概率论与数理统计

第一章 随机事件及其概率

1.1 随机现象及其统计规律

略

1.2 随机事件及其运算

略

1.3 概率的公理化定义及概率的加法公式

$P(\varnothing)=0$

（概率的有限可加性） $P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)$

（真差的概率公式） 若 $B\subset A$，$P(A-B)=P(A)-P(B)$

（概率的减法公式） 对任意两个事件 A 和 B 有，$P(A\overline{B})=P(A-B)=P(A)-P(AB)$

（关于两个事件的概率的加法公式） $P(A\bigcup B)=P(A)+P(B)-P(AB)$

（关于三个事件的概率的加法公式）

$$P(A\bigcup B\bigcup C)=[P(A)+P(B)+P(C)]-[P(AB)+P(AC)+P(BC)]+P(ABC)$$

1.4 古典概型和几何概型

略

1.5 条件概率与乘法公式

（条件概率） $P(B|A)=\dfrac{P(AB)}{P(A)}$

（一般乘法公式） $P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$

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

1.6 全概率公式与贝叶斯公式

（全概率公式） 若 $A_1, A_2, \cdots, A_n, \cdots$ 构成一个完备事件组，且 $P(A_i) > 0$，$i = 1,2,\cdots,n,\cdots$，则对任何事件 B，有 $P(B) = \sum P(A_i)P(B|A_i)$

（贝叶斯公式） 若 $A_1, A_2, \cdots, A_n, \cdots$ 构成一个完备事件组，且 $P(A_i) > 0$，$i = 1,2,\cdots,n,\cdots$，则对任何概率不为零的事件 B，$P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{\sum P(A_j)P(B|A_j)}$

1.7 事件的独立性与伯努利概型

略

第二章 随机变量及其分布

2.1 随机变量的概念及分布函数

略

2.2 离散型随机变量

略

2.3 几种重要的离散分布

分布名称	记号	分布列	数学期望	方差
0-1 分布	$Be(p)$	$P\{X=k\} = p^k(1-p)^{1-k}, k = 0,1$	p	$p(1-p)$
二项分布	$B(n,p)$	$P\{X=k\} = C_n^k p^k(1-p)^{1-k}, k = 0,1,\cdots$	np	$np(1-p)$
泊松分布	$P(\lambda)$	$P\{X=k\} = \dfrac{\lambda^k}{k!}e^{-\lambda}, k = 0,1,\cdots$	λ	λ
超几何分布	$H(n,M,N)$	$P\{X=k\} = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$		
几何分布	$Ge(p)$	$P\{X=k\} = p(1-p)^{k-1}, k = 1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

几何分布具有无记忆性（后续连续分布中的指数分布也具有无记忆性）

2.4 连续型随机变量

略

另附茆诗松版《概率与数理统计教程》，连续随机变量函数的分布：

设 X 是连续随机变量，其密度函数为 $p_X(x)$，$Y=g(X)$是另一个随机变量，若$Y=g(X)$严格单调，其反函数$h(y)$有连续导函数，则$Y=g(X)$的密度函数为

$$p_Y(y)=\begin{cases} p_X[h(y)]\bullet|h'(y)| & a<y<b \\ 0 & 其他 \end{cases}$$

a 为$Y=g(X)$的最小值，b 为$Y=g(X)$的最大值

2.5 几种重要的连续型分布

分布名称	记号	概率密度	数学期望	方差
均匀分布	$U[a,b]$	$f(x)=\begin{cases} \frac{1}{b-a} & a\le x\le b \\ 0 & 其他 \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布	$E(\lambda)$	$f(x)=\begin{cases} \lambda e^{-\lambda x} & x\ge 0 \\ 0 & x<0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布	$N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

（1）若随机变量 $X\sim N(\mu,\sigma^2)$，则 $\frac{X-\mu}{\sigma}\sim N(0,1)$。

（2）特别的，若 $X\sim N(0,1)$，则 $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$。

2.6 随机变量的函数分布

（1）一般地，服从均匀分布的随机变量的线性函数仍然服从均匀分布。

（2）若随机变量 $X\sim N(\mu,\sigma^2)$，a,b 为常数且 $a\neq 0$，则 $aX+b\sim N(a\mu+b,a^2\sigma^2)$，即正态随机变量的线性函数仍然服从正态分布。

第三章 二维随机变量及其分布

3.1 联合分布函数与边缘分布函数

略

3.2 二维离散型随机变量

略

3.3 二维连续型随机变量

（边缘密度函数） 若二维连续型随机变量$(X,Y)\sim f(x,y)$，则(X,Y)关于X和Y的边缘概率密度分别为：

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

3.4 随机变量的独立性

（X与Y独立）$\begin{cases}F(x,y)=F_X(x)F_Y(y)\\ f(x,y)=f_X(x)f_Y(y)\end{cases}$

可扩展到 n 个变量。

3.5 条件分布

离散型：

$Y=y_j$的条件下X的条件分布列 $P\{X=x_i|Y=y_j\}=\dfrac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\dfrac{p_{ij}}{p_{\cdot j}}$

$X=x_i$的条件下Y的条件分布列 $P\{Y=y_j|X=x_i\}=\dfrac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\dfrac{p_{ij}}{p_{i\cdot}}$

连续型：

$Y=y$的条件下X的条件分布列 $f_{X|Y=y}(x)=\dfrac{f(x,y)}{f_Y(y)}$

$X=x$的条件下Y的条件分布列 $f_{Y|X=x}(y)=\dfrac{f(x,y)}{f_X(x)}$

3.6 二维随机变量函数的分布

（二项分布的可加性） 若 $X \sim N(m,p)$，$Y \sim N(n,p)$，且 X 与 Y 相互独立，有 $X+Y \sim B(m+n,p)$。

（正态分布的可加性） 设随机变量 $X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(n,p)$，且 X 与 Y 相互独立，则 $X+Y \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

（泊松分布的可加性） 若 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，且 X 与 Y 相互独立，有 $X+Y \sim P(\lambda_1+\lambda_2)$。

（卷积公式） Z=X+Y 的密度：$f_Z(z)=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$

（最大值分布） $Y=\max(X_1,X_2,\cdots,X_n)$的分布为$F(y)=\prod_{i=1}^{n} F_i(y)$

（最小值分布） $Y=\min(X_1,X_2,\cdots,X_n)$的分布为$F(y)=1-\prod_{i=1}^{n}[1-F_i(y)]$

第四章 随机变量的数字特征和二维正态分布

4.1 数学期望

略

4.2 随机变量函数的数学期望

（1）若 C 是常数，则 $EC=C$。

（2）对任意常数 a 和 b，有 $E(aX-bY)=aEX+bEY$。

（3）若随机变量 X 与 Y 相互独立，有 $E(XY)=EX \cdot EY$。

4.3 方差

（定义） 设 X 是随机变量，若 $E(X-EX)^2$ 存在，则称 $DX=E(X-EX)^2=EX^2-(EX)^2$

（1）若 C 是常数，则 $DC=0$。

（2） $DC=0$ 的充要条件是 $P\{X=C\}=1$。

（3）对任意常数 C，有 $D(CX)=C^2DX$。

（4）对随机变量 X 与 Y，则有 $D(X\pm Y)=DX+DY\pm 2E[(X-EX)(Y-EY)]$。

（协方差相关）$D(X\pm Y)=DX+DY\pm 2Cov(X,Y)$

4.4 协方差与相关系数

（协方差 定义） $Cov(X,Y)=E[(X-EX)(Y-EY)]=E(XY)-EX\cdot EY$

（1） $Cov(aX,bY)=abCov(X,Y)$

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

（3）若 X 与 Y 相互独立，则有 $Cov(X,Y)=0$，特别的 $Cov(X,a)=0$

（相关系数）设 (X,Y) 是二维随机变量，且 $DX>0,DY>0$，称 $\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$

（1） $E(XY)=EX\cdot EY\Leftrightarrow D(X\pm Y)=DX+DY\Leftrightarrow Cov(X,Y)=0\Leftrightarrow X与Y不相关$

（2）若随机变量 X 与 Y 相互独立，则 X 与 Y 不相关

（现代证券组合理论） 马科维兹在 20 世纪 50 年代引进的均值-方差模型成为现代证券组合理论的基石。在这个理论中，假设有甲、乙两种证券可以投资，其收益率可以看做随机变量 X 与 Y，相应的均值（代表平均收益率）为 μ_1 与 μ_2，方差（代表风险）为 σ_1^2 与 σ_2^2。假定投资甲、乙证券的资金比例分别为 x 与 $1-x$，试求该证券组合的平均收益率和风险，并求风险最小的一种证券投资组合。

解： 设该证券组合的总收益率为 Z，则 $Z=xX+(1-x)Y$，其平均收益率为

$$EZ=xEX+(1-x)EY=x\mu_1+(1-x)\mu_2$$

风险为

$$DZ=x^2DX+(1-x)^2DY+x(1-x)Cox(X,Y)=x^2{\sigma_1}^2+(1-x)^2{\sigma_2}^2+x(1-x)\rho\sigma_1\sigma_2$$

令 $\frac{d}{dx}DZ=0$，解得

$$x=\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$$

4.5 随机变量的其他数字特征

（原点矩）$\mu_k=EX^k$

（中心矩）$v_k=E(X-EX)^k$

其他数字特征略

4.6 二维正态分布

二维正态分布 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$　　ρ是相关系数

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2}+\frac{(x-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

i：若二维随机变量 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X 与 Y 相互独立的充要条件是 $\rho=0$。

ii：$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=\int_{-\infty}^{+\infty}f_X(x)dx=1$

第五章　大数定律与中心极限定理

5.1 切比雪夫不等式

（定理）　若随机变量 X 的期望 EX 与方差 DX 都存在，对任意的 $\varepsilon>0$，有

$$P\{|X-EX|\geq\varepsilon\}\leq\frac{DX}{\varepsilon^2}$$

$$P\{|X-EX|\leq\varepsilon\}\geq 1-\frac{DX}{\varepsilon^2}$$

5.2 大数定律

1.切比雪夫大数定律

设随机变量序列 $\{X_n,n\geq1\}$满足：

（i）相互独立；

（ii）期望 $EX_1,EX_2,\cdots$ 和方差 $DX_1,DX_2,\cdots$ 都存在；

（iii）方差一致有界，即存在 $M>0$，使得 $DX_i\le M$，　$i=1,2,\cdots$

则当 $n\to\infty$ 时，$\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\xrightarrow{P}0$

2.伯努利大数定律

在成功概率为 p 的伯努利试验序列中，若用 μ_n 表示前 n 次试验中，当 $n\to\infty$ 时，

$\frac{\mu_n}{n}\xrightarrow{P}p$

3.辛钦大数定律

设随机变量序列 $\{X_n,n\ge1\}$ 满足：

（i）相互独立；

（ii）同分布；

（iii）期望 $EX_i=\mu$ 存在，$i=1,2,\cdots$，

则当 $n\to\infty$ 时，$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu$

5.3 中心极限定理

1.林德伯格-莱维中心极限定理（独立同分布中心极限定理）

设随机变量序列 $\{X_n,n\ge1\}$ 满足：

（i）相互独立；

（ii）同分布；

（iii）期望 $EX_i=\mu$ 和方差 $DX_i=\sigma^2$，$i=1,2,\cdots$ 都存在，

则对任意的 $x\in R$，有

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

2.棣莫弗-拉普拉斯中心极限定理

在每次成功概率为 p 的伯努利试验序列中，若用 μ_n 表示前 n 次试验中，则对任意的 $x \in R$，有

$$\lim_{n\to\infty} P\left\{\frac{\mu_n - np}{\sqrt{np(1-p)}} \le x\right\} = \Phi(x)$$

第六章 统计量及其分布

6.1 总体与样本

略

6.2 统计量与经验分布函数

（统计量） 设 $T(X_1, X_2,\cdots,X_n)$ 是样本 $X_1, X_2,\cdots,X_n$ 的函数，且不含任何未知参数，称 $T(X_1, X_2,\cdots,X_n)$ 为统计量。

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

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

（原点矩） $A_K = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

（中心矩） $B_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k$

（经验分布函数） $$\hat{F}_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \dfrac{k}{n}, & x_{(k)} \le x < x_{(k+1)}, k = 1,2,\cdots,n-1, \qquad x \in R \\ 1, & x \ge x_{(n)} \end{cases}$$

6.3 统计推断中的三大分布

1. χ^2 分布

（定义） 设 X_1，$X_2,\cdots,X_n$ 相互独立且都服从标准正态分布，记

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

称随机变量 χ^2 服从自由度为 n 的 χ^2 分布，记作 $\chi^2 \sim \chi^2(n)$。

i：设 X_1，$X_2,\cdots,X_n$ 是来自正态总体 $N(\mu，\sigma^2)$ 的样本，则 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(1)$。

ii：（可加性）若随机变量服从 $\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，且 χ_1^2 与 χ_2^2 相互独立，则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1+n_2)$$

iii：$E\chi^2 = n$，$D\chi^2 = 2n$

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，记

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

称随机变量 T 服从自由度为 n 的 t 分布，记作 $T \sim t(n)$。

i：$t_\alpha(n) = -t_{1-\alpha}(n)$。

3.F 分布

（定义） 设 $X \sim \chi^2(m)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，记

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

称随机变量 F 服从自由度为 m 和 n 的 F 分布，记作 $F \sim F(m,n)$，其中称 m 为分子自由度，n 为分母自由度。

$$F_{\alpha}(m,n)=\frac{1}{F_{1-\alpha}(n,m)}$$

6.4 正态总体下的抽样分布定理

（1）$\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)\Leftrightarrow\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$；

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

（3）$\overline{X}$ 和 S^2 相互独立。

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

（均值差的分布）设样本 X_1，$X_2,\cdots,X_{n_1}$ 和 Y_1，$Y_2,\cdots,Y_{n_2}$ 分别来自两个相互独立的正态总体 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$，则

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

i：若 $\sigma_1=\sigma_2=\sigma$，则

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$ （其中 $S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ ）

ii：设样本 X_1，$X_2,\cdots,X_{n_1}$ 和 Y_1，$Y_2,\cdots,Y_{n_2}$ 分别来自两个相互独立的正态总体 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$，则统计量

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

第七章 参数估计

7.1 点估计

（替换原理和矩法估计） 用样本矩替换对应的总体矩的思想称为替换原理，建立在替换原理基础上的估计称为矩法估计。

（例题）设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim f(x;\theta)=\begin{cases}\dfrac{2}{\theta^2}(\theta-x), & 0<x<\theta\\ 0, & \text{其他}\end{cases}$ 的样本，求 θ 的矩法估计量。

解：总体均值

$$EX=\int_0^\theta xf(x;\theta)dx=\int_0^\theta x\cdot\frac{2}{\theta^2}(\theta-x)dx=\frac{\theta}{3}$$

由替换原理，$\overline{X}=\dfrac{\theta}{3}$，解得 θ 的矩法估计量为 $\hat{\theta}=3\overline{X}$。

（最大似然估计） ①写似然函数→②求对数似然函数→③对对数似然函数求导得并令导函数为 0 得似然方程→④求出最大似然估计值→⑤检验是否满足

（例题）设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim P(\lambda)$ 的样本，求 λ 的最大似然估计量。

解：总体分布列为

$$P\{X=x\}=\frac{\lambda^x}{k!}e^{-\lambda},\qquad x=0,1,2,\cdots$$

对于给定的样本观测值 $x_1,x_2,\cdots,x_n$，似然函数

$$L(\lambda)=\prod_{i=1}^{n}P\{X_i=x_i\}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda}$$

取对数，得对数似然函数得

$$\ln L(\lambda)=\sum_{i=1}^{n}x_i\ln\lambda-\ln(x_1!x_2!\cdots x_n!)-n\lambda$$

对 λ 求导数并令其为 0，得对数似然方程

$$\frac{d}{d\lambda}\ln L(\lambda)=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0$$

解得 λ 的最大似然估计值为 $\hat{\lambda}=\overline{x}$，最大似然估计量为 $\hat{\lambda}=\overline{X}$

7.2 估计量评价的一般标准

（无偏估计）$E_\theta\left(\hat{\theta}\right)=\theta$

（有效性）$D\left(\hat{\theta}_1\right)\le D\left(\hat{\theta}_2\right)$[$\hat{\theta}_1$ 比有 $\hat{\theta}_2$ 效]

（相合性）$\lim\limits_{n\to\infty}P\left(\left|\hat{\theta}-\theta\right|\ge\varepsilon\right)=0$

7.3 单个正态总体的区间估计

1. 当 σ^2 已知时 μ 的区间估计

枢轴量 $U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，则 μ 的 $1-\alpha$ 置信区间为 $\left[x\pm u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}\right]$

2. 当 σ^2 未知时 μ 的区间估计

枢轴量 $T=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim t(n-1)$，则 μ 的 $1-\alpha$ 置信区间为 $\left[x\pm t_{1-\frac{\alpha}{2}}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right]$

3. 当 μ 未知时 σ^2 的区间估计

枢轴量 $\dfrac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$，则 μ 的 $1-\alpha$ 置信区间为 $\left[\dfrac{(n-1)S^2}{{\chi_{1-\frac{\alpha}{2}}}^2(n-1)},\dfrac{(n-1)S^2}{{\chi_{\frac{\alpha}{2}}}^2(n-1)}\right]$

7.4 两个正态总体的区间估计

1. 当 σ_1^2 和 σ_2^2 都已知时 μ_1 - μ_2 的区间估计

枢轴量 $U=\dfrac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$，则 μ 的 $1-\alpha$ 置信区间为 $\left[\left(\overline{X}-\overline{Y}\right)\pm u_{1-\frac{\alpha}{2}}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\right]$

2. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时 μ_1 - μ_2 的区间估计

枢 轴 量 $T=\dfrac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$ ， 则 μ 的 $1-\alpha$ 置 信 区 间 为

$$\left[\left(\overline{X}-\overline{Y}\right)\pm t_{1-\frac{\alpha}{2}}\left(n_1+n_2-2\right)\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

3. 方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计

枢轴量 $F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F\left(n_1-1,n_2-1\right)$，则 μ 的 $1-\alpha$ 置信区间为

$$\left[\frac{1}{F_{1-\frac{\alpha}{2}}\left(n_1-1,n_2-1\right)}\cdot\frac{S_1^2}{S_2^2},\frac{1}{F_{\frac{\alpha}{2}}\left(n_1-1,n_2-1\right)}\cdot\frac{S_1^2}{S_2^2}\right]$$

第八章 假设检验

8.1 假设检验的基本思想

（第一类错误） 如果原假设 H_0 成立，但是样本观测值落入拒绝域 W 因而拒绝 H_0，即拒绝了正确的结论，称此统计判决犯了**第一类错误**或**弃真错误**，其概率记为 α，即

$$\alpha=P_{H_0}\left(W\right)$$

（第二类错误） 如果原假设 H_0 不成立，但是样本观测值落入拒绝域 $\overline{W}$ 因而接受 H_0，即接受了错误的结论，称此统计判决犯了**第二类错误**或**纳伪错误**，其概率记为 β，即

$$\beta=P_{H_1}\left(\overline{W}\right)$$

8.2 单个正态总体的假设检验

表 8.2 正态总体均值的假设检验

H_0	H_1	方差 σ^2 已知（u 检验法） 检验统计量	方差 σ^2 已知（u 检验法） 拒绝域 W	方差 σ^2 未知（t 检验法） 检验统计量	方差 σ^2 未知（t 检验法） 拒绝域 W
$\mu=\mu_0$	$\mu\neq\mu_0$	$U=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\lvert u\rvert>u_{1-\alpha/2}$	$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\lvert t\rvert>t_{1-\alpha/2}(n-1)$
$\mu\leqslant\mu_0$	$\mu>\mu_0$		$u>u_{1-\alpha}$		$t>t_{1-\alpha}(n-1)$
$\mu\geqslant\mu_0$	$\mu<\mu_0$		$u<u_{\alpha}$		$t<t_{\alpha}(n-1)$

表 8.3 正态总体方差的假设检验

H_0	H_1	χ^2 检验法	
		检验统计量	拒绝域 W
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{n-1}{\sigma_0^2}S^2$	$\chi^2<\chi^2_{\alpha/2}(n-1)$ 或 $\chi^2>\chi^2_{1-\alpha/2}(n-1)$
$\sigma^2\leqslant\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2>\chi^2_{1-\alpha}(n-1)$
$\sigma^2\geqslant\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2<\chi^2_{\alpha}(n-1)$

8.3 两个正态总体的假设检验

表 8.4 两个正态总体均值差的假设检验

方差 σ_1^2,σ_2^2 都已知(u 检验法)			
H_0	H_1	检验统计量	拒绝域 W
$\mu_1=\mu_2$	$\mu_1\neq\mu_2$	$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\lvert u\rvert>u_{1-\alpha/2}$
$\mu_1\leqslant\mu_2$	$\mu_1>\mu_2$		$u>u_{1-\alpha}$
$\mu_1\geqslant\mu_2$	$\mu_1<\mu_2$		$u<u_{\alpha}$
方差 $\sigma_1^2=\sigma_2^2=:\sigma^2$ 未知(t 检验法)			
H_0	H_1	检验统计量	拒绝域 W
$\mu_1=\mu_2$	$\mu_1\neq\mu_2$	$T=\frac{\overline{X}-\overline{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$\lvert t\rvert>t_{1-\alpha/2}(n_1+n_2-2)$
$\mu_1\leqslant\mu_2$	$\mu_1>\mu_2$		$t>t_{1-\alpha}(n_1+n_2-2)$
$\mu_1\geqslant\mu_2$	$\mu_1<\mu_2$		$t<t_{\alpha}(n_1+n_2-2)$

表 8.5 两个正态总体方差比的假设检验(数学期望未知情形)

H_0	H_1	用 F 检验法	
		检验统计量	拒绝域 W
$\sigma_1^2=\sigma_2^2$	$\sigma_1^2\neq\sigma_2^2$	$F=\frac{S_1^2}{S_2^2}$	$f<F_{\alpha/2}(n_1-1,n_2-1)$, 或 $f>F_{1-\alpha/2}(n_1-1,n_2-1)$
$\sigma_1^2\leqslant\sigma_2^2$	$\sigma_1^2>\sigma_2^2$		$f>F_{1-\alpha}(n_1-1,n_2-1)$
$\sigma_1^2\geqslant\sigma_2^2$	$\sigma_1^2<\sigma_2^2$		$f<F_{\alpha}(n_1-1,n_2-1)$

8.4 非参数假设检验

1. χ^2 拟合优度检验

i. 分类数据的 χ^2 拟合优度检验

检验统计量：$\chi^2=\sum_{i=1}^{l}\frac{(n_i-np_i)^2}{np_i}$

拒绝域：$W=\left\{\chi^2>\chi^2_{1-\alpha}(l-1)\right\}$

ii. 分类数据的 χ^2 拟合优度检验

检验统计量：$\chi^2=\sum_{i=1}^{l}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$

拒绝域：$W=\left\{\chi^2>\chi^2_{1-\alpha}(l-k-1)\right\}$ （其中 k 为未知参数个数）

解题步骤

①求服从分布中的参数的最大似然估计

②计算分布的概率估计值

③各种计算结果列表

组号	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$
合计				

④查表求拒绝域，做出统计判断

2. K-S 检验

经验分布函数：$\hat{F}_n\left(x_{(i)}\right)=\frac{i}{n}$

记 $d_{1i}=\left|F_0\left(x_{(i)}\right)-\hat{F}_n\left(x_{(i-1)}\right)\right|$，$d_{2i}=\left|F_0\left(x_{(i)}\right)-\hat{F}_n\left(x_{(i)}\right)\right|$，则检验统计量为

$$D_n=\left|d_{1i},d_{2i},i=1,2,\cdots,n\right|$$

解题步骤

①列表计算

i	$x_{(i)}$	i/n	$f\left(x_{(i)}\right)$<概率>	d_{1i}	d_{2i}	$\max\{d_{1i},d_{2i}\}$

②查表求拒绝域，做出统计判断

3. 独立性检验

检验统计量：$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij}-n\hat{p}_{i\cdot}\hat{p}_{\cdot j}\right)^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}}$　　其中 r 为行，s 为列

拒绝域：$W=\left\{\chi^2>\chi^2_{1-\alpha}\left((r-1)(s-1)\right)\right\}$

解题步骤

①写出待检验假设

②查表给出拒绝域

③求出检验统计量的值，检验假设

第九章 线性回归分析与方差分析

9.1 一元线性回归分析

回归方程：$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x$

1. 回归系数的最小二乘估计

实测值 y_i 与回归值 $\hat{y}_i$ 之间的差 $e_i=y_i-\hat{y}_i,\quad i=1,2,\cdots,n$ 为残差。

按照最小二乘法的思想，希望回归方程中的 $\hat{\beta}_0,\hat{\beta}_1$ 使残差平方和达到最小，即要求

$$\sum_{i=1}^{n}\left[y_i-\left(\hat{\beta}_0+\hat{\beta}_1x\right)\right]^2=\min_{\hat{\beta}_0,\hat{\beta}_1}\sum_{i=1}^{n}\left[y_i-\left(\hat{\beta}_0+\hat{\beta}_1x\right)\right]^2$$

记 $Q(\beta_0,\beta_1)=\sum_{i=1}^{n}\left[y_i-\left(\hat{\beta}_0+\hat{\beta}_1x\right)\right]^2$ 称其为误差的平方和，分别对 $\hat{\beta}_0,\hat{\beta}_1$ 求偏导，并令偏导数等于零得到

$$\begin{cases}\frac{\partial Q}{\partial\beta_0}=-2\sum\left(y_i-\beta_0-\beta_1x_i\right)=0\\ \frac{\partial Q}{\partial\beta_1}=-2\sum x_i\left(y_i-\beta_0-\beta_1x_i\right)=0\end{cases}$$

整理得到一个关于 $\hat{\beta}_0,\hat{\beta}_1$ 的线性方程组

$$\begin{cases} n\beta_0+\beta_1\sum x_i=\sum y_i \\ \beta_0\sum x_i+\beta_1\sum x_i^2=\sum x_i y_i \end{cases}$$

以上两个方程组均为正规方程组，易得 $\hat{\beta}_0,\hat{\beta}_1$ 的最小二乘法估计

$$\hat{\beta}_1=\frac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum\left(x_i-\bar{x}\right)^2},\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

2. 求一元线性回归

$$l_{xx}=\sum\left(x_i-\bar{x}\right)^2=\sum x_i^2-n\bar{x}^2$$

$$l_{xy}=\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=\sum x_i y_i-n\bar{x}\bar{y}$$

$$l_{yy}=\sum\left(y_i-\bar{y}\right)^2=\sum y_i^2-n\bar{y}^2$$

解题步骤:

①列表计算：

$\sum x_i=$	$n=$	$\sum y_i=$
$\bar{x}=$		$\bar{y}=$
$\sum x_i^2=$	$\sum x_i y_i=$	$\sum y_i^2=$
$n\bar{x}^2=$	$n\bar{x}\bar{y}=$	$n\bar{y}^2=$
$l_{xx}=$	$l_{xy}=$	$l_{yy}=$

$$\hat{\beta}_1=\frac{l_{xy}}{l_{xx}}= \qquad ,\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=$$

②写出回归方程

3. 回归方程的显著性检验

总偏差平方和：$SST=l_{yy}$

回归平方和：$SSR=\frac{l_{xy}^2}{l_{xx}}$

残差平方和：$SSE=SST-SSR$

i：$\frac{SSE}{\sigma^2}\sim\chi^2(n-1)$，从而$\hat{\sigma}^2=\frac{SSE}{n-2}$

ii：SSE 与 $\hat{\beta}_0,\hat{\beta}_1$ 相互独立

iii：当 $\beta_1=0$ 时，$\frac{SSR}{\sigma^2}\sim\chi^2(1)$，且 SSR 与 SSE 相互独立。

解题步骤:

①列表计算：

来源	平方和	自由度	均方	F 比	临界值
回归	SSR	1	SSR	$\frac{SSR}{SSE/(n-2)}$	$F_{1-\alpha}(1,n-2)$
残差	SSE	$n-2$	$SSE/(n-2)$		
总和	SST	$n-1$			

②查表求出临界值，判断显著性

Ps：拒绝域$W=\{F\geq F_{1-\alpha}(1,n-2)\}$

4. 因变量的预测

自变量取 x_0 时的因变量 Y_0 的点预测：

$$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1x_0$$

Y_0 的置信水平 $1-\alpha$ 的预测区间是 $\left[\hat{Y}_0-\delta(x_0),\hat{Y}_0+\delta(x_0)\right]$

其中，$\delta(x_0)=t_{1-\frac{\alpha}{2}}(n-1)\cdot\hat{\sigma}\cdot\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})}{l_{xx}}}$

9.2 多元回归线性分析

略，类似一元线性回归。

9.3 方差分析

1. 单因素方差分析

$$\begin{cases} SST = \sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} - \dfrac{T^2}{n} \\ SSA = \sum_{i=1}^{r}\dfrac{T_{i\cdot}^{\ 2}}{n_1} - \dfrac{T^2}{n} \\ SSE = SST - SSA \end{cases}$$

统计量：$F = \dfrac{SSA/(r-1)}{SSE/(n-r)} \sim F(r-1, n-r)$

拒绝域：$W = \{f > F_{1-\alpha}(r-1, n-r)\}$

解题步骤：

①求出各偏差平方和。

②列表计算：

来源	平方和	自由度	均方	F 比	临界值
组间	SSA	$r-1$	$MSSA$	$f = \dfrac{MSSA}{MSSE}$	$F_{1-\alpha}(r-1, n-r)$
组内	SSE	$n-r$	$MSSE$		
总和	SST	$n-1$			

④查表求拒绝域，做出统计判断

2. 无交互作用的双因素方差分析

$$\begin{cases} SST = \sum_{i=1}^{r}\sum_{j=1}^{s} x_{ij} - \dfrac{T^2}{rs} \\ SSA = \dfrac{1}{s}\sum_{i=1}^{r} T_{i\cdot}^{\ 2} - \dfrac{T^2}{rs} \\ SSB = \dfrac{1}{r}\sum_{i=1}^{s} T_{\cdot j}^{\ 2} - \dfrac{T^2}{rs} \\ SSE = SST - SSA - SSB \end{cases}$$

解题步骤：

①求出各偏差平方和。

②列表计算：

来源	平方和	自由度	均方	F 比	临界值
因素 A	SSA	$r-1$	$MSSA$	$f_A=\dfrac{MSSA}{MSSE}$	$F_{A,1-\alpha}(r-1,(r-1)(s-1))$
因素 B	SSB	$s-1$	$MSSB$	$f_B=\dfrac{MSSB}{MSSE}$	$F_{B,1-\alpha}(s-1,(r-1)(s-1))$
误差	SSE	$(r-1)(s-1)$	$MSSE$		
总和	SST	$rs-1$			

④查表求拒绝域，做出统计判断