概率论与数理统计复习纲要

> ---NonoTion

第一章 随机事件及其概率

一. 随机事件及其运算

事件间的关系和运算：

- 包含 $A \in B$
- 并（和）
- 交
- 差 $A - B = A - AB$
- 互不相容 $(P(AB) = 0)$
- 对立 $(P(AB) = 0), A + B = \Omega$

运算律 P5

- 交换律
- 结合律
- 分配律
- 德摩根律

二.古典概型和几何概型

古典概型：

$$P(A) = \frac{\text{事件}A\text{包含的事件数}}{\text{事件的总数}}$$

集合概型：

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

测度：面积，体积等

Ω：全集

三.概率的公理化定义及其性质

三条公理:

- 非负性
- 规范性

- 可列可加性

概率的性质

1. $P(\text{空集}) = 0$
2. 有限可加性

若事件$A_1...A_n$两两互不相容，则

P(A_1+...+A_n)=$\sum_{i=1}^{n} P(A_i)$

3. 可减性

如果$A \in B$则$P(B-A) = P(B) - P(AB)$

4. 单调性

如果$A \in B$则$P(A) \leq P(B)$

5. P(A的补集)=1-P(A)
6. 加法公式

P(A+B)=P(A)+P(B)+P(AB)

还有上连续性和下连续性，感觉没啥用

四. 条件概率

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

性质：

- 非负性
- 规范性
- 可列可加性

乘法公式

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

事件的独立性

如果A,B独立

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

反之不一定

五.全概率公式和贝叶斯公式

全概率公式

设$A_1,...,A_n$是样本空间的一个分割，又叫做完备事件组，则B发生的概率

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

贝叶斯公式

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

六.伯努利概型

在n重伯努利试验中，事件A可能发生1,2...n次，每次实验中发生概率为p，求事件A恰好发生k次的概率

$$P_n(k)=C_n^k(1-p)^{n-k}p^k$$

第二章 随机变量及其分布

一. 随机变量与分布函数

分布函数:

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

性质:

- 单调不减
- F(+∞)=1,F(-∞)=0
- 右连续

二.离散型随机变量及其分布

分布律:

$$P(X=x_k)=p_k(k=1,2,3,...,)$$

常见的离散型随机变量:

- 0-1分布
- 二项分布
- 几何分布
- 超几何分布
- 泊松分布（n>10,p<0.1）近似计算二项分布的概率

泊松定理:

$$lim_{n->+\infty}np=\lambda$$

三.连续型随机变量

概率密度函数:

$$F(x)=\int_{-\infty}^{x}f(x)dx$$

常见分布:

- 均匀分布
- 指数分布
- 正态分布

四.随机变量函数的分布

离散型:

列表求解

连续型:

$$F_Y(y)=P(Y\leq y)=P(g(X)\leq y)$$

对上式求导就可以得到 $f_Y(y)$

公式法:

1. y=g(x)在区间(a.b)上严格单调
2. x=$g^{-1}(y)$具有连续导函数

则

$$f_Y(y)=f_X(h(y))|h^`(y)|,\alpha\leq y\leq\beta$$

其他时候为 0

$$\alpha=min(g(a),g(b))$$

$$\beta=max(g(a),g(b))$$

第三章 多维随机变量及其分布

一.二维随机变量及其分布

二维随机变量的分布函数

$$F(X,Y)=P(X\leq x,Y\leq y)$$

性质:

- 单调非减
- F(+∞，+∞)=1，F(+∞，-∞)=0，F(-∞，+∞)=0，F(-∞，-∞)=0
- 关于x,y都是右连续的

二维离散型随机变量

(X,Y)的联合概率分布 简称概率分布或分布律

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

二维连续性随机变量

(X,Y)的联合概率密度函数

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

几种特殊分布

- 均匀分布
- 二维正态分布

二.边缘分布

离散型:

关于X,Y的边缘概率函数

$$F_X(x)=\sum_{x_i\le x}p(X=x_i)$$

$$F_Y(y)=\sum_{y_i\le y}p(Y=y_i)$$

关于X,Y的边缘概率分布

$$P(X=x_i)=\sum_j p_{ij}=p_{i\cdot}$$

$$P(X=x_j)=\sum_i p_{ij}=p_{\cdot j}$$

连续型:

边缘密度函数

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

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

三.条件分布

条件分布=$\frac{\text{联合分布}}{\text{边缘分布}}$

掌握离散型即可，连续性不在期末考纲中

四.随机变量的独立性

二维随机变量(X,Y), 如果有

$$F(X,Y)=F_X(x)F_Y(y)$$

则称X,Y相互独立

- 离散型

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

- 连续性

$$f(x,y)=f_X(x)f_Y(y)$$

五.二维随机变量函数的分布

离散型

列表求解即可

连续型:

总体思想感觉类似高中线性规划的思想，涉及到二重积分，具体分下面几种情况

- 和分布(Z=X+Y)

卷积公式求解，建议看课本自己推导一遍，或许一通百通，剩下的几个也不用背了

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

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

- 商分布(Z=$\frac{X}{Y}$)

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

- 最大最小分布 比较简单 看课本即可

第四章 随机变量的数字特征与极限定理

一.数学期望

随机变量的数学期望

- 离散型:

$$E(X)=\sum_{k=1}^{+\infty}x_kp_k$$

- 连续性:

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$$

随机变量函数的数学期望

设y=g(x)是连续函数，Y是随机变量X的函数Y=g(X)

那么

- 离散型（级数收敛）：

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

- 连续型:（积分收敛）

$$E(Y)=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

数学期望的性质

1. C为常量，E(C)=C
2. E(CX)=CE(X)
3. E(X+Y)=E(X)+E(Y)
4. 若随机变量X,Y相互独立，E(XY)=E(X)E(Y)
5. $[E(XY)]^2\leq E(X^2)E(Y^2)$

二. 方差

方差的定义:

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

可利用 一 中的随机变量函数的期望计算方法计算

也可用简化公式计算

$$D(X)=E(X^2)-E(X)^2$$

$\sqrt{D(X)}$叫做标准差

方差的性质:

1. D(C)=0
2. D(CX)=$C^2D(X)$
3. X，Y为相互独立的随机变量，则

$$D(x+-y)=D(x)+D(Y)$$

三.协方差和相关系数

设二位变量(X,Y),它们的协方差为

$$Cov(X,Y)=E([X-EX][Y-EY])$$

相关系数为

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

含常用公式

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

计算Cov(X,Y)

协方差与相关系数的性质

1. $Cov(X,Y)=Cov(Y,X)$
2. $Cov(a_1X+b_1,a_2X+b_2)=a_1a_2Cov(X,Y)$
3. $Cov(X_1+X_2,Y)=Cov(X_1,Y)Cov(X_2,Y)$
4. $D(X+-Y)=D(X)+D(Y)+-2Cov(X,Y)$a_2
5. $|\rho_{XY}|\leq 1$
6. $|\rho_{XY}|=1$的充要条件是$P(Y=aX+b)=1(a,b$为常数，$a\neq 0$

四.大数定律

1.切比雪夫不等式

设随机变量的数学期望和方差都存在则对于任意的$\epsilon>0$,有

$$P(|X-EX|\geq\epsilon)\leq\frac{DX}{\epsilon^2}$$

2.大数定律

- 切比雪夫大数定律 设随机变量$X_1...X_n$,相互独立，且具有相同的期望和方差，则对于任意的$\epsilon > 0$都有

$$P(|\frac{\sum_{i=1}^{n} Xi}{n} - \mu| < \epsilon) = 1$$

- 辛钦大数定律 设随机变量$X_1...X_n$独立同分布，具有数学期望,则对于任意的$\epsilon > 0$都有

$$P(|\frac{\sum_{i=1}^{n} Xi}{n} - \mu| < \epsilon) = 1$$

- 伯努利大数定律，设k_n是n此独立重复实验中事件A发生的次数，而p是每次实验中A发生的概率，则对于任意的$\epsilon > 0$都有

$$P(|\frac{k_n}{n} - p| < \epsilon) = 1$$

五.中心极限定理

- 列维-林德伯格中心极限定理 （独立同分布的中心极限定理）

  设$X_1...X_n$是具有相同分布且相互独立的一系列随机变量，且EX=μ,DX=σ^2，则对任意x有

$$lim_{n->+\infty} P(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma/\sqrt{n}} < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\varPi}} e^{-\frac{t^2}{2}} dt = \varPhi(x)$$

- 棣莫弗-拉普拉斯中心极限定理 (二项分布中心极限定理)

设$Y_n(n = 1, 2, 3...)$服从二项分布，则对于任意x有

$$lim_{n->+\infty} P(\frac{\sum_{k=1}^{n} Y_k - np}{\sqrt{np(1-p)}} < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\varPi}} e^{-\frac{t^2}{2}} dt = \varPhi(x)$$

第五章 数理统计基本知识

一.总体与样本

简单随机样本：具有独立性和代表性

了解样本和总体的关系即可

二.统计量和三大分布

统计量：关于总体样本不含未知参数的实值函数

常用统计量:

- 样本均值
- 样本方差

- 样本标准差
- 样本k阶原点矩
- 样本k阶中心距

三大分布

- 卡方分布
- t分布
- F分布

了解图像和性质即可

上分位数（点）：设随机变量X的分布函数为F(x)，满足等式

$$P(X > x_\alpha) = 1 - F(x_\alpha) = \alpha$$

的实数α称为X的上α分位点（数）

三.正态总体的抽样分布

主要掌握单正态总体即可

那几条性质有点难打，在课本的P129

四.数据的整理

了解概念即可

第六章 参数估计和假设检验

一.参数的点估计

- 矩估计

用样本的k阶原点(中心)矩作为总体的k阶原点（中心）矩的估计

- 最大似然估计

总体X有独立同分布的样本$X_1...X_N$,求θ的值，使得事件{$X_1=x_1,...,X_n=x_n$}发生的概率最大，并以其作为参数θ的最大似然估计量

解决问题的步骤

1. 求最大似然函数L(θ)的表达式

离散型: L(θ)=P($X_1=x_1...X_n=x_n$)=$\prod_{i=1}^n P(X_i=x_i)$

连续性:L(θ)=$\prod_{i=1}^n f(x_i)$

2. 求最大似然函数的极值解

- 估计量优劣的评价标准

1. 无偏估计 估计量的期望=参数
2. 有效性 方差较小的估计量有效性更好
3. 相合性（一致性） 期末不考

二.参数的区间估计

概念：

- 置信区间
- 置信度

求置信区间

1. 找到一个与要估计参数θ有关的统计量T，一般是参数的良好估计
2. 设法找出T和θ之间的某一函数H，H叫做**枢轴变量**
3. 寻找适当的常数c,d 使得

$$P(c \le H(T,\theta)) \le d) = 1-\alpha$$

4. 将 3 中的不等式化为关于θ的不等式，不等式的两端便是参数的置信上线和置信下限

附：单个正态总体的区间估计表

它置信区间问题，也可讨论两个正态总体均值差、方差比的置信区间问题. 它们用到的枢轴变量 $H(T,\theta)$ 分别见定理 5.3.1 和定理 5.3.2，列表如表 6.2.1、表 6.2.2.

表 6.2.1　单个正态总体的参数区间估计表(置信水平为 1-α)

待估参数	条件	枢轴变量及其分布	置信区间
均值 μ	方差 σ^2 已知	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left[\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
	方差 σ^2 未知	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\left[\bar{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}},\ \bar{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]$
方差 σ^2	均值 μ 已知	$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$	$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}\right]$
	均值 μ 未知	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$

表 6.2.2　两个正态总体均值差、方差比的区间估计表(置信水平为 1-α)

置信区间

三.假设性检验

假设性检验的基本原理和步骤

1. 建立假设，H_0原假设, H_1备择假设

   假设检验的目的就是在二者中选择其一

2. 选取检验统计量

求H_0成立时检验统计量的分布，并与H_1成立时统计量的分布进行比较，从而得到对H_0不利的

3. 选取检验的显著性水平α与临界值，进而求出拒绝域
4. 作判断

原假设与备择假设的选取原则

通常，将研究者要证明的结论作为备择假设H_1

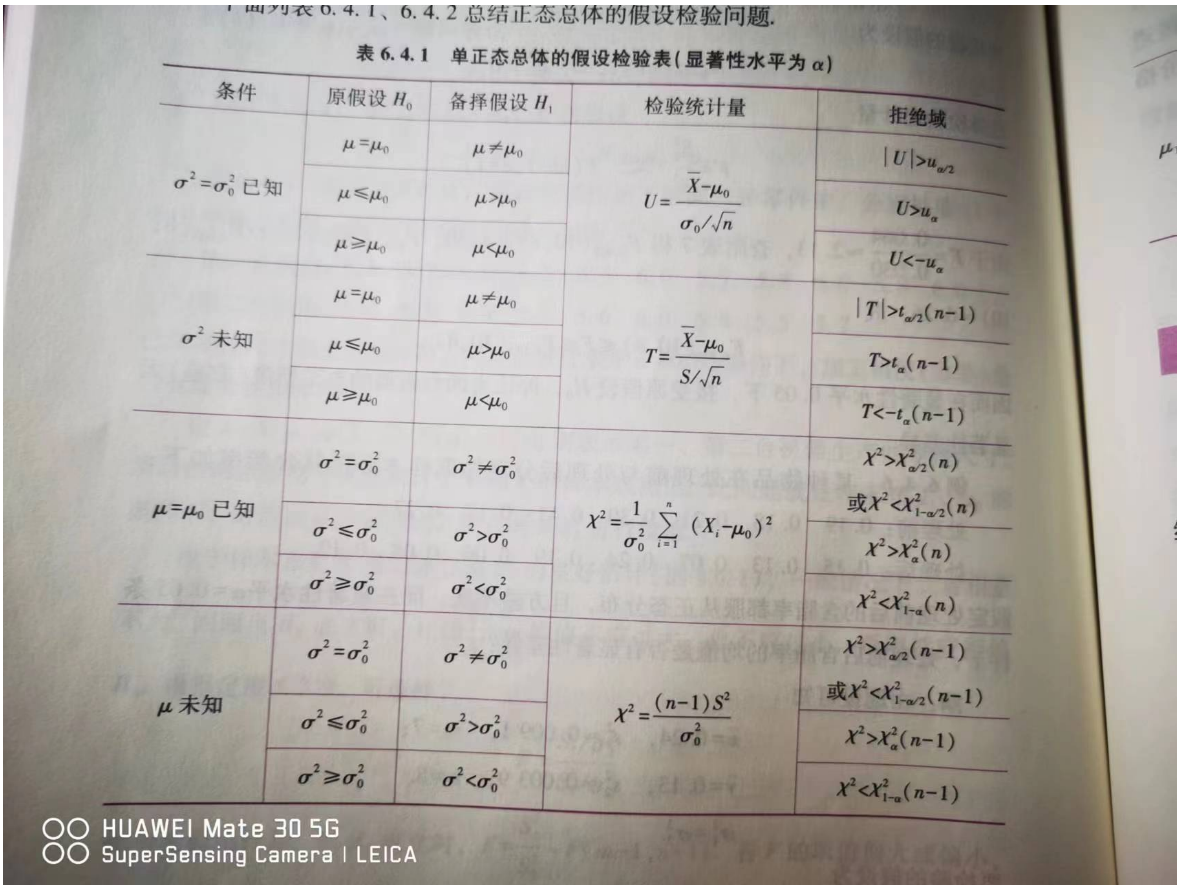
把研究者要反对的假设、现状和不能轻易否定的假设作为原假设H_0

假设性检验的两类错误

1. H_0成立拒绝H_0
2. H_1成立接受H_0

H_0, H_1给定后，α越小，u_α越大，放第二类错误的概率越大

单个正态总体的假设性检验



下面列表 6.4.1、6.4.2 总结正态总体的假设检验问题.

表 6.4.1 单正态总体的假设检验表(显著性水平为 α)

条件	原假设 H_0	备择假设 H_1	检验统计量	拒绝域
$\sigma^2=\sigma_0^2$ 已知	$\mu=\mu_0$	$\mu\neq\mu_0$	$U=\dfrac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}$	$\lvert U\rvert>u_{\alpha/2}$
	$\mu\leq\mu_0$	$\mu>\mu_0$		$U>u_\alpha$
	$\mu\geq\mu_0$	$\mu<\mu_0$		$U<-u_\alpha$
σ^2 未知	$\mu=\mu_0$	$\mu\neq\mu_0$	$T=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\lvert T\rvert>t_{\alpha/2}(n-1)$
	$\mu\leq\mu_0$	$\mu>\mu_0$		$T>t_\alpha(n-1)$
	$\mu\geq\mu_0$	$\mu<\mu_0$		$T<-t_\alpha(n-1)$
$\mu=\mu_0$ 已知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu_0)^2$	$\chi^2>\chi^2_{\alpha/2}(n)$ 或 $\chi^2<\chi^2_{1-\alpha/2}(n)$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2>\chi^2_\alpha(n)$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2<\chi^2_{1-\alpha}(n)$
μ 未知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$	$\chi^2>\chi^2_{\alpha/2}(n-1)$ 或 $\chi^2<\chi^2_{1-\alpha/2}(n-1)$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2>\chi^2_\alpha(n-1)$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2<\chi^2_{1-\alpha}(n-1)$

注:双正态总体的区间估计和假设性检验在期末考中都不做要求