假设H_0,即认为因素A影响显著.

若$F < F_\alpha(s-1,n-s)$,则接受H_0,即认为因素A影响不显著.

单因素方差分析问题的参数估计

$$\hat{\mu}=\overline{X},\hat{\mu}_j=\overline{X}_{\bullet j},\hat{\delta}_j=\hat{\mu}_j-\hat{\mu}=\overline{X}_{\bullet j}-\overline{X},$$

$$\hat{\sigma}^2=S_E/(n-s).$$

$\mu_j-\mu_k=\delta_j-\delta_k$ 的置信区间为

$$\left(\overline{X}_{\bullet j}-\overline{X}_{\bullet k}\pm t_{\alpha/2}(n-s)\sqrt{\overline{S}_E(\frac{1}{n_j}+\frac{1}{n_k})}\right),$$

其中 $\overline{S}_E=\dfrac{S_E}{n-s}$.

双因素等重复试验方差分析问题的假设检验

对检验水平α,

若 $F_A = \dfrac{S_A/(r-1)}{S_E/(rs(t-1))} \geq F_\alpha(r-1, rs(t-1))$

则拒绝假设H_{01},即因素A影响显著;

若 $F_B = \dfrac{S_B/(s-1)}{S_E/(rs(t-1))} \geq F_\alpha(s-1, rs(t-1))$

则拒绝假设H_{02},即因素B影响显著;

若 $F_{A\times B} = \dfrac{S_{A\times B}/((r-1)(s-1))}{S_E/(rs(t-1))}$

$\geq F_\alpha((r-1)(s-1), rs(t-1))$

则拒绝假设H_{03},即交互作用$A\times B$影响显著.

双因素无重复试验方差分析问题的假设检验

对检验水平α,

若 $$F_A=\frac{\overline{S}_A}{\overline{S}_E}\geq F_\alpha(r-1,(r-1)(s-1)),$$

则拒绝假设H_{01},即因素A影响显著;

若 $$F_B=\frac{\overline{S}_B}{\overline{S}_E}\geq F_\alpha(s-1,(r-1)(s-1)),$$

则拒绝假设H_{02},即因素B影响显著.

一元线性回归分析

(1)数学模型

设$(x_1, y_1), \cdots, (x_n, y_n)$满足一元线性模型

$$\begin{cases} Y_i = a + bx_i + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2) \end{cases} \qquad (i = 1, 2, \cdots, n),$$

a, b, σ^2为模型参数.

(2)线性回归方程

$$\hat{y} = \hat{a} + \hat{b}x, \quad 其中 \hat{a} = \overline{y} - \hat{b}\overline{x}, \hat{b} = \frac{S_{xy}}{S_{xx}}.$$

一元线性回归分析

(3)线性假设的显著性检验

检验假设： $H_0: b = 0, \quad H_1: b \neq 0.$

对检验水平α,

若 $$|t| = \frac{|\hat{b}|}{\hat{\sigma}}\sqrt{S_{xx}} \geq t_{\alpha/2}(n-2),$$

则拒绝假设 H_0, 认为回归方程是显著的.

反之则接受假设 H_0, 认为回归效果不显著.

一元线性回归分析

(4)系数b的置信区间

$$\left(\hat{b} \pm t_{\alpha/2}(n-2) \times \frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right).$$

(5)$y_0 = a + bx_0 + \varepsilon_0, \varepsilon_0 \sim N(0, \sigma^2)$的置信区间

$$\left(\hat{y}_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}\right).$$

多元线性回归分析

(1)数学模型

$$Y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + \varepsilon_i,$$

$\varepsilon_i \sim N(0,\sigma^2), i = 1,2,\cdots,n$且相互独立,

$$B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix}, \sigma^2 \text{为模型参数.}$$

多元线性回归分析

(2) 模型参数估计

记$X=\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$, $Y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$,

$$\hat{B}=(X'X)^{-1}X'Y,$$

回归方程为$\hat{y}=\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\cdots+\hat{b}_px_p.$