概率论与数理统计考前直播

注意事项　复习指导　章节概要　你问我答　转场答疑

7月9日课程期末考试注意事项

- 迟到15分钟不能"入场",到结束的半个小时以前不能提交;
- 题型只有单选题和填空题(全是客观题,不需要拍照上传);
- 填空题只能填数字,按惯例四舍五入保留到小数点后第四位;
- 分位数按幕课及习题册的上分位数,不需要查表(题中会给出);
- 允许"携带"自己整理的一张公式纸(惯例是半张A4纸);
- 允许使用计算器; 考试过程中遇到问题及时QQ联系任课教师

课程期末考试范围：第1至8章

期末考试的重要知识点:自己总结

模拟测试(样卷).pdf	朱坤平	316KB	2020-05-26	145

模拟测试的样卷有三套往年的笔试真题,考核的重点自己“悟”(这也是统计推断)

模拟测试的答案7月4号会发布在幕课平台,大家要自己亲手做,不要直接看答案

有同学在传一个非法公众号下载的历年考题, 需注意其中的分位数与幕课不同

期末总评中各项成绩所占比重

- 原则:总体上有利于学生(提高及格率),个体上尽可能公正合理(用数据说话)

  已讨论多种算法,如:按各项重要程度排序,根据及格率最大求权重,需利用最终数据再求

- 平台的学习数据有较大变化,视频滞后有改善,目前仍有近15%未完成;
- 问卷结果发生较大变化,平时成绩占比30%(开学初)→70%(学期末);
- 权重最大的肯定是卷面分,期末成绩最最重要! 大家一定要抓住这个关键;
- 卷面及格但欠缺平时数据可否及格? 理解同情,但目前看来本学期无法实现

  平台上提交成绩还只有按系数核算,暂时还无法实现提交max(卷面,总评)的功能

总之,权重的问题请大家放心,我们会尽可能给出一个专业的合理的解决方案

如何复习(仅供参考)

- 已完成视频学习的按“幕课版配套课件”复习(看视频太慢), 并订正各章单元测试;
- 做第一套样卷,对做错的相关知识点的题,再看对应课件,选做习题册部分相关习题;
  (目的找出薄弱环节. 考试不会涉及复杂的积分, 难度不超过习题册中的习题)
- 做第二套样卷,重点钻研做错的题目,总结一下重要知识点,整理公式纸;
- 做第三套样卷,订正错题(需再看课件及做习题册相关习题),就心中有数信心满满了

章节概要

知识点内容不懂或题目不会的,直播结束后转场平台个别答疑

如下是章节内容概要,帮大家理一理知识点的线索和导出关系

第一章核心内容：求随机事件的概率

随机试验

事件

概率

古典概型　几何概型　条件概率　公式

加法公式　乘法公式　全概率公式　贝叶斯公式

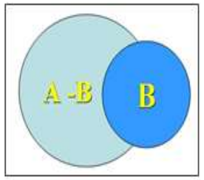
事件的关系、运算、性质

◇并运算　$A\bigcup B$ 表示A发生**或**B发生(即至少有一个发生)

◇交运算　$A\bigcap B$ 表示A发生**且**B发生(也称事件的积,简记为AB)

◇逆运算　$\bar{A}$ 表示**非A**即A不发生(一元运算)

◇差运算　$A-B$ 表示A发生而B不发生($A\bigcap\bar{B}$)



◇包含关系 $A\subseteq B$: 若A发生,则必然导致B发生

◇对立关系 事件A与B有且仅有一个发生(记为$B=\bar{A}$)

◇互斥关系 事件A与B不可能同时发生($AB=\varnothing$)

◇独立关系 $P(AB)=P(A)P(B)$ (事件的发生互不影响)

性质：1)交换率；2) 结合率；　3) 分配率；　4) **对偶率(德摩根率)**

$$\overline{A\bigcup B}=\bar{A}\bigcap\bar{B}\;;\qquad \overline{A\bigcap B}=\bar{A}\bigcup\bar{B}$$

$$\overline{\bigcup_{i=1}^{n} A_i}=\bigcap_{i=1}^{n}\overline{A_i},\qquad \overline{\bigcap_{i=1}^{n} A_i}=\bigcup_{i=1}^{n}\overline{A_i}$$

古典概型　　与　　几何概型

$$P(A)=\frac{|A|}{|\Omega|}$$

样本空间的选取不是唯一的

关键在于事件A的样本点的计数

计数方法很多(只需掌握一般方法）

$$P(A)=\frac{S_A}{S_\Omega}$$

几何尺度的计算可能涉及到积分

例: 袋中有a只黑球，b只白球，现把球一只一只摸出，求第k次摸出黑球的概率 （1≤k≤a+b）

解法1：考虑把球全部取出 $$p=\frac{a(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$$

解法2：只考虑前k次的取球 $$p=\frac{P_a^1 P_{a+b-1}^{k-1}}{P_{a+b}^{k}}=\frac{a}{a+b}$$

解法3：只考虑第k次的取球 $$p=\frac{a}{a+b}$$

注：三种解法样本空间是不同的. 抽签原理，结论需记住！！

概率的公理化定义

加法公式:

减法公式: $P(A-B)=P(A)-P(AB)$

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

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_iA_j)+\cdots+(-1)^{n-1}P\left(\prod_{i=1}^{n}A_i\right)$$

乘法公式 ← 条件概率 → 独立性

$$P(AB)=P(A\mid B)\,P(B)=P(B\mid A)\,P(A)$$

$$P(ABC)=P(A\mid BC)\,P(BC)=P(A\mid BC)\,P(B\mid C)\,P(C)$$

全概率公式

$P(A)=\sum_{i=1}^{n}P(A\mid B_i)\,P(B_i)$， $\Omega=\sum_{i=1}^{n}B_i$ 为 Ω 的一个分割

贝叶斯公式

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

例 盒中装有均匀的 m 只正品和 n 只次品硬币，正品硬币一面是国徽一面是麦穗，次品硬币两面均印有国徽，从盒中任取一只硬币并将它连抛 r 次，结果每次向上一面均为国徽，试利用贝叶斯公式求取出的这个硬币是正品的概率？

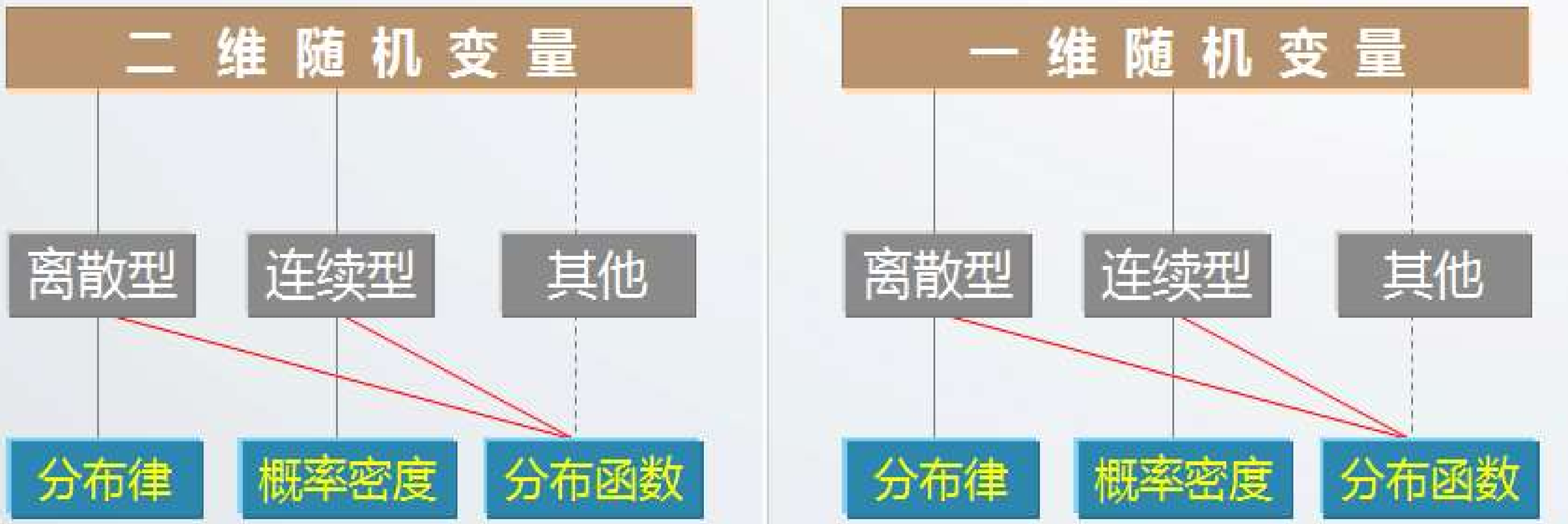
解：A 表示事件连抛 r 次每次向上一面均为国徽

B 表示事件取出的这个硬币是正品

则所求概率为

$$P(B\mid A)=\frac{P(B)P(A\mid B)}{P(A)}=\frac{\frac{m}{m+n}\left(\frac{1}{2}\right)^{r}}{\frac{m}{m+n}\left(\frac{1}{2}\right)^{r}+\frac{n}{m+n}\cdot 1}=\frac{m}{m+n\cdot 2^{r}}$$

对照　一维与二维随机变量的分布



随机变量的分布函数

一维随机变量的分布函数: $F(x)=P\{X\le x\}$

- 非降性：若$x_1<x_2$,则$F(x_1)\le F(x_2)$
- 0-1性：$\begin{cases} F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0 \\ F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1 \end{cases}$
- 右连续：$F(x+0)=\lim\limits_{t\to x^+}F(t)=F(x)$

$$\begin{cases} F(-\infty,y)=0 \\ F(x,-\infty)=0 \\ F(+\infty,+\infty)=1 \end{cases}$$

二维随机变量分布函数的定义:

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

对照 **一维及二维随机变量的概率密度**

二维连续型随机变量：

定义:设二维随机变量 (X,Y) 的分布函数为 $F(x,y)$,如果存在一个非负可积函数 $p(x,y)$,使得对任意 $(x,y)\in R^2$,都有:

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\,du\,dv$$

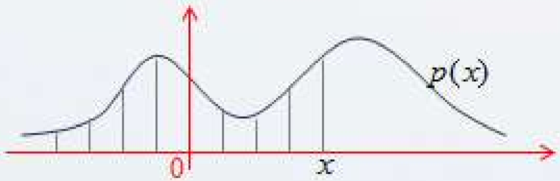
则称 (X,Y) 为二维连续型随机变量,此时称 $p(x,y)$ 为 (X,Y) 的概率密度(联合概率密度)

$F(x,y)=P\{X\le x,Y\le y\}=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\,du\,dv$ (体积)

性质：非负性；规范性

一维连续型随机变量：

定义：如果对于随机变量 X 的分布函数 $F(x)$，存在非负可积函数 $p(x)$,使得对任意 $x\in R$,有 $F(x)=\int_{-\infty}^{x}p(t)dt$，则称 X 为连续型随机变量，并称 $p(x)$ 为 X 的概率密度(或 X 的密度函数).



$F(x)=\int_{-\infty}^{x}p(t)dt=$ 阴影面积 $=P\{X\le x\}$

性质：非负性；规范性

定义: 相对二维随机变量的联合分布而言,其每个分量的分布都称为边缘分布.

$$F_X(x)=P\{X\le x\}=P\{(X\le x)\cap\Omega\}=P\{X\le x,Y\le +\infty\}=F(x,+\infty)$$

$$F_Y(y)=P\{Y\le y\}=P\{\Omega\cap(Y\le y)\}=P\{X\le +\infty,Y\le y\}=F(+\infty,y)$$

设二维连续型随机变量(X,Y)的概率密度为$p(x,y)$,则:

1) 随机变量X的边缘分布函数为: $F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}p(u,v)dvdu$

2) 随机变量X的边缘概率密度为:

$$\int_{-\infty}^{+\infty}p(u,v)dv=:f(u),\quad \left(\int_{-\infty}^{x}f(u)du\right)'=f(x)$$

$$p_X(x)=F_X{}'(x)=\left(\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}p(u,v)dv\right]du\right)'=\int_{-\infty}^{+\infty}p(x,v)dv=\int_{-\infty}^{+\infty}p(x,y)dy$$

3) 同理可求,随机变量Y的边缘概率密度为: $p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$

第三章: 二维随机变量及其边缘分布重要公式的导出关系

$$P\{a_1 < X \le a_2, b_1 < Y \le b_2\} = F(a_2,b_2) - F(a_1,b_2) - F(a_2,b_1) + F(a_1,b_1)$$

$$F(-\infty, y) = 0;\quad F(x,-\infty) = 0;\quad F(+\infty,+\infty) = 1$$?

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

离散

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

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

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

连续

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)\,du\,dv$$

$$p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

$$P\{(X,Y) \in G\} = \iint_G p(x,y)\,dx\,dy$$

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

$$F_X(x) = P\{X \le x\} = P\{X \le x, Y \le +\infty\} = F(x,+\infty)$$

独立: $F(x,y) = P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\} = F_X(x)F_Y(y) \Rightarrow p(x,y) = p_X(x)p_Y(y)$ -连续型

一维及多维随机变量函数的分布

离散型简单：考虑可能的取值及对应的概率,列表即可

连续型基本方法：无论一维还是多维, 都是从分布函数入手

- 一维严格单调可微型,直接代公式写出概率密度
- 极大极小值分布(写出公式?)
- 和的分布(写出卷积公式?)
- 一般情形(通用方法,多维类似):

把 $\eta = f(\xi)$ 的分布函数

$$F_\eta(y) = P\{\eta \le y\} = P\{f(\xi) \le y\}$$

表示为 ξ 的密度函数 $p_\xi(x)$

在对应区域内的积分

随机变量的数字特征(表示随机变量都某个特征的数,是数字)

表示随机变量取值“中心”的数学期望：

$$E\xi := \sum_{i=1}^{\infty} x_i p_i \qquad E\xi := \int_{-\infty}^{+\infty} xp(x)dx$$

表示随机变量取值分散程度的方差：

$$D\xi = E(\xi\text{-}E\xi)^2 = E(\xi^2) - (E\xi)^2$$

各种数字特征的计算都可归结为利用随机变量函数的期望公式

期望，方差，矩，协方差，相关系数等随机变量的数字特征,也统统都属于随机变量函数的期望

离散型：$Ef(\xi)=\sum_{i} f(x_i)p_i$

离散型：$Ef(\xi,\eta)=\sum_{i}\sum_{j} f(x_i,y_j)p_{ij}$

连续型：$Ef(\xi)=\int_{-\infty}^{+\infty} f(x)p(x)dx$

连续型：$Ef(\xi,\eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)p(x,y)dxdy$

令 $f(\xi)=\xi$ 得 $E\xi$

令 $f(\xi)=(\xi-E\xi)^2$ 得 $D\xi$

令 $f(\xi)=\xi^k$ 得 $E\xi^k$

令 $f(\xi,\eta)=(\xi-E\xi)(\eta-E\eta)$ 得 $Cov(\xi,\eta)$

令 $f(\xi,\eta)=\dfrac{(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$ 得 $\rho_{\xi,\eta}$

数字特征(期望,方差等)的性质：

数学期望的性质：$E(aX+bY+c)=aE(X)+bE(Y)+c;\qquad E(XY)\underset{\equiv}{\text{独立}}E(X)E(Y)$

方差的性质：$D(aX+bY+c)=a^2D(X)+b^2D(Y)+2ab\operatorname{cov}(X,Y)$

$$D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)\underset{\equiv}{\text{独立}}D(X)+D(Y)$$

切比雪夫不等式：$P\{|\xi-E\xi|\ge\varepsilon\}\le\frac{D\xi}{\varepsilon^2}$

相关系数的含义和公式

随机变量的标准化 $\xi^*=\dfrac{\xi-E\xi}{\sqrt{D\xi}}$,则 $E\xi^*=0,\quad D\xi^*=1$

常见分布(一维6个的分布律/概率密度;期望;方差;特殊性质. 二维2个的性质)

两点分布：______;________;________

二项分布：______;________;________

泊松分布：______;________;________； 可加性.

均匀分布：______;________;________

指数分布：______;________;________； 无记忆性.

正态分布：______;________;________

二维均匀分布

二维正态分布

1) 相互独立的服从正态分布的随机变量，其线性组合服从正态分布

2) 服从二维正态分布的随机变量，其线性组合服从正态分布

3) 二维正态分布的每个分量服从一维正态分布

4) 二维正态分布的两个分量,独立与不相关等价

第五章：大数定律与中心极限定理

- 大数定律:一定条件下随机变量的平均结果具有稳定性的一系列定理

$$\lim_{n\to\infty} P\{|\frac{\sum_{i=1}^{n}\xi_i}{n} - \frac{\sum_{i=1}^{n}E\xi_i}{n}| < \varepsilon\} = 1 \qquad (\forall\varepsilon > 0)$$

- 中心极限定理:一定条件下随机变量累加和的分布,近似正态分布的一系列定理(本课程涉及的条件:独立同分布)

$\sum_{i=1}^{n}\xi_i$ 近似服从正态分布 N(期望, 方差)

$$\frac{\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}E\xi_i}{\sqrt{D(\sum_{i=1}^{n}\xi_i)}} \sim N(0,1)$$

$$\sum_{i=1}^{n}\xi_i \sim N(n\mu, n\sigma^2)$$

$$\lim_{n\to\infty} P(\frac{\sum_{i=1}^{n}\xi_i - n\mu}{\sqrt{n}\sigma} \le x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

收敛与以概率收敛

$X_n \xrightarrow{P} c$

以概率收敛

$P\{|X_n - c| < \varepsilon\} \to 1, \qquad \forall \varepsilon > 0$

收敛

$\frac{1}{n}\sum \xi_i \xrightarrow{P} \frac{1}{n}\sum E(\xi_i)$

大数定律

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum \xi_i - \frac{1}{n}\sum E(\xi_i)\right| < \varepsilon\right\} = 1$$

大数定律

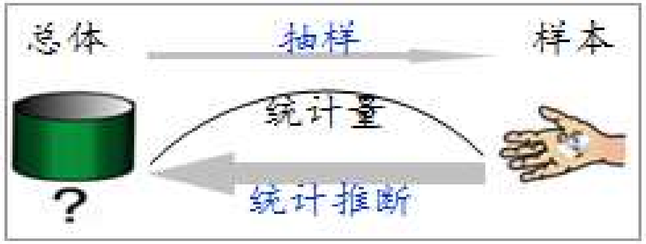
马尔可夫大数定律
切比雪夫大数定律
泊松大数定律

$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}, \qquad \forall \varepsilon > 0$

切比雪夫不等式

大数定律表示一定条件下，随机变量的平均结果以概率收敛于它的数学期望．直观含义是随机变量平均结果的取值,在其中心(数学期望)附近小幅波动．不同的条件就是不同的大数定律．除辛钦大数定律的证明需用到特征函数外，本课程其他的大数定律均可用切比雪夫不等式来证明

第六章：抽样分布



样本$(X_1, X_2, ..., X_n)$是n维的随机变量;具有独立性和同分布性; 样本观测值$x_1, x_2, ..., x_n$是数字

设总体ξ的数学期望$E\xi = \mu$和方差$D\xi = \sigma^2$都存在

$(X_1, X_2, \cdots, X_n)$是ξ的样本, $\overline{X}$是样本均值,

$S^2{}_{n-1}$是样本方差，则有

(1) $E\overline{X} = \mu$;　　(2) $D\overline{X} = \dfrac{\sigma^2}{n}$;

(3) $E(S_n{}^2) = \dfrac{n-1}{n}\sigma^2$　　(4) $E(S^2{}_{n-1}) = \sigma^2$

正态总体的若干抽样分布定理

三大抽样分布构造性定义

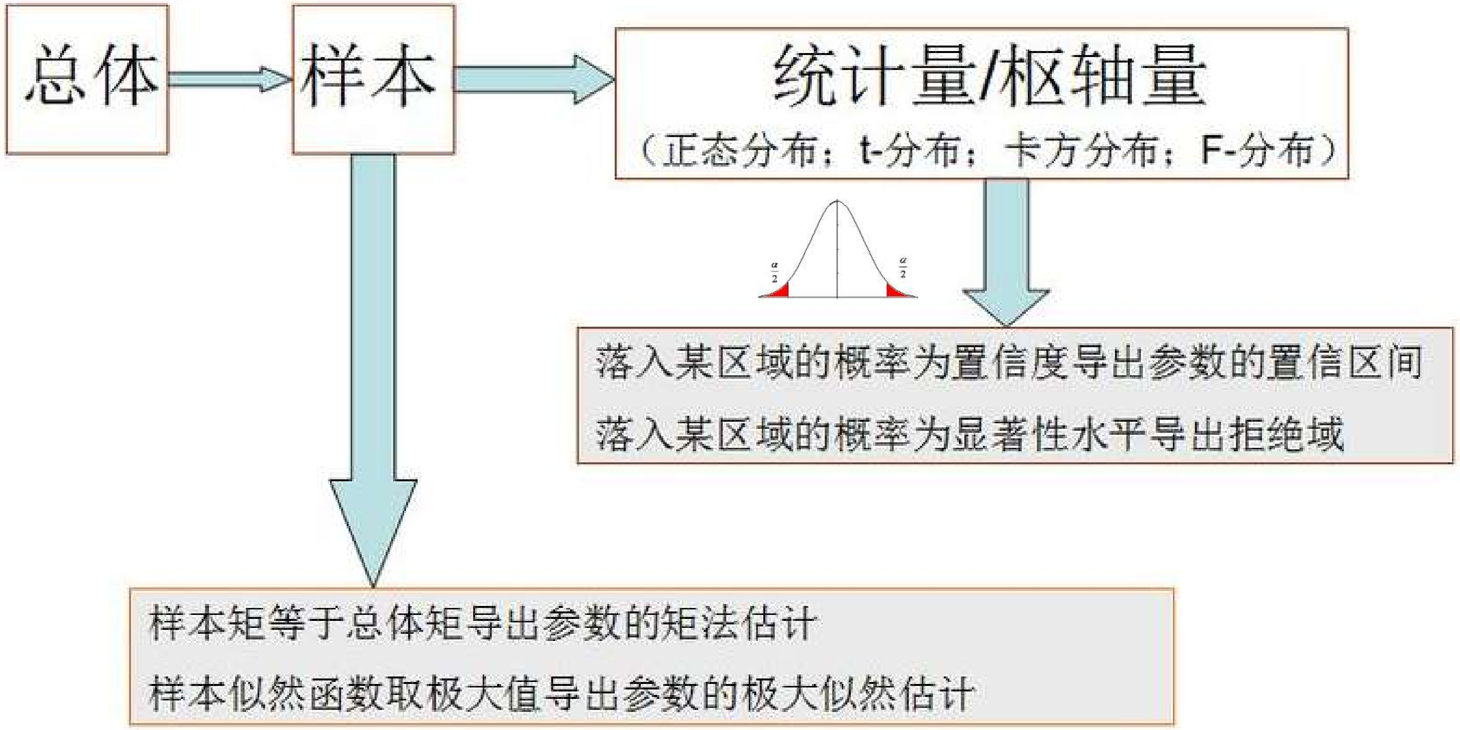
$$\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1) \qquad \frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1) \rightarrow \frac{\overline{X}-\mu}{S_{n-1}}\sqrt{n} \sim t(n-1)$$

$$\Downarrow$$

其他抽样分布定理

重要,很重要!

第七第八章：参数估计与假设检验



矩法估计：根据大数定律, $\overline{X^k} \xrightarrow{P} E(\xi^k)$, 总体矩含有未知参数, 而样本矩相当于已知

$$\begin{cases} E(\xi)=\overline{X} \\ E(\xi^2)=\overline{X^2} \\ \vdots \\ E(\xi^m)=\overline{X^m} \end{cases} \Rightarrow \begin{cases} \widehat{\theta_1}=? \\ \widehat{\theta_2}=? \\ \vdots \\ \widehat{\theta_m}=? \end{cases}$$

极大似然估计：根据是极大似然原理，即当未知参数取何值时能够使观测到的样本值出现的可能性达到最大：$maxL(\theta) \Rightarrow \widetilde{\theta_1}, \widetilde{\theta_2}, \dots, \widetilde{\theta_m}$

点估计的评价

无偏性

有效性

相合性

假设检验的两类错误

单侧检验与双侧检验的区别

以下进入现场问答环节，你问我答

直播到此结束，若大家还有问题，我们转平台讨论区继续答疑

最后，预祝大家考出好成绩！