人工智能综合基础

- 概率统计

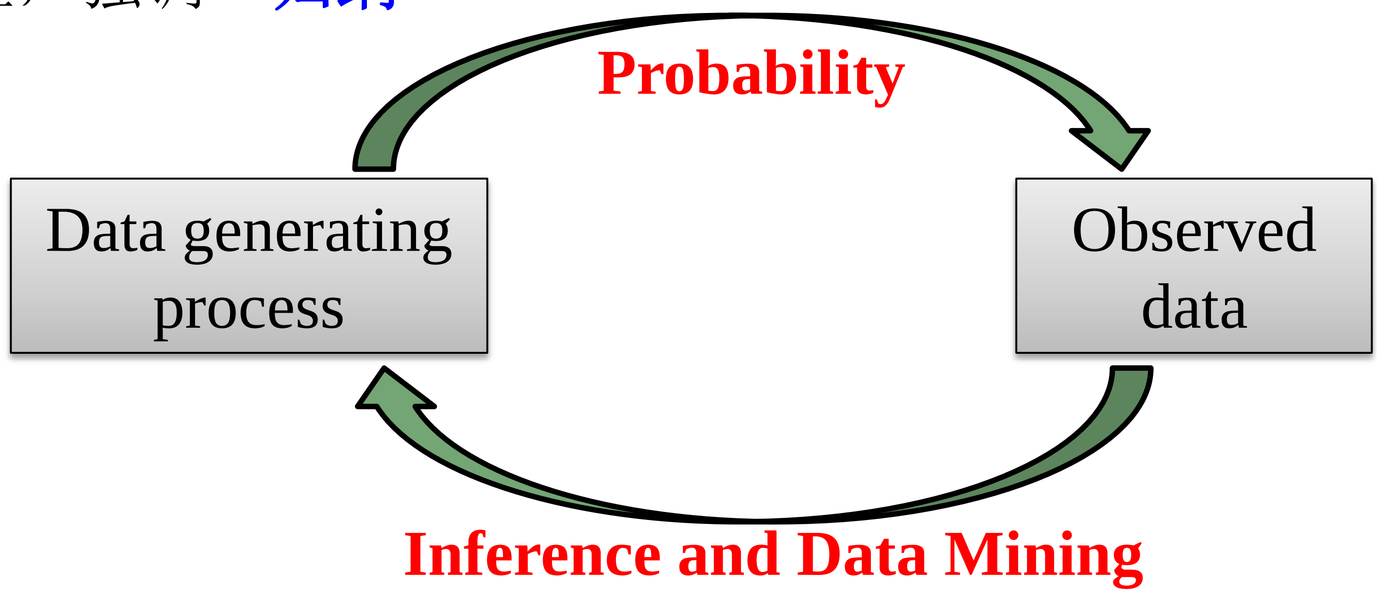
高 尉

问题

简答：概率与统计，最主要的区别

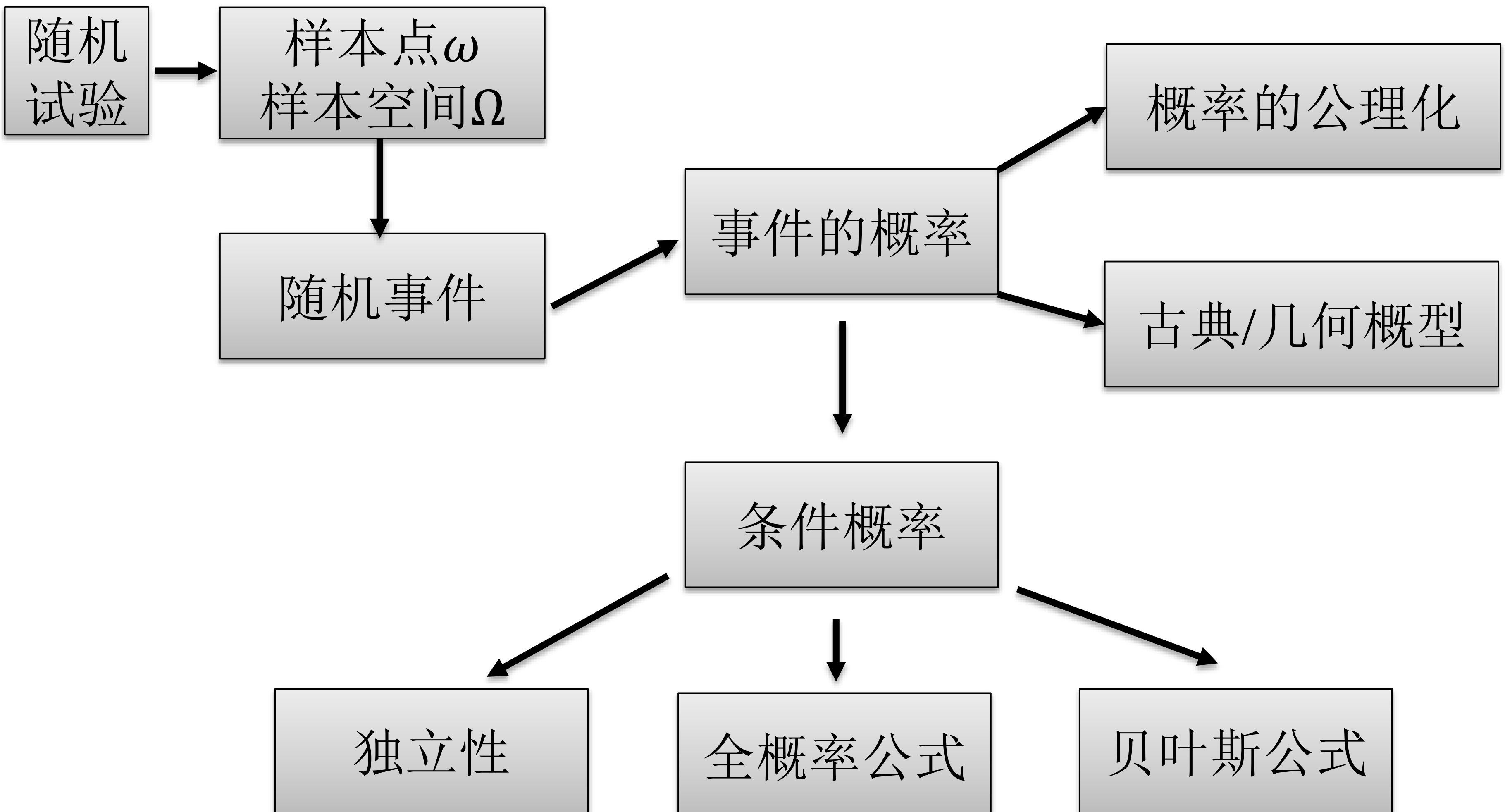
概率与统计

- **概率**：研究事件的不确定性，在给定数据生成过程中观察、研究数据的性质，强调**公理体系**、**推理**
- **统计**：收集与分析数据，根据观察的数据反思其数据生成过程，强调“**归纳**”



概率与统计重要的区别在于公理体系化

知识结构



随机试验
样本点ω
样本空间Ω
随机事件
事件的概率
概率的公理化
古典/几何概型
条件概率
独立性
全概率公式
贝叶斯公式

必然现象与随机现象

必然现象：**在一定条件下必然发生的现象**，其特征是**条件完全决定结果**

随机现象：在一定条件下可能出现、可能不出现的现象，其特征是**条件不能完全决定结果**

随机现象的二重属性：

- 偶然性：对随机现象做一次观察，观察结果不可预知
- 必然性：对随机现象做大量观察，观察结果具有一定的规律性，即**统计规律性**

随机试验(用E表示) 具备以下三个特点的试验：

– **可重复**：可在相同的条件下重复进行
– **多结果**：结果不止一个，所有可能的结果事先已知
– **不确定**：试验前无法预测/确定哪一种结果

随机试验

- **样本点**：试验的每一种可能的结果，记为ω
- **样本空间**：试验中所有可能的结果组成的集合，记为Ω

随机事件：样本空间Ω的子集，由单个或某些样本点ω的集合，**本质是集合**，一般用字母A、B、C等。称"**随机事件A发生**"当且仅当试验的结果是子集A中的元素

- **必然事件**：试验中必定发生的事件，记为 Ω
- **不可能事件**：试验中不可能发生的事件，用 $\emptyset$ 表示

记号	概率论	集合论
Ω	样本空间, 必然事件	全集
$\emptyset$	不可能事件	空集
ω	基本事件	元素
A	随机事件	子集

事件间的关系

$A \subset B$： 若A发生必然导致B发生，称事件B包含事件A

$A = B$： 若$A \supset B$且$B \supset A$

$A \cup B$： 事件A和B至少发生一个的事件

$A \cap B = AB$： 事件A和B同时发生的事件

$\bar{A}$: 事件A不发生的事件

互斥/互不相容：若事件A和B不可能同时发生

$A - B$: A发生，而B不发生的事件

$$A - B = A - AB = A\bar{B} = (A \cup B) - B$$

事件与集合的对应关系

记号	概率论	集合论
$\bar{A}$	A的对立事件	A的补集
$A \subset B$	A发生必然导致B发生	A是B的子集
$A = B$	事件A与事件B相等	集合A与集合B相等
$A \cup B$	事件A与事件B的和	集合A与集合B的并集
$A \cap B$	事件A与B的积事件	集合A与集合B的交集
$A - B$	事件A与事件B的差	集合A与集合B的差集
$A \cap B = \emptyset$	事件A与事件B互不相容	集合A与集合B中没有相同的元素

事件的运算规律

- 幂等律： $A \cup A = A$，$A \cap A = A$
- 交换律： $A \cup B = B \cup A$，$A \cap B = B \cap A$
- 结合律： $(A \cup B) \cup C = A \cup (B \cup C)$

  $(A \cap B) \cap C = A \cap (B \cap C)$
- 分配律： $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- 对偶律： $\overline{A \cup B} = \bar{A}\bar{B}$，$\overline{AB} = \bar{A} \cup \bar{B}$

上述规律可推广到多个事件

例题

设A、B、C为任意三个随机事件，证明

$$(\overline{\mathrm{A}} \cup B)(A \cup B)(\overline{\mathrm{A}} \cup \overline{\mathrm{B}})(A \cup \overline{\mathrm{B}}) = \emptyset$$

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

设A与B同时发生，则C必发生，则（）

A） $P(C) \leq \mathrm{P}(A) + P(B) - 1$

B） $P(C) \geq \mathrm{P}(A) + P(B) - 1$

C） $P(C) = \mathrm{P}(AB)$

D） $P(C) = \mathrm{P}(A \cup B)$

问题

简答：频率与概率的关系

概率

- 概率用于度量事件发生的可能性，是事件的固有属性
- 频率在一定程度上反映了事件发生的可能性
- 概率是恒定的，而频率在试验中具有随机性
- 若试验次数足够多，频率与概率非常接近
- 概率可以通过频率来“测量”，频率是概率的一个近似

概率的公理化定义 在随机试验的样本空间Ω上，对于每一个事件A 赋予一个实数，记为$P(A)$，若满足下列条件，称$P(A)$为事件A发生的概率：

- 非负性：$P(A) \geq 0$
- 规范性：$P(\Omega) = 1$
- 可列可加性：若$A_1, A_2, \dots$可列个两两互不相容的事件，则

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

概率的性质

- 对不可能事件∅有 **$P(\emptyset)=0$**, 对必然事件Ω有 **$P(\Omega)=1$**【事件可以推导出概率、但反之不成立】
- **有限可加性** 若$A_1, A_2 \cdots A_n$是两两不相容事件, 则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots P(A_n)$$

- 对任意事件A有$\mathrm{P}(\overline{\mathrm{A}}) = 1 - \mathrm{P}(\mathrm{A})$
- 若$\mathrm{B} \subset \mathrm{A}$，则$\mathrm{P}(\mathrm{A} - B) = \mathrm{P}(\mathrm{A}) - P(B)$和$\mathrm{P}(\mathrm{B}) \le \mathrm{P}(\mathrm{A})$
- 对任意事件A和B

$$\mathrm{P}(\mathrm{A} - \mathrm{B}) = \mathrm{P}(\mathrm{A}) - \mathrm{P}(\mathrm{AB}) = \mathrm{P}(A \cup B) - P(B)$$

- 容斥原理：

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

例题

设A与B满足P(A)=P(B)=1/2，P(A∪B)=1则(　)

A) $A\cup B=\Omega$　　B) $AB=\emptyset$

C) $\mathrm{P}(\bar{A}\cup\bar{B})=1$　　D) $\mathrm{P}(A-B)=0$

设A和B满足$P(AB)=\mathrm{P}(\bar{A}\bar{B})$, 且$P(A)=p$, 求$P(B)$

设$P(A)=P(B)=P(C)=1/4, P(AB)=0$, $P(AC)=P(BC)=1/8$, 求A, B, C全不发生的概率

古典概型

古典概型：试验结果只有有限种可能$\Omega = \{\omega_1, \omega_2, \ldots \omega_n\}$，每种结果发生的可能性相同$\mathrm{P}(\omega_i) = \mathrm{P}(\omega_j)$

若事件A包含k个基本事件，则事件A发生的概率为：

$$P(A) = k/n = |A|/|\Omega|$$

计数的两条基本原理：**加法原理、乘法原理**

排列、组合（组合计数十二路了解）

以前的经典例题：抽签、生日悖论、Matching问题

例题

一个盒子中装有n个球，编号为$1,2,\ldots,n,$ 有放回的取出k个球，求取出的球中最大编号为m的概率。

几何概型

在一个测度有限的区域Ω内等可能性投点, 落入Ω内的任意子区域A的可能性与A的测度成正比, 与A的位置与形状无关, 这样的概率模型称之为**几何概型**. 事件A发生的概率为

$$P(A)=\frac{A\text{的测度}}{\Omega\text{的测度}}=\frac{\mu(A)}{\mu(\Omega)}$$

例：假设一乘客到达汽车站的时间是任意的，客车间隔一段时间发班, 请规划最长的间隔发车时间, 才能确保乘客候车等待时间不超过20分钟的概率大于80%.

条件概率

条件概率：设A和B为同一样本空间下的随机事件, 且$P(A) > 0$,

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

为事件A发生的条件下事件B发生的概率, 简称 **条件概率**

两种计算方法：1) 定义　　2) 空间缩减法

条件概率是概率，具有概率的所有性质：

- 非负性、规范性、可列可加性、容斥原理
- $P(B_1 - B_2|A) = P(B_1|A) - P(B_1B_2|A)$
- 乘法公式：$P(AB) = P(A)P(B|A) = P(B)P(A|B)$

问题

简答：全概率公式、贝叶斯公式

全概率公式和贝叶斯公式

若事件$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 对任意事件B有

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

称之为 **全概率公式**

贝叶斯公式： 设$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 且事件B满足$P(B) > 0$. 对任意$1 \le i \le n$有

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

小概率原理： 若事件A在一次试验中发生的概率非常小, 但经过多次独立地重复试验, 事件A的发生是必然的

例题

事件A、B、C中A与C互不相容，$P(AB) = 1/2$，$P(C) = 1/3$，求$P(AB|\bar{C})$

例题

第一个袋子中有5个红球和3个白球，第二个袋子中有4个红球和3个白球，从第一个袋子中取3个球放入第二个袋子后，从第二个袋子任取一球，求：1）此球为红球的概率

2）若已知第二个袋子取出红球，求从第一个袋子取出3个球无红球的概率

例题

犯人a, b, c均被判为死刑, 法官随机赦免其中一人, 看守知道谁被赦免但不会说. 犯人a问看守: b和c谁会被执行死刑? 看守的策略: i) 若赦免b, 则说c; ii) 若赦免c, 则说b; iii) 若赦免a, 则以1/2的概率说b或c. 看守回答a: 犯人b会被执行死刑, 求在此信息下三人被赦免的概率?

两事件的独立性

事件A和B独立 $\Leftrightarrow P(AB) = P(A)P(B) \Leftrightarrow P(B|A) = P(B)$

A和B独立 $\Leftrightarrow$ A和$\bar{B}$独立$\Leftrightarrow$ $\bar{A}$和B独立$\Leftrightarrow$ $\bar{A}$和$\bar{B}$独立

概率为0或1的事件与任何事件独立

概念: 若事件A, B, C的相互独立与A, B, C的两两独立

性质: 若事件$P(B) \in (0,1)$，则

A和B独立 $\Leftrightarrow P(A|B) = P(A|\bar{B}) \Leftrightarrow P(A|B) + P(\bar{A}|\bar{B}) = 1$

问题

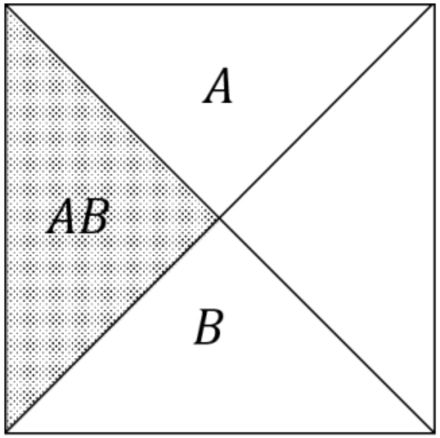
简答：独立性与互不相容性的关系

独立与互斥的关系

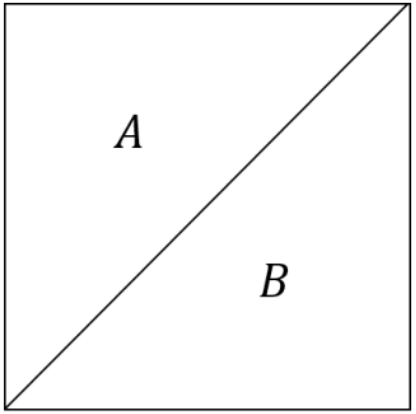
A与B相互独立： $P(AB) = P(A)P(B)$, 独立性与概率相关,
反映事件的概率属性

A与B互不相容：$AB = \emptyset$, 与事件运算关系相关, 与概率无关

独立与互斥反映事件不同的性质, 无必然联系



A与B独立, 但并不互斥



A与B互斥, 但并不独立

例题

事件A、B、C两两独立，且C与$A-B$独立, 则A、B、C相互独立.

案例分析：验证大矩阵乘法是否相等

给定矩阵$A, B, C \in \{0,1\}^{n\times n}$ $(n \geq 10000000)$, 验证$AB = C$?

独立随机产生一个向量$r \in \{0,1\}^n$, 判断

$$A(Br) = Cr?$$

计算$A(Br)$ 和Cr的复杂度均为$O(n^2)$. 若$A(Br) \neq Cr$则直接有$AB \neq C$; 若$A(Br) = Cr$并不能得出$AB = C$.

将上述过程独立进行K次, 可以证明以较大的概率有$AB = C$成立, 该过程被称为Freivalds算法

Freivalds算法分析

该算法的计算复杂度为$O(Kn^2)$, 若K比较小则显著降低了计算复杂度.

若返回No, 则必然有$AB \neq C$

若返回Yes, 然而并不一定有$AB = C$成立, 下面研究成立的概率.

设$D = AB - C \neq 0$, 则D中必存在一些元素不为0, 不妨令$d_{11} \neq 0$. 对任意一轮循环, 不妨设随机向量$r = \{r_1, r_2, \cdots, r_n\}$, 根据返回Yes可知$Dr = 0$, 进一步可得向量$Dr$的第一个元素等于0, 即

$$\sum_{j=1}^{n} d_{1j} r_j = 0 \quad \Longrightarrow \quad r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j$$

Freivalds算法分析

$$\sum_{j=1}^{n} d_{1j} r_j = 0 \quad \Longrightarrow \quad r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j$$

无论$r_2, \ldots, r_n$取何值, 等式$\sum_{j=1}^{n} d_{1j} r_j = 0$是否成立由$r_1$的值决定.

根据$P(r_1 = 0) = P(r_1 = 1) = 1/2$可知$\sum_{j=1}^{n} d_{1j} r_j = 0$成立的概率不超过1/2

在K轮独立循环中, 等式$\sum_{j=1}^{n} d_{1j} r_j = 0$成立的概率不超过$\frac{1}{2^K}$

取$K = \log_2 n$, 则算法Freivalds计算复杂度为$O(n^2 \log n)$, 若算法返回No, 则$AB \neq C$; 若返回Yes, 则有

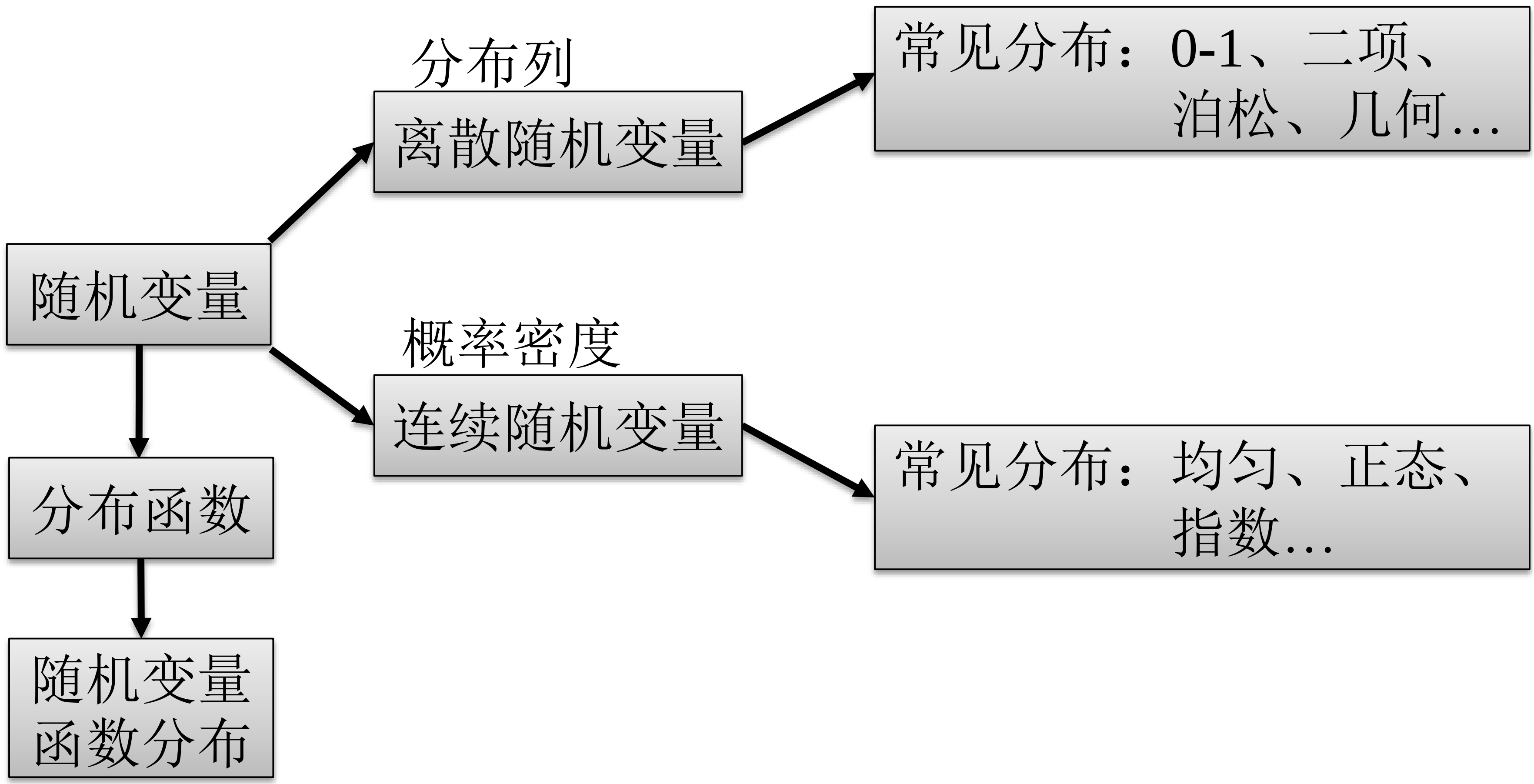
$$P(AB = C) > 1 - 1/n$$

即至少以$1 - 1/n$的概率有$AB = C$成立

第二章 随机变量及其分布

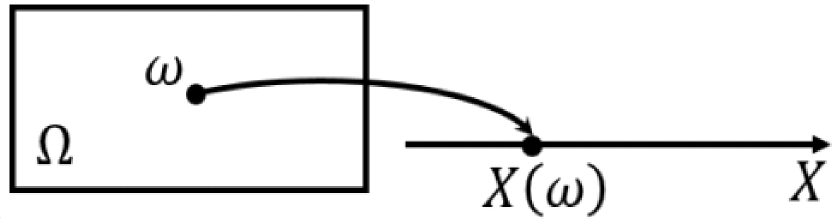
高 尉

知识结构


分布列
离散随机变量
常见分布：0-1、二项、
泊松、几何…
随机变量
概率密度
连续随机变量
常见分布：均匀、正态、
指数…
分布函数
随机变量
函数分布

随机变量

将样本空间Ω中每个样本点ω与一实数$X(\omega)$ 相对应，$X(\omega)$是ω的实值函数, 称实值函数 $X(\omega): \Omega \to \mathrm{R}$为随机变量, 简记$X$



给定任意随机变量X和实数x, 函数

$$F(x) = P(X \leq x)$$

称为随机变量X的**分布函数**, 分布函数的本质是概率

- 单调性: 若$x_1 < x_2$, 则$F(x_1) \leq F(x_2)$
- 规范性: $F(x) \in [0,1]$且$F(-\infty) = 0, F(+\infty) = 1$
- 右连续性: $F(x+0) = F(x)$

分布函数可由这三条性质完全刻画

- 对任意实数$x_1 < x_2$, $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

$$P(X = x_2) = F(x_2) - F(x_2 - 0)$$

例题

性质：若$F_1(x), F_2(x)$是分布函数，则$aF_1(x)+(1-a)F_2(x)$也是分布函数，其中$a \in (0,1)$

例：设随机变量X的分布函数为

$$F(x)=\begin{cases} 0 & x \le a \\ x^2-b & a < x \le \sqrt{2} \\ c & x > \sqrt{2} \end{cases}$$

求$\mathrm{P}(X=a)$和$\mathrm{P}[1<X<3]$

离散型随机变量

离散型随机变量：随机变量的取值是有限的、或无限可列的。

假设其取值为$x_1, x_2, \cdots, x_n, \cdots$, 事件$X = x_k$的概率记为

$$p_k = P(X = x_k) \qquad k = 1, 2, \cdots$$

称之为**随机变量*X*的分布列**，分布列包含随机变量的取值和概率，完全刻画其概率属性

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

性质：随机变量X的分布列$P(X = x_k) = p_k \ (k \geq 1)$满足$p_k \geq 0$且$\sum_k p_k = 1$

例:若随机变量X的分布列$P(X = k) = c/4^k \ (k \geq 0)$, 求$P(X = 1)$

均匀/0-1分布

设随机变量X的取值为$x_1, x_2, \cdots, x_n$, 且$P(X = x_i) = 1/n$, 称X服从离散 **均匀分布**

期望：$E(X) = \sum_{i=1}^{n} \frac{x_i}{n}$　　方差: $E(X) = \sum_{i=1}^{n} \frac{x_i^2}{n} - \left(\sum_{i=1}^{n} \frac{x_i}{n}\right)^2$

德国坦克数量问题： 假设德国生产N辆坦克, 编号为$1,2, \cdots, N$, 盟军战斗中随机击毁 k辆, 被随机击毁坦克编号分别为$x_1, x_2, \ldots, x_k$, 如何估计N的大小

随机变量X的取值为 $\{0, 1\}$, 其分布列$P(X = 1) = p$，称X服从参数为p的**$0 - 1$分布**, 或 **Bernoulli 分布**, 记 $\boldsymbol{X \sim \mathrm{Ber}(p)}$

若 $X \sim \mathrm{Ber}(p)$，则 $E(X) = p$ 和 $\mathrm{Var}(X) = p(1 - p)$

二项/几何分布

用随机变量X表示n重Bernoulli试验中事件A发生的次数, 则X的取值为$0,1,\cdots,n$, 其分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\quad k\in\{0,1,2,\ldots,n\}$$

称随机变量X服从参数为n和p的**二项分布 (binomial distribution)**, 记 $\boldsymbol{X\sim B(n,p)}$

对随机变量 $X\sim\ B(n,p)$ 有$E(X)=np$ 和 $\mathrm{Var}(X)=np(1-p)$

用随机变量X表示事件A首次发生时的试验次数, 则X的取值为$1,2,\cdots$, 其分布列为$P(X=k)=(1-p)^{k-1}p\ \ (k\geq 1)$

称**X服从参数为p的几何分布**, 记为 $\boldsymbol{X\sim G(p)}$

【无记忆性】$P(X>m+n\,|\,X\ >\ m)=P(X>n)$

若随机变量$X\sim G(p)$, 则有$\mathrm{E}(X)=1/p$和$\mathrm{Var}(X)=(1-p)/p^2$

负二项分布 (Pascal分布)

用X表示事件A第r次成功时发生的试验次数，则X取值$r, r+1, r+2, \cdots$, 其分布列为

$$P(X=n)=\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}\cdot p \quad (n\geq r)$$

称***X*服从参数为*r*和*p*的负二项分布，又称 Pascal分布**

设随机变量X服从参数为$p\in(0,1)$ 和$r>0$的负二项分布，则有$\mathrm{E}(X)=r/p$和$\mathrm{Var}(X)=r(1-p)/p^2$

泊松定理

若随机变量X的分布列为$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\ (k\geq 0)$其中$\lambda>0$是常数, 称随机变量$X$**服从参数为λ的泊松分布**, 记**为 $\boldsymbol{X\sim P(\lambda)}$**

若 $X\sim P(\lambda)$, 则$E(X)=\lambda$ 和$\mathrm{Var}\ (X)=\lambda$

泊松定理： 对任意常数$\lambda>0$, n为任意正整数, 设$np_n=\lambda$, 则对任意给定的非负整数k, 有 $\lim\limits_{n\to+\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$

泊松分布的应用: 若随机变量$X\sim B(n,p)$, 当n比较大而p比较小时, 令$\lambda=np$, 有$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\ \approx\ \frac{\lambda^k}{k!}e^{-\lambda}$

即利用**泊松分布近似计算二项分布**

常见分布

- 0/1分布：$X \sim Ber(p)$，$E(X) = p \quad Var(X) = p(1-p)$
- 二项分布：$X \sim B(n,p)$，$E(X) = np \quad Var(X) = np(1-p)$
- 几何分布：$X \sim G(p)$，$\mathrm{E}(X) = \frac{1}{p} \quad \mathrm{Var}(X) = \frac{1-p}{p^2}$

  \- 无记忆性：$P(X > m + n \mid X > m) = P(X > n)$
- 负二项分布：X服从参数为r和p的负二项分布

$$\mathrm{E}(X) = \frac{r}{p} \qquad 和 \qquad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

- 泊松分布：$X \sim P(\lambda)$, $E(X) = \lambda \quad \mathrm{Var}\,(X) = \lambda$

  \- 泊松定理：$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$

例题

设随机变量$X \sim \mathrm{U}(2,5)$, 若对X进行三次独立观测，求至少两次观察值大于3的概率。

概率密度函数

设随机变量X的分布函数为$F(x)$, 如果存在可积函数$f(x)$, 使得对任意实数x有$F(x) = \int_{-\infty}^{x} f(t)dt$成立, 则称$X$为连续型随机变量, 函数$f(x)$为随机变量$X$的**概率密度函数**, 简称**概率密度**

概率密度函数$f(x)$满足

$$\left.\begin{array}{l} \bullet\ \text{非负性：} f(x) \geq 0 \\ \bullet\ \text{规范性：} \int_{-\infty}^{+\infty} f(t)dt = 1 \end{array}\right\} \longleftrightarrow \textbf{概率密度}$$

- 任意$x_1 < x_2$，有$P(x_1 \leq X \leq x_2) = \int_{-\infty}^{x} f(t)dt, P(X = x_2) = 0$
- $F(x)$在$(-\infty, +\infty)$连续
- 若$f(x)$在x_0连续，则$F(x)$在x_0可导，且有$f(x_0) = F'(x_0)$
- 若$f(x)$为偶函数，则有$F(x) + F(-x) = 1$

例题

1) 若X的概率密度函数为$f(x)$，则以下为概率密度函数的有(   )

A) $f(2x)$ B) $f^2(x)$ C) $2xf(x^2)$ D) $3x^2f(x^3)$

2) 若X与$-X$具有相同的概率密度函数，则$F(x)+F(-x)=1$

3) 若X的概率密度函数为偶函数，则$F(-x)=\frac{1}{2}-\int_0^x f(t)dt$

均匀分布和指数分布

若随机变量X的概率密度为 $f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & \text{其它}\end{cases}$，称$\boldsymbol{X}$**服从区间**$[\boldsymbol{a},\boldsymbol{b}]$**上的均匀分布, 记**$\boldsymbol{X\sim U(a,b)}$

若$X\sim U(a,b)$,则$E(X)=(a+b)/2$, $\mathrm{Var}(X)=(b-a)^2/12$

给定常数$\lambda>0$, 若随机变量X的密度函数$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0\\ 0 & \text{其它}\end{cases}$

称X服从参数为λ的指数分布, 记 $X\sim e(\lambda)$

若随机变量$X\sim e(\lambda)$, 则$E(X)=1/\lambda$和$\mathrm{Var}(X)=1/\lambda^2$

唯一具有无记忆性的连续随机变量：若随机变量$X\sim e(\lambda)$, 则对任意$s>0,t>0$, 有$P(X>s+t|X>t)=P(X>s)$

例题

1）设随机变量$\xi \sim U(-3,6)$，试求方程$4x^2 + 4\xi x + \xi + 2 = 0$有实根的概率

2）已知随机变量$Y \sim e(1)$, 对任意$a > 0$求$P[Y \leq a + 1|Y > a]$

3）随机变量$X \sim e(\lambda)$, 且X落入$(1,3)$内的概率达到最大，求λ

正态分布

给定$u \in (-\infty, +\infty)$和$\sigma > 0$, 随机变量X的概率密度为

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

称X服从参数为μ, σ^2的正态分布，$N(0,1)$为标准正态分布

性质：

- 若$X \sim \mathrm{N}(\mu, \sigma^2)$, 则$\mathrm{E}(X) = \mu$和$\mathrm{Var}(X) = \sigma^2$
- 若$X \sim \mathrm{N}(\mu, \sigma^2)$, 则$\frac{X-\mu}{\sigma} \sim N(0,1)$
- 若$X \sim N(0,1)$, 则$\sigma X + \mu \sim N(\mu, \sigma^2)$
- 标准正态分布$N(0,1)$的分布函数$\Phi(x) = P(X \leq x)$，偶函数
- 标准正态分布的α分位数u_α满足$P(X > u_\alpha) = \alpha, u_{1-\alpha} = -u_\alpha$

例题

1) X服从正态分布$N(0,1)$，给定$\alpha \in (0,1)$，数u_α满足$P(X > u_\alpha) = \alpha$，若$P(|X| \le x) = \alpha$，则x等于 ( )

A) $u_{\alpha/2}$ B) $u_{1-\alpha/2}$ C) $u_{(1-\alpha)/2}$ D) $u_{1-\alpha}$

2) 设$f_1(x)$是$N(0,1)$的概率密度函数， $f_2(x)$是$[-1,3]$均匀分布的概率密度函数，若

$$f(x) = \begin{cases} \mathrm{a}f_1(x) & x \le 0 \\ \mathrm{b}f_2(x) & x > 0 \end{cases}$$

为概率密度函数，则a, b应满足什么条件

随机变量函数的概率分布

离散随机变量：X的分布列为$P(X=x_i)=p_i$，则$Y=g(X)$的分布列$P(\mathrm{Y}=g(x_i))=p_i$ (若值相同则需要合并)

X的分布列为$P(X=\mathrm{n})=1/2^n(n=1,2\ldots)$，求$Y=\sin(X\pi/2)$的分布函数

随机变量函数的概率分布

已知连续随机变量X的概率密度为$f_X(x)$，求新随机变量$Y = g(X)$的概率密度$f_Y(y)$？

求解步骤：1）求解Y=g(X)的分布函数

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = \int_{g(x)\le y} f_X(x)dx$$

2）利用分布函数和概率密度关系求解密度函数$f_Y(y) = F_Y'(y)$

数学工具 — 积分求导公式

$$F(y) = \int_{\psi(y)}^{\varphi(y)} f(x)dx$$

$$F'(y) = f\big(\varphi(y)\big)\varphi'(y) - f\big(\psi(y)\big)\psi'(y)$$

例题

已知X的分布函数$F_X(x)$，求$Y = 2X + 1$的分布函数

已知X服从均匀分布$U(-\pi/2, \pi/2)$，求$Y = \sin(x)$的概率密度

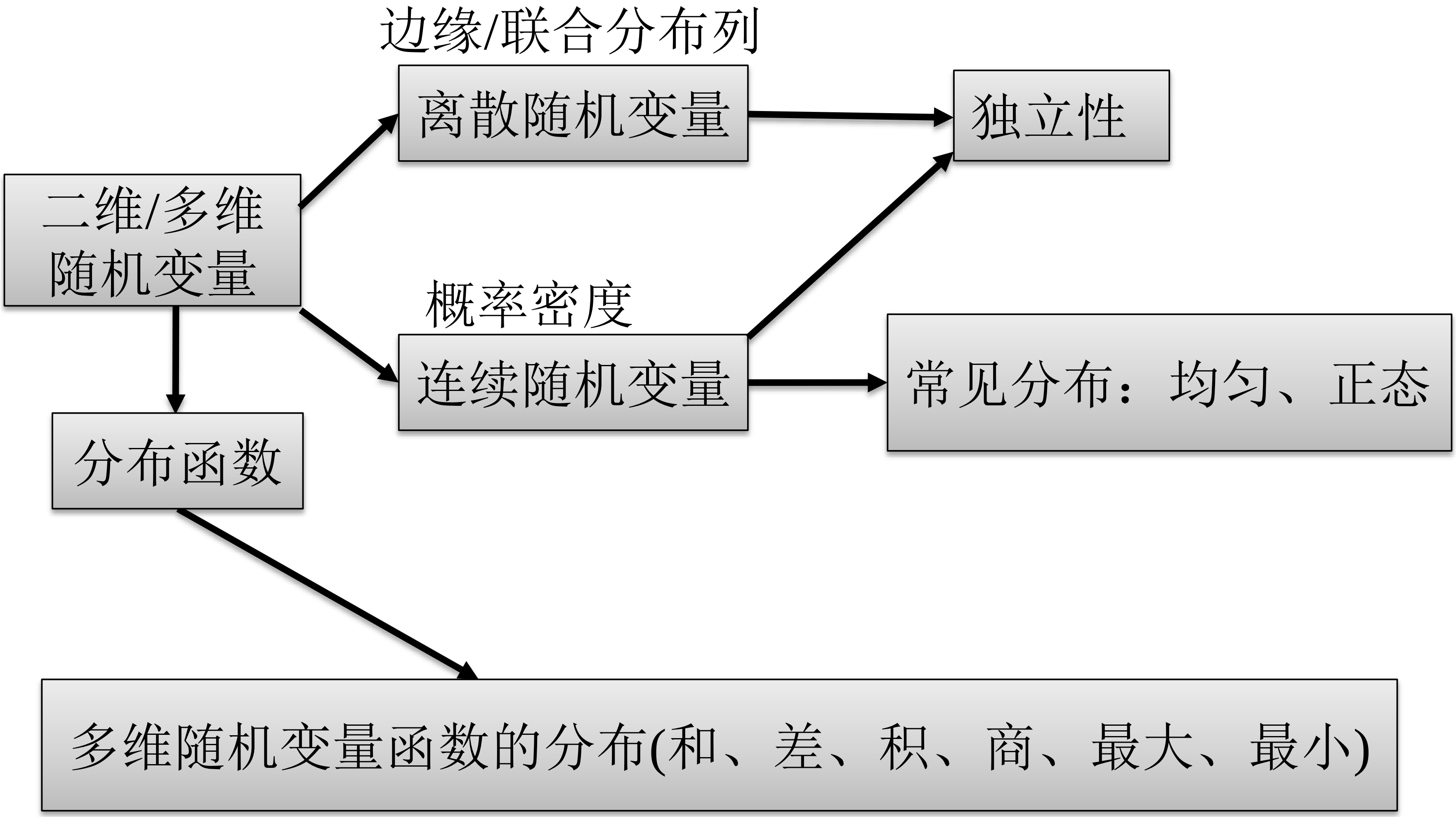
练习题：

1) 若$X \sim N(0,1)$，求$Y = e^X, Y = |X|, Y = X^2 + 1$的概率密度

2)若$X \sim e(1)$，求$Y = \mathrm{e}^{\mathrm{X}}$的概率密度

第三章 多维随机变量

知识结构



边缘/联合分布列
离散随机变量
独立性
二维/多维随机变量
概率密度
连续随机变量
常见分布：均匀、正态
分布函数
多维随机变量函数的分布(和、差、积、商、最大、最小)

二维随机变量

设$X = X(\omega)$和$Y = Y(\omega)$为定义在样本空间Ω上的随机变量，由它们构成的向量(X, Y)称为**二维随机向量**

设(X, Y)为二维随机变量，对任意实数$x \in (-\infty, +\infty)$和$y \in (-\infty, +\infty)$，称$F(x, y) = P(X \leq x, Y \leq y)$为**二维随机变量$(X, Y)$的分布函数**, 或称为**随机变量$X$和$Y$的联合分布函数**

- 分布函数$F(x, y)$对每个变量单调不减
  - 固定y, 当$x_1 > x_2$时有$F(x_1, y) \geq F(x_2, y)$
  - 固定x, 当$y_1 > y_2$时有$F(x, y_1) \geq F(x, y_2)$
- 对任意$x \in (-\infty, +\infty)$和$y \in (-\infty, +\infty)$，分布函数$F(x, y) \in [0,1]$, 且$F(+\infty, +\infty) = 1$和

$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$$

- 分布函数$F(x, y)$关于每个变量右连续

边缘分布与独立性

设二维随机变量(X,Y)的联合分布函数为$F(x,y)$, 称

$$F_X(x)=P(X\le x)=P(X\le x,y<+\infty)=\lim_{y\to+\infty}F(x,y)$$

为随机变量X的边缘分布函数.

同理定义随机变量Y的边缘分布函数为:

$$F_Y(y)=P(Y\le y)=P(Y\le y,x<+\infty)=\lim_{x\to+\infty}F(x,y)$$

$X\le x$和$Y\le y$相互独立, $P(X\le x,Y\le y)=P(X\le x)P(Y\le y)$,
等价于$F(x,y)=F_X(x)\,F_Y(y)$，则称随机变量***X*与*Y*相互独立**

随机事件的独立性：$P(AB)=P(A)P(B)$

随机变量的独立性

例题：设二维随机变量(X,Y)的联合分布函数为

$$F(x,y)=\begin{cases}A(B+\arctan x)(C-e^{-y}) & x\in R, y>0\\ 0 & \text{其它}\end{cases}$$

求$F(0,1)$

二维离散型随机变量

若二维随机变量 (X,Y) 的取值是有限个或无限可列的, 称 (X,Y) 为二维离散型随机变量

设离散型随机变量(X,Y)的取值分别为(x_i, y_j)，$i,j = 1,2,\cdots$，则称

$$p_{ij} = P(X = x_i, Y = y_j)$$

为**(X,Y)的联合分布列**

$X \backslash Y$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$

性质：$p_{ij} \geq 0$和$\sum_{i,j} p_{ij} = 1$

边缘分布列

根据二维随机变量(X,Y)的联合分布列p_{ij}

随机变量X的边缘分布列

$$P(X=x_i)=\sum_{j=1}^{+\infty}P(X=x_i,Y=y_j)=\sum_{j=1}^{+\infty}p_{ij}=p_{i\cdot}$$

随机变量Y的边缘分布列

$$P(Y=y_i)=\sum_{i=1}^{+\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{+\infty}p_{ij}=p_{\cdot j}$$

边缘分布列

二维随机变量联合和边缘分布表示在同一个表格

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

例题

有三个数1,2,3，随机变量X表示从这三个数中随机地抽取一个数，随机变量Y表示从1到X中随机抽取一个数. 求(X,Y)的联合分布列和边缘分布列

离散随机变量的独立性

对离散型随机变量(X,Y), 若对所有(x_i,y_j)有

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

即$p_{ij}=p_{i.}\,p_{.j}$，称**离散随机变量X与Y相互独立**

定理：对离散型随机变量(X,Y), 以下两种定义等价

$$p_{ij}=p_{i.}\,p_{.j} \leftrightarrow F(x_i,y_j)=F_X(x_i)\,F_Y(y_j)$$

离散随机变量的独立性

定理：设离散随机变量X和Y独立, 则对任意集合$A \in \mathrm{R}$, $B \in \mathrm{R}$, 有事件$X \in A$和$Y \in B$独立

例题：设离散型随机变量X, Y独立, 求解(X, Y)的联合分布列

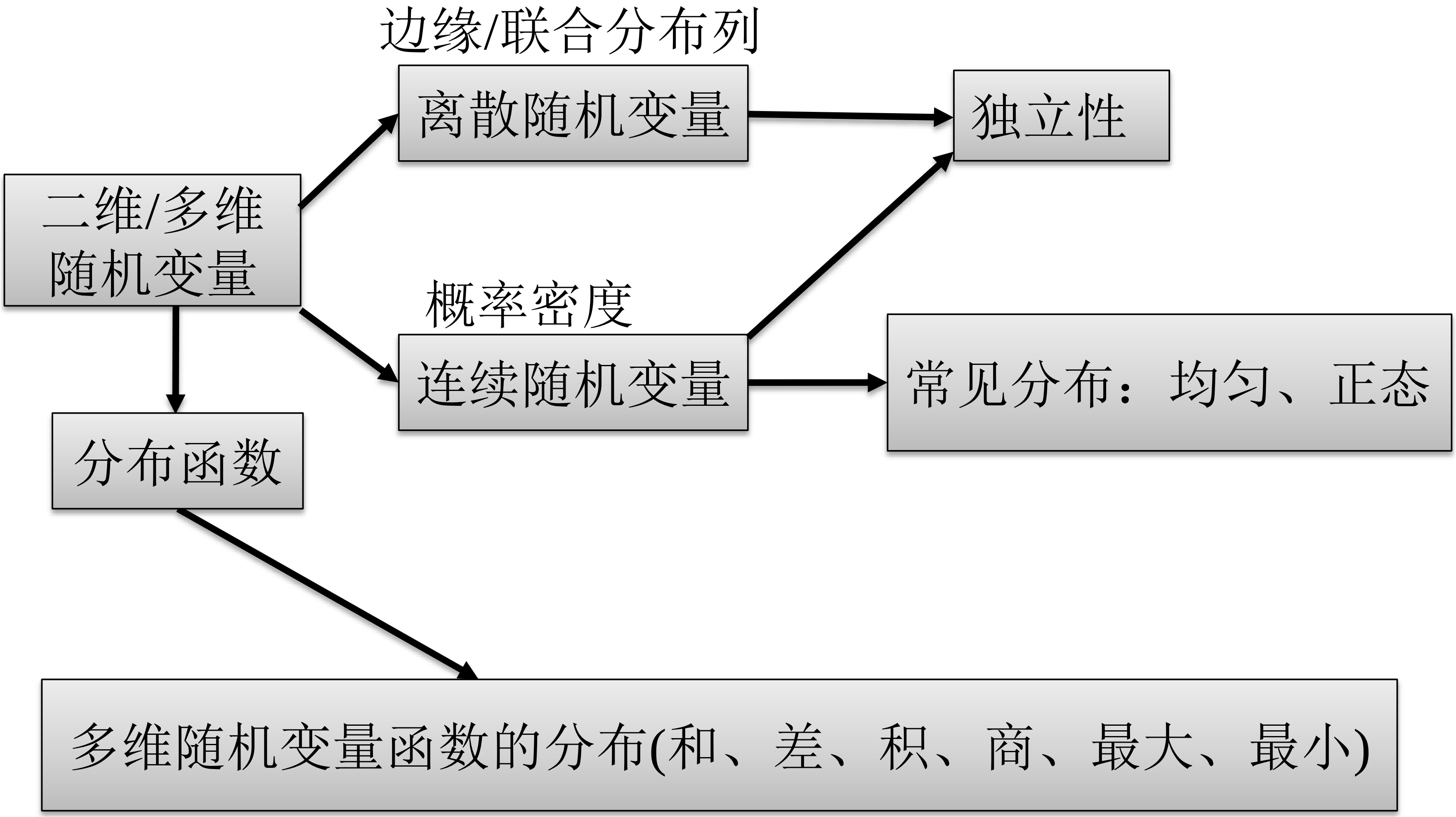
$X \backslash Y$	y_1	y_2	y_3	$p_{i\cdot}$
x_1		$1/8$		
x_2	$1/8$			
$p_{\cdot j}$	$1/6$			

例题

将两个球A,B放入编号为1,2,3的三个盒子中, 用随机变量X放入1号盒的球数, 用随机变量Y表示放入2号盒的球数, 判断X和Y是否独立

第三章 多维随机变量

知识结构


边缘/联合分布列
离散随机变量
独立性
二维/多维
随机变量
概率密度
连续随机变量
常见分布：均匀、正态
分布函数
多维随机变量函数的分布(和、差、积、商、最大、最小)

二维连续型随机变量

二维随机变量的分布函数为$F(x,y)$，如存在二元非负可积函数$f(x,y)$ 使 得 对 (x,y) 有 $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$ 则 称 **$f(x,y)$称为随机变量(X,Y)的概率密度**, 或称**联合概率密度**

- 非负性：$f(x,y)\geq 0$;
- 规范性：$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(u,v)dudv$
- 若$f(x,y)$ 在(x,y) 连续, 则$f(x,y)=\partial^2F(x,y)/\partial x\partial y$
- 若G为平面上的一个区域, 则点(X,Y)落入G的概率为

$$P((X,Y)\in G)=\iint_{(x,y)\in G}f(x,y)dxdy$$

例题

设X和Y的分布函数为

$$F(x,y)=\begin{cases}(a-e^{-x})(b-e^{-2y}) & x,y\in[0,+\infty)\\ 0 & \text{其他}\end{cases}$$

其中a,b 非负，求$F(1,1)$

例题

设X和Y的概率密度函数为

$$f(x,y)=\begin{cases}A(x+y) & x,y\in[0,1]\\ 0 & \text{其他}\end{cases}$$

求

- $P(X+Y\le 1)$
- $P(\max(X,Y)\le 1/2)$
- $P(\min(X,Y)\le 1/2)$
- $E[\max(X,Y)+\min(X,Y)]$
- $E[\max(X,Y)\times\min(X,Y)]$

边缘概率密度

设二维随机变量(X,Y)的概率密度为$f(x,y)$，则随机变量X和Y的边缘概率密度为

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

随机变量X和Y相互独立的等价条件：

- 分布函数$F(x,y)=F_X(x)F_Y(y)$
- 概率密度$f(x,y)=f_X(x)f_Y(y)$
- 条件概率$f_{Y|X}(y|x)=f_Y(y)$

例题

设二维随机变量(X,Y)的概率密度

$$f(x,y)=\begin{cases} c & 0<x<1, x^2<y<x \\ 0 & \text{其它} \end{cases}$$

求边缘概率密度

练习

1）设X和Y的分布函数为

$$F(x,y)=\begin{cases} 0 & x\le 0, y\le 0 \\ 2xy-x^2 & 0<x<y<1 \\ y^2 & 0<y\le x, y\le 1 \\ 2x-x^2 & 0<x\le 1, y\ge 1 \\ 1 & 1<y, 1<x \end{cases}$$

求X和Y的边缘分布函数

2）设(X,Y)的分布函数为$F(x,y)$，边缘分布函数分别为$F_X(x)$和$F_Y(y)$, 则Z=max(X,Y)的分布函数为(　)

A) $F_X(x)F_Y(y)$　B) $F_X(x)F_Y(x)$　C) $F(x,x)$　D) $F(x,y)$

二维正态分布

设$|\rho|<1$, 令$\mu=\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix}$ $\Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}$

若随机变量X和Y的联合概率密度函数为

$$f(x,y)=(2\pi)^{-2/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(\xi-\mu)^{\top}\Sigma^{-1}(\xi-\mu)}\qquad \xi=\begin{pmatrix}x\\ y\end{pmatrix}$$

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right]}$$

称$\boldsymbol{X}$**和**$\boldsymbol{Y}$**服从参数为**$\boldsymbol{\mu}$**和**$\boldsymbol{\Sigma}$**的正太分布**, 记 $(\boldsymbol{X},\boldsymbol{Y})\sim \boldsymbol{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$

- 参数的含义： μ,Σ,ρ
- X和Y的条件概率
- X与Y的独立性
- 联合分布可以推出边缘分布，反只不成立

条件概率

条件分布列： 二维离散型随机变量(X,Y)的分布列为$\{p_{ij}\}$，称

$P\left(X=x_i \middle| Y=y_j\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(Y=y_j\right)}=\frac{p_{ij}}{p_{\cdot j}}$ 在$Y=y_j$条件下随机变量X的条件分布列

条件概率密度： 随机变量(X,Y)的联合概率密度为$f(x,y)$，以及Y的边缘概率密度为$f_Y(y)>0$，称$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$在$\boldsymbol{Y=y}$**条件下随机变量$\boldsymbol{X}$的条件概率密度**

性质与例题

性质：非负性、规范性

条件概率的计算

$$P(a < X < b|Y = y_0) = \int_a^b f_{X|Y}(x|y_0)dx = \int_a^b f(x,y)/f_Y(y_0)dx$$

例题：已知随机变量$X \sim U(0,1)$，当观察到$X = x$的条件下随机变量$Y \sim U(x,1)$，求Y的概率密度以及概率$P(X + Y \leq 1)$

随机变量函数的分布

离散：$p_{ij} = P(X = x_i, Y = y_j), Z = g(X, Y)$ 将相同值合并

连续：i) 给出分布函数$f(x, y)$，求和、差、积

ii) 两个独立随机变量的和、差、积

方法：公式讨论独立随机变量的和、差、积、商、最大、最小

分布函数法

- Step 1: (X, Y) 的联合概率密度$f(x, y)$，则Z的分布函数为

$$F_Z(z) = P(g(X, Y) \leq z) = \int_{g(x,y)\leq z} f(x, y)dxdy$$

- Step 2: 求概率密度$f_Z(z) = F_z'(z)$

难点：分类讨论

随机变量函数的结论

多维随机变量函数$\max(X,Y)$ 和 $\min(X,Y)$

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

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

$$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|f(x,xz)dx$$

$X\sim B(n_1,p)$和$Y\sim B(n_2,p)$独立, $X+Y\sim B(n_1+n_2,p)$

$X\sim P(\lambda_1)$ 和$Y\sim P(\lambda_2)$ 独立, 则$X+Y\sim P(\lambda_1+\lambda_2)$

$X\sim N(\mu_1,\sigma_1^2)$和$Y\sim N(\mu_2,\sigma_2^2)$ 独立, 则

$$X+Y\sim \mathrm{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

例题

设(X,Y)的联合概率密度为

$$f(x,y)=\begin{cases}1 & x,y\in[0,1]\\ 0 & \text{其他}\end{cases}$$

求$Z=|X-Y|$的概率密度

综合题

设随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}Ae^{-y} & y>x>0\\ 0 & \text{其它}\end{cases}$$

求1) 常数A

2) 求X与Y的边缘概率密度

3) 求X与Y是否独立?

4) 求$Z=X+Y$的概率密度

5) 求条件概率密度$f_{X|Y}(x|y)$, 或条件概率$P(1<X<2|Y=1)$

第四章 随机变量的数字特征

知识结构

定义、性质

期望

常见分布期望

随机变量函数的期望

定义、性质

方差

常见分布方差

定义、性质

协方差

正态分布协方差

相关系数

独立与不相关

期望及其性质

离散随机变量**期望** $\boldsymbol{E(X)=\sum_k p_k x_k}$

连续随机变量$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$

若随机变量$X\equiv c$, 则$E(c)=c$

对随机变量X和常数$a,b\in \mathrm{R}$, 有$E(aX+b)=aE(X)+b$

对随机变量X,Y和常数$a,b\in \mathrm{R}$, 有$E(aX+bY)=aE(X)+bE(Y)$

对随机变量X和连续凸函数$g:[a,b]\to \mathrm{R}$, 有$g\big(E(X)\big)\le E\big(g(X)\big)$

离散变量X, 及连续函数$g: R\to R$, 有$E[g(X)]=\sum_{k\ge 1} g(x_k)p_k$

连续变量X, 及连续函数$g: R\to R$, 有$E\big(g(X)\big)=\int_{-\infty}^{+\infty} g(t)f(t)dt$

例题

- 设随机变量$X \sim \mathrm{E}(\lambda)$, 则$P\left(X > \sqrt{\mathrm{Var}(X)}\right) =$
- X与Y相互独立且期望存在，记$U = \max(X, Y), V = \min(X, Y)$, 则$E(UV) =$
- 设X与Y服从正态分布，则下列服从正太分布的有( )

  A) $X + Y$ B) $X - Y$ C) (X, Y) D) $2X + 1$

- 设随机变量X的分布函数$F(x) = 0.3\Phi(x) + 0.7\Phi((x-1)/2)$，其中$\Phi(x)$为标准正态分布的分布函数，则$E(X) =$

  A) 0 B) 0.3 C) 0.7 D) 1

方差的定义及其性质

随机变量方差：$Var\ (X) = E\big(X - E(X)\big)^2 = E(X)^2 - \big(E(X)\big)^2$

- 若随机变量$X \equiv c$, 则$\mathrm{Var}(X) = 0$
- 对随机变量X和常数$a, b \in\ R$, 有

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}\ (X)$$

- 一般情况下方差不具有线性性, 即

$$\mathrm{Var}\big(f(X) + g(X)\big) \neq\ \mathrm{Var}\big(f(X)\big) + \mathrm{Var}\big(g(X)\big)$$

常见分布的期望和方差

- 0/1分布：$X \sim Ber(p)$，$E(X) = p$ $\quad Var(X) = p(1-p)$
- 二项分布：$X \sim B(n,p)$，$E(X) = np$ $\quad Var(X) = np(1-p)$
- 几何分布：$X \sim G(p)$，$\mathrm{E}(X) = \frac{1}{p}$ $\quad \mathrm{Var}(X) = \frac{1-p}{p^2}$
- 负二项分布：X服从参数为r和p的负二项分布

$$\mathrm{E}(X) = \frac{r}{p} \qquad \text{和} \qquad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

- 泊松分布：$X \sim P(\lambda)$, $E(X) = \lambda$ $\quad \mathrm{Var}\ (X) = \lambda$

均匀分布 $X \sim U(a,b)$: $E(X) = \frac{a+b}{2}$ $\quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$

指数分布 $X \sim e(\lambda)$: $E(X) = \frac{1}{\lambda}$ $\quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$

正态分布 $X \sim N(\mu,\sigma^2)$：$E(X) = \mu$ $\quad \mathrm{Var}(X) = \sigma^2$

多元随机变量的期望

离散随机变量 $E[Z] = E[g(X,Y)] = \sum_{i,j} g\left(x_i, y_j\right)p_{ij}$

连续随机变量$E[Z] = E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$

- 对任意随机变量X, Y和常数a, b有$E[aX + bY] = aE[X] + bE[Y]$
- 对**独立**随机变量X和Y, 以及任意函数$h(x)$和$g(y)$, 有

$$E[XY] = E[X]E[Y] \text{ 和 } E[h(X)g(Y)] = E[h(X)]E[g(Y)]$$

- 对任意随机变量X和Y, 有Cauchy-Schwartz不等式

$$E[XY] \leq \sqrt{E[X^2]E[Y^2]}$$

协方差及其性质

协方差: $\mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y)$

对随机变量X和常数c, 有$\mathrm{Cov}(X,c) = 0$.

交换律 $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

对任意常数a和b, 随机变量X和Y, 有

$$\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y)\ ,$$

$$\mathrm{Cov}(X+a,Y+b) = \mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(X_1+X_2,Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$$

$$Cov\left(\sum_{i=1}^{n} X_i\,, \sum_{j}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i,Y_j)$$

$$Var\left(\sum_{i}^{n} X_i\right) = \sum_{i}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i,X_j)$$

相关系数的定义

X与Y的相关系数：$\rho_{XY} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$

- $|\rho_{XY}| = 1$的充要条件为X与Y有线性关系$Y = aX + b$

- **本质上$\boldsymbol{\rho_{XY}}$刻画了$\boldsymbol{X}$与$\boldsymbol{Y}$的线性相关程度, 又称为“线性相关系数”**

若相关系数$|\rho_{XY}| = 0$，则随机变量***X*和*Y*不相关**

X和Y不相关　　　　X和Y独立

例题

随机变量X和Y在以点$(0,1),(1,0),(1,1)$为顶点的三角形区域服从均匀分布，求$U=X-Y$的方差

第五章 大数定律与中心极限定理

问题

什么是大数定律？

什么是中心极限定理？

其局限是什么，可以采用什么方式弥补

知识结构

大数定律 → **依概率收敛** → 四个大数定律

中心极限定理 → **依分布收敛** → 两个中心极限定理

大数定律

设$X_1, X_2, \cdots, X_n, \cdots$是一随机变量序列，$a$是一常数，如果对任意$\epsilon > 0$有$\lim_{n\to\infty} P[|X_n - a| < \epsilon] = 1$, or $\lim_{n\to\infty} P[|X_n - a| > \epsilon] = 0$
则称随机变量序列$X_1, X_2, \cdots, X_n, \cdots$依概率收敛于$a$, 记$X_n \xrightarrow{P} a$

若随机变量序列$X_1, X_2, \cdots, X_n, \cdots$满足

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E[X_i],$$

则称$\{\boldsymbol{X_n}\}$**服从大数定律**

大数定理刻画了随机变量的均值 (算术平均值) 依概率收敛于期望的均值 (算术平均值)

大数定律总结

Markov大数定律: 若随机变量序列$\{X_i\}$满足$\mathrm{Var}(\sum_{i=1}^{n} X_n)/n^2 \to 0$, 则满足大数定律

Chebyshev大数定律: 若独立随机变量序列$\{X_i\}$满足$\mathrm{Var}(X_i) \leq c$, 则满足大数定律

Khintchine大数定律: 若独立同分布随机变量序列$\{X_i\}$期望存在, 则满足大数定律

Bernoulli大数定律: 对二项分布$X_n \sim \mathrm{B}(n, p)$, 有$X_n/n \xrightarrow{P} p$

依分布收敛

设随机变量Y的分布函数为$F_Y(y)=P(Y\leq y)$，以及随机变量序列$Y_1,Y_2,\cdots,Y_n,\cdots$的分布函数分别为$F_{Y_n}(y)=P(Y_n\leq y)$, 如果

$$\lim_{n\to\infty} P[Y_n\leq y]=P[Y\leq y] \text{ 即 } \lim_{n\to\infty} F_{Y_n}(y)=F_Y(y)$$

则称随机变量序列$\boldsymbol{Y_1,Y_2,\cdots,Y_n,\cdots}$**依分布收敛于**$\boldsymbol{Y}$, 记$\boldsymbol{Y_n\xrightarrow{d}Y}$.

- 林德贝格-勒维中心极限定理：独立同分布随机变量，若$E[X_k]=\mu$和$\mathrm{Var}(X_k)=\sigma^2$, 则

$$\frac{\sum_{k=1}^{n}X_k-n\mu}{\sigma\sqrt{n}}\xrightarrow{d}N(0,1)$$

- 李雅普诺夫定理: 独立不同分布中心极限定理

第六章 集中不等式

知识结构

基础不等式： Markov不等式、Chebyshev不等式、Hölder不等式、Cauchy-Schwartz不等式、单边Chebyshev不等式

Chernoff 方法

Chernoff 引理

0/1 Chernoff 界

有界Chernoff 界

Sub-Gaussian随机变量: 有界随机变量、Gaussian随机变量

基础不等式

Markov不等式: $\boldsymbol{P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}}$

Chebyshev不等式: $\boldsymbol{P(|X-\mu| > \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}}$

单边Chebyshev不等式: $P(X-\mu \geq \epsilon) \leq \frac{\sigma^2}{\sigma^2+\epsilon^2}$

Hölder不等式: $E(|XY|) \leq \left(E(|X|^p)\right)^{\frac{1}{p}}\left(E(|Y|^q)\right)^{\frac{1}{q}}$

例题

随机变量X和Y满足$E(X) = 2$，$E(Y) = 2$，$\text{Var}(X) = 1, \text{Var}(Y) = 4, \rho_{XY} = -1/2$. 利用Chebyshev不等式估计$\text{P}(|X - Y| \geq 6) \leq ???$

设随机变量$X \sim N(-1,2)$和$Y \sim E(1)$，且X和Y的相关系数为-1/2，利用Chebyshev不等式估计$P(|X + Y| \geq 6) \leq ???$

Chernoff方法

对任意$\epsilon > 0$和$t < 0$有

$$P[X \leq -\epsilon] = P[tX \geq -t\epsilon] \leq e^{t\epsilon}E[e^{tX}]$$

同理有

$$P[X \leq -\epsilon] \leq \min_{t<0}\{e^{t\epsilon}E[e^{tX}]\}.$$

称为Chernoff方法, 是证明集中不等式最重要的方法之一

有界的Chernoff不等式

Chernoff 引理: 随机变量$X \in [0,1]$的期望$\mu = E[X]$. 对$\forall t > 0$有

$$E[e^{tX}] \le \mathrm{e}^{t\mu + \frac{t^2}{8}}$$

推论: $X \in [a, b]$期望$\mu = E[X]$, 对$\forall t > 0$有$E[e^{tX}] \le \mathrm{e}^{t\mu + \frac{t^2(b-a)^2}{8}}$

Chernoff不等式：假设$X_1, \ldots, X_n$是n独立的随机变量、且满足$X_i \in [a, b]$. 对任意$\epsilon > 0$有

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \ge \epsilon\right] \le e^{-2n\epsilon^2/(b-a)^2}$$

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \le -\epsilon\right] \le e^{-2n\epsilon^2/(b-a)^2}$$

例题

分类器f在集合$S_n = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$ 的错误率为$\hat{p} > 0$，请估计分类器f在分布$\mathcal{D}$上的错误率，求f在分布$\mathcal{D}$上的错误率为$(9\hat{p}/10, 11\hat{p}/10)$ 之间的概率.

第七章 统计的基本概念

知识结构

统计学基本概念 → 三大统计分布 → 五大抽样定理

统计学基本概念 ↓

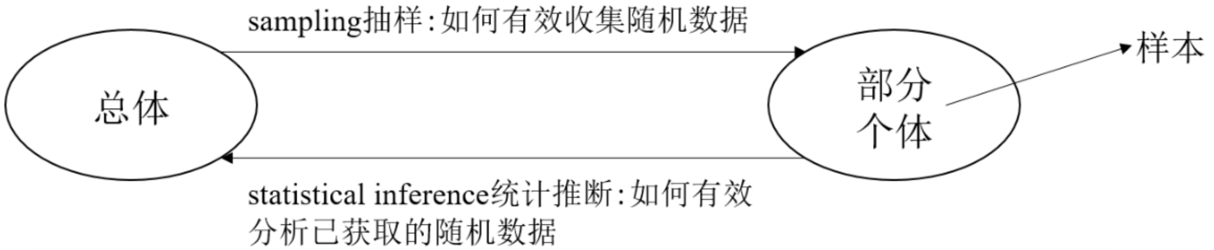
概念：分位点

统计学

统计学：以概率论为基础，研究如何有效收集研究对象的随机数据，以及如何运用所获得的数据揭示统计规律的一门学科.

统计学的研究内容包括：

- **抽样**
- **参数估计**
- **假设检验**



统计学基本概念

总体与样本

统计量： $g(X_1,\cdots,X_n)$ 关于$X_1,\cdots,X_n$连续、且不含任意参数

常见统计量：

- **样本均值**
- **样本方差**
- **样本标准差**
- **修正后的样本方差**
- k阶原点矩/中心矩
- 次序统计量

欧拉积分函数及其分布(**了解**)

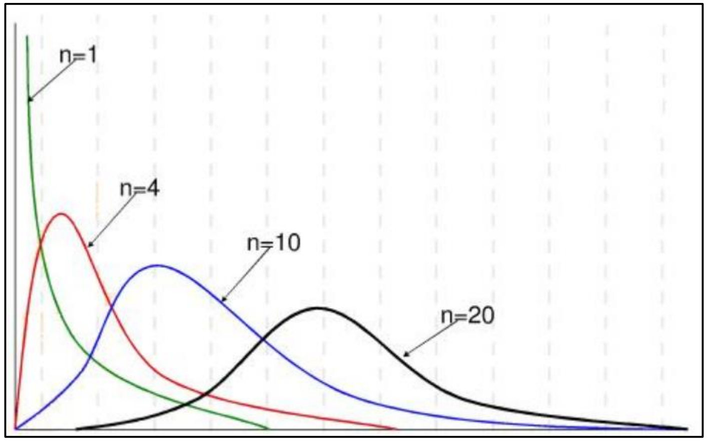
- Beta函数：$\text{Beta}\,(\alpha_1,\alpha_2)=\int_0^1 x^{\alpha_1-1}\,(1-x)^{\alpha_2-1}dx$
- Γ-函数为$\Gamma(\alpha)=\int_0^{+\infty} x^{\alpha-1}e^{-x}dx$
- 两类函数的关系

- Beta分布
- Γ-分布、性质、独立可加性
- 标准正态分布的平方Γ(1/2,1/2)
- Dirichlet分布、性质

χ^2分布

若$X_1, X_2, \cdots, X_n$是来自总体$X \sim \mathrm{N}(0,1)$ 的一个样本, 称$Y = X_1^2 + X_2^2 + \cdots + X_n^2$为服从自由度为$n$的$\chi^2$分布, 记$Y \sim \chi^2(n)$

根据$X_1^2 \sim \Gamma(1/2,1/2)$和$\Gamma$分布的独立可加性可得$Y \sim \Gamma(n/2,1/2)$



定理：随机变量$X \sim \chi^2(n)$, 则$E(X) = n$和$\mathrm{Var}(X) = 2n$;

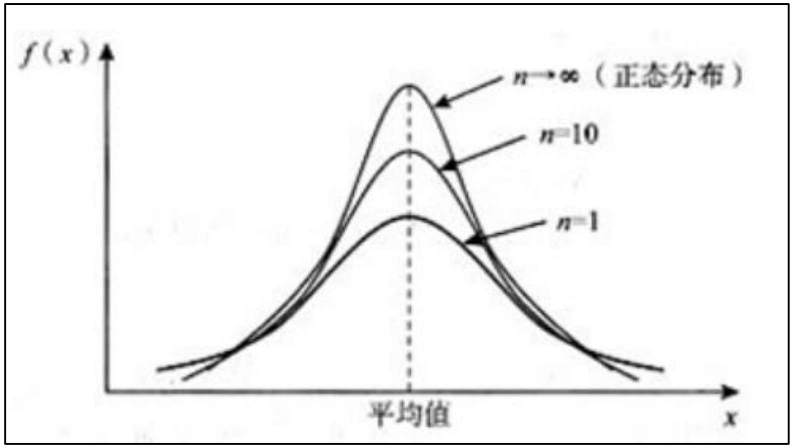
若随机变量$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$相互独立，则$X + Y \sim \chi^2(m+n)$

t分布(student distribution)

随机变量$X \sim \mathrm{N}(0,1)$和$Y \sim \chi^2(n)$ 相互独立, 则随机变量

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

服从自由度为n的t-分布, 记$T \sim t(n)$.



定理： 当$n \to \infty$时, 随机变量$T \sim t(n)$的概率密度

$$f(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

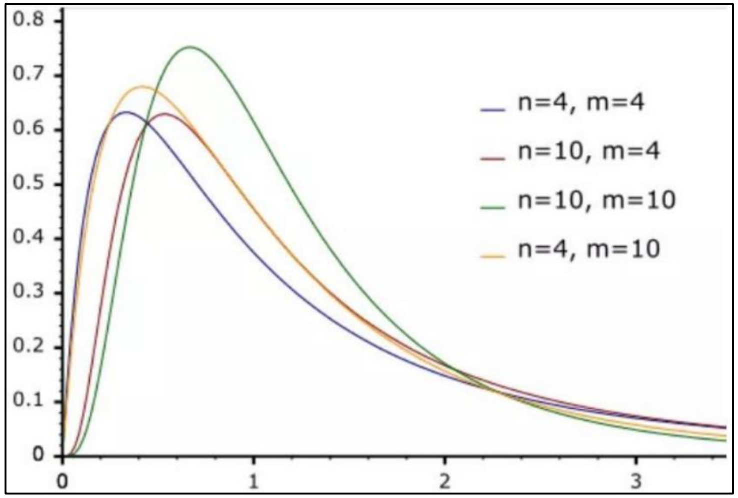
因此当n足够大时, $f(x)$ 被近似为$N(0,1)$ 的密度函数.

F分布

随机变量$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$相互独立, 称随机变量

$$F = \frac{X/m}{Y/n}$$

服从自由度为(m, n) 的F-分布，记$F \sim F(m, n)$.



定理： 若随机变量$F \sim F(m, n)$, 则$1/F = F(n, m)$.

例题

- 设 X_1, X_2, X_3, X_4 是来自于总体 $N(0,4)$ 的样本，以及 $Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$. 求$a, b$取何值时，$Y$服从$\chi^2$分布，并求其自由度.
- 独立同分布随机变量 $X_1, X_2, \cdots, X_n$ 满足 $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$，求 $\sum_{i=1}^{n}(X_i - \mu_i)^2/\sigma_i^2$的分布
- $X_1, X_2, \cdots, X_9$和$Y_1, Y_2, \cdots, Y_9$是来自总体$N(0,9)$两样本，求$(X_1 + X_2 + \cdots + X_9)/\sqrt{Y_1^2 + Y_1^2 + \cdots + Y_9^2}$的分布
- 设$X_1, X_2, \ldots, X_{2n}$来自总体$N(0, \sigma_2^2)$的样本，求 $(X_1^2 + X_3^2 + \cdots + X_{2n-1}^2)/(X_2^2 + X_4^2 + \cdots + X_{2n}^2)$的分布

例题

- 若随机变量$X \sim t(n)$, 求$Y = X^2$的分布.

- 设$X_1, X_2, \cdots, X_5$是来自总体$N(0,1)$的样本，令$Y = c_1(X_1 +$

抽样分布定理

定理一： 设$X_1, X_2, \cdots, X_n$是来自总体$N(\mu, \sigma^2)$的样本, 则有

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}(\mu, \sigma^2/n) \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

定理二： 设$X_1, X_2, \cdots, X_n$是来自总体$N(\mu, \sigma^2)$的样本,其样本均值和修正样本方差分别为

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1\}}^{n} (X_i - \bar{X})^2$$

则**有$\bar{X}$和S^2相互独立**, 且

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

抽样分布定理

定理三： 设$X_1, X_2, \cdots, X_m$是来自总体$N(\mu, \sigma^2)$的样本,其样本均值和修正样本方差分别为

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1\}}^{n} (X_i - \bar{X})^2$$

则

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t(n-1)$$

定理四： 设$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$分别来自总体$N(\mu_X, \sigma_X^2)$和$N(\mu_Y, \sigma_Y^2)$的两个独立样本, 其修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

正态分布的抽样分布定理五

定理五： 设$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$分别来自总体$N(\mu_X, \sigma^2)$和$N(\mu_Y, \sigma^2)$的两个独立样本，令其样本均值分别$\bar{X}$和$\bar{Y}$，修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

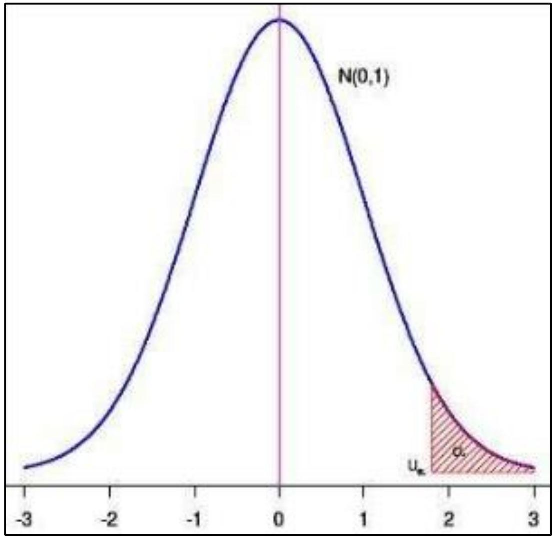
分位数

对正态分布$X \sim N(0,1)$, 给定$\alpha \in (0,1)$, 满足

$$P(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$$

的点 $\boldsymbol{\mu_\alpha}$**称为正态分布上侧**$\boldsymbol{\alpha}$**分位点**

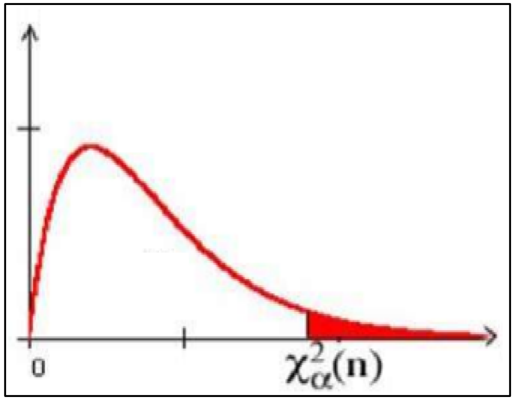
$\alpha = 1 - \Phi(\mu_\alpha)$， $\mu_{1-\alpha} = -\mu_\alpha$



对χ^2分布$X \sim \chi^2(n)$, 给定$\alpha \in (0,1)$, 满足

$$P\left(X \geq \chi^2_\alpha(n)\right) = \alpha$$

的点$\boldsymbol{\chi^2_\alpha(n)}$**称为**$\boldsymbol{\chi^2(n)}$**分布上侧**$\boldsymbol{\alpha}$**分位点**



$t-$分布和$F-$分布的分位数

对t-分布$X \sim t(n)$, 给定$\alpha \in (0,1)$, 满足$P\big(X > t_\alpha(n)\big) = \alpha$

的点$\boldsymbol{t_\alpha(n)}$**称为**$\boldsymbol{t(n)}$**-分布上侧**$\boldsymbol{\alpha}$**分位点**

由对称性可知$t_{1-\alpha}(n) = -t_\alpha(n)$

对F-分布$X \sim F(m,n)$, 给定$\alpha \in (0,1)$, 满足

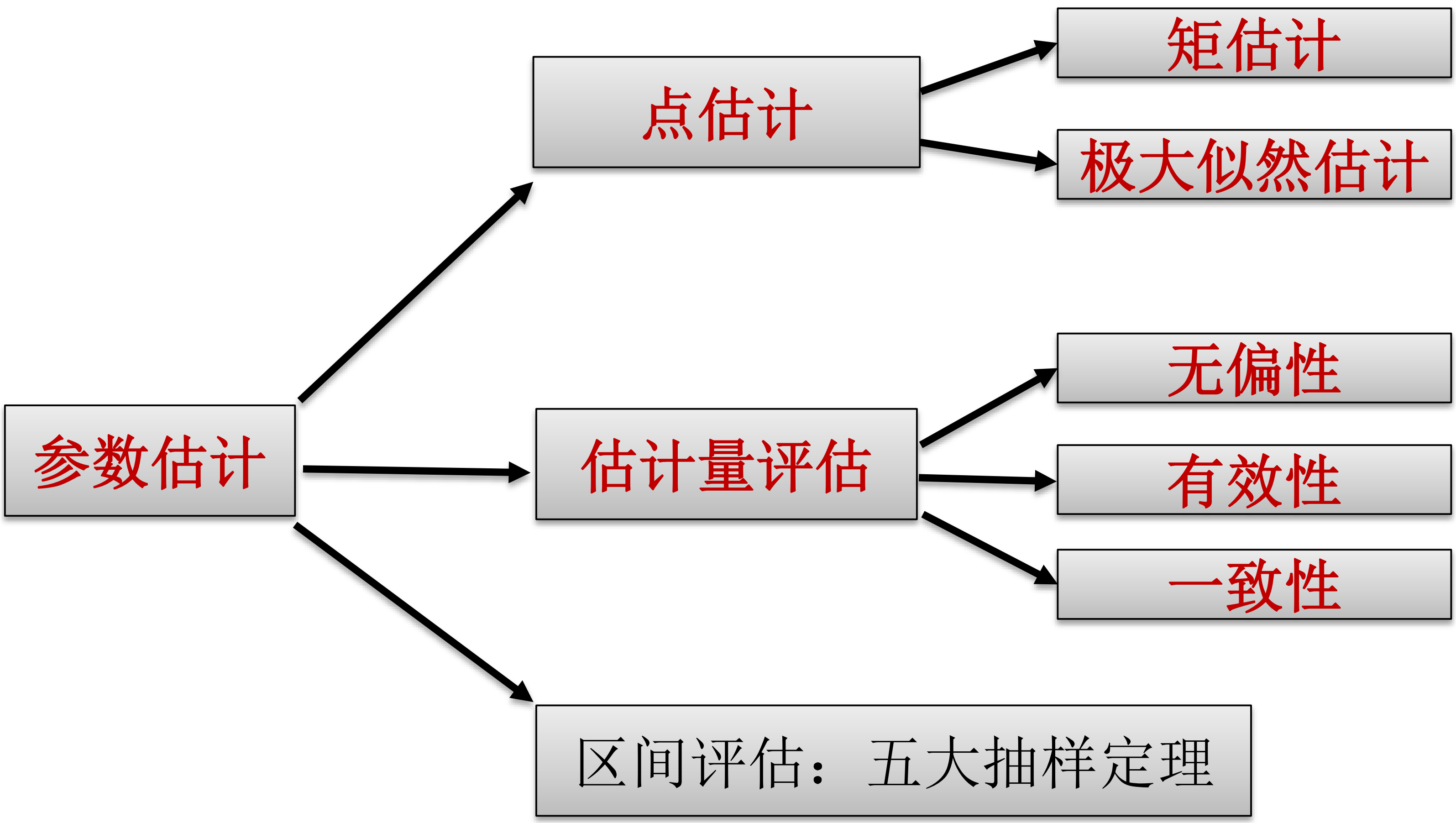
$$P[X > F_\alpha(m,n)] = \alpha$$

的点$\boldsymbol{F_\alpha(m,n)}$**称为**$\boldsymbol{F(m,n)}$ **分布上侧**$\boldsymbol{\alpha}$**分位点**

对F-分布的分位点有$F_{1-\alpha}(m,n) = \frac{1}{F_\alpha(n,m)}$.

第八章 参数估计

知识结构



参数估计

设总体X的分布/密度函数为$F(X,\theta)$, 其中θ为未知参数

现从总体中抽取一样本 $X_1,X_2,\cdots,X_n$

问题：如何依据样本$X_1,X_2,\cdots,X_n$估计参数θ，或θ的函数$g(\theta)$,
此类问题称为 **参数估计问题**

- 研究内容包括: 点估计, 估计量标准, 区间估计
- 点估计包括：矩估计法、极大似然估计法

矩估计法

总体X的k阶矩: $a_k = E[X^k]$　　样本k阶矩: $A_k = \sum_{i=1}^{n} X_i^k$

用**样本矩去估计总体矩**求参数θ的方法称为 **矩估计法**

总体X的k阶中心矩: $b_k = E\left[\left(X - E(X)\right)^k\right]$

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

用**样本中心矩去估计总体中心矩**求参数θ的方法亦称为 **矩估计法**

基础： 独立同分布变量$X_1, \cdots, X_n$, 若$E(X) = \mu$, 则$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$

矩估计方法

总体X的分布函数F包含m个未知参数$\theta_1, \theta_2, \cdots, \theta_m$

- 计算总体X的k阶矩：$a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) = E[X^k] \;\; k \in [m]$
  (a_k一般为$\theta_1, \theta_2, \cdots, \theta_m$的函数)

- 计算样本的k阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

- 令样本矩等于总体矩：

$$A_k = a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) \; k \in [m]$$

  得到m个关于$\theta_1, \theta_2, \cdots, \theta_m$的方程组

- 求解方程组得到估计量$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$

最大似然估计（离散）

设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本.

若总体X为离散型随机变量，其分布列为$P(X=x)=P(X=x;\theta)$, 则样本$X_1, X_2, \cdots, X_n$的分布列为

$$L(\theta)=L(x_1, x_2, \cdots, x_n; \theta)=\prod_{i=1}^{n} P(x_i; \theta)$$

这里$L(\theta)$表示样本$X_1=x_1, \cdots, X_n=x_n$发生的概率

称$\boldsymbol{L(\theta)}$**为样本**$\boldsymbol{X_1, X_2, \cdots, X_n}$**的似然函数**

最大似然估计（连续）

若总体X为连续型随机变量，其概率密度为$f(x;\theta)$，则$X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$的联合概率密度为

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

根据概率密度定义可知$L(\theta)$ 越大，样本$X_1, X_2, \cdots, X_n$落入$x_1, x_2, \cdots, x_n$的邻域内概率越大

称**$L(\theta)$为样本$X_1, X_2, \cdots, X_n$的似然函数**

最大似然函数及不变性

综合离散和连续随机变量，统称$\boldsymbol{L(\theta)}$**为样本**$\boldsymbol{X_1, X_2, \cdots, X_n}$**的似然函数**，若

$$\hat{\theta} = \operatorname{argmax}_{\theta} L(x_1, x_2, \cdots, x_m; \theta)$$

则称 $\boldsymbol{\hat{\theta}}$**为**$\boldsymbol{\theta}$**的最大似然估计量**

最大似然估计量$\boldsymbol{\hat{\theta}}$**是使观测值**$\boldsymbol{X_1 = x_1, \cdots, X_n = x_n}$**出现概率最大**

定理：设$\mu(\theta)$为θ的函数且存在反函数$\theta = \theta(\mu)$. 若$\hat{\theta}$是θ的最大似然估计，则$\hat{\mu} = \mu(\hat{\theta})$是$\mu$的最大似然估计，

称为**最大似然估计的不变性**

求解与例题

求解步骤

- 计算对数似然函数$\ln\left(L(x_1, x_2, \cdots, x_m; \theta)\right)$
- 求对数似然函数中参数θ的一阶偏导, 令其等于零
- 求解方程组得到最大似然估计量$\hat{\theta}$

设总体X的概率密度函数

$$f(x) = \begin{cases} (\alpha + 1)x^{\alpha} & x \in (0,1) \\ 0 & \text{其它} \end{cases}$$

设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 求α的矩估计和极大似然估计

例题

设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 以及总体X的密度函数为

$$f(x) = \begin{cases} \frac{1}{\theta}\, e^{-(x-\mu)/\theta} & x \geq \mu \\ 0 & \text{其它} \end{cases}$$

其中$\theta > 0$, 求μ和θ的极大似然估计

设$X_1, \cdots, X_n$是取自总体$X \sim B(1, p)$的样本, 求p的最大似然估计

估计量的评价标准

不同的估计方法可能得到不同的估计值，哪一种估计量更好，或更好的标准是什么呢?

无偏性、有效性、一致性

设$X_1, X_2, \cdots, X_n$来自总体X的样本，令$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量, 若

$$E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}\right] = E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}(X_1, X_2, \cdots, X_n)\right] = \theta$$

则称**$\hat{\theta}$为θ的无偏估计**

无偏估计要求在期望的情形下有$E(\hat{\theta}) = \theta$成立,无系统性偏差

无偏估计不具有函数不变性: $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的无偏估计, 但并不一定有$g(\hat{\theta})$是$g(\theta)$的无偏估计

有效性

设$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$是$\theta$的两个无偏估计, 若

$$\mathrm{Var}\left(\hat{\theta}_1\right) \leq \mathrm{Var}\left(\hat{\theta}_2\right)$$

则称$\boldsymbol{\theta_1}$**比**$\boldsymbol{\theta_2}$**有效**

有效统计量(了解)

随机变量X的概率密度为$f(x;\theta)$或分布函数为$F(x;\theta)$, 令

$$\mathrm{Var}_0(\theta)=\left[nE\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2\right]^{-1}$$

$$\mathrm{Var}_0(\theta)=\left[nE\left(\frac{\partial\ln F(X;\theta)}{\partial\theta}\right)^2\right]^{-1}$$

对任意的无偏估计量$\hat{\theta}$有$\mathrm{Var}\left(\hat{\theta}\right)\geq\mathrm{Var}_0(\theta)$,称$Var_0(\theta)$为估计量$\hat{\theta}$方差的下界.

当$Var\left(\hat{\theta}\right)=Var_0(\theta)$时称$\hat{\theta}$为达到方差下界的无偏估计量，此时$\hat{\theta}$为最有效估计量, 简称**有效估计量**

一致性

设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量，若当$n \to \infty$时有$\hat{\theta}_n \xrightarrow{P} \theta$成立，即对任意$\epsilon > 0$有$\lim\limits_{n\to\infty} P[|\hat{\theta}_n - \theta| > \epsilon] = 0$，则称$\hat{\theta}_n$为$\theta$的一致估计量

定理：设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量, 若满足:

$$\lim_{n\to\infty} E\left[\hat{\theta}_n\right] = \theta \qquad \lim_{n\to\infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0$$

则$\hat{\theta}_n$为θ的一致估计量.

综合例题

设$X_1,\cdots,X_n$是取自总体X的一个样本，且总体X的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x\geq 0\\ 0, & x<0\end{cases}$$

1) 求θ的极大似然函数估计$\hat{\theta}$

2) 设$Z=n\min\{X_1,X_2,\ldots,X_n\}$，求证$Z$和$\hat{\theta}$均为$\theta$的无偏估计量

3) 请问$\hat{\theta}$和Z哪个估计量更有效

4) $\hat{\theta}$是否为一致统计量

第九章 假设检验

知识结构（概念+原理）

假设检验基本思想: 小概率原理(小概率事件在一次实验中不应该发生)

假设检验概念

两类错误

显著性检验

显著性检验基本步骤

假设检验

根据样本信息来检验关于总体的某个假设是否正确，此类问题称为 **假设检验问题**, 可分为参数检验问题和非参数检验问题

假设检验方法(反证)： 先假设所做的假设H_0成立，然后从总体中取样, 根据样本来判断是否有 **不合理** 的现象出现, 最后做出接受或者拒绝所做假设的决定.

不合理的现象: 小概率事件在一次事件中几乎不会发生

假设检验的两类错误

- 第一类错误：拒绝实际真的假设H_0（弃真）
- 第二类错误：接收实际不真的假设H_0（存伪）

小事件

假设检验中需要对 **不合理** 的小事件给出一个定性描述，通常给出一上界α, 当一事件发生的概率小于α时则成为小概率事件

通常取α =0.05,0.1,0.01，其具体取值根据实际问题而定．在假定H_0成立下，根据样本提供的信息判断出不合理的现象 (概率小于α的事件发生), 则认为假设H_0不显著, α被称为显著水平.

注意：不否定假设H_0并不是肯定假设H_0一定成立，只能说差异不够显著, 没达到否定的程度, 所以假设检验被称为``**显著性检验**“

假设检验的一般步骤

- 根据实际问题提出原假设H_0和备择假设H_0
- 确定检验统计量 (分布已知)
- 确定显著性水平α, 并给出拒绝域
- 由样本计算统计量的实测值, 判断是否接受原假设H_0

问题：正态总体情形下的期望与方差检验 + 五大抽样定理