目录

第一章 概率论的基本概念......1
第二章 随机变量及其分布......1
★第三章 多为随机变量及其分布......3
★第四章 随机变量的数字特征......3
第五章 大数定律和中心极限定理......4
第六章 样本及抽样分布......4
★第七章 参数估计......6
第八章 假设检验......7

第一章 概率论的基本概念

1. 互斥(互不相容)：$AB=\emptyset$　　对立(互逆)：$\bar{A}=B$

完备(完全)事件组：$A_1,A_2,\dots,A_n$，满足$A_iA_j=\emptyset, i\neq j$，且$\bigcup_{i=1}^{n}A_i=\Omega$

2. 德摩根律：$\overline{A\cup B}=\bar{A}\cap\bar{B}$　　$\overline{A\cap B}=\bar{A}\cup\bar{B}$　　$\overline{\mathrm{A}-\mathrm{B}}=\overline{\mathrm{A}\bar{\mathrm{B}}}=\bar{\mathrm{A}}\cup\mathrm{B}$

3. $A\subset B, AB=A$

4. $P(A\cup B)=P(A)+P(B)-P(AB)$　　$P(A-B)=P(A)-P(AB)=P(A\bar{B})$

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

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

5. 古典型概率计算公式：$P(A)=\frac{k}{n}=\frac{\text{事件 A 中的基本事件数}}{\text{样本空间}S\text{中的基本事件总数}}$

6. 不放回抽取，连续取 n 次每次取 1 个⇔一次取 n 个

7. 条件概率：$P(B|A)=\frac{P(AB)}{P(A)}$　　乘法定理：$P(AB)=P(B|A)P(A)$

缩减样本空间解法　　$P(ABC)=P(C|AB)P(B|A)P(A)$

8. 全概率公式：离散：$P(A)=\sum_{i=1}^{n}P(AB_i)=\sum_{i=1}^{n}P(A|B_i)P(B_i)$

连续：$p_X(x)=\int_{-\infty}^{+\infty}p(x|y)p_Y(y)\,dy$　　$p_Y(y)$同理

9. 贝叶斯公式：离散：$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{P(A)}=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}$

连续：$p(y|x)=\frac{p(x|y)p_Y(y)}{p_X(x)}=\frac{p(x|y)p_Y(y)}{\int_{-\infty}^{+\infty}p(x|y)p_Y(y)\,dy}$　　$p(x|y)$同理

10. A、B 独立：$P(AB)=P(A)P(B)$

定理：①一列独立事件中任一部分改为对立事件，所得事件列仍相互独立

②事件 A、B、C，任取两个事件都独立，则

两两独立：$P(ABC)\neq P(A)P(B)P(C)$

相互独立：$P(ABC)=P(A)P(B)P(C)$

第二章 随机变量及其分布

1. 概率密度函数**充要条件**①$f(x)>0$；②$\int_{-\infty}^{+\infty}f(x)\,dx=1$

2. 伯努利分布$\xrightarrow{+\text{伯努利分布}}$二项分布$\xrightarrow{+\text{二项分布}}$二项分布$\xrightarrow{\text{n 很大}}$正态分布$\xrightarrow{+\text{正太分布}}$正态分布

3. 不重要结论：

a) 正态分布：$P(\mu-\sigma<X<\mu+\sigma)\approx 68.26\%$,

$P(\mu-3\sigma<X<\mu+3\sigma)\approx 99.72\%$

b) 指数分布：无记忆性：①$P\{X>t\}=\int_{t}^{+\infty}\lambda e^{-\lambda t}\,dt=e^{-\lambda t}, t>0$

②$P\{X>t+s|X>s\}=P\{X>t\}$

离散型		$P\{X=k\}/p(x)$	$E(X)$	$D(X)$	
0-1 分布		$p^k(1-p)^{1-k}$	p	$p(1-p)$	抛硬币，二选一
二项分布	$X\sim B(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	n 重伯努利，出现 k 次“是”
松柏分布	$X\sim P(\lambda)$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ	二项分布的**极限**，$p=\frac{\lambda}{n}$ 结论：$X_1+X_2\sim P(2\lambda)$ 意义：单位时间内随机事件发生的次数；例：汽车站台的候客人数
几何分布	$X\sim Ge(p)$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	n 重伯努利，第 k 次**首次**出现”是” 无记忆性 为负二项分布的特例 r=1
负二项分布	$X\sim Nb(r,p)$	$C_{k-1}^{r-1}p^r(1-p)^{k-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	几何分布的**和** X=第 k 次实验，正好发生 r 次”是”
超几何分布	$X\sim h(n,N,M)$	$\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$	$n\frac{M}{N}$	$n\frac{M}{N}(1-\frac{M}{N})(1-\frac{n-1}{N-1})$	**不放回**抽样的二项分布 若 **N 巨大**，近似为二项分布 意义：N 件物品，有 M 件次品，抽 n 件（不放回）有 k 件次品概率 分子：k 件从 M 中抽取，剩下的在 N 中抽取；分母：从 N 件中随便抽取 n 件

连续型		$P\{X=k\}/p(x)$	$F(x)$	$E(X)$	$D(X)$	
均匀分布	$X\sim U(a,b)$	$\begin{cases}\frac{1}{b-a}, a<x<b\\ 0, 其他\end{cases}$	$\begin{cases}0, x<a\\ \frac{x-a}{b-a}, a\le x\le b\\ 1, x>b\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	古典派的几何概型
正态分布	$X\sim N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\Phi\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$	μ	σ^2	二项分布的另一种极限 a$X+b\sim N(a\mu+b,a^2\sigma^2)$; $\boldsymbol{Z=\frac{X-\mu}{\sigma}\sim N(0,1)}$ 上α分位点：$Z\sim N(0,1)$，$P(Z>z_\alpha)=\alpha$
标准正态分布	$X\sim N(0,1)$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$	0	1	$\Phi(-x)=1-\Phi(x)$，$\Phi(0)=0.5$，$\int_{-\infty}^{+\infty}\varphi(t)\,dt=1$
指数分布	$X\sim Exp(\lambda)$	$\begin{cases}\lambda e^{-\lambda x}, x\ge 0\\ 0, x<0\end{cases}$	$\begin{cases}1-e^{-\lambda x}, x\ge 0\\ 0, x<0\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	泊松分布的间隔，连续的几何分布
二维均匀分布		$p(x,y)=\begin{cases}1/S\\ 0, 其他\end{cases}$				
二维正态分布	$(X,Y)\sim$ $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$	$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho)^2}\left[\frac{(x-u_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-u_2)}{\sigma_1\sigma_2}+\frac{(y-u_2)^2}{\sigma_2^2}\right]}$				X,Y相互独立$\Leftrightarrow\rho=0$； $aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+2ab\sigma_1\sigma_2\rho+b^2\sigma_2^2)$

4. (累积)分布函数 CDF: $F(x)=P(X\le x)$

离散: $F(x)=P(X\le x)=\Sigma_{a\le x}p(a)$

连续: $F(x)=P(X\le x)=\int_{-\infty}^{x}p(t)\,dt$ $\frac{d}{dx}F(x)=p(x)$

$F(x)$为分布函数的**充分必要条件**: ①$F(x)$单调非减; ②$F(x)$右连续;

③$\lim\limits_{x\to-\infty}F(x)=0$, $\lim\limits_{x\to+\infty}F(x)=1$;

$P\{a<X\le b\}=P\{X\le b\}-P\{X\le a\}=F(b)-F(a)$

$P(X<a)=1-P(X\ge a)=F(a-0)$ ⇐"a-0"为 a 左极限, 离散时有意义

$P(x=a)=P(X\le a)-P(X<a)=F(a)-F(a-0)$

5. 中心极限定理: 正态分布是所有分布的最终归宿

6. 泊松过程: $P(X=k,t)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$ 时间可变的泊松分布(t=1)

7. 唯二无记忆性的分布: 几何分布、指数分布

8. 随机变量的函数分布: PDF 为$p_X(x)$的X, PDF 为$p_Y(y)$的$Y=g(X)$

$h(y)$是单调函数$y=g(x)$的反函数, 求$p_Y(y)$

①写出$p_X(x)$; ②$F_Y(y)=P\{Y\le y\}=P\{g(X)\le y\}=P\{X\le h(y)\}=F_X(h(y))$;

③$\boldsymbol{p_Y(y)=F'_Y(y)=\begin{cases}p_X[h(y)]\cdot|h'(y)|, a<y<b\\ 0,\text{其他}\end{cases}}$

其中 a,b 为函数$g(X)$在 X 可能取值区间上的值域

9. $\boldsymbol{\int_{-\infty}^{+\infty}e^{-t^2}dt=\sqrt{\pi}}$; $\boldsymbol{\int_{-\infty}^{0}e^{-t^2}dt=\int_{0}^{+\infty}e^{-t^2}dt=\frac{\sqrt{\pi}}{2}}$; $\sum_{k=0}^{\infty}\frac{1}{k!}=e$

★第三章 多为随机变量及其分布

1. 离散: **联合**概率质量函数 JPMF

边缘概率质量函数 MPMF（边缘分布）:

$p_{i\cdot}=P\{X=x_i\}=\sum_{j=1}^{\infty}P\{X=x_i,Y=y_i\}=\sum_{j=1}^{\infty}p_{ij}, i\in N$ Y 同理

条件概率质量函数: $P\{X=x_i|Y=y_i\}=\frac{P(X=x_i,Y=y_i)}{P(Y=y_i)}$ Y 同理

2. 连续: **联合**概率密度函数 JPDF

边缘概率密度函数 MPDF（边缘密度）: $p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\,dy$ Y 同理

条件概率密度函数: $p_{X|Y}(x|y)=\frac{p(x,y)}{p_Y(y)}$ X 为条件同理

3. 条件概率组成条件分布

4. 联合累积分布函数 JCDF: $F(x,y)=P(\{X\le x\}\text{且}\{Y\le y\})=P(X\le x,Y\le y)$

边缘累积分布函数 MCDF: $F_X(x)=F(x,+\infty)=\lim\limits_{y\to+\infty}F(x,y)=P(X\le x)$

$F_Y(y)=F(+\infty,y)$ **$F_X(x)F_Y(y)$也是分布函数**

条件累积分布函数: 连续: $F_{X|Y}(x|y)=\int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}du$ X 为条件同理

5. 相互**独立**: CDF: $F(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}F_i(x_i)$

PMF: $P(X=x_1,X=x_2,\dots,X=x_n)=\prod_{i=1}^{n}P(X=x_i)$

PDF: $p(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}p_i(x_i)$

离散: $p_{ij}=p_{i\cdot}p_{\cdot j}$ 连续: $p(x,y)=p_X(x)p_Y(y)$

6. $Z=(X,Y)$的分布:

a) X,Y离散: $P(X+Y=k)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)$ 卷积公式

b) X,Y连续: $Z=Y+X$ $p_Z(z)=\int_{-\infty}^{+\infty}p(x,z-x)\,dx=\int_{-\infty}^{+\infty}p(z-y,y)\,dy$

$\overset{X,Y\text{相互独立}}{=}\int_{-\infty}^{+\infty}p_X(x)p_Y(z-x)\,dx=\int_{-\infty}^{+\infty}p_X(z-y)p_Y(y)\,dy$

$Z=\frac{Y}{X}$ $p_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|P(x,xz)\,dx$ $|x|$意义: $|\frac{dy}{dz}|$

$Z=YX$ $p_{YX}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}P\left(x,\frac{z}{x}\right)dx$

c) X离散, Y连续: $F_Z(z)=P\{Z\le z\}=\sum_i P\{X=x_i\}P\{g(x_i,Y)\le z|X=x_i\}$

7. $X_i\sim F_i(x)$ $Y=max(X_1,X_2,\dots,X_n)$ $F_Y(y)=\prod_{i=1}^{n}F_i(y)$

$Z=min(X_1,X_2,\dots,X_n)$ $F_Z(z)=1-\prod_{i=1}^{n}(1-F_i(z))$

★第四章 随机变量的数字特征

1. **数学期望**（随机变量的一阶矩）意义: ①对不确定性的计量; ②加权平均（重心）

离散: $\mu=\mu_X=\boldsymbol{E(X)=\sum_{i=1}^{\infty}x_ip(x_i)}$ 前提: $\sum_{i=1}^{\infty}|x_i|p(x_i)<\infty$

连续: $\boldsymbol{E(X)}=\int_{-\infty}^{\infty}xp(x)\,dx$

性质：①$E[g(X)]=\sum_{i=1}^{\infty}g(x_i)p(x_i)=\int_{-\infty}^{\infty}g(x)p(x)\,dx$；

$$E[g(X,Y)]=\sum_j\sum_i g(x_i,y_j)p(x_i,y_j)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)p(x,y)\,dxdy$$

②$\boldsymbol{E(c)=c}\to E(E(X))=E(X)$　　⑥X,Y不相关：$E(XY)=E(X)E(Y)$

③齐次性：$\boldsymbol{E(aX+b)=aE(X)+b}$

④可加性：$E(X\pm Y)=E(X)\pm E(Y)$

⑤施瓦茨不等式：$\boldsymbol{[E(XY)]^2\le E(X^2)E(Y^2)}$

2. 示性函数：$X=\begin{cases}1,事件A发生\\-1,事件A没发生\end{cases},Y=\begin{cases}1,事件B发生\\-1,事件B没发生\end{cases},XY=\begin{cases}1,AB\cup\bar{A}\bar{B}\\-1,A\bar{B}\cup\bar{A}B\end{cases}$

3. **方差**（二阶矩）：衡量集中程度

$$Var(X)=\sigma^2=\boldsymbol{D(X)=E\left[(X-E(X))^2\right]}=\sum_{i=1}^{\infty}(x_i-\mu)^2p(x_i)$$

性质：①$\boldsymbol{D(X)=E(X^2)-E(X)^2}$；　　③$D(c)=0\to D(D(x))=0$

②$\boldsymbol{D(aX+b)=a^2D(X)}$；　b 的几何意义为平移量

④$D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$

⑤$D(X)=0\Leftrightarrow$存在常数 c 使得$P(X=c)=1$　$X=c$与$P(X=c)=1$不同

4. **标准差**：解决方差单位不一致　　$\sigma=\sigma_x=\sqrt{Var(X)}$

5. 协方差：$Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]=\boldsymbol{E(XY)-E(X)E(Y)}\cancel{=Cov(Y,X)}$

$Cov(X,Y)>0$，**正相关**；<0，**负相关**；$=0$，**不相关**　　不相关⇐独立

性质：①$Cov(aX,bY)=abCov(X,Y)$　　③$Cov(X,X)=D(X)$

②$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

对称性求解协方差：例如 X、Y、Z 独立，且 X+Y+Z=2，

则$cov(X,Y)=cov(X,2-X-Z)=cov(X,2)-cov(X,X)-cov(X,Z)$

因为$cov(X,Y)=cov(X,Z)$，所以$cov(X,Y)=D(X)/2$

6. 相关系数：$\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$　　因为标准差有单位　　$\rho_{XY}\in[-1,1]$

$\rho>0$，**正相关**；　　$\rho<0$，**负相关**；

$\rho=1$，**完全正相关**；$\rho=-1$，**完全负相关**；$\rho=0$，**（线性）不相关**；

$|\rho_{XY}|=1\Leftrightarrow$存在常数$a(\ne 0),b$使得$P(Y=aX+b)=1$

7. 满足二维正态分布的 X,Y 独立 ⇔ $\rho=0$，即不相关

第五章 大数定律和中心极限定理

1. 马尔可夫不等式：$P(X\ge a)\le\frac{E(X)}{a}$　　**切比雪夫不等式：**$P(|X-u|\ge k)\le\frac{\sigma^2}{k^2}$

2. 概率收敛：记为$Y_n\xrightarrow{P}a$　　$\lim_{n\to\infty}P\{|Y_n-a|<\varepsilon\}=1$

3. 弱大数定律　　统计存在的基础　　结论：切比雪夫 or 辛钦能推出伯努利

a) 伯努利大数定律：条件：$X_n\sim B(n,p)$　　$\lim_{n\to\infty}P\left\{\left|\frac{X_n}{n}-p\right|<\varepsilon\right\}=1$

b) 辛钦大数定律：　条件：$X_1,X_2,\dots,X_n$独立同分布，期望存在

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1$$

c) 切比雪夫大数定律：记为$\bar{X}\xrightarrow{P}\mu,n\to\infty$

条件：$X_1,X_2,\dots,X_n$两两不相关，$E(X_i)$与$D(X_i)$存在，$D(X_i)\le c$ (常数)

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

4. 中心极限定理　　解释了为什么生活中正态分布处处可见

a) 棣莫弗-拉普拉斯定理　　理解: 伯努利分布的和的极限是正态分布

条件：$X_n\sim B(n,p)$

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

b) 列维-林德伯格定理：　条件：$X_1,X_2,\dots,X_n$独立同分布，$E(X_i)$与$D(X_i)$存在

$$\lim_{n\to\infty}P\left(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le x\right)=\Phi(x)\qquad\sum_{i=1}^{n}X_i\sim N(n\mu,n\sigma^2)$$

c) 计算方法：例如$P\{X>k\}=P\left\{\frac{X_n-np}{\sqrt{np(1-p)}}>\frac{k-np}{\sqrt{np(1-p)}}\right\}=1-\Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$

第六章 样本及抽样分布

1. 样本$X_1,X_2,\dots,X_n$

a) 若X的分布为$F(x)$，则样本的分布为$F_n(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}F(x_i)$

b) 若X的密度为$f(x)$，则样本的密度为$f_n(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}f(x_i)$

c) $X_1,X_2,\dots,X_n\sim N(\mu,\sigma^2)$，其线性组合也服从正态分布

2. 统计量(样本数字特征)：不含未知参数

a) 样本均值：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 样本标准差：$S=\sqrt{S^2}$

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

c) k(原点)阶矩：$E(X^k)$ k 阶中心矩：$E\{[X-E(X)]^k\}$

k+l 阶混合矩：$E(X^kY^l)$ k+l 阶混合中心矩$E\{[X-E(X)]^k[Y-E(Y)]^l\}$

性质：①$E(\bar{X})=E\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n}\sum E(X_i)=\frac{1}{n}*n*E(X)=E(X)$；

②$D(\bar{X})=D\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n^2}\sum D(X_i)=\frac{1}{n}D(X)$； ③$E(S^2)=D(X)$

3. 抽样分布：统计量的分布

a) **一个正态总体的抽样分布**：总体$X\sim N(\mu,\sigma^2)$

i. 样本均值的分布：$\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ 或 $U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ 量化$\bar{X}$逼近 μ的靠谱程度

ii. 样本方差的分布：量化S^2逼近σ^2的靠谱程度

$\bar{X}$与 S^2相互独立，且$\chi^2=\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n-1)$

$\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$

b) **两个正太总体的抽样分布**：总体$X\sim N(\mu_1,\sigma_1^2)$和$Y\sim N(\mu_2,\sigma_2^2)$，且相互独立

i. 样本均值的差：

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

若$\sigma_1^2=\sigma_2^2$，$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$ $S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

ii. 样本方差的比例：$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

三大分布	统计量	性质
卡方分布	$X_i\sim N(0,1),\chi^2=\sum_{i=1}^{n}X_i^2\sim\chi^2(n)$	①χ_1^2,χ_2^2独立，$\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$； ②$E(X)=n$ $D(X)=2n$ ②上α分位点：$P\{\chi^2>\chi_\alpha^2(n)\}=\int_{\chi_\alpha^2(n)}^{\infty}f(y)\,dy=\alpha$
t 分布	$\begin{cases}X\sim N(0,1)\\Y\sim\chi^2(n)\end{cases},X,Y$独立，$T=\frac{X}{\sqrt{Y/n}}\sim t(n)$	①上α分位点：$P\{T>t_\alpha(n)\}=\int_{t_\alpha(n)}^{\infty}h(t)\,dt=\alpha$；②$h(t)$为偶函数；③n 充分大，$t(n)$近似于$N(0,1)$ ④$P\{\lvert T\rvert>t_{\alpha/2}(n)\}=\alpha$； ⑤$t_{1-\alpha}(n)=-t_\alpha(n)$ ⑥$E(X)=0(n>1)$ $D(X)=\frac{n}{n-2}(n>2)$
F 分布	$\begin{cases}X\sim\chi^2(n_1)\\Y\sim\chi^2(n_2)\end{cases},X,Y$独立，$F=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$	①上α分位点：$P\{F>F_\alpha(n_1,n_2)\}=\int_{F_\alpha(n_1,n_2)}^{\infty}\psi(y)\,dy=\alpha$； $E(X)=\frac{n_2}{n_2-2}(n_2>2)$ ②$F\sim F(n_1,n_2)$，则$\frac{1}{F}\sim F(n_2,n_1)$，$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_1,n_2)}$ $D(X)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}(n_2>4)$

★*第七章 参数估计*

1. 参数估计意义：分布函数已知，部分参数未知

2. 点估计：样本构造估计量$\hat{\theta}(X_1,X_2,\dots,X_n)$，未知参数$\theta$，估计值 $\hat{\theta}(x_1,x_2,\dots,x_3)$

种类：①一致估计量：$\lim\limits_{n\to\infty}P(|\hat{\theta}-\theta|<\varepsilon)=1$ $\hat{\theta}\xrightarrow{P}\theta$ 大样本容量

②无偏估计量：$E(\hat{\theta})=\theta$ 小样本容量

③更有效估计量：$D(\hat{\theta}_1)<D(\hat{\theta}_2)$，$\hat{\theta}_1$更有效

a) **矩估计法**：

理论基础：辛钦大数定律$\lim\limits_{n\to\infty}P(\left|\frac{1}{n}\sum_{i=1}^{n}X_i^k-E(X^k)\right|<\varepsilon)=1$

计算方法：$E(X^l)=\frac{1}{n}\sum_{i=1}^{n}X_i^l, l=1,2,\dots,n$ 考研中l最多为 2

b) **最大似然估计法**：可能性最大的就是事实

最大似然函数：$L(x_1,x_2,\dots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta}L(x_1,x_2,\dots,x_n;\theta)$

最大似然估计值：$\hat{\theta}(x_1,x_2,\dots,x_n)$； 最大似然估计量：$\hat{\theta}(X_1,X_2,\dots,X_n)$

计算方法：$L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)=\prod_{i=1}^{n}f(x_i;\theta)$ $\frac{d}{d\theta}\ln L(\theta)=0$

	待估参数	其他参数	枢轴量的分布	置信区间(置信水平$1-\alpha$)
一个正态总体	μ	σ^2已知	$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}\pm\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$
		σ^2未知	$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left(\bar{X}\pm\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$
	σ^2	μ已知	$\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n)$	$(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)})$
		μ未知	$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)})$
两个正态总体	$\mu_1-\mu_2$	σ_1^2,σ_2^2已知	$U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2)/n_1+(\sigma_2^2)/n_2}}\sim N(0,1)$	$(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{(\sigma_1^2)/n_1+(\sigma_2^2)/n_2})$
		$\sigma_1^2=\sigma_2^2=$未知	$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2),S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$(\bar{X}-\bar{Y}\pm t_{\frac{\alpha}{2}}(n_1+n_2-2)S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$
	σ_1^2/σ_2^2	μ_1,μ_2未知	$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$(\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)})$

第八章 假设检验

1. 第一类错误：H_0是对的，但我们拒绝了它（弃真）

   第二类错误：H_0 错 ， 接受 （纳伪）

2. 显著性水平 $\alpha = P\{当H_0为真时拒绝H_0\} = P_{\mu_0}\{\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| > k\}$

$$\beta(\theta) = P_{\mu}(接收H_0) = P_{\mu}\{\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_{\alpha}\}$$

3. 显著性检测：只控制第一类错误

步骤：①提出H_0（**必须带等号**）； ②给出显著性水平α；

③确定检验统计量及拒绝域形式； ④求出拒绝域W

检验参数	其他参数	H_0	H_1	检验统计量的分布	拒绝域
μ	σ^2已知	$\mu = \mu_0$ $\mu \le \mu_0$ $\mu \ge \mu_0$	$\mu \ne \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$U = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$\lvert U\rvert \ge u_{\alpha/2}$ $U \ge u_{\alpha}$ $U \le -u_{\alpha}$
	σ^2未知			$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$	$\lvert T\rvert \ge t_{\alpha/2}(n-1)$ $T \ge t_{\alpha}(n-1)$ $T \le -t_{\alpha}(n-1)$
σ^2	~~μ已知~~	$\sigma^2 = \sigma_0^2$ $\sigma^2 \le \sigma_0^2$ $\sigma^2 \ge \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	~~$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$~~	$\chi^2 \le \chi^2_{1-\alpha/2}(n)$ or $\chi^2 \ge \chi^2_{\alpha/2}(n)$ $\chi^2 \ge \chi^2_{\alpha}(n)$ $\chi^2 \le \chi^2_{1-\alpha}(n)$
	μ未知			$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sim \chi^2(n-1)$	$\chi^2 \le \chi^2_{1-\alpha/2}(n-1)$ or $\chi^2 \ge \chi^2_{\alpha/2}(n-1)$ $\chi^2 \ge \chi^2_{\alpha}(n-1)$ $\chi^2 \le \chi^2_{1-\alpha}(n-1)$
$\mu_1-\mu_2$	σ_1^2,σ_2^2已知	$\mu_1-\mu_2 = \mu_0$ $\mu_1-\mu_2 \le \mu_0$ $\mu_1-\mu_2 \ge \mu_0$	$\mu_1-\mu_2 \ne \mu_0$ $\mu_1-\mu_2 > \mu_0$ $\mu_1-\mu_2 < \mu_0$	$U = \frac{(\bar{X}-\bar{Y})-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$\lvert U\rvert \ge u_{\alpha/2}$ $U \ge u_{\alpha}$ $U \le -u_{\alpha}$
	$\sigma_1^2 = \sigma_2^2$ =未知			$T = \frac{(\bar{X}-\bar{Y})-\mu_0}{S_{\omega}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$	$\lvert T\rvert \ge t_{\alpha/2}(n_1+n_2-2)$ $T \ge t_{\alpha}(n_1+n_2-2)$ $T \le -t_{\alpha}(n_1+n_2-2)$
σ_1^2/σ_2^2	μ_1,μ_2已知	$\sigma_1^2 = \sigma_2^2$ $\sigma_1^2 \le \sigma_2^2$ $\sigma_1^2 \ge \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$	$F = \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum_{j=1}^{n_2}(X_j-\mu_2)^2/n_2} \sim F(n_1,n_2)$	$F \le F_{1-\alpha/2}(n_1,n_2)$ or $F \ge F_{\alpha/2}(n_1,n_2)$ $F \ge F_{\alpha}(n_1,n_2)$ $F \le F_{1-\alpha}(n_1,n_2)$
	μ_1,μ_2未知			$F = S_1^2/S_2^2 \sim F(n_1-1,n_2-1)$	$F \le F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)$ or $F \ge F_{\frac{\alpha}{2}}(n_1-1,n_2-1)$ $F \ge F_{\alpha}(n_1-1,n_2-1)$ $F \le F_{1-\alpha}(n_1-1,n_2-1)$