# 习题解答*
# 习题解答

†

南开大学统计研究院, 天津, 300071

2014 年 2 月 3 日

## 1 习题一

1. 设$X_1,\ldots,X_n$为来自总体分布$F(x)$的iid样本，记其经验分布函数为$F_n(x)$，试证对于任意给定的$x\in R$，有

$$E(F_n(x))=F(x),\quad \mathrm{Var}(F_n(x))=\frac{1}{n}F(x)(1-F(x)).$$

解：由经验分布函数的定义

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(X_i<x)$$

其中$\mathrm{I}(x)$是示性函数。因为$X_1,\ldots,X_n$是独立的，而$\mathrm{I}(X_i<x)$ 以$F(x)$的概率取1，$1-F(x)$的概率取0，故$nF(x)$服从二项分布$\mathrm{B}(n,F(x))$, 则

$$\begin{aligned}\mathrm{E}(F_n(x)) &= F(x)\\ \mathrm{Var}(F_n(x)) &= \frac{1}{n}F(x)(1-F(x))\end{aligned}$$

2. 设$X_1,\ldots,X_n$为来自总体分布$F(x)$的iid样本，记其经验分布函数为$F_n(x)$，试证对于任意给定的$x\in R$，$F_n(x)\xrightarrow{P}F(x)$.

解：由切比雪夫不等式，对任意给的$\varepsilon$

$$\mathrm{P}(|F_n(x)-E(F_n(x))|>\varepsilon)\le\frac{\mathrm{Var}(F_n(x))}{\varepsilon^2}$$

*王兆军, 邹长亮 (2014), 数理统计讲义, 高等教育出版社

†wangguanghui@mail.nankai.edu.cn

由第一题的结论有

$$\mathrm{P}(|F_n(x)-F(x)|>\varepsilon)\leq\frac{F(x)(1-F(x))}{n\varepsilon^2}$$

而$0\leq F(x)\leq 1$,则$F(x)(1-F(x))\leq\frac{1}{4}$,所以对于任意$\varepsilon$,在$n\to\infty$的条件下，

$$\mathrm{P}(|F_n(x)-F(x)|>\varepsilon)\to 0$$

得证。

3. 设$X_1,\ldots,X_{n+1}$为来自总体$X$的 iid样本，试证明

$$ {S^*_{n+1}}^2=\frac{n}{n+1}\left[S^{*2}_n+\frac{1}{n+1}(X_{n+1}-\bar{X}_n)^2\right.$$

证明：

$$\begin{aligned}
S^{*2}_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1}(X_i-\bar{X}_{n+1})^2\\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1}X_i^2-(n+1)\bar{X}^2_{n+1})\\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1}X_i^2-\frac{(n\bar{X}_n+X_{n+1})^2}{n+1})\\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1}X_i^2-\frac{n^2\bar{X}_n^2}{n+1}-\frac{X^2_{n+1}}{n+1}-\frac{2n\bar{X}_nX_{n+1}}{n+1})\\
&= \frac{1}{n+1}(\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2+\frac{n}{n+1}(\bar{X}_n^2+X^2_{n+1}-2n\bar{X}_nX_{n+1}))\\
&= \frac{1}{n+1}(nS^{*2}_n+\frac{n}{n+1}(X_{n+1}-\bar{X}_n)^2)\\
&= \frac{n}{n+1}(S^{*2}_n+\frac{1}{n+1}(X_{n+1}-\bar{X}_n)^2)
\end{aligned}$$

4. 设$X_1,\ldots,X_n$为来自$N(0,\sigma^2)$的iid样本， $A=(a_{ij})$是一$n$阶正交阵，$\mathbf{c}=(c_1,\ldots,c_n)^{\mathrm{T}}$为一$n$维常向量，记$\mathbf{X}=(X_1,\ldots,X_n)^{\mathrm{T}}$，则$\mathbf{Y}=A\mathbf{X}+\mathbf{c}$ 的各分量相互独立，且分别服从$N(c_i,\sigma^2)$.

解：由$X_1,\ldots,X_n$是独立的正态分布知，$X$ 的联合密度函数为

$$f(X)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{X'X}{2\sigma^2}}$$

由密度变换公式知
$$f(Y)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}|A|e^{-\frac{(A'(Y-c))'(A'(Y-c))}{2\sigma^2}}$$
$A$是正交阵，故$|A|=1, A'A=I$，则$Y$的密度函数为
$$f(Y)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{(Y-c)'(Y-c)}{2\sigma^2}}$$
易知$Y_i$的分布密度函数为
$$f(Y_i)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(Y_i-c_i)^2}{2\sigma^2}}$$
故$Y_1,\ldots,Y_n$的联合密度函数等于密度函数的乘积，得证。

5. 设$X_1,\ldots,X_8$为来自$N(0,1)$的iid样本，$Y=(X_1+\cdots+X_4)^2+(X_5+\cdots+X_8)^2$，试求常数$c$，使$cY$服从$\chi^2$分布.
解：$c=\frac{1}{4}$.

6. 设$X_1,\ldots,X_4$为来自$N(0,1)$的iid样本，$X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$，试求$a,b$，使 $X$服从$\chi^2$分布.

解：$a=\frac{1}{5}, b=\frac{1}{25}$. 7. 设$X_1,\ldots,X_n$为来自$N(\mu,\sigma^2)$的iid样本，请问当$k$为多少时，
$$P\{\bar{X}>\mu+kS_n\}=0.95?$$
解：$K=\frac{t_{0.95}(n-1)}{\sqrt{n}}$

8. 设$X_1,\ldots,X_n$为来自某总体$X\sim F$的iid样本，且$EX=\mu, \mathrm{Var}(X)=\sigma^2$. 记$T_n=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$，则在某些连续性假设下，证明：$T_n$的分布函数可展成如下形式：
$$\Phi(x)-\phi(x)\left[\frac{\gamma_1}{6\sqrt{n}}H_2(x)+\frac{1}{n}\left(\frac{\gamma_2}{24}H_3(x)+\frac{\gamma_1^2}{72}H_5(x)\right)+r_n\right],$$
其中$\gamma_1,\gamma_2$分别为总体$X$的偏度与峰度，$r_n=o(n^{-1})$, $H_r(\cdot)$为$r$阶Hermite多项式，其定义如下：$\frac{\mathrm{d}^r(e^{-x^2/2})}{\mathrm{d}x^r}=(-1)^rH_r(x)e^{-x^2/2}$. 上述展式被称为Edgeworth展开. 由上述Edgeworth展开，请验证：$T_n$的$\alpha$分位数可近似为
$$u_\alpha+\frac{\gamma_1}{6\sqrt{n}}(u_\alpha^2-1)+\frac{\gamma_2}{24n}(u_\alpha^3-3u_\alpha)-\frac{\gamma_1^2}{36n}(2u_\alpha^3-5u_\alpha)+r_n,$$
其中$u_\alpha$为标准正态的$\alpha$分位数，且上述展式被称为Fisher-Cornish展开.
解：易知下列Hermite函数为$H_2(x)=x^2-1, H_3(x)=x^3-3x, H_5(x)=x^5-10x^3+15x$.

令$x = u_\alpha + \frac{r_1}{6\sqrt{n}}(u_\alpha^2 - 1) + \frac{r_2}{24n}(u_\alpha^3 - 3u_\alpha) - \frac{r_1^2}{36n}(2u_\alpha^3 - 5u_\alpha + r_n)$. 将$\Phi(x)$在$u_\alpha$点展开，则有

$$\begin{aligned}\Phi(x) &= \alpha + \phi(u_\alpha)\frac{r_1}{6\sqrt{n}}(u_\alpha^2 - 1) + \phi(u_\alpha)\frac{r_2}{24n}(u_\alpha^3 - 3u_\alpha) \\ &- \phi(u_\alpha)\frac{r_1^2}{72n}(u_\alpha^5 + 2u_\alpha^3 - 9u_\alpha) + o(n^{-1})\end{aligned}$$

而$\phi(x) = \phi(u_\alpha) - \phi(u_\alpha)u_\alpha\frac{r_1}{6\sqrt{n}}(u_\alpha^2 - 1) + o(n^{-1})$, 且

$$\begin{aligned}\frac{r_1}{6\sqrt{n}}H_2(x) &+ \frac{1}{n}(\frac{r_2}{24}H_3(x) + \frac{r_1^2}{72}H_5(x)) = \frac{r_1}{6\sqrt{n}}(u_\alpha^2 - 1) + \frac{r_1^2}{18n}(u_\alpha^3 - u_\alpha) \\ &+ \frac{r_2}{24}(u_\alpha^3 - 3u_\alpha) + \frac{r_1^2}{72n}(u_\alpha^5 + 15u_\alpha - 10u_\alpha^3) \\ &+ o(n^{-1})\end{aligned}$$

代入上面结果知

$$\Phi(x) - \phi(x)[\frac{r_1}{6\sqrt{n}}H_2(x) + \frac{1}{n}(\frac{r_2}{24}H_3(x) + \frac{r_1^2}{72}H_5(x)) + r_n = \alpha$$

则因左式是分布函数的近似知，$x$是$\alpha$分位点。

10. 设$X_1, \ldots, X_n$为来自总体$X \sim F$的iid样本，总体概率密度函数为$f(x)$, 则对于给定的$1 \le j \le k \le n$，统计量$X_{(j)}$与$X_{(k)}$的联合概率密度函数为

$$f_{j,k}(x, y) = \begin{cases} \frac{n!}{(j-1)!(k-j-1)!(n-k)!}F^{j-1}(x)[F(y) - F(x)^{k-j-1}[1 - F(y)^{n-k}f(x)f(y), & x < y, \\ 0, & \text{否则}. \end{cases}$$

解：对任意给定的$x < y$,$f_{j,k}(x, y)$等于这样一个事件：有$j - 1$个样本小于$x$, 有一个样本等于$x$, 有$k - j - 1$个样本位于$y$和$x$之间，有一个样本等于$y$,有$n - k$个样本大于$y$,而上述事件对应的概率分别为$F(x), f(x), F(y) - F(x), f(y), 1 - F(y)$,由多项分布的密度函数知

$$f_{j,k}(x, y) = \frac{n!f(x)f(y)}{(j-1)!(k-j-1)!(n-k)!}F^{j-1}[F(y) - F(x)^{k-j-1}[1 - F(y)^{n-k}$$

而对任意的$x \ge y$,$f_{j,k}(x, y) = 0$,得证。

11. 设$X_1, \ldots, X_{n+1}$为来自$N(\mu, \sigma^2)$的iid样本，试求统计量

$$T = \frac{X_{n+1} - \bar{X}_n}{S_n^*}\sqrt{\frac{n-1}{n+1}}$$

的抽样分布.

解：由$X_1,\dots,X_{n+1}$独立正态，$X_{n+1}-\bar{X}_n$为正态分布，而

$$\mathrm{E}(X_{n+1}-\bar{X}_n)=0,\mathrm{Var}(X_{n+1}-\bar{X}_n)=\frac{n+1}{n}\sigma^2$$

则$\xi=\frac{X_{n+1}-\bar{X}_n}{\sqrt{\frac{n+1}{n}}\sigma}$服从标准正态分布，有$\eta=\frac{nS_n^{*2}}{\sigma^2}$服从自由度为$n-1$的卡方分布，且$\xi,\eta$独立，则$T=\frac{\xi}{\sqrt{\eta/(n-1)}}$服从自由度为$n-1$的t分布。

12. 设$X_1,\dots,X_{m+n}$为来自$N(0,\sigma^2)$的iid样本，试求统计量

$$Z=\sqrt{\frac{m}{n}}\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=n+1}^{n+m}X_i^2}},\quad F=\frac{m}{n}\frac{\sum_{i=1}^n X_i^2}{\sum_{i=n+1}^{n+m}X_i^2}$$

的抽样分布.

解：因$X_1,\dots,X_{n+m}$独立服从$N(0,\sigma^2)$,令

$$\begin{aligned}\xi &= \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n X_i}{\sigma}\\ \eta &= \sum_{i=n+1}^{n+m}\frac{X_i^2}{\sigma^2}\\ \theta &= \sum_{i=1}^{n}\frac{X_i^2}{\sigma^2}\end{aligned}$$

则$\xi$服从标准正态分布，$\eta$服从自由度为$m$ 的卡方分布，$\theta$服从自由度为$n$ 的卡方分布，且$\xi,\theta$均与$\eta$独立，而$Z=\frac{\xi}{\sqrt{\eta/m}},F=\frac{\theta/n}{\eta/m}$，故$Z,F$分别服从自由度为$m$的t分布和自由度为$(n,m)$的F 分布。

13. 设随机变量$X,Y$相互独立且均服从$N(0,\sigma^2)$，则

(1) $X^2+Y^2$与$X/\sqrt{X^2+Y^2}$相互独立；

(2) $\theta=\sin^{-1}\frac{X}{\sqrt{X^2+Y^2}}$服从$(-\pi/2,\pi/2)$上的均匀分布；

(3) $X/Y$服从Cauchy 分布.

解：由$X,Y$独立且分别服从正态分布易知，$X,Y$的联合密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

做变换 $x=r\sin(\theta),y=r\cos(\theta)$,其中$\theta\in[-\pi/2,\pi/2$,则变换的雅可比行列式为$|r|$,由密度函数变换公式，$r,\theta$的联合密度函数为

$$f(r,\theta)=\frac{1}{2\pi\sigma^2}|r|e^{-\frac{r^2}{2\sigma^2}}$$

则$r,\theta$是独立的，而$\frac{X}{\sqrt{X^2+Y^2}}=\sin(\theta), X^2+Y^2=r^2$, 故(1)得证。
(2) 由$r,\theta$联合密度函数知$\theta$服从$(-\pi/2,\pi/2)$上的均匀分布，得证。
(3)因$\frac{X}{Y}=\tan(\theta)$, 而$\theta$ 的密度函数为$f(\theta)=\frac{1}{\pi}$，故$Z=\tan(\theta)$ 的密度函数为 $f(z)=\frac{1}{\pi}(\arctan(z))'=\frac{1}{\pi(1+z^2)}$，即$\frac{X}{Y}$服从Cauchy 分布。

14. 设$X_1,\ldots,X_4$为来自$N(0,\sigma^2)$的iid样本，$U=\frac{\sqrt{3}X_1}{\sqrt{X_2^2+X_3^2+X_4^2}}$, 试求$EU$, Var$U$.
解：容易知道$U$服从自由度为3的t分布，因此$\mathrm{E}(U)=0, \mathrm{Var}(U)=3$.

15. 设$X_1,\ldots,X_{n+1}$为来自$N(\mu,\sigma^2)$的iid样本，记$V_i=X_i-\frac{1}{n+1}\sum_{i=1}^{n+1}X_i, i=1,\cdots,n+1$. 试求$V_i$的分布.
解：因为$X_1,\ldots,X_{n+1}$独立正态，故$V_i=\frac{n}{n+1}X_i-\frac{1}{n+1}\sum_{j\neq i}X_j$服从正态分布，又

$$\mathrm{E}(V_i)=0, \mathrm{Var}(V_i)=\frac{n}{n+1}\sigma^2$$

故$V_i$服从均值0,方差$\frac{n}{n+1}\sigma^2$ 的正态分布。

16. 设$X_1,\ldots,X_m$为来自$N(\mu_1,\sigma^2)$的iid样本，$Y_1,\ldots,Y_n$为来自 $N(\mu_2,\sigma^2)$的iid样本，且全样本独立. 又设$a,b$为两个常数，$Z=\frac{a(\bar{X}-\mu_1)+b(\bar{Y}-\mu_2)}{\sqrt{(m-1)S_{1m}^2+(n-1)S_{2n}^2}}$，试求常数$c$，使得$cZ$服从$t$分布，并给出其自由度.
解：$c=\sqrt{\frac{m+n-2}{\frac{a^2}{m}+\frac{b^2}{n}}}$, 自由度为$n+m-2$.

17. 设$X_1,\ldots,X_n$为来自指数分布$E(\lambda)$的iid样本，记$T=2\lambda\sum_{i=1}^n X_i$，则证明$T$服从$\chi^2(2n)$.
解：因$X_1,\ldots,X_n$服从$E(\lambda)=\Gamma(1,\lambda)$,由$\Gamma$ 分布可加性，$\sum_{i=1}^n X_i$服从$\Gamma(n,\lambda)$,又$2\lambda\sum_{i=1}^n X_i$服从$\Gamma(n,\frac{1}{2})$ 即为自由度为$2n$的卡方分布。

18. 设随机变量$X$服从自由度为$n,m$的$F$分布，试求$Y=1/X$的分布，并试明

$$Z=\frac{nX}{m+nX}\sim\beta(n/2,m/2).$$

解：记$X=\frac{\xi/n}{\eta/m}$,则$Z=\frac{\xi}{\xi+\eta}$.由卡方分布与$\Gamma$分布关系知，$\xi,\eta$分别服从自由度为$n/2,m/2$的$\Gamma$分布，又由$\Gamma$ 分布和 $\beta$分布的关系知,$Z$服从自由度为$(\frac{n}{2},\frac{m}{2})$的$\beta$分布。

19. 设$X_1\sim\beta(a_1,b_1)$, $X_2\sim\beta(a_2,b_2)$, $a_1=a_2+b_2$，且$X_1$与$X_2$独立，则$Y=X_1X_2\sim\beta(a_2,b_1+b_2)$.
解：因$X_1$服从$\beta(a_1,b_1)$,$X_2$服从$\beta(a_2,b_2)$，而$X_1$ 与$X_2$ 独立，则$X_1,X_2$的联合密度函数为

$$f_1(x,y)=\Gamma(a_1+b_1)\Gamma(a_1)\Gamma(b_1)\frac{\Gamma(a_2+b_2)}{\Gamma(a_2)\Gamma(b_2)}x^{a_1-1}(1-x)^{b_2-1}y^{a_2-1}(1-y)^{b_2-1}$$

令$U=X_1X_2$,则$X_1$与$U$的联合密度函数为

$$
\begin{aligned}
f_2(x,u) &= \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\frac{\Gamma(a_2+b_2)}{\Gamma(a_2)\Gamma(b_2)}x^{a_1-1}(1-x)^{b_1-1}(\frac{u}{x})^{a_2-1}(1-\frac{u}{x})^{b_2-1}|J| \\
&= \frac{\Gamma(a_1+b_1)}{\Gamma(a_2)\Gamma(b_2)}x^{a_1-1}(1-x)^{b_1-1}(\frac{u}{x})^{a_2-1}(1-\frac{u}{x})^{b_2-1}\frac{1}{x} \\
&= \frac{\Gamma(b_1+b_2)}{\Gamma(b_1)\Gamma(b_2)}\frac{\Gamma(a_2+b_1+b_2)}{\Gamma(b_1+b_2)\Gamma(a_2)}(1-x)^{b_1-1}(x-u)^{b_2-1}u^{a_2-1} \\
&= \frac{\Gamma(b_1+b_2)}{\Gamma(b_1)\Gamma(b_2)}\frac{1}{1-u}(\frac{1-x}{1-u})^{b_1-1}(\frac{x-u}{1-u})^{b_2-1} \\
&\times \frac{\Gamma(a_2+b_1+b_2)}{\Gamma(b_1+b_2)\Gamma(a_2)}u^{a_2-1}(1-u)^{b_1+b_2-1}
\end{aligned}
$$

则$U$的边缘密度函数为(令$t=\frac{1-x}{1-u}$)

$$
\begin{aligned}
f_3(u) &= \int_u^1 \frac{\Gamma(b_1+b_2)}{\Gamma(b_1)\Gamma(b_2)}\frac{1}{1-u}(\frac{1-x}{1-u})^{b_1-1}(\frac{x-u}{1-u})^{b_2-1}dx \\
&\times \frac{\Gamma(a_2+b_1+b_2)}{\Gamma(a_2)\Gamma(b_1+b_2)}u^{a_2-1}(1-u)^{b_1+b_2-1} \\
&= \int_0^1 \frac{\Gamma(b_1+b_2)}{\Gamma(b_1)\Gamma(b_2)}t^{b_1-1}(1-t)^{b_2-1}dt \\
&\times \frac{\Gamma(a_2+b_1+b_2)}{\Gamma(a_2)\Gamma(b_1+b_2)}u^{a_2-1}(1-u)^{b_1+b_2-1} \\
&= \frac{\Gamma(a_2+b_1+b_2)}{\Gamma(a_2)\Gamma(b_1+b_2)}u^{a_2-1}(1-u)^{b_1+b_2-1}
\end{aligned}
$$

即$U=X_1X_2$服从$\beta(a_2,b_1+b_2)$, 得证。

20. 设$X\sim N(0,1),T\sim t(n)$，则存在一个正数$t_0$，使得

$$P\{|T|\geq t_0\}\geq P\{|X|\geq t_0\}.$$

解：标准正态的分布密度函数是$f_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, 自由度为$n$ 的分布密度函数为

$$f_2(x)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$$

由于$f_1(x)f_2(x)$ 关于y轴对称，故

$$\mathrm{P}\{|T|\geq t_0\}\geq \mathrm{P}\{|X|\geq t_0\}\Longleftrightarrow \mathrm{P}(T\geq t_0)\geq \mathrm{P}(X\geq t_0)$$

而易知$\frac{f_2(x)}{f_1(x)}\to\infty, x\to\infty$,故存在$x_0>0$当$x>x_0$时$f_2(x)>f_1(x)$, 取$t=x_0$,则有$\mathrm{P}\{T\geq t_0\}\geq \mathrm{P}\{X\geq t_0\}$.

21. 检验下列分布族是否是指数型分布族:

(1) Poisson分布族；

(2) 双参数指数分布族;

(3) $\Gamma$分布族$\Gamma(\alpha,\lambda)$;

(4) Cauchy分布族：$\left\{f(x,\lambda)=\frac{\lambda}{\pi(\lambda^2+x^2)}:\lambda>0\right\}$.

解：(1)Poisson分布族$f(x,\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}=e^{-\lambda}e^{x\ln\lambda}\frac{1}{x!}$, 为指数分布族。

(2)双参数分布族$f(x,u,\sigma)=\frac{1}{\sigma}\exp(-\frac{x-u}{\sigma}), x>u$, 支撑集与未知参数有关，不是指数分布族。

(3)$\Gamma$分布族$\Gamma(\alpha,\lambda)$,$f(x,\alpha,\lambda)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}=e^{(\alpha-1)\ln x-\lambda x}\frac{\lambda^\alpha}{\Gamma(\alpha)}$ 为指数分布族。

(4)Cauchy分布族$f(x,\lambda)=\frac{\lambda}{\pi(\lambda^2+x^2)}$ 不是指数分布族。

22. 设$X_1,\ldots,X_n$为来自PDF为

$$f(x,\theta)=\theta^{-1}\exp\{-|x|/\theta\},\ x\in R,\theta>0$$

的总体的iid样本，证明$T=\sum_{i=1}^n|X_i|$是$\theta$的充分统计量.

解：因为$X_1,\ldots,X_n$是独立的，故联合密度函数为

$$f(x,\theta)=\theta^{-n}\exp(-\sum_{i=1}^n|x_i|/\theta),$$

由因子分解定理知，$T=\sum_{i=1}^n|X_i|$ 为$\theta$的充分统计量。

23. 设$X_1,\ldots,X_n$是来自均匀分布族$\{U(\theta_1,\theta_2):-\infty<\theta_1<\theta_2\infty\}$的iid样本，证明：$(X_{(1)},X_{(n)})$ 是 $(\theta_1,\theta_2)$ 的充分统计量.

解：因为$X_1,\ldots,X_n$是独立的，故联合密度函数为$f(X,\theta_1,\theta_2)=(\frac{1}{\theta_2-\theta_1})^n\mathrm{I}\{\theta_1<X_{(1)}<X_{(n)}<\theta\}$, 由因子分解定理，$(X_{(1)},X_{(n)})$为 $(\theta_1,\theta_2)$充分统计量。

24. 设$Y_i\sim N(\alpha+\beta x_i,\sigma^2)$, $i=1,\cdots,n$, 且诸$Y_i$相互独立，其中$\alpha,\beta\in R$, $\sigma>0$是未知参数，$x_1,\ldots,x_n$为已知的常数. 试证明统计量

$$\left(\sum_{i=1}^n Y_i,\sum_{i=1}^n x_iY_i,\sum_{i=1}^n Y_i^2\right)$$

是$(\alpha,\beta,\sigma^2)$的充分统计量.

解：因为$Y_1,\ldots,Y_n$是独立的，故联合密度函数为

$$\begin{aligned}f(Y,\alpha,\beta,\sigma^2)&=(\frac{1}{\sqrt{2\pi\sigma^2}})^{2n}\exp(-\frac{\sum_{i=1}^n(Y_i-(\alpha+\beta X_i))^2}{2\sigma^2})\\&=(\frac{1}{\sqrt{2\pi\sigma^2}})^{2n}\exp(-\frac{\sum_{i=1}^nY_i^2+\beta^2\sum_{i=1}^nX_i^2+2\alpha\beta\sum_{i=1}^nX_i}{2\sigma^2})\\&\times\exp(-\frac{-2\alpha\sum_{i=1}^nY_i-2\beta\sum_{i=1}^nX_iY_i+n\alpha^2}{2\sigma^2})\end{aligned}$$

由因子分解定理，$(\sum_{i=1}^n Y_i, \sum_{i=1}^n X_iY_i, \sum_{i=1}^n Y_i^2)$ 为$(\alpha,\beta,\sigma^2)$的充分统计量。

25. 设$X_1,\ldots,X_n$为来自PDF为

$$f(x,\theta)=\theta/x^2,\ 0<\theta<x<\infty$$

的总体的iid样本，试证明$X_{(1)}$是充分统计量.

解：易知样本联合密度函数为，

$$f(X,\theta)=\prod_{i=1}^n \frac{\theta}{X_i^2}\prod_{j=1}^n \mathrm{I}\{\theta<X_j\}=\frac{\theta^n}{\prod_{i=1}^n X_i^2}\mathrm{I}\{\theta<X_{(1)}\}$$

, 由因子分解定理知 $X_{(1)}$为$\theta$的充分统计量。

26. 设$X_1,\ldots,X_n$为来自两点分布$b(1,p)$的iid样本，试写出$X_1,\ldots,X_n$的联合分布，并指出$X_1+X_2, X_{(n)}, X_n+2p, (X_n-X_1)^2$中哪些是统计量，为什么？

解：$X_1+X_2, X_{(n)}, (X_n-X_1)^2$ 为统计量，而$X_n+2p$依赖于未知参数$p$，不是统计量.

27. 设$X_1,\ldots,X_n$为来自概率密度函数为

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

的总体的iid样本，试求$X_{(1)}, X_{(n)}$的PDF.

解：易知此时分布函数为$F(x)=x^2$,而

$$\mathrm{P}(X_{(1)}\le x)=1-\mathrm{P}(X_{(1)}>x)=1-\prod_{i=1}^n \mathrm{P}(X_i>x)=1-\prod_{i=1}^n(1-x^2)=1-(1-x^2)^n$$

故$X_{(1)}$的pdf为$f_1(x)=(1-(1-x^2)^2)'=2nx(1-x^2)^{n-1}$. 又

$$\mathrm{P}(X_{(n)}<x)=\prod_{i=1}^n \mathrm{P}(X_i\le x)=x^{2n}$$

,故$X_n$ 的pdf为$f_n(x)=(x^{2n})'=2nx^{2n-1}$.

28. 设$X_1,\ldots,X_n$为来自具有PDF为$f(x)$的总体的iid样本，试证明: 样本极差$R=X_{(n)}-X_{(1)}$ 的 PDF为

$$f_R(x)=n(n-1)\int_{-\infty}^{\infty}[F(t+x)-F(t)]^{n-2}f(x+t)f(t)\mathrm{d}t,$$

其中$F$为总体的CDF.

解：由第10题知，$X_{(1)}, X_{(n)}$的联合pdf 为

$$f_{1,n}(x,y)=n(n-1)(F(y)-F(x))^{n-2}f(x)f(y), x<y.$$

令$R = X_{(n)} - X_{(1)}, S = X_{(1)} + X_{(n)}$,则$X_{(1)} = \frac{S-R}{2}, X_{(n)} = \frac{S+R}{2}$ 于是$|J| = \frac{1}{2}$. 从而，R,S的联合pdf为

$$f_{R,S}(u,v) = \frac{1}{2}n(n-1)(F(\frac{u+v}{2}) - F(\frac{v-u}{2}))^{n-2}f(\frac{v-u}{2})f(\frac{v+u}{2})$$

故$R$的边缘pdf为（令$t = \frac{v-u}{2}$）

$$\begin{aligned} f_R(u) &= \int_{-\infty}^{\infty} \frac{1}{2}n(n-1)(F(\frac{u+v}{2}) - F(\frac{v-u}{2}))^{n-2}f(\frac{v-u}{2})f(\frac{v+u}{2})dv \\ &= \int_{-\infty}^{\infty} n(n-1)(F(t+u) - F(t))^{n-2}f(u+t)f(t)dt \end{aligned}$$

则$f_R(x) = n(n-1)\int_{-\infty}^{\infty}(F(t+x) - F(t))^{n-2}f(x+t)f(t)dt$

29. 设$X_1, \ldots, X_m$为来自$N(\mu, \sigma_1^2)$的iid样本，$Y_1, \ldots, Y_n$为来自$N(\mu, \sigma_2^2)$ 的iid样本，且全样本独立，记$\alpha_1 = S_{1m}^2/(S_{1m}^2 + S_{2n}^2), \alpha_2 = S_{2n}^2/(S_{1m}^2 + S_{2n}^2)$. 试求$Z = \alpha_1\bar{X} + \alpha_2\bar{Y}$的期望.

解：因为量总体都是正态分布且样本独立，故$\bar{X}$ 与$S_{1m}^2, S_{2n}^2$ 独立,$\bar{Y}$ 与$S_{1m}^2, S_{2n}^2$独立，从而$\bar{X}$与$\alpha_1 = S_{1m}^2/(S_{1m}^2 + S_{2n}^2)$,$\bar{Y}$ 与 $\alpha_2 = S_{2n}^2/(S_{1m}^2 + S_{2n}^2)$ 独立，故$E(Y) = E(\alpha_1)E(\bar{X}) + E(\alpha_2)E(\bar{Y}) = \mu E(\alpha_1 + \alpha_2) = \mu$.

30. 设$X \sim B(n,p)$, 对$k \geq 1$，请证明

$$P\{X \geq k\} = \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} = kC_n^k \int_0^p t^{k-1}(1-t)^{n-k}\mathrm{d}t.$$

解：因为$X$服从二项分布B$(n,p)$,所以$P(X \geq k) = \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i}$, 记$f(p) = \sum_{i=1}^{n} C_n^i p^i (1-p)^{n-i}, g(p) = kC_n^k \int_0^p t^{k-1}(1-t)^{n-k}dt$, 则

$$\begin{aligned} & f'(p) = \sum_{i=k}^{n}[iC_n^i p^{i-1}(1-p)^{n-i} - (n-i)C_n^i p^i (1-p)^{n-i-1} \\ = \; & kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k+1}^{n} iC_n^i p^{i-1}(1-p)^{n-i} - \sum_{i=k}^{n}(n-i)C_n^i p^i (1-p)^{n-i-1} \\ = \; & kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k}^{n-1}(i+1)C_n^{i+1} p^i (1-p)^{n-i-1} - \sum_{i=k}^{n-1}(n-i)C_n^i p^i (1-p)^{n-i-1} \\ = \; & kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k}^{n-1}[(i+1)C_n^{i+1} - (n-i)C_n^i p^i (1-p)^{n-i-1} \\ = \; & kC_n^k p^{k-1}(1-p)^{n-k} = g'(p) \end{aligned}$$

且$f(0)=g(0)$,故$f(p)=g(p)$,即

$$\sum_{i=k}^{n} C_n^k p^i (1-p)^{n-i} = kC_n^k \int_0^p t^{k-1}(1-t)^{n-k}dt.$$

31. 设$X_1,\ldots,X_n$为来自总体分布为$F$的iid样本， $X_{(1)} \le \cdots \le X_{(n)}$为其次序统计量. 如果$F(x)$连续，则

$$E[F(X_{(i)} = \frac{i}{n+1}, \quad \mathrm{Var}[F(X_{(i)} = \frac{i(n+1-i)}{(n+1)^2(n+2)}.$$

解：易知$X_{(i)}$的pdf为$f_i(x) = iC_n^i F^{i-1}(x)(1-F(x))^{n-i}f(x)$, 故

$$\begin{aligned}
E(F(X_{(1)})) &= \int_{-\infty}^{\infty} F(x) iC_n^i F^{i-1}(x)(1-F(x))^{n-i} dF(x) \\
&= iC_n^i \int_{-\infty}^{\infty} F^i(x)(1-F(x))^{n-i} dF(x) \\
&= iC_n^i \int_0^1 t^i (1-t)^{n-i} dt \\
&= iC_n^i \beta(i+1, n-i+1) = \frac{i}{n+1}
\end{aligned}$$

而

$$\begin{aligned}
E(F^2(X_{(i)})) &= \int_{-\infty}^{\infty} F^2(x) iC_n^i F^{i-1}(x)(1-F(x))^{n-i} dF(x) \\
&= iC_n^i \int_{-\infty}^{\infty} F^{i+1}(x)(1-F(x))^{n-i} dF(x) \\
&= iC_n^i \int_0^1 t^{i+1}(1-t)^{n-i} dt \\
&= iC_n^i \beta(i+2, n-i+1) = \frac{i(i+1)}{(n+2)(n+1)}
\end{aligned}$$

从而，$\mathrm{Var}(F(X_{(i)})) = E(F(X_{(i)})^2) - (E(F(X_{(i)})))^2 = \frac{i(n+1-i)}{(n+1)^2(n+2)}$.

32. 设$X_1,\ldots,X_n$是$n$个相互独立的随机变量，且$X_i \sim N(0,\sigma_i^2)$. 记

$$Y = \frac{\sum_{i=1}^n X_i/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}, \quad Z = \sum_{i=1}^n \frac{(X_i - Y)^2}{\sigma_i^2},$$

试证明$Z \sim \chi^2(n-1)$.

解： 由$Z = \sum_{i=1}^n \frac{(X_i-Y)^2}{\sigma_i^2} = \sum_{i=1}^n (\frac{X_i}{\sigma_i})^2 - (2Y\sum_{i=1}^n \frac{X_i^2}{\sigma_i^2} - Y^2 \sum_{i=1}^n \frac{1}{\sigma_i^2})$ 做正交变

换$(Y_1,\ldots,Y_n)'=A(\frac{X_1}{\sigma_1},\ldots,\frac{X_n}{\sigma_n})'$, 其中$A$ 的第一行为

$$(\frac{1/\sigma_1}{\sqrt{\sum_{i=1}^n 1/\sigma_i^2}},\frac{1/\sigma_2}{\sqrt{\sum_{i=1}^n 1/\sigma_i^2}},\ldots,\frac{1/\sigma_n}{\sqrt{\sum_{i=1}^n 1/\sigma_i^2}}).$$

因为$X_i$服从正态分布$N(0,\sigma_i^2)$,$\frac{X_i}{\sigma_i}$ 服从标准正态分布。而根据正交变换的性质,$Y_1,\ldots,Y_n$独立且服从标准正态分布且$\sum_{i=1}^n Y_i^2=\sum_{i=1}^n(\frac{X_i}{\sigma_i})^2$, 又

$$Y_1^2=(\sum_{i=1}^n\frac{X_i/\sigma_i^2}{\sqrt{\sum_{i=1}^n\frac{1}{\sigma_i^2}}})^2=2Y\sum_{i=1}^n\frac{X_i}{\sigma_i^2}-Y^2\sum_{i=1}^n\frac{1}{\sigma_i^2}$$

从而$Z=\sum_{i=2}^n Y_i^2$服从自由度为$n-1$ 的卡方分布。

33. 设$X_1,\ldots,X_n$为来自正态总体$N(\mu,\sigma^2)$的iid样本，记

$$\tau=\frac{X_1-\bar{X}}{S}\sqrt{\frac{n}{n-1}},$$

试证明$t=\frac{\tau\sqrt{n-2}}{\sqrt{n-1-\tau^2}}\sim t(n-2)$.

解：由$\tau=\frac{X_1-\bar{X}}{S}\sqrt{\frac{n}{n-1}}$, 故

$$\begin{aligned}t=\frac{\tau\sqrt{n-2}}{n-1-\tau^2}&=\frac{\frac{X_1-\bar{X}}{S}\sqrt{\frac{n}{n-1}}\sqrt{n-2}}{n-1-\frac{n}{n-1}(\frac{X_1-\bar{X}}{S})^2}\\&=\frac{(X_1-\bar{X})\sqrt{n(n-2)/(n-1)}}{\sqrt{[(n-1)^2S^2-n(X_1-\bar{X})^2/(n-1)}}\end{aligned}$$

又

$$\begin{aligned}&\frac{1}{n-1}((n-1)^2S^2-n(X_1-\bar{X})^2)=\frac{1}{n-1}((n-1)\sum_{i=1}^n(X_i-\bar{X})^2-n(X_1-\bar{X})^2)\\=&\frac{1}{n-1}((n-1)(\sum_{i=1}^n X_i^2-n\bar{X}^2)-n(X_1^2-2X_1\bar{X}+\bar{X}^2))\\=&\frac{1}{n-1}((n-1)\sum_{i=2}^n X_i^2-X_1^2+(X_1+\ldots+X_n)(X_1-\ldots-X_n))\\=&\frac{1}{n-1}((n-1)\sum_{i=2}^n X_i^2-(n-1)^2\bar{X}_{n-1}^2)\\=&\sum_{i=2}^n(X_i-\bar{X}_{n-1})^2\end{aligned}$$

其中$\bar{X}_{n-1}=\frac{1}{n-1}\sum_{i=2}^n X_i$. 由$X_2,\ldots,X_n$独立服从$N(\mu,\sigma^2)$,故$\frac{\sum_{i=2}^n(X_i-\bar{X})^2}{\sigma^2}$ 服从自由度为 $n-2$的卡方分布，又$X_1,\ldots,X_n$独立服从$N(\mu,\sigma^2)$,故$\frac{X_1-\bar{X}}{\sqrt{n/(n-1)}}$服从标准正态分布，而且,$\sum_{i=2}^n(X_i-\bar{X}_{n-1})^2$ 与 $\bar{X}_{n-1}$,$\sum_{i=2}^n(X_i-\bar{X}_{n-1})^2$ 与$X_1$分别独立，有$\sum_{i=2}^n(X_i-\bar{X}_{n-1})^2$ 与$X_1-\bar{X}=\frac{n-1}{n}X_1-\frac{n-1}{n}\bar{X}_{n-1}$ 独立. 从而，

$$t=(\frac{X_1-\bar{X})/\sqrt{\frac{n-1}{n}}\sigma}{\sqrt{[(n-1)^2S^2-n(X_1-\bar{X})^2/(n-1(n-2))}}$$

服从自由度为$n-2$的t分布。

34. 设$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix},\cdots,\begin{pmatrix} X_n \\ Y_n \end{pmatrix}$为来自二维正态$N(\mu,\Sigma)$的iid样本，其中 $\mu=(\mu_1,\mu_2)^{\mathrm{T}}$, $\Sigma=\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. 定义二样本间的相关系数为

$$r(X,Y)=\frac{\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\sum_{i=1}^n(X_i-\bar{X})^2\sum_{i=1}^n(Y_i-\bar{Y})^2\right]^{1/2}},$$

试证明：如果$\rho=0$, 则统计量

$$t=\sqrt{n-2}\frac{r(X,Y)}{\sqrt{1-r^2(X,Y)}}\sim t(n-2).$$

解：先给定$Y_i,i=1,\ldots,n$，令$Z_i=\frac{Y_i-\bar{Y}}{\sqrt{\sum_{i=1}^n(Y_i-\bar{Y})^2}}$, $Z=(Z_1,\ldots,Z_n)'$, 易知

$$\sum_{i=1}^n(X_i-\bar{X})Z_i=0$$

,故$r(X,Y)=\frac{\sum_{i=1}^n X_iZ_i}{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2}}$ 带入有

$$t=\frac{\sqrt{n-2}\sum_{i=1}^n X_iZ_i}{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2-(\sum_{i=1}^n X_iZ_i)^2}},$$

即$t=\frac{\sqrt{n-2}Z'X}{\sqrt{X'(I-\frac{1}{n}11')X-X'ZZ'X}}$, 记$A=I-\frac{1}{n}11'-ZZ'$ 则易验证$A^2=A,A'=A,Z'A=0,\mathrm{rank}(A)=n-2$, 则$Z'X$ 与$X'AX$独立，且$X'AX/\sigma_1^2$服从自由度为$n-2$的卡方分布，又$Z'X/\sigma_1$服从标准正态分布，故$t=\frac{\sqrt{n-2}Z'X}{\sqrt{X'AX}}$ 服从自由度为$n-2$的t分布，由于该结论对任意的$i=1,\ldots,n$均成立故$t$服从自由度为$n-2$的t分布。

35. 证明服从非中心$t$分布的随机变量的平方服从非中心$F$分布.

解：取$\xi$服从正态分布$N(\mu,\sigma^2)$, $\eta/\sigma^2$服从自由度为$n$的卡方分布，且二者独立，$\mu\neq$

0，则$T=\frac{\xi}{\sqrt{\eta/n}}$服从非中心t 分布$t(n,\delta)$,其中$\delta=\mu/\sigma$ 则$T^2=\frac{\xi^2}{\eta/n}$服从非中心F分布$F(1,n,\delta)$即非中心t分布随机变量的平方服从非中心F分布。

36. 一个总体有$N$个元素，其指标值分别为$a_1>a_2>\cdots>a_N$. 指定自然数$M<N,n<N$，并设$m=nM/N$为整数. 在$(a_1,\ldots,a_M)$中不放回的抽取$m$个，在$(a_{M+1},\ldots,a_N)$ 中不放回的抽取$n-m$个. 写出所得样本的分布.
解：样本分布密度函数为

$$f(X_{i_1},\ldots,X_{i_n})=\begin{cases}\frac{C_M^mC_{N-M}^{n-m}}{C_N^M} & \#\{k,i_k\le M\}=m\\ 0 & 否则\end{cases}$$

37. 设$X_1,\ldots,X_n$为来自区间$(a,b)$上均匀分布的IID样本， $-\infty<a<b<\infty$. 试证对于任何 $a$ 和 $b$，$(X_{(i)}-X_{(1)})/(X_{(n)}-X_{(1)}),i=2,\ldots,n-1$ 是相互独立的.、、
解：

38. 设$X_1,\ldots,X_n$为来自总体$P_\theta\in\{P_\theta:\theta\in\Theta\}$的IID样本. 在下列的几种情况下，找出一个与$\theta\in\Theta$相同维数的充分统计量.
(1) $P_\theta$是泊松分布$P(\theta)\theta\in(0,\infty)$;
(2) $P_\theta$是负二项分布$NB(\theta,r)$，且$r$已知，$\theta\in(0,1)$;
(3) $P_\theta$是指数分布$E(\theta),\theta\in(0,\infty)$.
解：(1)样本均值$\bar{X}$
(2)最小次序统计量$X_{(1)}$
(3)样本均值$\bar{X}$.

39. 设$X_1,\ldots,X_n$为来自$N(\theta,\theta^2)(\theta>0)$的iid样本. 问$\bar{X}$ 是否仍为充分统计量？
解：样本的联合密度函数为

$$f(X,\theta)=\frac{1}{(2\pi\theta^2)^{1/2}}\exp(-\frac{\sum_{i=1}^n X_i^2}{2\theta^2}+\frac{\sum_{i=1}^n X_i}{\theta}-\frac{n}{2})$$

. 由因子分解定理知$(\sum_{i=1}^n X_i^2,\sum_{i=1}^n X_i)$ 为$\theta$的充分统计量。而易知$\sum_{i=1}^n X_i^2$不能由$\sum_{i=1}^n X_i$ 表出，故$\bar{X}=\sum_{i=1}^n X_i$ 不是$\theta$的充分统计量。

40. 设 $X_1,\ldots,X_n\sim N(0,\sigma^2)$, $\bar{X}$ 为样本均值， $\xi=f(X_1,\ldots,X_n)$ 满足条件$f(X_1+c,\ldots,X_n+c)=f(X_1,\ldots,X_n)$ (对任何常数$c$). 证明：$\xi$与$\bar{X}$独立.
解：令$Y_i=X_i+c,i=1,\ldots,n$,则$Y_i$来自正态分布$N(c,\sigma^2)$.易知样本均值$\bar{X}$ 是未知参数$c$的充分完备统计量，而由条件$f(X_1,\ldots,X_n)=f(X_1+c,\ldots,X_n+c)$ 知，统计量$\xi$关于未知参数$c$是无关的，故由Basu定理$\bar{X}$和$\xi$是独立的。

41. 非中心$\chi^2$变量$\xi=\sum_{i=1}^n(X_i+a_i)^2$， $X_1,\ldots,X_n\sim N(0,1)$, $a_1,\ldots,a_n$为常数. 证明：$\xi$的分布只依赖于$n$和$\delta=\sqrt{\sum_{i=1}^n a_i^2}$ (提示：作正交变换，使$Y_1=$

$\sum_{j=1}^n a_j X_j/\delta)$.
解：

## 2 习题二

1. 设总体$X$的CDF为

$$P\{X=x\}=p(1-p)^{x-1},\ x=1,2,\cdots,$$

$X_1,\ldots,X_n$为来自此总体的iid样本，试求 $p$ 的矩估计和 MLE.
解: 几何分布期望为 $1/p$, 所以 $p$ 的矩估计 $\hat{p}_{\mathrm{M}}$ (下同) 为 $1/\bar{X}$. 对数似然为 $l=n\log p+\sum(X_i-1)\log(1-p)$, 易得其 MLE 为 $\hat{p}_{\mathrm{ML}}$ 为 $1/\bar{X}$.

2. 设$X_1,\ldots,X_n$为来自总体分布

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

的iid样本，试求$\alpha$的矩估计和MLE.
解: $X$ 来自 $\beta(\alpha+1,1)$, 其期望为 $(\alpha+1)/(\alpha+2)$, 则 $\hat{\alpha}_{\mathrm{M}}=(2\bar{X}-1)/(1-\bar{X})$. 对数似然为 $l=n\log(\alpha+1)+\alpha\sum\log X_i$, 易得其 MLE 为 $\hat{\alpha}_{\mathrm{ML}}=-\frac{n}{\sum\log X_i}-1$.

3. 设$X_1,\ldots,X_n$为来自总体$X$的iid样本，$n$个正常数$\alpha_i$满足$\sum_{i=1}^n\alpha_i=1$. 试证在$EX$的所有形如$\sum_{i=1}^n\alpha_i X_i$的UE中，以$\bar{X}$为最优. (最优的标准为方差最小)
证: 问题可转化为 $\min\sum\alpha_i^2$ s.t. $\sum\alpha_i=1$. 显然最优解为 $\alpha_i=1/n$, 即 $\bar{X}$ 为最优.

4. 设$X_1,\ldots,X_n$为来自双指数分布

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

的iid样本. 试求$\theta,\lambda$的MLE.
解: 对数似然为 $l=(-n\log\lambda-\frac{\sum X_i-n\theta}{\lambda})\mathbf{1}_{\{X_{(1)}\geq\theta\}}$. $l$ 关于 $\theta$ 单增, 则 $\hat{\theta}_{\mathrm{ML}}=X_{(1)}$, $\hat{\lambda}_{\mathrm{ML}}=\bar{X}-X_{(1)}$.

5. 设$X_1,\ldots,X_n$为来自总体$N(\mu,\sigma^2)$的iid样本，求满足

$$P\{X>a\}=\int_a^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=0.05$$

的点$a$的MLE.
解: 易知 $\frac{a-\mu}{\sigma}=\Phi^{-1}(0.95)=1.64$, 由 MLE 的不变性知: $\hat{a}_{\mathrm{ML}}=\bar{X}+1.64S_n^*$.

6. 设$X_1,\ldots,X_n$为来自均匀分布$U(a,b)$的iid样本，试求$a^2$的MLE.

解: 似然为 $L=(b-a)^{-n}\mathbf{1}_{\{a<X_{(1)}<X_{(n)}<b\}}$. 显然 $\hat{a}_{\mathrm{ML}}=X_{(1)}$, 从而 $\hat{a}^2_{\mathrm{ML}}=X^2_{(1)}$.

7. 设某种产品的寿命$X$服从指数分布$E(\lambda)$, $t_0$为给定常数，$X_1,\ldots,X_n$是独立观测到的$n$件产品的寿命. 求产品在$t_0$前失效的概率$P\{X\le t_0\}=1-e^{-\lambda t_0}=g(\lambda)$的UMVUE.

解: 显然 $\mathbf{1}_{\{X_1\le t_0\}}$ 是 $g(\lambda)$ 的一个 UE, 而 $\bar{X}$ 为 $\lambda$ 的充分完备统计量, 因而所求 UMVUE 为

$$\mathrm{E}(\mathbf{1}_{\{X_1\le t_0\}}|\bar{X})=\mathrm{P}(X_1\le t_0|\bar{X})=\mathrm{P}(X_1/\bar{X}\le t_0/\bar{x}|\bar{X}).$$

由 Basu 定理知, $X_1/\bar{X}$ 与 $\bar{X}$ 独立. 又 $X_1/\bar{X}\sim\beta(1,n-1)$, 则

$$\mathrm{E}(\mathbf{1}_{\{X_1\le t_0\}}|\bar{X})=\mathrm{P}(X_1/\bar{X}\le t_0/\bar{x})=1-(1-\frac{t_0}{n\bar{X}})^{n-1},\quad t_0<\bar{X}.$$

8. 设$X_1,\ldots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中$\theta>0$为未知参数.

(1) 试证明$\hat{\theta}_1=\frac{n+1}{n}X_{(n)}$, $\hat{\theta}_2=(n+1)X_{(1)}$均是$\theta$的UE.

(2) 上述两个估计中哪个方差最小?

(3) 试证明$\hat{\theta}_1$是$\theta$的UMVUE.

解: 因为 $X/\theta\sim U(0,1)$, 所以 $X_{(1)}/\theta\sim\beta(1,n)$, $X_{(n)}/\theta\sim\beta(n,1)$. 而 $\beta(a,b)$ 分布的期望、方差分别为 $\frac{a}{a+b}$ 和 $\frac{ab}{(a+b+1)(a+b)^2}$. 易证 (1), 且易算得 (2) 中方差分别为 $\mathrm{Var}(\hat{\theta}_1)=\frac{\theta^2}{n(n+2)}$ 和 $\mathrm{Var}(\hat{\theta}_2)=\frac{n\theta^2}{n+2}$. 证明 (3) 时, 只需注意到 $X_{(n)}$ 为 $\theta$ 的充分完备统计量即可.

9. 设$X_1,\ldots,X_n$为来自$N(0,\sigma^2)$的iid样本，试求$\sigma^2$的充分完备统计量，$\sigma$和$3\sigma^4$的UMVUE，并证明$\frac{1}{n}\sum_{i=1}^n X_i^2$是$\sigma^2$的有效估计和相合估计.

解: 显然 $S_n^2\doteq\frac{1}{n}\sum X_i^2$ 是 $\sigma^2$ 的充分完备统计量. 因为 $\frac{nS_n^2}{\sigma^2}\sim\chi^2_n$, 对于 $k>0$, 计算

$$\begin{aligned}\mathrm{E}(\frac{nS_n^2}{\sigma^2})^k &= \frac{1}{\Gamma(n/2)2^{n/2}}\int_0^\infty x^k x^{n/2-1}e^{-x/2}\mathrm{d}x\\ &= \frac{2^k\Gamma(k+n/2)}{\Gamma(n/2)}.\end{aligned}$$

取 $k=1/2$, $k=2$ 分别得到如下估计:

$$\begin{aligned}\hat{\sigma} &= \frac{\sqrt{n/2}\Gamma(n/2)}{\Gamma((n+1)/2)}S_n,\\ \widehat{3\sigma^4} &= \frac{n}{n+2}S_n^4.\end{aligned}$$

易知 $S_n^2$ 是 $\sigma^2$ 的 UE 且 $\mathrm{Var}S_n^2=\frac{2\sigma^4}{n}$. 另一方面 $I(\sigma)=\frac{2n}{\sigma^2}$, 则 C-R 下界为 $\frac{2\sigma^4}{n}$. 有效估计得证. 而相合性可以简单的由大数定律得证.

10. 设$a_1,\ldots,a_n$是$n$个实数，定义函数$h(a)=\sum_{i=1}^n|a_i-a|$.
(1) 证明当$a$为$a_1,\ldots,a_n$的样本中位数时，$h(a)$达到最小值；
(2) 设$X_1,\ldots,X_n$为来自具有PDF为$e^{-|x-\theta|}/2$的总体(此分布称为Laplace分布)的iid样本，求参数$\theta$ 的矩估计与MLE.
解: (1) 注意到 $h(a)$ 的导数为: $h'(a)=-\sum\mathrm{sgn}(a_i-a)$ 即可得证.
(2) $\hat\theta_{\mathrm{M}}=\bar X$, $\hat\theta_{\mathrm{ML}}=X_{(\mathrm{med})}$.

11. (1)证明:

$$f(x,a,\sigma)=(\sqrt{2\pi}\sigma^3)^{-1}(x-a)^2\exp\{-(x-a)^2/2\sigma^2\},\ x\in R$$

作为$x$的函数是一PDF，其中$a\in R,\sigma>0$为参数.
(2) 设$X_1,\ldots,X_n$为来自此上述总体的iid样本，求$a,\sigma^2$的矩估计.
(3) 给出$a,\sigma^2$的MLE所满足的方程，并给出一种迭代求解方法.
解: 令 $Y=\frac{X-a}{\sigma}$, 考虑 $k\geq 0$ 时,

$$\mathrm{E}Y^k=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^{k+2}e^{-\frac{y^2}{2}}\mathrm{d}y.$$

显然 $k$ 为奇数时, $\mathrm{E}Y^k=0$; 当 $k$ 为偶数时,

$$\begin{aligned}\mathrm{E}Y^k &= \sqrt{\frac{2}{\pi}}\int_0^{\infty}y^{k+2}e^{-\frac{y^2}{2}}\mathrm{d}y\\ &= \frac{2^{k/2+1}}{\sqrt{\pi}}\Gamma(\frac{k+3}{2}).\end{aligned}$$

(1) 取 $k=0$, $\mathrm{E}Y^0=1$, 得证.
(2) 取 $k=0,1$ 知: $\mathrm{E}Y=0$, $\mathrm{E}Y^2=3$, 则 $\mathrm{Var}Y=3$. 从而 $\mathrm{E}X=a$, 且 $\mathrm{Var}X=3\sigma^2$. 据此我们得到如下矩估计: $\hat a_{\mathrm{M}}=\bar X$ 和 $\hat\sigma^2_{\mathrm{M}}=S_n^{*2}/3$.
(3) 参考习题 47, 我们可以采用 Newton-Raphson 迭代算法或者 Fisher-scoring 迭代算法. 取初始估计为上述矩估计即可.

12. 设$X$为来自Poisson分布$P(\lambda)$的一个样本，参数$\lambda$的先验密度为$\pi(\lambda)=e^{-\lambda}$, $\lambda>0$(当$\lambda\leq 0$时，$\pi(\lambda)=0$). 试求$\lambda$的Bayes估计.
解: 略.

13. 设$X_1,\ldots,X_n$为来自均匀分布$U(0,\theta)$的iid样本. 证明
(1) $\hat\theta_1=X_{(n)}+X_{(1)}$是$\theta$的一个UE；
(2) 对适当选择的常数$c_n$，$\hat\theta_2=c_nX_{(1)}$是$\theta$的UE，但这个估计的方差比另外两个UE $\hat\theta_3=2\bar X$和$\hat\theta_4=\frac{n+1}{n}X_{(n)}$都大(除非$n=1$).
证: (1) 见 8 题 (1).

(2) 由 8 题 (1) 知 $c_n = n+1$.
(3) 参见 8 题 (2). 并且 $\mathrm{Var}(\hat{\theta}_3) = 4\mathrm{Var}(\bar{X}) = \frac{4}{n}\mathrm{Var}X_1 = \frac{\theta^2}{3n}$.

14. 设$X_1, \ldots, X_n$为来自PDF为

$$f(x,\theta) = \begin{cases} 2\sqrt{\theta/\pi}\exp\{-\theta x^2\}, & x > 0 \\ 0, & x \le 0, \end{cases}$$

的总体的iid样本，证明：对适当选取的常数$c$，$\hat{\theta} = c\sum_{i=1}^n X_i^2/n$是$1/\theta$的UMVUE.
证: 易证 $\sum X_i^2$ 为 $\theta$ 的充分完备统计量, 只需计算

$$\mathrm{E}X_1^2 = 2\sqrt{\theta/\pi}\int_0^\infty x^2 e^{-\theta x^2}\mathrm{d}x.$$

令 $\theta x^2 = y$, 则

$$\mathrm{E}X_1^2 = \frac{1}{\theta\sqrt{\pi}}\int_0^\infty y^{1/2}e^{-y}\mathrm{d}y = \frac{1}{2\theta}.$$

从而 $c = 2$, $\hat{\theta}$ 为 $1/\theta$ 的 UE.

15. (1)若$\hat{\theta}_1, \hat{\theta}_2$是$\theta$的UMVUE，则$(\hat{\theta}_1 + \hat{\theta}_2)/2$也是.

(2)如果$\hat{\theta}$是$\theta$的UMVUE，而$a \ne 0$和$b$都是已知常数，则$a\hat{\theta} + b$是$a\theta + b$的UMVUE.
证: 由定义易证.

16. 设$X_1, \ldots, X_n$为来自PDF为

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

的总体的iid样本.
(1)求$\theta$的矩估计$\hat{\theta}_M$;
(2)求$\theta$的MLE $\hat{\theta}_{ML}$;
(3)$\hat{\theta}_M$是否是$\theta$的相合估计？为什么？
解: (1)$X \sim E(\theta, 1)$, 所以 $\mathrm{E}X = \theta + 1$. 从而 $\hat{\theta}_{\mathrm{M}} = \bar{X} - 1$.
(2)似然为 $L = \exp - \sum(X_i - \theta)\mathbf{1}_{\{X_{(1)} > \theta\}}$. 从而 $\hat{\theta}_{\mathrm{ML}} = X_{(1)}$.
(3)由大数定律知: $\bar{X} \xrightarrow{p} \mathrm{E}X = \theta + 1$, 从而 $\hat{\theta}_{\mathrm{M}} = \bar{X} - 1 \xrightarrow{p} \theta$.

17. 设$X_1, \ldots, X_n$为来自总体

$$X \sim \begin{pmatrix} -1 & 0 & 2 \\ 2\theta & \theta & 1 - 3\theta \end{pmatrix}$$

的iid样本，其中$0 < \theta < 1/3$. 试求参数$\theta$的MLE.
解: 记样本中取值 -1, 0, 2 的个数分别为 $a, b, c$. 则似然为 $L = (2\theta)^a\theta^b(1-3\theta)^c$. 易知 $\hat{\theta}_{\mathrm{ML}} = \frac{a+b}{3n}$.

18. 在买面包作早点的男、女消费者中，男性购买者的比例$p$未知，但知道$\frac{1}{3} \le p \le \frac{1}{2}$. 设在70个购买者中发现12个是男性，58个是女性，试求$p$的MLE. 如果对$p$没有限制，试求$p$的MLE.
解: 假设男性购买者的个数为 $X$, 则 $X \sim B(n,p)$. 考虑对数似然

$$l(p) \propto X\log p + (n-X)\log(1-p).$$

若对 $p$ 没有限制, 则 $\hat{p}_{\mathrm{ML}} = X/n = 12/70$.
若限制 $1/3 \le p \le 1/2$, 而 $X/n = 12/70 < 1/3$, 所以 $\hat{p}_{\mathrm{ML}} = 1/3$.

19. 设$X_1, \ldots, X_n$为来自PDF为 $f(x,\theta) = \frac{1}{\sigma}\exp\{-(x-\mu)/\sigma\}, x \ge \mu$的iid样本，其中$\mu \in R, \sigma^2 > 0$为未知参数.
(1)求$\mu, \sigma^2$的MLE;
(2)对给定的$t > \mu$，求$P\{X_1 \ge t\}$的MLE.
解: (1)参见 4 题, $\hat{\mu}_{\mathrm{ML}} = X_{(1)}$, $\hat{\sigma}_{\mathrm{ML}} = \bar{X} - X_{(1)}$.
(2)$\mathrm{P}(X_1 > t) = e^{-\frac{t-\mu}{\sigma}}$, 由 MLE 的不变性知其 MLE 为 $e^{-\frac{t-\hat{\mu}_{\mathrm{ML}}}{\hat{\sigma}_{\mathrm{ML}}}}$.

20. 设$X_1, \ldots, X_n$为来自概率密度函数为 $f(x,\theta) = \frac{1}{\pi(1+(x-\theta)^2)}$的Cauchy分布的iid样本，试证：
(1)如果$n = 1$, 则$\theta$的MLE为$X_1$;
(2)如果$n = 2$，则$\theta$的MLE存在并唯一，但似然方程有多个根.
证: (1)若 $n = 1$, 要最大化 $L(\theta; x)$, 必须最小化 $(X_1 - \theta)^2$, 所以 $\hat{\theta}_{\mathrm{ML}} = X_1$.
(2)若 $n = 2$ 时, 似然方程为:

$$s(\theta) = 2\Big[\frac{X_1-\theta}{1+(X_1-\theta)^2} + \frac{X_2-\theta}{1+(X_2-\theta)^2}\Big] = 0.$$

此方程有 3 个根(包括复根). 进一步的, 当 $(X_1 - X_2)^2 < 4$ 时, 方程有一个实根 $\theta_1 = \frac{X_1+X_2}{2}$ 和两个复根, 而 $\frac{\partial^2 \ln L}{\partial\theta^2}\mid_{\theta=\theta_1} < 0$, 从而 $\hat{\theta}_{\mathrm{ML}} = \theta_1$. 当 $(X_1 - X_2)^2 > 4$ 时, 方程有三个实根 $\theta_1$ 和 $\theta_{2,3} = \frac{X_1+X_2\pm\sqrt{(X_1-X_2)^2-4}}{2}$. 易知, 此时 $\hat{\theta}_{\mathrm{ML}} = \theta_1$. 所以, 虽然 $n = 2$ 时, 似然方程有多个根, 但是 $\theta$ 的 MLE 存在并唯一.

22. 设$X_1, \ldots, X_n$为来自正态总体$N(\mu, \sigma^2)$ 的iid样本，求常数$c$使$c\sum_{i=1}^n |X_i - \bar{X}|$是$\sigma$的UE.
解: 因为 $X_i - \bar{X} \sim N(0, \frac{n-1}{n})$. 而易证若 $Y \sim N(0, \sigma^2)$, 则 $\mathrm{E}|Y| = \frac{\sqrt{2}\sigma}{\pi}$. 简单计算可知: $c = \sqrt{\frac{\pi}{2n(n-1)}}$.

23. 设$X_1, X_2$为来自PDF为

$$f(x,\theta)=\begin{cases}3x^2/\theta^3, & 0<x<\theta,\\ 0, & \text{其它},\end{cases}$$

的总体的iid样本，其中$\theta>0$为参数.

(1) 证明$T_1=\frac{2}{3}(X_1+X_2)$和$T_2=\frac{7}{6}\max\{X_1,X_2\}$是$\theta$的UE;

(2) 计算$T_1, T_2$的方差，并指出何者更有效;

(3) 证明在均方误差意义下，在形为$T_c=c\max\{X_1,X_2\}$的估计中，$T_{\frac{8}{7}}$最有效.

解: 令 $Y=\left(\frac{X}{\theta}\right)^3$, 易证 $Y\sim U(0,1)$. 令 $Z=\max\{Y_1,Y_2\}$, 则 $Z\sim 2z,\ 0<z<1$.

(1) $\mathrm{E}X_1=\theta\mathrm{E}Y^{1/3}=3/4\theta$, 则 $\mathrm{E}T_1=\theta$;

$\mathrm{E}\max\{X_1,X_2\}=\theta\mathrm{E}Z^{1/3}=6/7\theta$, 则 $\mathrm{E}T_2=\theta$.

(2) $\mathrm{E}X_1^2=\theta^2\mathrm{E}Y^{2/3}=3/5\theta^2$, 则 $\mathrm{E}T_1^2=8/9\mathrm{E}X_1^2=5/18\theta^2$;

$\mathrm{E}\max\{X_1,X_2\}^2=\theta^2\mathrm{E}Z^{2/3}=3/4\theta^2$, 则 $\mathrm{E}T_2^2=49/48\theta^2$.

从而 $T_1$ 更有效.

(3) 由 (1)、(2) 知: $\mathrm{E}T_c=6/7c\theta$, $\mathrm{E}^2T_c=3/4c^2\theta^2$, 则 $\mathrm{Var}T_c=(3/4-36/49)c^2\theta^2$. 从而 $\mathrm{MSE}T_c=\left[(3/4-36/49)c^2+(1-6/7c)^2\right]\theta^2$.

$\min g(c)=(3/4-36/49)c^2+(1-6/7c)^2=3/4c^2-12/7c+1$, 当 $c=8/7$ 时, $g(c)$ 取得最小值, 证毕.

24. 设$X_1,\ldots,X_n$为来自PDF为

$$f(x,\theta)=\begin{cases}\theta^2xe^{-\theta x}, & x\geq 0,\\ 0, & x<0,\end{cases}$$

的总体的iid样本，其中$\theta>0$.

(1) 求$\theta$的MLE $\hat\theta_{ML}$;

(2) 求$\theta$的UMVUE;

(3) $\hat\theta_{ML}$是$\theta$的相合估计吗？为什么？

解: (1) 对数似然为 $l=2n\log\theta+\sum\log X_i-\theta\sum X_i$. 易得 $\hat\theta_{\mathrm{ML}}=\frac{2}{\bar X}$.

(2) 事实上, $X\sim\Gamma(2,\theta)$, 则 $\sum X_i\sim\Gamma(2n,\theta)$. 易证: 若 $Y\sim\Gamma(\alpha,\lambda)$, 则 $\mathrm{E}\frac{1}{Y}=\frac{\lambda}{\alpha-1}$. 从而易知 $\hat\theta=(2n-1)\frac{1}{\sum X_i}$ 为 $\theta$ 的 UE. 又 $\bar X$ 为 $\theta$ 的充分完备统计量, 所以 $\hat\theta$ 也是 $\theta$ 的 UMVUE.

(3) 由大数定律知: $\bar X\xrightarrow{p}\mathrm{E}X=\frac{2}{\theta}$. 由连续映射定理知: $\hat\theta^{\mathrm{ML}}\xrightarrow{p}\theta$.

25. 设$X_1,\ldots,X_n$为来自均匀分布$U(\theta-0.5,\theta+0.5)$的iid样本($n\geq 2$)，$-\infty<\theta<\infty$为未知参数.

(1) 求$\theta$的矩估计$\hat\theta_M$;

(2)$\hat\theta_M$和$\hat\theta_1=(X_{(1)}+X_{(n)})/2$是$\theta$的UE吗？为什么？

(3) $\hat{\theta}_M$和$\hat{\theta}_1$哪个方差较小？为什么？
解: (1) 因为 $\mathrm{E}X = \theta$, 所以 $\hat{\theta}_{\mathrm{M}} = \bar{X}$.
(2) 显然 $\hat{\theta}_{\mathrm{M}}$ 是 $\theta$ 的 UE. 因为 $X_{(1)} - (\theta - 0.5) \sim \beta(1, n)$ 且 $X_{(n)} - (\theta - 0.5) \sim \beta(n, 1)$. 所以 $\mathrm{E}[X_{(1)} - (\theta - 0.5) + X_{(n)} - (\theta - 0.5) = 1$. 从而 $\hat{\theta}_1$ 也是 $\theta$ 的 UE.
(3) $\mathrm{Var}(\hat{\theta}_{\mathrm{M}}) = \frac{1}{12n}$. 下面求 $\hat{\theta}_1$ 的方差. 记 $Y_i = X_{(i)} - (\theta - 0.5)$, $i = 1, n$.
则 $\mathrm{Var}(\hat{\theta}_1) = \frac{1}{4}\mathrm{Var}(Y_1 + Y_n)$. 而 $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_n) = \frac{n}{(n+2)(n+1)^2}$. 简单但是略微繁琐的计算可知 $\mathrm{E}Y_1Y_n = \frac{1}{n+2}$, 从而 $\mathrm{Cov}(Y_1, Y_n) = \frac{1}{(n+1)^2(n+2)}$. 所以 $\mathrm{Var}(\hat{\theta}_1) = \frac{1}{2(n+1)(n+2)}$. 那么我们可以得出结论:

$$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_{\mathrm{M}}), \quad \text{当} \quad n > 2.$$

26. 设$X_1, \ldots, X_n$为来自均匀分布$U(\theta_1 - \theta_2, \theta_1 + \theta_2)$的IID样本, 其中$\theta_1 \in \mathbb{R}, \theta_2 > 0$为未知参数.
(1)试求$\theta_1, \theta_2$的矩估计;
(2)试求$\theta_1, \theta_2$的UMVUE;
(3)试求$\frac{\theta_1}{\theta_2}$ 的UMVUE.
解: (1) 由 $\mathrm{E}X = \theta_1$ 和 $\mathrm{Var}X = \theta_2^2/3$ 可得出:

$$\hat{\theta}_{1\mathrm{M}} = \bar{X}, \quad \text{和} \quad \hat{\theta}_{2\mathrm{M}} = \sqrt{3}S_n^*.$$

(2) 易证 $(X_{(1)}, X_{(n)})$ 为 $(\theta_1, \theta_2)$ 的充分完备统计量. 令 $Y_i = \frac{X_i - (\theta_1 - \theta_2)}{2\theta_2} \sim U(0, 1)$. 同前面习题, 可证明 $\theta_1, \theta_2$ 的 UE 分别为 $\frac{X_{(n)} + X_{(1)}}{2}$ 和 $\frac{(n+1)(X_{(n)} - X_{(1)})}{2(n-1)}$. 则其为相应的UMVUE.
(3) 只需找到 $\frac{\theta_1}{\theta_2}$ 的一个 UE 即可. 易证:

$$\begin{aligned} \mathrm{E}\frac{X_n + X_{(1)}}{X_n - X_1} &= \mathrm{E}\frac{Y_n + Y_{(1)}}{Y_n - Y_1} + \frac{\theta_1 - \theta_2}{\theta_2}\mathrm{E}\frac{1}{Y_{(n)} - Y_{(1)}} \\ &= \frac{n}{n-2}\frac{\theta_1}{\theta_2}. \end{aligned}$$

从而, $\frac{\theta_1}{\theta_2}$ 的 UMVUE 为 $\frac{n-2}{n}\frac{X_n + X_{(1)}}{X_n - X_1}$.

27. 验证(1.4.1)式给出的统计量$D_n$是$\sigma$的渐近UE，并求其渐近效率.
解: 参考 22 题.

28. 设$X_1, \ldots, X_n$为来自Pareto分布

$$f(x, \theta) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}}, & 0 < x < \infty, \\ 0, & x \leq 0, \end{cases}$$

的iid样本，其中$\theta > 1$为未知参数，试求$1/\theta$的无偏估计.
解: 易证 $Y = \log(1+X) \sim E(\theta) = \Gamma(1,\theta)$. 从而 $\mathrm{E}Y = \frac{1}{\theta}$,则$\mathrm{E}\bar{Y} = \frac{1}{\theta}$.因此$1/\theta$的UE为$\frac{1}{n}\sum \log(1+X_i)$.

30. 设$X_1,\ldots,X_n$为来自对数正态总体

$$f(x,\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}x^{-1}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\},\ x>0$$

的iid样本，其中$\mu \in R, \sigma > 0$为未知参数，试求$\mu, \sigma^2$的矩估计.
解: 对数正态分布的期望和方差分别为:

$$\mathrm{E}X = e^{\mu+\sigma^2/2} \quad \text{和} \quad \mathrm{Var}X = e^{2\mu+\sigma^2}(e^{\sigma^2}-1).$$

联立 $\mathrm{E}X = \bar{X}$ 和 $\mathrm{Var}X = S_n^{*2}$, 解得:

$$\hat{\mu}_{\mathrm{M}} = \ln\bar{X} - \hat{\sigma}^2_{\mathrm{M}}/2 \quad \text{和} \quad \hat{\sigma}^2_{\mathrm{M}} = \ln(\frac{S_n^{*2}}{\bar{X}^2}-1).$$

31. 设$\beta = a\theta + b$，$\hat{\theta}$是$\theta$的有效估计，试证明$\hat{\beta} = a\hat{\theta} + b$ 是$\beta$的有效估计.
证: 由定义易证.

32. 设$X_1,\ldots,X_n$为来自均匀分布$U(\theta_1,\theta_2)$ 的iid样本，其中$\theta_1,\theta_2$未知.
(1) 证明：$T(\mathbf{X}) = (X_{(1)}, X_{(n)})$是参数$(\theta_1,\theta_2)$的充分统计量;
(2) 如果$T(\mathbf{X})$也是完备的(其证明可见陈希孺(1997))，求$(\theta_1+\theta_2)/2$的UMVUE.
解: (1) 样本联合密度为:

$$(\theta_2-\theta_1)^{-n}\mathbf{1}_{\{\theta_1<X_{(1)}<X_{(n)}<\theta_2\}}.$$

由因子分解定理知: $(X_{(1)}, X_{(n)})$ 是 $(\theta_1,\theta_2)$ 的充分统计量.
(2) 因为 $(X_{(1)}, X_{(n)})$ 是 $(\theta_1,\theta_2)$ 的充分完备统计量. 因此只需要找 $(\theta_1+\theta_2)/2$ 的一个基于充分完备统计量的 UE. 因为 $X \sim U(\theta_1,\theta_2)$, 所以 $Y = \frac{X-\theta_1}{\theta_2-\theta_1} \sim U(0,1)$. 从而 $\mathrm{E}(Y_{(1)}+Y_{(n)}) = 1$, 即 $\mathrm{E}(\frac{X_{(1)}+X_{(n)}}{2}) = \frac{\theta_1+\theta_2}{2}$. 因而所求 UMVUE 为 $(X_{(1)}+X_{(n)})/2$.

33. 设$X_1,\ldots,X_n$为来自均匀分布$U(-\theta/2,\theta/2)$的iid样本，其中$\theta > 0$为参数. 试证明 $(X_{(1)}, X_{(n)})$是
$\theta$ 的充分统计量. 它是完备的吗？为什么？
解: 同 32 题, 显然 $(X_{(1)}, X_{(n)})$ 是 $\theta$ 的充分统计量. 但是它不是完备的. 取 $g(x) = \sin(x) \neq 0$, 但是因为 $\sin(x)$ 关于原点中心对称, 所以 $\mathrm{E}g(x) = \frac{1}{\theta}\int_{-\theta/2}^{\theta/2}\sin(x) = 0$.

34. 设随机变量$X$服从负二项分布$NB(r,p)$，其中$p \in (0,1)$为未知参数，$r$为已知参数.

(1) 试求$p^t$ 的UMVUE，其中$t < r$;

(2) 试求$\mathrm{Var}(X)$ 的UMVUE;

(3) 试求$\ln p$ 的UMVUE.

解: (1) 只有一个样本 $X$, 所以 $X$ 是 $p$ 的充分完备统计量. 我们要找一个 $h(x)$ 满足 $\mathrm{E}(h(X)) = p^t$, $\forall p$. 即:

$$\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}p^r(1-p)^{x-r} = p^t, \quad \forall p.$$

令 $p = 1-q$, 则:

$$\begin{aligned}
\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x &= \frac{q^r}{(1-q)^{r-t}} \\
&= \frac{q^{r-t}}{(1-q)^{r-t}}q^t \\
&= \sum_{x=r-t}^{\infty}\binom{x-1}{r-t-1}q^x q^t \\
&= \sum_{y=r}^{\infty}\binom{y-t-1}{r-t-1}q^y,
\end{aligned}$$

从而 $h(x) = \binom{x-t-1}{r-t-1}/\binom{x-1}{r-1}$, $x = r, r+1, \ldots$.

(2) 易得 $\mathrm{Var}(X) = r(1-p)/p^2 = rq/(1-q)^2$. 同上问, 我们需要找到 $h(x)$ 满足如下关系:

$$\begin{aligned}
\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x &= \frac{rq^{r+1}}{(1-q)^{r+2}} \\
&= \frac{q^{r+2}}{(1-q)^{r+2}}\frac{r}{q} \\
&= \sum_{x=r+2}^{\infty}\binom{x-1}{r+2-1}q^x\frac{r}{q} \\
&= \sum_{y=r+1}^{\infty} r\binom{y}{r+1}q^y,
\end{aligned}$$

从而 $h(x) = r\binom{x}{r+1}/\binom{x-1}{r-1}$, $x = r+1, \ldots$; 且 $h(r) = 0$.

(3) 注意到 $\log(1-p) = \sum_{i=1}^{\infty}\frac{q^i}{i}$ 和上述两问中用到的恒等式

$$\sum_{x=j}^{\infty}\binom{x-1}{j-1}q^x = \frac{q^j}{(1-q)^j}, \quad \forall j > 0$$

可得 $h(r)=0$ 且 $h(x)=\frac{1}{\binom{x-1}{r-1}}\sum_{k=0}^{x-r-1}\binom{r+k-1}{k}\frac{1}{k+r-x}$.

35. 设$X_1,\ldots,X_n$为来自Poisson分布$P(\lambda)$的iid样本，试求

(1) $\lambda^k$的UMVUE.

(2) $P\{X_1=k\}$的UMVUE，其中$k$为大于0的整数.

解: (1) $X_i\sim P(\lambda)$, 则 $T=\sum X_i\sim P(n\lambda)$. 易知 $T$ 是 $\lambda$ 的充分完备统计量. 我们只需要找到 $h(t)$ 满足: $\mathrm{E}(h(t))=\lambda^k$. 简单计算比较可知: $h(t)=\frac{t!}{n^k(t-k)!}$, 当 $t\ge k$; $h(t)=0$, 当 $t<k$.

(2)利用 7 题原理, 可知所求 UMVUE 为

$$\mathrm{P}(X_1=k|T)=\binom{T}{K}(\frac{1}{n})^k(1-\frac{1}{n})^{T-k},\ \ T\ge k.$$

36. 对于线性模型

$$y_i=\beta_0+\beta_1x_i+\epsilon_i,\ i=1,\cdots,n$$

其中$\epsilon_1,\ldots,\epsilon_n$独立同分布，且$E(\epsilon_i)=0,\ \mathrm{Var}(\epsilon_i)=\sigma^2$. 证明：$\beta_0,\beta_1$的LSE不相关的充要条件是$\bar{x}=\sum_{i=1}^n x_i/n=0$.

证: 见例 2.9.1.

37. 设$X_1,\ldots,X_n$为来自$\Gamma(\alpha,\lambda)$的iid样本，其中$\alpha,\lambda$均未知，试求$\alpha/\lambda$ 的UMVUE.

解: 易知 $\bar{X}$ 为 $(\alpha,\lambda)$ 的充分完备统计量. 而 $\mathrm{E}X=\frac{\alpha}{\lambda}$. 从而 $\alpha/\lambda$ 的 UMVUE 为 $\bar{X}$.

38. 设$X_1,\ldots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中$\theta>0$. 记$\theta$的MLE和UMVUE分别为$\hat{\theta}_m=X_{(n)}$和$\hat{\theta}_u=(n+1)\hat{\theta}_m/n$. 如记$T=(n+2)\hat{\theta}_m/(n+1)$，则请证$MSE(T)<\min\{MSE(\hat{\theta}_m),MSE(\hat{\theta}_u)\}$.

解: 注意到 $X/\theta\sim U(0,1)$ 从而 $X_{(n)}/\theta\sim\beta(n,1)$. 由

$$\mathrm{E}X_{(n)}=\frac{n}{n+1}\theta\qquad \mathrm{Var}X_{(n)}=\frac{n}{(n+1)^2(n+2)}\theta^2$$

易证该结论.

39. 设$X$服从Poisson分布$P(\theta)$, 令统计量$T(X)=\begin{cases}1, & X=0,\\ 0, & \text{其它},\end{cases}$

(1) 计算估计$ET$的信息不等式下界，并证明它严格小于$T$的方差；

(2) 证明$T$是其期望的UMVUE.

解: (1) $\mathrm{E}T=\mathrm{P}(X=0)=e^{-\theta}$, $\mathrm{Var}T=\mathrm{E}T^2-(\mathrm{E}T)^2=e^{-\theta}-e^{-2\theta}$. 又 $P(\theta)$ 是正则分布族, 且 $I(\theta)=1/\theta$, 所以信息不等式下界为 $\frac{((e^{-\theta})')^2}{I(\theta)}=\theta e^{-2\theta}$. 因为 $e^\theta>1+\theta$, 易证该下界严格小于 $\mathrm{Var}T$.

(2) 因为 $T$ 是其期望的无偏估计且 $X$ 为 $\theta$ 的充分完备统计量, 所以 $T$ 是其期望 $e^{-\theta}$ 的 UMVUE.

40. 设$X_1,\ldots,X_n$为来自总体PDF为$f(x,\theta)=\theta x^{\theta-1},0<x<1$的iid样本，其中$\theta>0$为未知参数，试求$1/\theta$的有效估计.
解: 易知该分布族是正则的. 由下题知: $I(\theta)=-\mathrm{E}\frac{\partial^2\ln f(x,\theta)}{\partial\theta^2}=-\mathrm{E}(-1/\theta^2)=1/\theta^2$. 所以信息不等式下界为 $\frac{((1/\theta)')^2}{nI(\theta)}=1/(n\theta^2)$. 另一方面, 令 $Y_i=-\ln X_i$, 易证 $Y_i\sim E(\theta)$. 所以 $\bar{Y}$ 是 $1/\theta$ 的无偏估计且 $\mathrm{Var}\bar{Y}=1/(n\theta^2)$. 从而我们要找的估计为 $\bar{Y}=-\frac{1}{n}\sum\ln X_i$, 其方差达到信息不等式下界, 为 $1/\theta$ 的有效估计.

41. 在正则条件下，请证明Fisher信息量$I(\theta)=-E_\theta\frac{\xrightarrow{p}^2\mathrm{inf}(X,\theta)}{\xrightarrow{p}\theta^2}$.
证: 由密度函数的性质知:

$$\int f(x,\theta)\mathrm{d}x=1.$$

两边同时关于 $\theta$ 求导可得:

$$\int\frac{\partial\ln f(x,\theta)}{\partial\theta}f(x,\theta)\mathrm{d}x=0.$$

两边再同时关于 $\theta$ 求导可得:

$$\int\Big[\frac{\partial^2\ln f(x,\theta)}{\partial\theta^2}f(x,\theta)+\frac{\partial\ln f(x,\theta)}{\partial\theta}\frac{\partial\ln f(x,\theta)}{\partial\theta}f(x,\theta)\Big]\mathrm{d}x=0.$$

即证得:

$$I(\theta)=\mathrm{E}\Big[(\frac{\partial\ln f(x,\theta)}{\partial\theta})^2\Big]=-\mathrm{E}\Big[\frac{\partial^2\ln f(x,\theta)}{\partial\theta^2}\Big].$$

43. 证明：对于标准的单参数指数型分布族，定理2.6.2的条件成立.
证: 容易验证.

44. 如果样本分布族为单参数指数型分布族：

$$C(\theta)\exp\{Q(\theta)T(x)\}h(x),$$

则当且仅当$g(\theta)=E_\theta[aT(X)+b$时，$g(\theta)$的一个无偏估计$aT(X)+b$的方差达到C-R下界，其中$a,b$为常数.
证: 注意到

$$\frac{\partial\ln f(x,\theta)}{\partial\theta}=\frac{C'(\theta)}{C(\theta)}+Q'(\theta)T(x),$$

并且对于 $g(\theta)$ 的一个无偏估计 $\hat{g}(\theta)$, 要达到 C-R 下界的充分必要条件为: 存在 $a(\theta)$ 满足

$$\frac{\partial\ln f(x,\theta)}{\partial\theta}=a(\theta)(\hat{g}(\theta)-g(\theta))$$

即可得证.

45. 验证例2.6.9中的UMVUE以及其方差 .

解: UMVUE 的验证参见习题 7.

46. 设$X_1,\ldots,X_n$为来自逆Gauss分布

$$f(x;\mu,\lambda)=(\frac{\lambda}{2\pi x^3})^{1/2}e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}},\ 0<x<\infty,$$

的IID样本，试证明统计量 $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$ 和 $T=\frac{n}{\sum_{i=1}^n\frac{1}{X_i}-\frac{1}{\bar{X}}}$ 是充分且完备的统计量.

证: 似然函数为

$$\begin{aligned} L &= (\frac{\lambda}{2\pi})^{\frac{n}{2}}\prod_{i=1}^n(X_i^{-\frac{3}{2}})\exp\{-\frac{\lambda}{2\mu^2}\sum_{i=1}^n\frac{(X_i-\mu)^2}{X_i}\} \\ &= C(\mu,\lambda)\exp\{-\frac{\lambda}{2\mu^2}\sum X_i-\frac{\lambda}{2}\sum\frac{1}{X_i}\}h(X_1,\ldots,X_n). \end{aligned}$$

由因子分解定理知, $(\sum X_i,\sum\frac{1}{X_i})$ 为 $(\mu,\lambda)$ 的充分统计量; 而自然参数空间为 $(-\frac{\lambda}{2\mu^2},-\frac{\lambda}{2})=R^-\times R^-$ 是开集, 包含内点, 则 $(\sum X_i,\sum\frac{1}{X_i})$ 是完备的. 而 $(\bar{X},T)$ 是上述充分完备统计量的一一变换, 所以也是充分完备的. 进一步的, $\hat{\mu}_{\mathrm{ML}}=\bar{X}$ 且 $\hat{\lambda}_{\mathrm{ML}}=T$.

47. 设$X_1,\ldots,X_n$为来自Logistic分布

$$f(x;\mu,\sigma)=\frac{e^{-(x-\mu)/\sigma}}{\sigma[1+e^{-(x-\mu)/\sigma}]^2}$$

的IID样本，

(1) 当$\sigma$已知，$\mu\in\mathbb{R}$时，试求$\mu$的MLE;

(2) 当$\mu$已知，$\sigma>0$时，试求$\sigma$的MLE.

解: 此题间接考查 Newton-Raphson 迭代算法或者 Fisher-scoring 迭代算法, 因为这种情况下, 我们无法得到 $\mu$ 和 $\sigma$ 的 MLE 的显式解.

(1) 当 $\sigma$ 已知时, 记 $L(\mu)$ 为似然函数, 则得分函数及其一阶导数为:

$$S_\sigma(\mu)=\frac{\partial\log L(\mu)}{\partial\mu}=n/\sigma-2/\sigma\sum_{i=1}^n(e^{-(X_i-\mu)/\sigma})/(1+e^{-(X_i-\mu)/\sigma}),$$

$$S_\sigma'(\mu)=\frac{\partial^2\log L(\mu)}{\partial\mu^2}=-2/\sigma^2\sum_{i=1}^n(e^{-(X_i-\mu)/\sigma})/(1+e^{-(X_i-\mu)/\sigma})^2.$$

另外, 我们可以计算得 Fisher 信息量 $I(\mu)=\frac{n}{3\sigma^2}$. 由 Newton-Raphson 法或者 Fisher-scoring 法, 我们可以得到 $\mu$ 的一步迭代估计为:

$$\begin{aligned} \hat{\mu}^{(1)} &= \hat{\mu}^{(0)}-(S_\sigma'(\hat{\mu}^{(0)}))^{-1}S_\sigma(\hat{\mu}^{(0)}), \\ \hat{\mu}^{(1)} &= \hat{\mu}^{(0)}+3\sigma^3/nS_\sigma(\hat{\mu}^{(0)}). \end{aligned}$$

给定 $\hat{\mu}^{(0)}$, 比如矩估计 $\bar{X}$, 通过迭代至算法收敛, 我们可以得到 $\mu$ 的 MLE 的数值解.
(2) 同样的方法, 当 $\mu$ 已知时, 也可以采用上述两种迭代方法求得 $\sigma$ 的 MLE 的数值解. 在这里我们给出 Fisher 信息量 $I(\sigma)=\frac{1}{\sigma^2}(\frac{1}{3}+\frac{\pi^2}{9})$ 作为参考. 此时可取 $\sigma$ 的一个初始估计为 $\sqrt{\frac{3S_n^{*2}}{\pi^2}}$.

48. 设$X_1,\ldots,X_n$为来自正态分布$N(\mu,\sigma^2)$ 的IID样本，其中$\mu\in\mathbb{R}$为未知参数，$\sigma^2>0$为已知参数.
(1) 试求$\mu^3,\mu^4$的UMVUE;
(2) 试求概率$P(X_1\le t)$ 以及$\frac{d}{dt}P(X_1\le t)$的UMVUE，其中$t\in\mathbb{R}$为给定的常数.
解: (1) 易知 $\bar{X}$ 是 $\mu$ 的充分完备统计量. 由正态分布的期望、偏度和峰度知:

$$\mathrm{E}(\bar{X})=\mu,\ \ \mathrm{E}(\bar{X}-\mu)^3=0\ \ 且\ \ \mathrm{E}(\bar{X}-\mu)^4=3\sigma^4.$$

展开后两式移项可得 $\mu^3$ 和 $\mu^4$ 的 UE 分别为:

$$\bar{X}^3-(3\sigma^2/n)\bar{X}\ \ 和\ \ (\bar{X}^4-(6\sigma^2/n)(\bar{X}^2-\sigma^2/n)-3\sigma^4)/4,$$

同时也是 $\mu^3$ 和 $\mu^4$ 的 UMVUE's.
(2) 因为 $\mathrm{E}(\mathbf{1}_{\{X_1\le t\}})=\mathrm{P}(X_1\le t)$, 从而 $\mathrm{P}(X_1\le t)$ 的 UMVUE 为:

$$\begin{aligned}\mathrm{E}(\mathbf{1}_{\{X_1\le t\}}|\bar{X}) &= \mathrm{P}(X_1\le t|\bar{X})\\ &= \mathrm{P}(X_1-\bar{X}\le t-\bar{x}|\bar{X}).\end{aligned}$$

由 Basu 定理易知: $X_1-\bar{X}$ 与 $\bar{X}$ 独立, 且 $X_1-\bar{X}\sim N(0,(1-n^{-1})\sigma^2)$, 所以

$$\begin{aligned}\mathrm{E}(\mathbf{1}_{\{X_1\le t\}}|\bar{X}) &= \mathrm{P}(X_1-\bar{X}\le t-\bar{x})\\ &= \Phi(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}).\end{aligned}$$

由控制收敛定理知:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}(X_1\le t)=\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[\Phi(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}})=\mathrm{E}[\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}),$$

从而 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}(X_1\le t)$ 的 UMVUE 为:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}})=\frac{1}{\sigma\sqrt{1-n^{-1}}}\phi(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}).$$

49. 设$X$是来自PDF为

$$f(x;\theta)=\left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|},x=-1,0,1;0\le\theta\le1$$

的一个观测.
(1) 试求：$\theta$的MLE;
(2) 定义一个估计量$T(X)$为：

$$T(X)=\begin{cases}2, & x=1;\\ 0, & \text{否则}.\end{cases}$$

试证：$T(X)$是$\theta$ 的一个无偏估计量.
解: (1) 定义 $0^0=1$, 且

$$f(x,\theta)=\begin{cases}1-\theta & \text{如果} \ \ x=0;\\ \theta/2 & \text{如果} \ \ x=-1,1.\end{cases}$$

如果 $X=0$, $\hat{\theta}_{\mathrm{ML}}=0$; 如果 $X=-1$ 或 1, $\hat{\theta}_{\mathrm{ML}}=1$. 从而 $\hat{\theta}_{\mathrm{ML}}=|X|$.
(2) $\mathrm{E}T(X)=2\mathrm{P}(X=1)=2\frac{\theta}{2}=\theta$.

50. 设$X_1,\ldots,X_n$为分别来自下列PDF的IID样本，问是否存在$\theta$的一个函数$g(\theta)$，对它存在一个方差达到Cramér-Rao下界的无偏估计量？若存在，则求之; 若不存在，试说明为什么.
(1) $f(x;\theta)=\theta x^{\theta-1}$, $0<x<1,\theta>0$;
(2) $f(x;\theta)=\frac{\ln(\theta)}{\theta-1}\theta^x$, $0<x<1,\theta>1$.
解: (1) 见 40 题. 取 $g(\theta)=1/\theta$. 所求估计为 $-\frac{1}{n}\sum\ln X_i$.
(2) 事实上, 这是第 44 题的一个应用. 显然易见, 我们要找的估计为 $\bar{X}$. 而对应的

$$g(\theta)=\mathrm{E}X=\int_0^1\frac{\log\theta}{\theta-1}x\theta^x\mathrm{d}x=\frac{\theta}{\theta-1}-\frac{1}{\log\theta}.$$

# 3 第三章

1. 假设在105次独立射击中有60次命中目标，试求命中率$p$的95%的置信区间.
解:命中率$p$的置信区间为[0.4759, 0.6619]

2. 设$X_1,\ldots,X_n$为来自Poisson分布$P(\lambda)$的iid样本，试求$\lambda$的置信水平近似为$1-\alpha$的大样本置信区间.
解:由泊松分布的性质知，$E(\bar{X})=\lambda, Var(\bar{X})=\frac{\lambda}{n}$. 故由中心极限定理在$n\to\infty$时,$\frac{\bar{X}-\lambda}{\sqrt{\frac{\lambda}{n}}}$ 服从标准正太分布，而由于$\bar{X}$是$\lambda$的相合估计，可构造水平为$1-\alpha$的大样本的置信区间为$[\bar{X}-\sqrt{\frac{\bar{X}}{n}}u_{\alpha/2},\bar{X}-\sqrt{\frac{\bar{X}}{n}}u_{\alpha/2}]$.

3. 某通讯公司随机抽查了1万个手机短信的字符长度，得到其平均值为$\bar{X}=9$, 样本标准差为$S=18$. 在置信水平95%下，求该公司用户手机短信均值$\mu$的置信区间.
解:该公司用户手机短信均值$\mu$的置信区间为$[8.6472, 9.3528]$

4. 现在南开大学随机调查了1000名学生，发现其中651名同学有手机. 请求我校有手机同学比例$p$的置信水平为0.95的置信区间.
解:我校有手机同学比例$p$的置信水平为0.95的置信区间为$[0.6209, 0.6799]$

5. 随机抽取某大学60位正教授的住房面积，得知他们的平均住房面积为$115.4m^2$，标准差是$15.8m^2$. 又随机地调查了该大学60位副教授的住房面积，得知他们的平均住房面积为$89.3m^2$，标准差为$21.3m^2$. 试在置信水平0.95下，
(1)求该大学正教授平均住房面积的置信区间；
(2)求该大学副教授平均住房面积的置信区间；
(3)如认为两个总体的方差等于样本方差时，求该校正副教授住房面积差的置信区间.
解:(1)$[111.32, 1119.48]$(2)$[83.8, 94.8]$(3)$[-32.811, -19.389]$

6. 科学中的伟大发现往往是由比较年轻的人提出的. 下表是16世纪中叶到20世纪的12个重大科学突破的情况：

| 科学发现 | 科学家 | 年份 | 年龄 |
|---|---|---|---|
| 日心说 | 哥白尼 | 1543 | 70 |
| 望远镜及天文学的基本定律 | 伽利略 | 1609 | 45 |
| 动力学、万有引力、微积分 | 牛顿 | 1665 | 23 |
| 电的实质 | 富兰克林 | 1746 | 40 |
| 燃烧即氧化 | 拉瓦锡 | 1774 | 31 |
| 地球的演变 | 莱尔 | 1830 | 33 |
| 进化论 | 达尔文 | 1858 | 49 |
| 光的电磁场特性 | 麦克斯韦尔 | 1864 | 33 |
| 放射性 | 居里 | 1896 | 29 |
| 量子力学 | 普朗克 | 1901 | 43 |
| 狭义相对论 | 爱因斯坦 | 1905 | 26 |
| 量子力学的数学基础 | 薛定谔 | 1926 | 39 |

设提出重大科学发现时科学家的年龄服从$N(\mu, \sigma^2)$分布($\mu, \sigma$未知)，构造$\mu$ 的置信水平为0.95的置信区间.
解:$[31.39, 40.95]$

7. 设$X_1, \ldots, X_n$为来自分布为$F(x)$的总体的iid样本，对于固定的$x \in R$，试求总体分布$F(x)$的置信水平近似为$1-\alpha$的大样本置信区间.

解:对任意的$x$,由第一章的结论知，样本经验分布函数$F_n(x)$的$n$倍服从二项分布$B(n, F(x))$.故$E(F_n(x)$
$F(x), Var(F_n(x)) = \frac{F(x)(1-F(x))}{n}$. 则由中心极限定理知，在$n \to \infty$时，$\frac{F_n(x)-F(x)}{\sqrt{\frac{F(x)(1-F(x)}{n}}}$ 服从标准正太分布，而由于$F_n(x)$是$F(x)$ 的相合估计，故可构造水平为$1-\alpha$的置信区间为$[F_n(x) - \sqrt{\frac{F_n(x)(1-F_n(x))}{n}}u_{\frac{\alpha}{2}}, F_n(x) - \sqrt{\frac{F_n(x)(1-F_n(x))}{n}}u_{\frac{\alpha}{2}}]$

8. 假设院领导想考察两位老师给学生的期末考试成绩是否有显著性的区别. 假设本学期共有50位学生选修老师甲的课程，且其期末平均成绩为84.5，标准差为14.5，而选修老师乙所讲授课程的学生人数为 115，期末平均成绩为88.3，标准差为6.8. 如假设学生成绩服从正态分布，则请给出甲乙两位老师所教课程的平均成绩的置信水平为0.95 的置信区间以及两位老师所教课程平均成绩差的置信水平为0.95的置信区间.
解:甲老师所教课程平时成绩$[80.38, 88.62]$,乙老师所教课程平时成绩$[87.04, 89.56]$,两位老师所教课程平均成绩差$[-0.41, 8.01]$.

9. 设有来自对数正态分布$LN(\mu, 0.8^2)$的iid样本为： 4.95，1.73，10.07，9.78，14.15，1.52，11.7
10.38. 试求
(1) 总体期望$\beta = EX$;
(2) $\mu$的置信水平为0.95的置信区间；
(3) $\beta$的置信水平为0.95的置信区间.
解:(1)$\beta = e^{\mu+0.32}$(2)$[1.31, 2.35]$(3)$[5.10, 14.44]$

10. 为制定高中学生的体育锻炼标准，某区教育局在该区高中随机抽选了30名男生，测验100米跑成绩，结果平均为13.5秒，标准差为0.1秒. 假定高中男生100米跑成绩服从正态分布，试建立全区高中男生100米跑成绩的置信水平为0.95 的置信区间.
解:$[13.465, 13.535$

11. 为估计每个家庭月平均收入，现抽取一定容量的简单随机样本，假设样本标准差为350元. 问为以95%的概率使区间的长度为360元，需要抽取多少家庭?
解:15

12. 设$X_1, \ldots, X_n$为来自正态总体$N(\mu, \sigma_0^2)$的iid样本，其中$\sigma_0$为已知常数. 又设介于0与1间的三个常数$\alpha_1, \alpha_2, \alpha$满足: $\alpha_1 + \alpha_2 = \alpha$. 又设$\mu$的置信水平$1-\alpha$的置信区间为

$$\left(\bar{X} - u_{\alpha_1}\frac{\sigma_0}{\sqrt{n}},\ \bar{X} + u_{\alpha_2}\frac{\sigma_0}{\sqrt{n}}\right).$$

试证明当$\alpha_1 = \alpha_2$时，上述置信区间的长度最小.
解:若$u_{\alpha_1} < u_{\alpha/2} < u_{\alpha_2}$,则由于$\Phi(-u_{\alpha_1}) - \Phi(-u_{\alpha/2}) = \Phi(u_{\alpha_2}) - \Phi(u_{\alpha/2})$, 且

$$\phi(x) > \phi(y), \forall x \in [-u_{\alpha/2}, -u_{\alpha_1}], y \in [u_{\alpha/2}, u_{\alpha_2}],$$

其中$\Phi, \phi$分别表示标准正太分布的分布函数和密度函数，则易知$u_{\alpha_1}+u_{\alpha_2} > 2u_{\alpha/2}$,当$u_{\alpha_2} < u_{\alpha/2} < u_{\alpha_1}$时同理可证。

13. 设$X_1,\ldots,X_n$和$Y_1,\ldots,Y_m$分别是来自参数为$\lambda_1$和$\lambda_2$的指数分布的iid样本，且全样本独立，试求$\lambda_2/\lambda_1$的置信水平为$1-\alpha$ 的置信区间.
解:易知$2\lambda_1\sum_{i=1}^n X_i$服从自由度为$2n$的卡方分布，$2\lambda_2\sum_{i=1}^m Y_i$服从自由度为$2m$的卡方分布。因此$\frac{\lambda_2\sum_{i=1}^m Y_i}{\lambda_1\sum_{i=1}^n X_i}$ 服从自由度为$(m,n)$的F 分布。故可构造水平为$1-\alpha$的置信区间为$[\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^m Y_i}F_{1-\alpha/2}(m.n),\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^m Y_i}F_{\alpha/2}(m,n)]$.

14. 对方差$\sigma^2$为已知的正态总体，问需要抽取容量$n$为多大的子样，才能使总体均值$\mu$ 的置信度为$1-\alpha$的置信区间的长度不大于$L$?
解:易知$\bar{X}$服从$N(\mu,\frac{\sigma^2}{n})$,则水平为$1-\alpha$的置信区间为$[\bar{X}-\sqrt{\frac{\sigma^2}{n}}u_{\alpha/2},\bar{X}+\sqrt{\frac{\sigma^2}{n}}u_{\alpha/2}]$, 其长度为$2\sqrt{\frac{\sigma^2}{n}}u_{\alpha/2}$, 令$2\sqrt{\frac{\sigma^2}{n}}u_{\alpha/2}\le L$可解出$n\ge\frac{4\sigma^2u_{\alpha/2}^2}{L^2}$.

15. 设$X_i=\frac{\theta}{2}t_i^2+\epsilon_i, i=1,\cdots,n$, 其中$\epsilon_i$为来自$N(0,\sigma^2)$的iid样本，试求基于$\theta$的UMVUE的水平为$1-\alpha$ 的置信区间.
解:易知关于$\theta$的充分完备统计量为$\sum_{i=1}^n t_i^2X_i$,而$E(\sum_{i=1}^n t_i^2X_i)=\sum_{i=1}^n t_i^4\theta/2$, 故$\theta$ 的UMVUE为 $\hat{\theta}=\frac{2\sum_{i=1}^n t_i^2X_i}{\sum_{i=1}^n t_i^4}$, 而$\mathrm{Var}(\hat{\theta})=\frac{4\sigma^2}{\sum_{i=1}^n t_i^4}$, 故水平为$1-\alpha$的置信区间为$[\hat{\theta}-\sqrt{\frac{4\sigma^2}{\sum_{i=1}^n t_i^4}}u_{\alpha/2},\hat{\theta}+\sqrt{\frac{4\sigma^2}{\sum_{i=1}^n t_i^4}}u_{\alpha/2}]$

16. 设$X_1,\ldots,X_n$为分别来自下列密度函数的IID样本，试分别求出$\theta$的置信水平为$1-\alpha$的置信区间.
(1) $f(x;\theta)=1,\theta-\frac{1}{2}<x<\theta+\frac{1}{2}$;
(2) $f(x;\theta)=2x/\theta^2,0<x<\theta,\theta>0$;
(3) $f(x;\theta)=\theta/x^2,0<\theta<x$.
解:(1)易知关于$\theta$的充分完备统计量为$(X_{(1)},X_{(n)})$，令$U=X_{(1)}+X_{(n)}-2\theta$,则$U$ 的分布密度函数为

$$f(u)=\left\{\begin{array}{ll}\frac{n}{2}(u+1)^{n-1} & -1<u\le 0\\ \frac{n}{2}(1-u)^{n-1} & 0<u<1\end{array}\right.$$

故可构造水平为$1-\alpha$的置信区间为$[\frac{X_{(1)}+X_{(n)}}{2}+\alpha^{\frac{1}{n}}-1,\frac{X_{(1)}+X_{(n)}}{2}-\alpha^{\frac{1}{n}}+1]$
(2)充分完备统计量为$X_{(n)}$,易知分布函数为$F(x)=\frac{t^2}{\theta^2}$,则$P(X_{(n)}<t)=(\frac{t^2}{\theta^2})^n$, 令$t=\theta(1-\alpha)^{\frac{1}{n}}$,则$\mathrm{P}(X_{(n)}<\theta(1-\alpha)^{\frac{1}{n}})=1-\alpha$, 故可构造水平为$1-\alpha$的置信区间为$(\frac{X_{(n)}}{(1-\alpha)^{\frac{1}{n}}},\infty)$
(3)充分完备统计量为$X_{(1)}$,已知分布函数为$F(x)=1-\frac{\theta}{x}$,则$\mathrm{P}(X_{(1)}>t)=\frac{\theta^n}{t^n}$,令$t=\frac{\theta}{(1-\alpha)^{\frac{1}{n}}}$, 则$\mathrm{P}(X_{(1)}>\theta(1-\alpha)^{\frac{1}{n}})=1-\alpha$, 故可构造水平为$1-\alpha$的置信区间为，$[0,\frac{X_{(1)}}{(1-\alpha)^{\frac{1}{n}}}]$

17. 设$X_1,\ldots,X_n$为来自均匀分布$U(\theta_1,\theta_2)$的IID样本.
(1) 试构造$\theta_2-\theta_1$的置信水平为$1-\alpha$的置信区间;
(2) 试构造$\frac{\theta_1+\theta_2}{2}$的置信水平为$1-\alpha$的置信区间.
解:易知$(X_{(1)},X_{(n)})$的联合密度函数为，$f(x,y)=n(n-1)\frac{(y-x)^{n-2}}{(\theta_2-\theta_1)^n}$，令$U=X_{(1)}+$

$X_{(n)}, V = X_{(n)} - X_{(1)}$,则由密度变换公式知，$(U,V)$的联合密度函数为，$f(u,v) = \frac{n(n-1)}{2}\frac{v^{n-2}}{(\theta_2-\theta_1)^n}$. 注意到$U,V$间的相关性，可求得他们各自的分布密度函数为

$$f(u) = \begin{cases} \frac{n}{2}\frac{(u-2\theta_1)^{n-1}}{(\theta_2-\theta_1)^n} & 2\theta_1 < u \le \theta_1+\theta_2 \\ \frac{n}{2}\frac{(2\theta_2-u)^{n-1}}{(\theta_2-\theta_1)^n} & \theta_1+\theta_2 < u < 2\theta_2 \end{cases}$$

$$f(v) = n(n-1)(\theta_2-\theta_1-v)\frac{v^{n-2}}{(\theta_2-\theta_1)^n}$$

由此可知$v$的分布函数为$F_v(t) = \frac{nt^{n-1}}{(\theta_2-\theta_1)^{n-1}} - (n-1)\frac{t^n}{(\theta_2-\theta_1)^n}$，令$t = (\theta_2-\theta_1)t'$,其中$t'$满足方程$nt^{n-1}-(n-1)t^n = 1-\alpha$,则可构造$\theta_2-\theta_1$水平为$1-\alpha$的置信区间为$[\frac{X_{(n)}-X_{(1)}}{t'}, \infty)$. 令$U' = U-(\theta_2+\theta_1)$,则$U'$的分布函数为$F_u(t) = \frac{(t-\theta_1+\theta_2)^n}{2(\theta_2-\theta_1)^n}, t<0$, 令$t = (X_{(n)}-X_{(1)})\alpha^{\frac{1}{n}} - X_{(n)} + X_{(1)}$, 则由于$X_{(n)}-X_{(1)}$是$\theta_2-\theta_1$的相合估计，故可构造$\frac{\theta_1+\theta_2}{2}$水平为$1-\alpha$的置信区间为$[X_{(1)} + \frac{X_{(n)}-X_{(1)}}{2}\alpha^{\frac{1}{n}}, X_{(n)} - \frac{X_{(n)}-X_{(1)}}{2}\alpha^{\frac{1}{n}}]$

18. 若$L(X)$和$U(X)$满足$P_\theta(L(X)\le\theta) = 1-\alpha_1$和$P_\theta(U(X)\ge\theta) = 1-\alpha_2$，且对于所有的$x$，$L(X)\le U(X)$，试证明：

$$P_\theta(L(X)\le\theta\le U(X)) = 1-\alpha_1-\alpha_2.$$

解:因对任意的$x$有$L(x)\le U(x)$,易知$\mathrm{P}_\theta(L(x)\le\theta\le U(x)) = 1-\mathrm{P}_\theta(L(X)\ge\theta, U(x)\ge\theta) - \mathrm{P}_\theta(L(x)\le\theta, U(x)\le\theta)$, 而$\mathrm{P}_\theta(L(X)\ge\theta, U(x)\ge\theta) = \mathrm{P}_\theta(L(x)\ge\theta) = \alpha_1$,$\mathrm{P}_\theta(L(x)\le\theta, U(x)\le\theta) = \mathrm{P}_\theta(U(x)\le\theta) = \alpha_2$, 故$\mathrm{P}_\theta(L(x)\le\theta\le U(x)) = 1-\alpha_1-\alpha_2$.

19. 设$X_1,\dots,X_n$为来自PDF的形式为$\sigma^{-1}f((x-\theta)/\sigma)$的一组IID样本，试列出至少三种不同的枢轴量.

解:(1)$\frac{X_{(n)}-X_{(1)}}{S_n}$,其中$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$(2)$\frac{X_{\mathrm{med}}-\bar{X}}{S_n}$, 其中$X_{\mathrm{med}}$为样本中位数（3）$\frac{(X_{(n)}+X_{(1)})/2-\bar{X}}{S_n}$.

20. 设$X_1,\dots,X_n$为来自PDF为

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

的IID样本，试基于最小次序统计量$X_{(1)} = \min(x_1,\dots,x_n)$构造$\theta$的置信水平为$1-\alpha$的置信区间.

解:易知$X_{(1)}$的分布函数为$F(t) = 1-e^{nt-n\theta}$,令$t = \theta - \frac{1}{n}\ln\alpha$,则易知$\mathrm{P}(X_{(1)} < \theta - \frac{1}{n}\ln\alpha) = 1-\alpha$, 则可构造水平为$1-\alpha$ 的置信区间为$[X_{(1)} + \frac{1}{n}\ln\alpha, \infty)$

# 4 第四章

1. 在正常情况下，如乘坐夏利出租车从南开大学东门到天津火车站需要13元左右. 为检验校内出租车运营是否规范，现随机地乘座其中15辆出租车从东门到火车站（假设路况正常），其平均花费为15.4，标准差为2.3，请在水平0.05下，检验校内出租车运营是否规范.

解：$H_0: \mu \le \mu_0 \leftrightarrow H_1: \mu > \mu_0$

则有，$\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t(n-1)$,

由题意，$n = 15, \mu = 13, \bar{X} = 15.4, S = 2.3$,即有$\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} = 4.04 > t_{0.05}(14)$。

则有校内出租车运营不规范。

2. 为研究矽肺病患者肺功能的变化情况，某医院对I，II期矽肺病患者各33名测定其肺活量，得到二者的平均值分别为2710ml和2830ml，其标准差分别为147ml和118ml. 请问：大水平0.05下，第I，II期矽肺病患者的肺活量有无显著差异？

解：$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \ne \mu_2$

则有，$|\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{m}}}| \sim N(0,1)$,

由题意，$m = n = 33, \bar{X} = 2710, \bar{Y} = 2830, S_1 = 147, S_2 = 118$, 即有$|\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{m}}}| = 3.657 > u_{0.05/2}$。

则第I、II期矽肺病患者的肺活量有显著性差异。

3. 设$X_1, \ldots, X_n$为来自两点分布$b(1,p)$的iid样本，试求假设 $H_0: p = p_0$的似然比检验.

解：$X_1, \cdots, X_n$的联合pdf为

$$f(x) = p^{\sum X_i}(1-p)^{n-\sum X_i} = (1-p)^n \exp\{\ln \frac{p}{1-p} \sum X_i\}$$

在$H_0: p = p_0$下，$f(x) = (1-p_0)^n \exp\{\ln \frac{p_0}{1-p_0} \sum X_i\}$

在$H_1$下,$f(x) = (1-\frac{\sum X_i}{n})^n \exp\{\ln \frac{\frac{\sum X_i}{n}}{1-\frac{\sum X_i}{n}} \sum X_i\}$

则有，

$$\lambda(x) = (\frac{1-p_0}{1-\frac{\sum X_i}{n}})^n \exp\{\sum X_i(\ln \frac{p_0}{1-p_0} - \ln \frac{\sum X_i}{n-\sum X_i})\}$$

$$\partial\lambda(x)/\partial(\sum X_i) = (\frac{n(1-p_0)}{n-\sum X_i})^n (\ln \frac{p_0}{1-p_0} - \ln \frac{\sum X_i}{n-\sum X_i})$$

即$\lambda(x)$关于$\sum X_i$先减后增，故检验$W = \{\lambda(x) \le c\}$等价于$w = \{\sum X_i \ge c_1\}$或$w = \{\sum X_i \le c_2\}$。

4. 设$X_1,\dots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中参数$\theta>0$. 如对于假设

$$H_0:\theta\geq 2\longleftrightarrow H_1:\theta<2,$$

取检验统计量为最大次序统计量$X_{(n)}$，且拒绝域为$W=\{\mathbf{x}:x_{(n)}\leq 1.5\}$，则求此检验犯第一类错误概率的最大值.

解：$X_{(n)}$的pdf为$f(x,\theta)=\frac{nx^{n-1}}{\theta^n}\mathrm{I}(x\in(0,\theta))$, $\mathrm{P}_{H_0}(x\in W)=\mathrm{P}\{x_{(n)}\leq 1.5\}=\int_0^{1.5}\frac{nx^{n-1}}{\theta^n}dx=(\frac{1.5}{\theta})^n$, 又在$H_0$:$\theta\geq 2$, 则此检验犯第一类错误概率的最大值为$(\frac{1.5}{\theta})^n=(\frac{3}{4})^n$。

5. 设$X_1,\dots,X_n$为来自两点分布$b(1,p)$的iid样本.

(1) 试求假设$H_0:p\leq 0.01\longleftrightarrow H_1:p>0.01$的水平为0.05的显著性检验；

(2) 如要求这个检验在$p=0.08$时的第二类错误概率不超过0.1，样本容量$n$应为多少？

解：(1)$X_1,\cdots,X_n$的联合pdf为

$$f(x)=p^{\sum X_i}(1-p)^{n-\sum X_i}=(1-p)^n\exp\{\ln\frac{p}{1-p}\sum X_i\}$$

f(x)为单参数指数型分布族，且$\ln\frac{p}{1-p}$ 关于p单调递增，则有$T(x)=\sum X_i$ 为检验统计量。又$T(x)\sim \mathrm{B}(n,p)$，则否定域$W=\{T>c\}$,且$\mathrm{P}_{H_0}(x\in W)\leq 0.05$，即$c=\min\{k|\sum_{i=k}n(0.01)^i(1-0.01)^{n-i}\leq 0.05\}$。

(2)当$p=0.08$时，$1-\mathrm{P}_{H_1}(x\in W)<0.1$,，即有，

$$\sum_{i=c}^{n}C_n^i(0.08)^i(1-0.08)^{n-i}>0.9$$

且由(1)$\sum_{i=c}n(0.01)^i(1-0.01)^{n-i}\leq 0.05$，即可求出n。

6. 为了检测甲乙两种小麦品种的好坏，现在十块地上同时试种这两个品种. 收获之后测得二者的平均产量分别为670kg和800kg，样本标准差分别为120kg和90kg. 试在水平0.01下，检验这两个品种的小麦有无显著性差异.

解：$H_0:\mu_1=\mu_2\leftrightarrow H_1:\mu_1\neq\mu_2$

在$H_0$下则有，$|\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{m}}}|\sim N(0,1)$,

由题意，$m=n=10,\bar{X}=670,\bar{Y}=800,S_1=120,S_2=90$, 即有

$|\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{m}}}|=2.74>u_{0.01/2}=2.58$。

则这两个品种的小麦有显著性差异。

7. 下列哪些假设是简单的，哪些是复合的？

(1) $X$服从正态分布；

(2) $X$服从正态分布$N(2, 0.5^2)$;
(3) $EX = 2, \mathrm{Var}X = 0.5^2$;
(4) $X$不服从正态分布；
(5) $X$服从参数为0.5的两点分布；
(6) $X$服从参数为0.5的二项分布；
(7) $X$服从指数分布$E(3)$.
解：(1)复合，(2)简单，(3)复合，(4)复合，(5)简单，(6) 简单，(7)简单。

8. 设$X$是来自均匀分布$U(\theta - 0.5, \theta + 0.5)$的一个样本，对于假设

$$H_0 : \theta \leq 3 \longleftrightarrow H_1 : \theta \geq 4,$$

构造一个检验法，使其势函数$\beta(\theta)$满足：

$$\beta(\theta) = 0,\ 当\theta \leq 3;\ \ \beta(\theta) = 1, 当\theta \geq 4.$$

解：拒绝域为$W = \{x : x \geq 3.5\}$，
当$\theta \leq 3$时，$\beta(\theta) = P_{\theta \leq 3}\{x \geq 3.5\} = 0$;
当$\theta \geq 4$时，$\beta(\theta) = P_{\theta \geq 4}\{x \geq 3.5\} = 1$;

9. 设$X_1, \ldots, X_n$为来自指数分布$E(\lambda_1)$的iid样本， $Y_1, \ldots, Y_m$为来自指数分布$E(\lambda_2)$的iid样本，且两组样本独立，其中$\lambda_1, \lambda_2$是未知的正参数.
(1) 求假设$H_0 : \lambda_1 = \lambda_2 \longleftrightarrow H_1 : \lambda_1 \neq \lambda_2$的似然比检验；
(2) 证明上述检验法的拒绝域仅依赖于比值$\sum_{i=1}^n X_i / \sum_{i=1}^m Y_i$;
(3) 求统计量$\sum_{i=1}^n X_i / \sum_{i=1}^m Y_i$的零分布.
解：(1)$X_1, \cdots, X_n, Y_1, \cdots, Y_m$的联合分布为
$f(X, Y) = \lambda_1^n \exp\{-\lambda_1(X_1 + \cdots + X_n)\}\lambda_2^m \exp\{-\lambda_2(Y_1 + \cdots + Y_m)\}$。
在$H_0 : \lambda_1 = \lambda_2$下，

$$\sup_{\lambda_1, \lambda_2 \in \Theta_0} f(X, Y) = (\frac{m+n}{X_1 + \cdots + X_n + Y_1 + \cdots + Y_m})^{m+n} \exp\{-(m+n)\}$$

在$\Theta$下，

$$\sup_{\lambda_1, \lambda_2 \in \Theta} f(X, Y) = \bar{X}^{-n}\bar{Y}^{-m} \exp\{-(m+n)\}$$

令$T = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^m Y_i}$，则似然比检验统计量为

$$\lambda(x) = \frac{\sup_{\lambda_1, \lambda_2 \in \Theta_0} f(X, Y)}{\sup_{\lambda_1, \lambda_2 \in \Theta} f(X, Y)} = \frac{m^m n^n}{(n+m)^{(m+n)}} \frac{T^n}{(T+1)^{m+n}}$$

(2)由(1)知，$\partial\lambda(x)/\partial T=0$，有$T=\frac{n}{m}$, 则有$\lambda(x)$关于$T$先增后减，即有$w=\{\lambda(x)<c\}$等价于$w=\{T<c_1\}\cup\{T>c_2\}$，即证。
(3)在$H_0:\lambda_1=\lambda_2$下，$X_i\sim\exp(\lambda_1)=\Gamma(1,\lambda_1),\sum X_i\sim\Gamma(n,\lambda_1)$，$Y_i\sim\exp(\lambda_1)=\Gamma(1,\lambda_1),\sum X_i\sim\Gamma(m,\lambda_1)$，由$\sum X_i,\sum Y_i$独立， $T\sim F(2n,2m)$。

10. 用甲乙两种材料的灯丝制造灯泡，现随机地从中抽取若干个进行寿命试验，得到如下数据（单位：小时）：
甲材料：1610，1650，1680，1710，1720，1800.
乙材料：1580，1600，1640，1630，1700.
如假设寿命服从指数分布，则在水平0.05下检验这两种材料生产的灯泡寿命有无显著性差异.
解：由题意，$X_1\sim\Gamma(1,\lambda_1),Y_1\sim\Gamma(1,\lambda_2)$,
则有，$2\lambda_1\Sigma_{i=1}^6X_i\sim\Gamma(6,1/2),2\lambda_2\Sigma_{i=1}^5Y_i\sim\Gamma(5,1/2)$, $\frac{2\lambda_1\Sigma_{i=1}^6X_i/12}{2\lambda_2\Sigma_{i=1}^5Y_i/10}\sim F(12,10)$,
即，$\frac{\lambda_1\bar{X}}{\lambda_2\bar{Y}}\sim F(12,10)$，在$H_0$ 下，$\frac{\bar{X}}{\bar{Y}}\sim F(12,10)$
又$\frac{\bar{X}}{\bar{Y}}=1.04$，因此无显著性差异。

11. 设$X_1,\ldots,X_m$和$Y_1,\ldots,Y_n$分别来自正态总体$N(\mu_1,\sigma^2)$和$N(\mu_2,\sigma^2)$的iid样本，且全样本独立，其中 $\mu_1,\mu_2,\sigma^2$均未知. 请仿照两样本$t$检验，构造假设$H_0:\mu_1=c\mu_2$的水平$\alpha$的显著性检验，其中$c\neq 0$为常数.
解：$\bar{X}-c\bar{Y}\sim N(\mu_1-c\mu_2,(\frac{1}{m}+\frac{c^2}{n})\sigma^2)$,
$\Sigma_{i=1}^m(X_i-\bar{X})^2/\sigma^2\sim\chi^2(n-1),\Sigma_{i=1}^n(Y_i-\bar{Y})^2/\sigma^2\sim\chi^2(m-1)$,
$\frac{(\bar{X}-c\bar{Y})/\sqrt{(\frac{1}{m}+\frac{c^2}{n})\sigma^2}}{\sqrt{\frac{\Sigma_{i=1}^m(X_i-\bar{X})^2+\Sigma_{i=1}^n(Y_i-\bar{Y})^2}{m+n-2}}}\sim t(m+n-2)$,
令$T=\frac{(\bar{X}-c\bar{Y})}{\sqrt{\Sigma_{i=1}^m(X_i-\bar{X})^2+\Sigma_{i=1}^n(Y_i-\bar{Y})^2}}\cdot\sqrt{\frac{(m+n-2)mn}{n+mc^2}}$,
则拒绝域为$W=\{|T|>t_{\alpha/2}(m+n-2)\}$。

12. 设$X_1,\ldots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中$\theta>0$为参数. 对于给定的$\theta_0>0$，请给出假设$H_0:\theta\leq\theta_0$的水平$\alpha$的显著性检验.
解：$\frac{n+1}{n}X_{(n)}$为$\theta$的UMVUE，且$\frac{X_{(n)}}{\theta_0}$的密度函数为$f(x)=nx^{n-1}$,
则拒绝域的形式为$W=\{\frac{X_{(n)}}{\theta_0}>c\}$, 即有，$P(\frac{X_{(n)}}{\theta_0}>c)\leq\alpha,\int_c^1 nx^{n-1}dx\leq\alpha$, 即拒绝域为$W=\{\frac{X_{(n)}}{\theta_0}>\sqrt[n]{1-\alpha}\}$。

13. 设$X_1,\ldots,X_n$为来自正态总体$N(\mu,\sigma^2)$的iid样本，其中$\mu,\sigma^2$为未知参数. 试证明关于假设

$$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\neq\sigma_0^2$$

的似然比检验就是检验统计量为$\chi^2=(n-1)S_n^2/\sigma_0^2$的$\chi^2$检验，其中$S_n^2$为样本方差.

解：

$$f(x,\mu,\sigma^2)=(\frac{1}{2\pi\sigma^2})^{n/2}\exp\{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}\},$$

在$H_0:\sigma^2=\sigma_0^2$下，

$$\sup_{\theta\in\Theta_0}f(x,\mu,\sigma^2)=(\frac{1}{2\pi\sigma_0^2})^{n/2}\exp\{-\frac{\sum_{i=1}^n(x_i-\bar{x})^2}{2\sigma_0^2}\};$$

而在$\Theta$下，

$$\sup_{\theta\in\Theta}f(x,\mu,\sigma^2)=(2\pi)^{-n/2}[\frac{\sum_{i=1}^n(x_i-\bar{x})^2}{n}]^{-n/2}\exp\{-\frac{n}{2}\};$$

则似然比检验统计量为：

$$\lambda(x)=\frac{\sup_{\theta\in\Theta_0}f(x,\mu,\sigma^2)}{\sup_{\theta\in\Theta}f(x,\mu,\sigma^2)}=[\frac{\sigma_0^2}{\sum_{i=1}^n(x_i-\bar{x})^2/n}]^{-\frac{n}{2}}\exp\{\frac{n}{2}-\frac{\sum_{i=1}^n(x_i-\bar{x})^2}{2\sigma_0^2}\}$$

令$\chi^2=\frac{(n-1)S_n^2}{\sigma_0^2}$，即有，

$$\lambda(x)=(\frac{\chi^2}{n})^{n/2}\exp\{\frac{n}{2}-\frac{\chi^2}{2}\}$$

又$\lambda(x)$是$\chi^2$的先增后减函数，则否定域为$W=\{\lambda(x)\leq c\}$，即$W=\{\chi^2\leq c_1\}\cup\{\chi^2\geq c_2\}$，即证。

14. 在对一种新的流感疫苗进行人体实验时，为实验组的900位志愿者注射了疫苗，在6 个月内他们中有9个人得了流感. 为对照组的900位志愿者注射了老疫苗，在6个月内他们中有19个人得到流感. 请问新疫苗是否更有效？

解：令$p_1,p_2$分别为新疫苗和旧疫苗的发病概率。假设为$H_0:p_1\leq p_2\leftrightarrow H_1:p_1<p_2$，其$X_1,\cdots,X_n,Y_1,\cdots,Y_n$的联合分布为

$$f(X,Y)=p_1^9(1-p_1)^{900-9}p_2^{19}(1-p_2)^{900-19}$$

在$\Theta$下，

$$\sup_{\theta\in\Theta}f(X,Y)=(9/900)^9(1-9/900)^{900-9}(19/900)^{19}(1-19/900)^{900-19}$$

则$\lambda(X,Y)=1$,且又$-2\log\lambda(X,Y)\sim\chi^2(1)$,即否定域为$w=\{-2\log\lambda(X,Y)\geq\chi^2_{1-\alpha}(1)\}$,则新疫苗是更有效。

15. 在村A随机调查了90位男村民，其中有45人对现任村委会主任表示满意，又随机调查了100位女村民，有69人对现任村委会主任表示满意. 在显著性水平0.05下，

(1) 能否认为男女村民的态度有明显的差异？
(2) 求村中满意村委会主任的男女村民的比例$p_1$和$p_2$的置信水平为95%的置信区间;
(3) 由(2)的区间估计能否认为$p_1 > p_2$?
解：(1)令$p_1, p_2$分别为男村民和女村民满意的概率。假设为$H_0: p_1 = p_2 \leftrightarrow H_1: p_1 \neq p_2$, 其$X_1, \cdots, X_n, Y_1, \cdots, Y_n$的联合分布为

$$f(X, Y) = p_1^{45}(1-p_1)^{45} p_2^{69}(1-p_2)^{31}$$

在$\Theta$下，

$$\sup_{\theta \in \Theta} f(X, Y) = (1/2)^{45}(1/2)^{45}(23/30)^{69}(1-23/30)^{31}$$

在$\Theta_0$下，

$$\sup_{\theta \in \Theta_0} f(X, Y) = (114/190)^{114}(1-114/190)^{190-114}$$

则$\lambda(X,Y) = 4.083$,且又$-2\log\lambda(X,Y) \sim \chi^2(1)$,即否定域为$w = \{-2\log\lambda(X,Y) \geq \chi^2_{1-\alpha}(1)\}$,则男女满意度有差异。
(2)$X_i = \{\begin{smallmatrix} 0 & \text{不满意} \\ 1 & \text{满意} \end{smallmatrix}$，$X_i \sim b(1,p)$，则$\sum_{i=1}^n X_i \sim B(1,p)$，有大样本理论，$\frac{T_n - np}{\sqrt{np(1-p)}} \to^L N(0,1)$，按(1)假设，则$p_1$的置信水平为95%的置信区间为$[0.399, 0.601]$,$p_2$的置信水平为95%的置信区间为$[0.594, 0.772]$。
(3)不能。

16. 以$46^0$仰角发射了9颗库存了1个月的某型号炮弹，射程分别是(单位：km):

30.89, 31.74, 33.82, 32.79, 31.87, 31.79, 31.7, 32.23, 31.85.

又以相同的仰角发射了8颗库存了两年的同型号炮弹，其射程分别是(单位：km):

32.84, 31.46, 32.31, 31.75, 30.15, 31.51, 31.43, 31.74.

如果射程都服从正态分布，则在显著性水平0.05下，
(1) 能否认为这两批炮弹射程的标准差有显著性差异？
(2) 在(1)的基础上，能否认为这两批炮弹的平均射程有显著性差异？
解：(1)$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$，且$\mu_1, \mu_2$ 未知，由两样本正态总体方差的显著性检验，

$$F = \frac{\sum_{i=1}^m (X_i - \bar{X})^2/(m-1)}{\sum_{i=1}^n (Y_i - \bar{Y})^2/(n-1)} = 1.1667$$

故两批炮弹射程的标准差无显著性差异。
(2)$\sigma_1^2 = \sigma_2^2$未知，$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$,由两样本正态总体均值的显著性检验，

$$T = \sqrt{\frac{mn}{m+n}} \frac{\bar{X} - \bar{Y}}{S_{mn}^*} = 1.085$$

则两批炮弹射程的平均射程无显著性差异。

17. 设$X_1,\ldots,X_n$为来自概率密度函数为

$$f(x,\theta)=2\frac{x}{\theta^2}\exp\left\{-\frac{x^2}{\theta^2}\right\},\quad x>0$$

的Rayleigh分布的iid样本，其中$\theta>0$为未知参数.
(1) 利用此总体的充分完备统计量，建立一个关于假设

$$H_0:\theta=1\longleftrightarrow H_1:\theta>1$$

的水平近似为$\alpha$的检验;
(2) 验证此检验统计量在$H_1$下比在$H_0$下具有更大的均值.
解：(1)f(x)是指数型分布族，其充分完备统计量为$T=\sum_1^n X_i^2$，又$X\sim 2\frac{x}{\theta^2}\exp\{-\frac{x^2}{\theta^2}\}$，则$T\sim\Gamma(n,\theta^2)$。在$H_0$ 下，$2T\sim\chi^2_{2n}$，即检验为$w=\{T(X)\geq\chi^2_{2n}(\alpha)\}$。
(2)由$T\sim\Gamma(n,\theta^2)$,$\mathrm{E}_{H_0}T=n$,$\mathrm{E}_{H_1}T=n\theta^2$, 又$\theta>1$,故可证得结论。

18. 设$X_1,X_2$独立，且分别服从$N(\mu_1,\sigma^2),N(\mu_2,\sigma^2)$. 当$\sigma^2$已知时，求关于假设

$$H_0:\mu_1=\mu_2=0\longleftrightarrow H_1:\mu_1^2+\mu_2^2>0$$

的水平为$\alpha$的检验.
解：$X_1,X_2$的联合分布为$f(X_1,X_2)=\frac{1}{2\pi\sigma^2}\exp\{-\frac{\sum_1^2(X_i-\mu_i)^2}{2\sigma^2}\}$,
在$H_0:\mu_1=\mu_2=0$下，$\sup_{\Theta_0}f(X_1,X_2)=\frac{1}{2\pi\sigma^2}\exp\{-\frac{X_1^2+X_2^2}{2\sigma^2}\}$，在$\Theta$下，$\sup_{\Theta}f(X_1,X_2)=\frac{1}{2\pi\sigma^2}$，则

$$\lambda(x)=\frac{\sup_{\Theta_0}f(x)}{\sup_{\Theta}f(x)}=\exp\{-\frac{X_1^2+X_2^2}{2\sigma^2}\}$$

则拒绝域为$W=\{\lambda(x)\leq c\}$等价于$w=\{X_1^2+X_2^2\geq c_1\}$。在$H_0$下，$X_1/\sigma\sim N(0,1),X_2/\sigma\sim N(0,1),(X_1/\sigma)^2\sim\chi_1^2,(X_2/\sigma)^2\sim\chi_1$,即有$(X_1^2+X_2^2)/\sigma^2\sim\chi_2^2$，则取$c_1$ 保证第一类错误概率小于$\alpha$即可。

19. 设$X,Y$相互独立，且$X\sim N(\mu_1,1),Y\sim N(\mu_2,1)$. 设$\rho=\mu_1/\mu_2,\theta=(\rho,\mu_2)$. 证明

$$T(X,Y,\rho_0)=\begin{cases}1, & |X-\rho_0Y|>\sqrt{1+\rho_0^2}u_{\alpha/2},\\ 0, & \text{否则}\end{cases}$$

是假设$H_0:\rho=\rho_0$的水平为$\alpha$的检验.
解：$X,Y$的联合分布为$f(X,Y)=\frac{1}{2\pi}\exp\{-\frac{(X-\mu_1)^2+(Y-\mu_2)^2}{2}\}$, 且$\rho=\frac{\mu_1}{mu_2}$, 则$f(X,Y)=\frac{1}{2\pi}\exp\{-\frac{(X-\rho\mu_2)^2+(Y-\mu_2)^2}{2}\}$
在$H_0:\rho=\rho_0$和$\Theta$下，则可得$\lambda(x)=\exp\{-\frac{(\frac{X-\rho_0Y}{1+\rho_0^2})^2+(\frac{\rho_0(\rho_0Y-X)}{1+\rho_0^2})^2}{2}\}$.

令$T(X,Y)=\frac{|X-\rho_0 Y|}{\sqrt{1+\rho_0^2}}$,则$\lambda(x)$关于T递减，则拒绝域为$W=\{\lambda(x)\le c\}$等价于$w=\{T\ge c_1\}$。而又$T$服从标准正态分布，即可证得$T(X,Y,\rho_0)$ 是$H_0$的水平为$\alpha$的检验。

20. 设$X$服从二项分布$B(n,p)$, 证明关于假设$H_0:p=1/2\longleftrightarrow H_1:p\neq 1/2$的似然比检验统计量等价于$|2X-n|$.

解：$f(x,p)=C_n^x p^x(1-p)^{n-x}$;

在$H_0:p=1/2$下，$\sup_{p\in\Theta_0}f(x,p)=C_n^x(\frac{1}{2})^n$,

而$\sup_{p\in\Theta}f(x,p)=C_n^x(\frac{x}{n})^x(1-\frac{x}{n})^{n-x}$,

则似然比检验统计量为

$$\lambda(x)=\frac{\sup_{p\in\Theta_0}f(x,p)}{\sup_{p\in\Theta}f(x,p)}=\frac{(1/2)^n}{(x/n)^x(1-x/n)^{n-x}}=\frac{n^n}{(2x)^x(2n-2x)^{n-x}}$$

令$T=2x-n$，即$x=\frac{T+n}{2}$，即有

$$\lambda(x)=\frac{n^n}{(T+n)^x(2n-T-n)^{n-x}}=\frac{n^n}{(T+n)^{(T+n)/2}(n-T)^{(n-T)/2}}$$

则有$\ln\lambda(x)=\frac{T+n}{2}\ln\frac{n}{T+n}+\frac{n-T}{2}\ln\frac{n}{n-T}$,

当$T<0,\ln T$关于T递减，当$T>0,\ln T$关于T递增，即$\lambda(x)$关于$|T|$递减，其否定域$W=\{\lambda(x)<c\}$等价于$W=\{|\lambda|>c_1\}$，即证。

21. 设$X_1,\ldots,X_n$为来自正态总体$N(\mu,\sigma^2)$的iid样本，其中$\mu,\sigma^2$未知. 证明关于假设$H_0:\mu\le\mu_0\longleftrightarrow H_1:\mu>\mu_0$的单边$t$检验是检验比检验(显著性水平$\alpha<1/2$).

$X_1,\cdots,X_n$的联合密度函数为

$$f(X)=(\frac{1}{2\pi\sigma^2})^{-n/2}\exp\{-\frac{\sum_{i=1}^n(X_i-\mu)^2}{2\sigma^2}\}$$

则可推出

$$\lambda(x)=\{_{1\qquad\qquad \mu_0\ge\bar{X}}^{(\frac{\sum(X_i-\bar{X})^2}{\sum(X_i-\mu_0)^2})^{n/2}\quad \mu_0<\bar{X}}$$

令$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{S_n}$，则$\lambda(X)=(1+\frac{T^2}{n-1})^{-n/2},\mu_0<\bar{X}$，即单边t检验是似然比检验。

22. 设$X_1,\ldots,X_n$为来自正态总体$N(\mu,\sigma^2)$的iid样本，其中$\mu,\sigma^2$未知. 证明关于假设$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\neq\sigma_0^2$的水平为$\alpha$的似然比检验的拒绝域的补集为

$$W=\left\{c_1\le\frac{1}{\sigma_0^2}\sum_{i=1}^n(X_i-\bar{X})^2\le c_2\right\},$$

其中$c_1,c_2$满足

$$\int_{c_1}^{c_2}\chi^2(x;n-1)\mathrm{d}x=1-\alpha;\quad c_1-c_2=n\ln(c_1/c_2).$$

解：在$H_0:\sigma^2=\sigma_0^2$下，$\sup_{\Theta_0} f(X)=(\frac{1}{\sqrt{2\pi}\sigma_0})^n\exp\{-\frac{\sum(X_i-\mu)^2}{2\sigma_0^2}\}$
在$H_1:\sigma^2\neq\sigma_0^2$下，$\sup_{\Theta} f(X)=(\frac{1}{\sqrt{2\pi}(\frac{\sum_1 n(X_i-\bar{X})^2}{n-1})})^n\exp\{-\frac{\sum(X_i-\bar{X})^2}{2(\frac{\sum_1 n(X_i-\bar{X})^2}{n-1})}\}$
则$\lambda(x)=\frac{\sup_{\Theta_0} f(X)}{\sup_{\Theta} f(X)}=(\frac{T}{n-1})^n\exp\{-\frac{T}{2}+(n-1)\}$，其中$T=\frac{\sum_{i=1}^n(X_i-\bar{X})^2}{\sigma_0^2}$。
$\lambda(x)$ 关于T先增后减，则否定域为$W=\{T<c_1\}\cup\{T>c_2\}$,即否定域的补集为$W^c=\{c_1\leq\frac{\sum_{i=1}^n(X_i-\bar{X})^2}{\sigma_0^2}\leq c_2\}$。又$T$ 服从$\chi^2_{n-1}$分布,由$\mathrm{P}_{H_0}(x\in W^c)$可得，$c_1,c_2$满足$\int_{c_1}^{c_2}\chi^2(x;n-1)dx=1-\alpha$,即$c_1-c_2=n\ln(c_1/c_2)$。

23. 设$X_1,\ldots,X_m$和$Y_1,\ldots,Y_n$ 为分别来自正态总体$N(\mu_1,\sigma^2)$和$N(\mu_2,\sigma^2)$的iid样本，其中$\mu_1$, $\mu_2$, $\sigma^2$未知. 证明关于假设$H_0:\mu_1\leq\mu_2\longleftrightarrow H_1:\mu_1>\mu_2$的似然比检验统计量等价于两样本$t$检验量$|T|$(统计量$T$的定义见(4.3.6) 式).
解：由式(4.3.6)，$|T|=\sqrt{\frac{mn}{m+n}}|\frac{\bar{X}-\bar{Y}}{S^*_{m,n}}|, S^*_{m,n}=\frac{\sum_1^m(X_i-\bar{X})^2+\sum_1^n(Y_i-\bar{Y})^2}{m+n-2}$
$X_1,\cdots,X_m,Y_1,\cdots,Y_n$联合分布如下，

$$f(X,Y)=(\frac{1}{\sqrt{2\pi\sigma}})^{m+n}\exp\{-\frac{\sum_1^m(X_i-\bar{X})^2+\sum_1^n(Y_i-\bar{Y})^2}{2\sigma^2}\}$$

在$\Theta$下，$\hat{\mu}_1=\bar{X},\hat{\mu}_2=\bar{Y},\hat{\sigma}_m^2=\frac{\sum_1^m(X_i-\bar{X})^2+\sum_1^n(Y_i-\bar{Y})^2}{m+n}$
在$H_0:\mu_1\leq\mu_2$下， (1)$\bar{X}=\bar{Y},\hat{\mu}=\frac{\sum_1^m X_i+\sum_1^n Y_i}{m+n},\hat{\sigma}=\frac{\sum_1^m(X_i-\hat{\mu})^2+\sum_1^n(Y_i-\hat{\mu})^2}{m+n}$

$$\sup_{\theta\in\Theta} f(X,Y)=(2\pi\hat{\sigma}^2)^{-(m+n)/2}\exp\{-(m+n)/2\}$$

(2)$\bar{X}<\bar{Y},\hat{\mu}_1=\bar{X},\hat{\mu}_2=\bar{Y},\hat{\sigma}=\frac{\sum_1^m(X_i-\hat{\mu})^2+\sum_1^n(Y_i-\hat{\mu})^2}{m+n}$

$$\sup_{\theta\in\Theta} f(X,Y)=(2\pi\hat{\sigma}^2)^{-(m+n)/2}\exp\{-(m+n)/2\}$$

(3)$\bar{X}>\bar{Y},\hat{\mu}_1=\hat{\mu}_2=\frac{m\bar{X}+n\bar{Y}}{m+n},\hat{\sigma}=\frac{\sum_1^m(X_i-\hat{\mu})^2+\sum_1^n(Y_i-\hat{\mu})^2}{m+n}$

$$\sup_{\theta\in\Theta} f(X,Y)=(2\pi\hat{\sigma}^2)^{-(m+n)/2}\exp\{-(m+n)/2\}$$

则$\lambda(X,Y)=\{^{1 \qquad\qquad \bar{X}\leq\bar{Y}}_{(\hat{\sigma}^2/\hat{\sigma}_m^2)^{-(m+n)/2}\quad \bar{X}>\bar{Y}}$ 当$\bar{X}>\bar{Y}$时，

$$\begin{aligned}\lambda(X,Y) &= (\frac{\sum_1^m(X_i-\frac{m\bar{X}+n\bar{Y}}{m+n})^2+\sum_1^n(Y_i-\frac{m\bar{X}+n\bar{Y}}{m+n})^2}{\sum_1^m(X_i-\bar{X})^2+\sum_1^n(Y_i-\bar{Y})^2})^{-(m+n)/2}\\ &= (1+\frac{mn(\bar{X}-\bar{Y})^2/(m+n)}{\sum_1^m(X_i-\bar{X})^2+\sum_1^n(Y_i-\bar{Y})^2})^{-(m+n)/2}\\ &= (1+\frac{T^2}{m+n-2})^{-(m+n)/2}\end{aligned}$$

综上，$\lambda(X,Y)$关于$T^2$递减，检验的拒绝域$W=\{\lambda(X,Y)<c\}$等价于$W=\{|T|>c_1\}$，即证。

24. 称正态分布随机变量列$X_1,\ldots,X_n$为一个自回归(autoregressive)序列，如果$X_i=\theta X_{i-1}+\epsilon_i$，$i=1,2\cdots,n$, 其中$X_0=0$, $\epsilon_1,\ldots,\epsilon_n$是独立的$N(0,\sigma^2)$的iid序列.
证明关于假设$H_0:\theta=0\longleftrightarrow H_1:\theta\neq 0$的似然比检验统计量等价于$-\left(\sum_{i=2}^n X_iX_{i-1}\right)^2/\sum_{i=1}^{n-1}X_i^2$.
解：$\varepsilon_i=X_i-\theta X_{i-1}$,且$\varepsilon_i\sim N(0,\sigma^2)$, 则$\varepsilon_1,\cdots,\varepsilon_n$的联合分布为,

$$f(\varepsilon)=(\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{\sum(X_i-\theta X_{i-1})^2}{2\sigma^2}}$$

则其似然比检验量为，

$$\lambda(\varepsilon)=\frac{sup_{\theta=0}f(\varepsilon)}{sup_{\Theta}f(\varepsilon)}=\frac{\sum(X_i-\frac{\sum X_iX_{i-1}}{\sum_{i=1}^{n-1}X_{i-1}^2}X_{i-1})^2}{\sum_{i=1}^n X_i^2}=1-T$$

其中，$T=\frac{(\sum X_iX_{i-1})^2}{\sum_{i=1}^{n-1}X_{i-1}^2}$，即可证得题意。

25. 投1000次硬币，560次正面向上，440次反面向上，假设硬币质地均匀是否合理？证明你的结果.
解：令$X_i=\begin{cases}0 & \text{第i次出现反面}\\ 1 & \text{第i次出现正面}\end{cases}$，则$X_1,\cdots,X_n$ 来自B$(1,p)$ 的独立同分布样本。且假设为$H_0:p=0.5\leftrightarrow H_1:p\neq 0.5$。则检验统计量为$T=\sum_{i=1}^n X_i$，拒绝域为$\{T\leq c_1\}\cup\{T\geq c_2\}$，且$c_1,c_2$ 满足，假设给定检验水平$\alpha=0.05$，则有
$\sum_{i=0}^{c_1}C_{1000}^i(0.05)^{1000}\leq 0.025,\sum_{i=c_2}^{1000}C_{1000}^i(0.05)^{1000}\leq 0.025$ 用正态分布近似，$\Phi(\frac{c_1-1000*0.5}{\sqrt{1000*0.5*0.5}})\leq 0.025, 1-\Phi(\frac{c_2-1000*0.5}{\sqrt{1000*0.5*0.5}})\leq 0.025$,则可求得$c_1\leq 402, c_2\geq 598$。则在水平$\alpha=0.05$，此硬币均匀。

26. 假设我们有两个独立的随机样本：$X_1,\ldots,X_n$和$Y_1,\ldots,Y_m$为指数分布，参数分别为$\theta$和$\mu$.
(1) 求$H_0:\theta=\mu\longleftrightarrow H_1:\theta\geq\mu$的似然比检验;
(2) 证明(1)中的检验可以基于统计量$T=\frac{\sum X_i}{\sum X_i+\sum Y_i}$;
(3) 当$H_0$为真时，写出$T$的分布.
解：(1)$X_1,\cdots,X_n,Y_1,\cdots,Y_m$的联合分布为

$$f(X,Y)=\theta^n\exp\{-\theta(X_1+\cdots+X_n)\}\mu^m\exp\{-\mu(Y_1+\cdots+Y_m)\}$$

在$H_0:\theta=\mu$下,

$$\sup_{\theta,\mu\in\Theta_0}f(X,Y,\theta,\mu)=\frac{X_1+\cdots+X_n+Y_1+\cdots+Y_m}{m+n}^{-(m+n)}\exp\{-(m+n)\}$$

在$\Theta$下，

$$\sup_{\theta,\mu\in\Theta} f(X,Y,\theta,\mu)=\bar{X}^{-n}\bar{Y}^{-m}\exp\{-(m+n)\}$$

则似然比检验统计量为

$$\lambda(x)=\frac{sup_{\theta,\mu\in\Theta_0}f(X,Y,\theta,\mu)}{sup_{\theta,\mu\in\Theta}f(X,Y,\theta,\mu)}=\frac{(\bar{X}^n\bar{Y}^m)(m+n)^{(m+n)}}{(n\bar{X}+m\bar{Y})^{(m+n)}}$$

(2)令$T=\frac{\sum X_i}{\sum X_i+\sum Y_i}$，则有，

$$\lambda(x)=\frac{(\sum X_i)^n(\sum Y_i)^m}{(\sum X_i+\sum Y_i)^{(m+n)}}\frac{(m+n)^{m+n}}{n^n m^m}=\frac{(m+n)^{m+n}}{n^n m^m}T^n(1-T)^m$$

则有$\lambda(x)$关于$T$先增后减，即有$w=\{\lambda(x)<c\}$等价于$w=\{T<c_1\}\cup\{T>c_2\}$，即证。

(3)在$H_0:\theta=\mu$下，$X_i\sim\exp(\theta)=\Gamma(1,\theta),\sum X_i\sim\Gamma(n,\theta)$，$Y_i\sim\exp(\theta)=\Gamma(1,\theta),\sum X_i\sim\Gamma(m,\theta)$，令$X=\sum X_i,Y=\sum Y_i,U=\frac{X}{X+Y},V=X+Y$，则由X、Y 独立，

$$f(U,V)=\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}U^{n-1}(1-U)^{m-1}V^{n+m-1}\exp\{-\theta V\}\frac{\theta^{m+n}}{\Gamma(m+n)}$$

则U的pdf为$\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}U^{n-1}(1-U)^{m-1}$，即$T\sim\beta(n,m)$。

27. 设$X_1,\ldots,X_n$为来自$\{1,\ldots,\theta\}$上的离散均匀分布的IID样本，这里$\theta$为整数且$\theta\geq 2$，求水平为$\alpha$的以下假设的似然比检验.

(1) $H_0:\theta\leq\theta_0\longleftrightarrow H_1:\theta>\theta_0$，其中$\theta_0\geq 2$的已知整数;

(2) $H_0:\theta=\theta_0\longleftrightarrow H_1:\theta\neq\theta_0$.

解：(1)似然函数为$l(\theta)=\theta^{-n}I_{X_{(n)},X_{(n+1)},\cdots}(\theta),X_{(n)}$ 为其最大的次序统计量。

在$\Theta_0$下，$\sup_{\theta=2,\cdots,\theta_0}l(\theta)=\{^{X_{(n)}^{-n}}_{0}\quad{}^{X_{(n)}\leq\theta_0}_{X_{(n)}>\theta_0}$

在$\Theta$下，$\sup_{\theta=2,3,\cdots}l(\theta)=X_{(n)}^{-n}$

则，似然比检验统计量为$\lambda(x)=\{^{1}_{0}\quad{}^{X_{(n)}\leq\theta_0}_{X_{(n)}>\theta_0}$，所以当$X_{(n)}\leq\theta_0$拒绝$H_0$。

(2)同(1)，$\lambda(x)=\{^{(\frac{X_n}{\theta_0})^n}_{0}\quad{}^{X_{(n)}\leq\theta_0}_{X_{(n)}>\theta_0}$

则,$W=\{\lambda(X)<c\}$等价于 $\{X_{(n)}>\theta_0\}$或 $\{X_{(n)}<\theta_0 c1^{1/n}\}$。

28. 假设$X_{i1},\ldots,X_{in_i},i=1,2$，为两个独立的IID样本，分布服从均匀分布$U(0,\theta_i),i=1,2$，这里$\theta_1>0,\theta_2>0$均未知. (1) 求水平为$\alpha$的假设$H_0:\theta_1=\theta_2\longleftrightarrow H_1:\theta_1\neq\theta_2$的似然比检验;

(2) 求$-2\ln\lambda$的极限分布，这里$\lambda$为(1)中的似然比.

解：(1)令$Y_i=\max_{j=1,\cdots,n}X_{ij},i=1,2,Y=\max\{Y_1,Y_2\}$，则有

$$l(\theta_1,\theta_2)=\theta_1^{-n_1}\theta_2^{-n_2}\mathrm{I}_{(0,\theta_1)}(Y_1)\mathrm{I}_{(0,\theta_2)}(Y_2)$$

在$H_0:\theta_1=\theta_2$下，$\sup_{\Theta=_0} l(\theta_1,\theta_2)=Y^{-n_1-n_2}$
在$\Theta$下，$\sup_{\Theta} l(\theta_1,\theta_2)={Y_1}^{-n_1}{Y_2}^{-n_2}$
在$H_0:\theta_1=\theta_2$下,对任意$t\in(0,1)$,

$$\begin{aligned}
\mathrm{P}(\lambda<t) &= \mathrm{P}(\lambda<t,Y_1\geq Y_2)+\mathrm{P}(\lambda<t,Y_1<Y_2)\\
&= \mathrm{P}(Y_2<t^{1/n_2}Y_1)+\mathrm{P}(Y_1<t^{1/n_1}Y_2)\\
&= t
\end{aligned}$$

故水平为$\alpha$的似然比检验为$w=\{\lambda(x)<\alpha\}$拒绝$H_0$。
(2)在$H_0$下,由(1)知，$\lambda(x)$在$(0,1)$是均匀分布，不依赖于$(n_1,n_2)$。$-2\ln\lambda$的分布是在区间$(0,\infty)$,尺度参数为$1/2$的指数分布，即$\chi_1^2$。
当$n_1/n_2\to\kappa\in(0,\infty)$,假设$\theta_1<\theta_2$，由$Y_2-Y_1\xrightarrow{p}\theta_2-\theta_1$. 则

$$\mathrm{P}(Y_1>Y_2)=\mathrm{P}(Y_2-Y_1-(\theta_2-\theta_1)<-(\theta_2-\theta_1))\to 0$$

假设$Y_1<Y_2$，则$-2\ln\lambda=2n_1(\ln Y_2-\ln Y_1)$. 又$n_i(\theta_i-Y_i)\xrightarrow{d}\theta_i Z_i, i=1,2,$, $Z_i$是在区间$(0,\infty)$,尺度参数为1的指数分布。由Delta定理，$n_i(\ln\theta_i-\ln Y_i)\xrightarrow{d}Z_i, i=1,2$。$Y_1,Y_2$独立，

$$-2\ln\lambda+2\ln(\theta_1/\theta_2)n_1=2[n_1(\ln\theta_1-\ln Y_1)-\frac{n_1}{n_2}n_2(\ln\theta_2-\ln Y_2)\to 2(Z_1-\kappa Z_2)$$

其中$Z_1,Z_2$独立. 同理当$\theta_1>\theta_2$的极限分布可以类似得到。

29. 假设$(X_{11},X_{12}),\ldots,(X_{n1},X_{n2})$为来自二元正态分布的IID样本，均值和方差矩阵均未知. 假设$H_0:\rho=0\longleftrightarrow H_1:\rho\neq 0$，这里$\rho$为相关系数. 证明：$|W|>c$为一个似然比检验，其中

$$W=\sum_{i=1}^n(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)/[\sum_{i=1}^n(X_{i1}-\bar{X}_1)^2+\sum_{i=1}^n(X_{i2}-\bar{X}_2)^2$$

并求$W$的分布.
解：(1)二元正态分布的$\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$的似然估计分别为$\bar{X}_1,\bar{X}_2,\frac{\sum_{i=1}^n(X_{i1}-\bar{X}_1)}{n},\frac{\sum_{i=1}^n(X_{i2}-\bar{X}_2)}{n}$。而又$H_0:\rho=0$，可得似然比统计量为$\lambda=(1-W^2)^{n/2}$。故当$\{|W|>c\}$ 时拒绝原假设。
(2)W的密度函数为

$$\frac{\Gamma((n-1)/2)}{\sqrt{\pi}\Gamma((n-2)/2)}(1-t^2)^{(n-4)/2}\mathrm{I}_{(-1,1)}(t)$$

30. ($p$-值)假设$X$的分布为$P_\theta$，其中$\theta\in\mathbb{R}$为未知参数，对于拒绝域为$W_\alpha$的原假设$H_0:\theta=\theta_0$（或$\theta<\theta_0$）满足$P_{\theta_0}(X\in W_\alpha)=\alpha, 0<\alpha<1$和$W_{\alpha_1}=\cap_{\alpha>\alpha_1}W_\alpha, 0<$

$\alpha_1 < 0$，考虑$H_0$的一族非随机化水平检验：
(1) 证明$p$-值为$\hat{\alpha}(x) = \inf\{\alpha : x \in W_\alpha\}$;
(2)证明当$\theta = \theta_0$时，$\hat{\alpha}(x)$服从均匀分布$U(0,1)$;
(3)如果拒绝域为$W_\alpha$的检验是无偏的，证明在$H_1$下，$P_\theta(\hat{\alpha}(x) \le \alpha) \ge \alpha$.

31. 假设$X_1,\dots,X_n$是来自正态总体$N(\theta,\sigma^2)$的一组随机样本，$\sigma^2$已知. 假设$H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0$的一个似然比检验为如果$|\bar{X} - \theta_0|/(\sigma/\sqrt{n}) > w$则拒绝$H_0$.
(1) 求这个检验的势函数，用标准正态的概率写出这个表达式;
(2) 试验者希望在$\theta = \theta_0 + \sigma$点犯第一类错误的概率是0.05，犯第二类错误的最大概率是0.25. 求为达到这些要求$n$和$w$的值.
解：(1)$\mathrm{P}_\theta(X \in W) = \mathrm{P}_\theta(\frac{|\bar{X}-\theta_0|}{\sigma/\sqrt{n}} > w)$，而又$\bar{X} \sim N(\theta,\sigma^2/n)$，即$\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}} \sim N(\theta-\theta_0,1)$，即

$$\begin{aligned}\mathrm{P}_\theta(X \in W) &= \mathrm{P}(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}} - \frac{\theta_0-\theta}{\sigma/\sqrt{n}} > w - \frac{\theta_0-\theta}{\sigma/\sqrt{n}}) + \mathrm{P}(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}} - \frac{\theta_0-\theta}{\sigma/\sqrt{n}} < -w - \frac{\theta_0-\theta}{\sigma/\sqrt{n}}) \\ &= 1 - \Phi(w - \frac{\theta_0-\theta}{\sigma/\sqrt{n}}) + \Phi(-w - \frac{\theta_0-\theta}{\sigma/\sqrt{n}})\end{aligned}$$

(2)由(1)可得，则$w = 1.96, n = 4$.

# 5 习题五

1. 设$X_1,\dots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中$\theta > 0$为未知参数，取检验函数

$$\phi(\mathbf{X}) = \begin{cases} 1, & X_{(n)} \ge c, \\ 0, & \text{否则}, \end{cases}$$

(1) 计算检验$\phi(\mathbf{X})$的势函数，并证明它关于$\theta$单增；
(2) 当$c$取何值时，用$\phi(\mathbf{X})$来检验假设$H_0 : \theta \le 1/2 \longleftrightarrow H_1 : \theta > 1/2$时的水平正好是0.05?
(3) $n$取多大才能使(2)中的检验对$\theta = 3/4$有势0.98?
解：(1) $x_{(n)}$ 的 pdf 为:

$$f_n(x,n) = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & \text{if } 0 < x_{(n)} < 0; \\ 0 & \text{otherwise}. \end{cases}$$

(分情况讨论)
当 $x > \theta$ 时,

$$\beta_\Phi(\theta) = E_\theta \Phi(x) = P_\theta\{x_{(n)} \ge c\} = 1 - (\frac{c}{\theta})^n > 0$$

当 $x \le \theta$ 时, $\beta_\Phi(\theta)=0$ ，所以，

$$\beta_\Phi(\theta)\begin{cases}1-(\frac{c}{\theta})^n, & \text{if } \theta>c;\\ 0 & \text{if } \theta\le c.\end{cases}$$

对函数求导，

$$\frac{\mathrm{d}\beta_\Phi(\theta)}{\mathrm{d}\theta}=-n(\frac{c}{n})^{n-1}(-\frac{c}{(\theta)^2})=\frac{nc^n}{\theta^{n+1}}>0$$

说明 $\beta_\Phi(\theta)$ 在 $\theta>c$ 时单增.
又因为 $\theta<c$ 时, $\beta_\Phi(\theta)=0<1-(\frac{c}{\theta})^n$, 故 $\beta_\Phi(\theta)$ 单增, 对 $\forall\theta>0$.
(2)因 $\beta_\Phi(\theta)$ 单增,

$$\beta_\Phi(\theta)\le\beta_\Phi(\frac{1}{2})=1-(2c)^n=0.05$$

所以 $c=\frac{1}{2}\sqrt[n]{0.95}$.
(3)

$$\beta_\Phi(\frac{3}{4})=1-(\frac{4}{3}\frac{1}{2}\sqrt[n]{0.95})^n=0.98$$

所以 $n\approx 9.5$, 取整 $n=10$.

2. 设$X_1,X_2,X_3$为来自两点分布$b(1,p)$的iid样本，对于假设

$$H_0:p=\frac{1}{2}\longleftrightarrow H_1:p=\frac{3}{4},$$

如取一个检验的拒绝域为

$$W=\{(x_1,x_2,x_3):x_1+x_2+x_3\ge 2\},$$

则求此检验的第一、二类错误概率及$p=3/4$时的势.
解：

$$\Phi(x)=\begin{cases}1, & \text{if } x\in W;\\ 0 & \text{otherwise.}\end{cases}$$

$$x_1+x_2+x_3\sim B(3,p)$$

所以 $\beta_\Phi(P)=Px_1+x_2+x_3\ge 2=3P^2-2P^3$, 因此第一类错误概率为:

$$\beta_\Phi(P_1)=3(\tfrac{1}{2})^2\times 2(\tfrac{1}{2})^3=\tfrac{1}{2}.$$

第二类错误概率为:

$$1-\beta_\Phi(P_2)=1-3(\tfrac{3}{4})^2+2(\tfrac{3}{4})^3=\tfrac{5}{32}$$

$P=\frac{3}{4}$ 的势为:

$$\beta_\Phi(\tfrac{3}{4})=1-(1-\beta_\Phi(P_2))=\tfrac{27}{32}.$$

3. 设总体$X$的密度函数可取下面的$f_0(x)$或$f_1(x)$:

$$f_0(x)=\begin{cases}1, & 0\le x\le 1,\\ 0, & \text{其它},\end{cases}\qquad f_1(x)=\begin{cases}2x, & 0\le x\le 1,\\ 0, & \text{其它},\end{cases}$$

现基于一个样本$X_1$，考虑如下假设

$$H_0: f(x)=f_0(x)\longleftrightarrow H_1: f(x)=f_1(x),$$

试在水平0.1下，求出使第二类错误概率最小的检验法.
解：$\lambda(x)=\frac{f_1(x)}{f_0(x)}=2X\cdot I_{\{0\le x\le 1\}}$, 关于 x 单调上升, 由 N-P 引理, 水平为 $\lambda$ 的 MPT为

$$\Phi(x)=\begin{cases}1, & \text{if } x\ge k;\\ 0 & \text{if } x<k.\end{cases}$$

$$E_{H_0}\Phi(x)=\textstyle\int_k^1 1\mathrm{d}x=0.1\implies k=0.9$$

$$\Phi(x)=\begin{cases}1, & \text{if } x\ge 0.9;\\ 0 & \text{if } x<0.9.\end{cases}$$

4. 设$X_1,\dots,X_n$为来自Weibull分布的iid样本，其总体PDF为

$$f(x;\lambda)=\lambda cx^{c-1}\exp\{-\lambda x^c\}, x>0$$

其中$\lambda>0$为未知参数，$c>0$为已知参数.
(1) 证明$\sum_{i=1}^n X_i^c$是检验假设$H_0:\frac{1}{\lambda}\le\frac{1}{\lambda_0}\longleftrightarrow H_1:\frac{1}{\lambda}>\frac{1}{\lambda_0}$的一个最优检验统计量；
(2) 如取上述检验的拒绝域为$W=\{\sum_{i=1}^n X_i^c\ge k\}$，则当其水平为$\alpha$时，求临界值$k$及其势函数.
解：(1)$\lambda(x)=\frac{f(x,\lambda_1)}{f(x,\lambda_0)}=(\frac{\lambda_1}{\lambda_0})^n exp\{(\lambda_0-\lambda_1)\sum_{i=1}^n x_i^n\}$
因为 $\lambda_0-\lambda_1<0$, 那么 $\lambda(x)$ 关于 $\sum_{i=1}^n x_i^n$ 单调下降, 该分布是关于统计量 $T(x)=\sum_{i=1}^n x_i^n$ 的单调似然比函数, 故 $T(x)$ 是其一个最优检验.
(2)对假设 $H_0:\frac{1}{\lambda}\le\frac{1}{\lambda_0}\longleftrightarrow H_1:\frac{1}{\lambda}>\frac{1}{\lambda_0}$, 由定理 5.3.4, 其水平 $\alpha$ 的 UMPT 为

$\Phi(x)=\begin{cases}1, & \text{if } T(x)\ge k;\\ 0 & \text{if } T(x)<k.\end{cases}$ 且 $E_{\frac{1}{\lambda_0}}\Phi(x)=\alpha.$

令 $y_i=x_i^n$, 则 $y_i$ 的 Pdf: $f(y,\lambda)=\lambda c y^{\frac{c-1}{c}}\cdot exp\{-\lambda y\}\cdot\frac{1}{c}\cdot y^{-\frac{c-1}{c}}=\lambda e^{-\lambda y}\ \ (y>0)$ 是一个指数分布, $y_i\sim E(\lambda)=\Gamma(a,\lambda)$, 故

$$T(x)=\textstyle\sum_{i=1}^n x_i^c=\sum_{i=1}^n y_i\sim\Gamma(n,\lambda)\implies 2\lambda T(y)\sim\Gamma(n,\frac{1}{2})$$

即$2\lambda T(y)\sim X^2(2n)$.

检验函数为 $\Phi(x)=\begin{cases}1, & \text{if } T\ge\frac{1}{2\lambda_0}X_\alpha^2(2n);\\ 0 & \text{if } T<\frac{1}{2\lambda_0}X_\alpha^2(2n).\end{cases}$

势函数 $\beta_\Phi(\lambda)=P_{\frac{1}{\lambda}}(T\ge\frac{1}{2\lambda_0}X_\alpha^2(2n))=P_{\frac{1}{\lambda}}\{(2\lambda T)\ge X_\alpha^2(2n)\}=\int_{\frac{\lambda}{\lambda_0}X_\alpha^2(2n)}^{x}\frac{\rho^{-\frac{x}{2}}\cdot x^{n-1}}{2^n\Gamma(n)}\mathrm{d}x.$

5. 设$X_1,\dots,X_n$为来自Poisson分布$P(\lambda)$的iid样本，试求假设

$$H_0:\lambda\le\lambda_0\longleftrightarrow H_1:\lambda>\lambda_0\ (\lambda_0\text{已知})$$

的水平为$\alpha$的UMP检验.

解：同 P155 页例 5.3.6.

6. 设$X_1,\dots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，试求假设

$$H_0:\theta\le\theta_0\longleftrightarrow H_1:\theta>\theta_0\ (\theta_0>0\text{已知})$$

的水平$\alpha$的UMP检验.

解：$x_1,\cdots,x_n$ 的联合 PDF 为 $f(x,n)=\frac{1}{\theta^n}I\{0<X_{(1)}<X_{(n)}<\theta\}$

$\forall\,\theta_2>\theta_1,\lambda(x)=\frac{f(x,\theta_2)}{f(x,theta_1)}\begin{cases}\frac{\theta_1^n}{\theta_2^n}, & \text{if } 0<X_{(n)}<\theta;\\ \infty & \text{otherwise.}\end{cases}$

则知 $\lambda(x)$ 关于 $X_{(n)}$ 非降, 故该分布为关于 $T=X_{(n)}$ 的非降 MLR.

由定理知, 其水平 $\alpha$ 的 UMP 为

$$\Phi(x)=\begin{cases}1, & \text{if } X_{(n)}>c;\\ 0 & \text{if } X_{(n)}\le c.\end{cases}\ \text{且}\ E_{\theta_0}\Phi(x)=\alpha$$

则 $P_{\theta_0}(X_{(n)}>c)=\int_c^{\theta_0}1\cdot\frac{nt^{n-1}}{\theta_0^n}\mathrm{d}t=1-\frac{c^n}{\theta_0^n}=\alpha,\implies\ c=(1-\alpha)^{\frac{1}{n}}\theta_0$

检验函数为: $\Phi(x)=\begin{cases}1, & \text{if } X_{(n)}>(1-\alpha)^{\frac{1}{n}}\theta_0;\\ 0 & \text{if } X_{(n)}\le(1-\alpha)^{\frac{1}{n}}\theta_0\end{cases}.$

7. 设$X_1,\ldots,X_n$为来自正态总体$N(\mu,1)$的iid样本，对于假设

$$H_0:\mu\le 0\longleftrightarrow H_1:\mu>0,$$

(1) 求水平为0.025的UMP检验，并求出其势函数$\beta(\mu)$;
(2) 为了使上述检验在$\mu\ge 0.5$时的势函数$\beta(\mu)\ge 0.9$，样本容量$n$应至少取多大？
解：$X_1,\cdots,X_n$样本联合pdf为

$$\begin{aligned} f(x) &= (2/\pi)^{-\frac{n}{2}}exp\{-\frac{1}{2}\sum(X_i-\mu)^2\} \\ &= (2/\pi)^{-\frac{n}{2}}\exp\{-n\mu^2/2\}\exp\{n\mu\bar{X}\}\exp\{-\frac{\sum_{i=1}^n X_i}{2}\} \end{aligned}$$

是单参数指数型分布族，且$Q(\mu)=n\mu$关于$\mu$严格单调递增，所以它关于统计量$T(X)=\bar{X}$是单增MLR分布族，则水平为$\alpha$的UMP检验为

$$\phi(T(X))=\{^{1\qquad T(X)>c}_{0\qquad T(X)<c}$$

其中$c$满足$E_{\theta_0}\phi(X)=\alpha$。又因为当$H_0:\mu=0$时，$\bar{X}\sim N(0,1/n)$，故$c=\mu_\alpha/\sqrt{n}$。
(1)$\beta(\mu)=E_\mu\phi(X)=P_\mu(T(X)>\mu_\alpha/\sqrt{n})$，其中$\alpha=0.025,T(X)\sim N(\mu,\frac{1}{n})$
(2)$0.9\le\beta(\mu)\le\beta_{\mu=0.5}=P\{\bar{X}>\frac{\mu_\alpha}{\sqrt{n}}\}=P\{\sqrt{n}(\bar{X}-0.5)>\sqrt{n}(\frac{\mu_\alpha}{\sqrt{n}}-0.5)\}$
，即有$\sqrt{n}(\frac{\mu_\alpha}{\sqrt{n}}-0.5)=\mu_{0.9}$，即$n=42$。

8. 现从某正态总体中随机抽取了9个iid样本，测得其均值为$\bar{X}=0.4$，样本标准差为$S_n=1.0$. 在水平0.05和0.01下分别检验假设
(1) $H_0:\mu\le 0\longleftrightarrow H_1:\mu>0$,
(2) $H_0:\mu\ge 0\longleftrightarrow H_1:\mu<0$,
并解释检验结果的意义. 如果样本容量增加到25，而样本均值与标准差不变，则再检验上述假设，并解释结果的意义.
解：(1)由单样本正态总体的优势检验，水平为$\alpha$的UMPU检验为$\phi(W)=\{^{1\qquad W(X)<t_\alpha(n-1)}_{0\qquad W(X)>t_\alpha(n-1)}$，其中$W(X)=\sqrt{n}(\bar{X}-\mu_0)/S_n$。
即由题意，$W(X)=2.8>t_{0.05}(8)=1.86$，在水平为0.05时，应拒绝原假设，该正态总体均值大于0。$W(X)=2.8<t_{0.01}(8)=2.90$，在水平为0.01时，应接受原假设，该正态总体均值小于等于0。
(2)水平为$\alpha$的UMPU检验为$\phi(W)=\{^{1\qquad W(X)>t_\alpha(n-1)}_{0\qquad W(X)<t_\alpha(n-1)}$。
即又题意，$W(X)=2.8>t_{0.05}(8)=1.86$，在水平为0.05时，应接受原假设，该正态总体均值小于等于0。$W(X)=2.8<t_{0.01}(8)=2.90$，在水平为0.01时，应拒绝原假设，该正态总体均值大于0。
(3)n=25,同上所述。

9. 设$X_1,\ldots,X_n$为来自均匀分布$U(0,\theta)$的iid样本，其中$\theta>0$为参数. 试求假设
(1) $H_0:\theta=1\longleftrightarrow H_1:\theta=2$,
(2) $H_0:\theta=1\longleftrightarrow H_1:\theta=1/2$
的水平为$\alpha$的MP检验.
解：(1)$x_1,\cdots,x_n$ 的联合 pdf 为:

$$f(x,n)=\begin{cases}\frac{1}{\theta^n}, & \text{if } 0<x_{(n)}<0;\\ 0 & \text{otherwise.}\end{cases}$$

所以,

$$\lambda(x)=\frac{f(x,2)}{f(x,1)}=\begin{cases}\frac{1}{2^n}, & \text{if } 0<x_{(n)}<0;\\ 0 & \text{if } 1\le x_{(n)}<2.\end{cases}$$

$\lambda(x)$ 的分布函数 $h(c)=P_{H_0}\{\lambda(x)\le c\}$ 是个退化分布.
当 $c<\frac{1}{2^n}$ 时, $h(c)=0$; 当 $c\ge\frac{1}{2^n}$ 时, $h(c)=XX$ 根据 N-P 引理可以证明: 我们取 $k=\frac{1}{2^n}$ 时有: $0=h(k-0)<1-\alpha<h(k)=1$, 再由 N-P 引理, 此时的 MPT 为:

$$\Phi(x)=\begin{cases}1, & \text{if } 1\le x_{(n)}<2;\\ \alpha, & \text{if } 0<x_{(n)}<1;\\ 0 & \text{otherwise.}\end{cases}$$

(2)

$$\lambda(x)=\frac{f(x,\frac{1}{2})}{f(x,1)}=\begin{cases}2^n, & \text{if } 0<x_{(n)}<\frac{1}{2};\\ 0 & \text{if } \frac{1}{2}\le x_{(n)}<1.\end{cases}$$

$\lambda$ 的分布函数 $h(c)=P_{H_0}\{\lambda(x)\le c\}$
当 $0\le c<2^n$ 时, $h(c)=\int_{\frac{1}{2}}^{1}nt^{n-1}\mathrm{d}t=1-\frac{1}{2^n}$;
当 $c<0$ 时, $h(c)=0$;
当 $c\ge 2^n$ 时, $h(c)=1$.
这样, 根据 N-P 引理:
① 当 $h(0)=1-\frac{1}{2^n}<1-\alpha$ 时, 取 $k=2^n$, 则 $h(k-0)=1-\frac{1}{2^n}<1-\alpha<h(k)=1$, 此时的 MPT 为:

$$\Phi(x)=\begin{cases}1, & \text{if } \lambda(x)>k;\\ \frac{h(k)-(1-\alpha)}{h(k)-h(k-0)}, & \text{if } \lambda(x)=k;\\ 0 & \text{if } \lambda(x)<k.\end{cases}$$

即

$$\Phi(x)=\begin{cases}1, & \text{otherwise;}\\ 2^n\alpha, & \text{if } 0<x_{(n)}<\frac{1}{2};\\ 0 & \text{if } \frac{1}{2}\le x_{(n)}<1.\end{cases}$$

② 当 $h(0)=1-\frac{1}{2^n}>1-\alpha$ 时,

$$\Phi(x)=\begin{cases}1, & \text{if } \lambda(x)>0;\\ \frac{h(0)-(1-\alpha)}{h(0)}, & \text{if } \lambda(x)=0;\\ 0 & \text{if } \lambda(x)<0.\end{cases}$$

即

$$\Phi(x)=\begin{cases}1, & \text{if } 0<x_{(n)}<\frac{1}{2};\\ \frac{\alpha-\frac{1}{2^n}}{1-\frac{1}{2^n}}, & \text{if } \frac{1}{2}\le x_{(n)}<1;\\ 0 & \text{otherwise.}\end{cases}$$

③ 当 $h(0)=1-\frac{1}{2^n}=1-\alpha$ 时,

$$\Phi(x)=\begin{cases}1, & \text{if } \lambda(x)>0;\\ 0 & \text{if } \lambda(x)<0.\end{cases}$$

10. 设$\alpha\in(0,1)$，且$H_0,H_1$是两个简单假设. 则$\phi$是假设$H_0\longleftrightarrow H_1$的水平$\alpha$的MP检验，且$\beta=E_{H_1}\phi(X)<1$，则$1-\phi$是假设$H_1\longleftrightarrow H_0$的水平为$1-\beta$的MP检验.

解：令$f_j,j=1,2$分别是样本在$H_0,H_1$下对应的概率分布，则检验$\phi$满足，

$$\phi(X)=\{_{0\qquad f_1(X)<cf_0(X)}^{1\qquad f_1(X)>cf_0(X)},$$

又$\beta=E_{H_1}\phi(X)<1$, 则c一定为一个正常量。则，

$$1-\phi(X)=\{_{0\qquad f_0(X)<c^{-1}f_1(X)}^{1\qquad f_0(X)>c^{-1}f_1(X)},$$

且$1-\phi(X)$的水平为$1-\beta$，则有N-P引理可证结论。

11. 设$X_1,\ldots,X_n$为来自两点分布$b(1,p)$的iid样本，试求假设

$$H_0:p=1/2\longleftrightarrow H_1:p<1/2$$

的水平$\alpha$的UMP检验.

解：$X_1,\cdots,X_n$ 的联合分布为:

$$f(X)=p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum X_i}=(1-p)^n exp\{ln\frac{p}{1-p}(\sum_{i=1}^n X_i)\}$$

由 5.4.2 可知, $Q = ln\frac{p}{1-p}$, $Q' = \frac{1}{p(1-p)} > 0$, 是关于 P 的严格单增函数, 故是关于$T(X) = \sum_{i=1}^{n} X_i$单增的 MLR, 有水平 $\alpha$ 的 UMPT:

$$\phi(T(X)) = \begin{cases} 1, & T(X) > c \\ \delta, & T(X) = c \\ 0 & T(X) < c \end{cases}$$

且$E_iH_0\phi(T(X)) = \alpha$，因$H_0$下，$T \sim B(n, 1/2)$，则，

$$E_{H_0}\phi(T(X)) = P_{H_0}(T(X) > c) + \delta \cdot P_{H_0}(T(X) = c) = \sum_{k>c} C_n^k(1/2)^n + \delta C_n^c(1/2)^n = \alpha$$

即有$\delta = \frac{\alpha - \sum_{k>c} C_n^k(1/2)^n}{C_n^c(1/2)^n}$，即$\phi(T(X))$为UMPT。

12.设$X_1, \ldots, X_n$为来自$\Gamma$分布$\Gamma(1, p)$的iid样本，试求假设

$$H_0 : p = p_0 \longleftrightarrow H_1 : p > p_0$$

的水平$\alpha$的UMP检验.

解：$X_1, \cdots, X_n$ 的联合分布为:

$$f(X) = \frac{1}{(\Gamma(p))^n} exp\{-\sum X_i + (p-1)\ln X_1 \cdots X_n\}$$

为单参数指数型分布族，且$T(X) = \sum \ln X_i$,且$Q(\theta) = \theta - 1$为$\theta$的严格单调增函数，则f(X)关于$T(X) = \sum \ln X_i$的MLR，则UMPT为，

$$\phi(T(X)) = \{ {}^{1}_{0} \quad {}^{T \geq c}_{T < c}$$

且c由$E_{H_0}(\phi(T(X))) = P\{T \geq c | p = p_0\}$决定。

13. 设$X_1, \ldots, X_n$为来自两点分布$b(1, p)$的iid样本，试求假设

$$H_0 : p = p_0 \longleftrightarrow H_1 : p \neq p_0$$

的水平$\alpha$的UMPU检验.

解：$x_1, \cdots, x_n$ 的联合分布为:

$$\begin{aligned} P\{X_1 = x_1, \cdots, X_n = x_n\} &= p^{\sum_{i=1}^{n} X_i}(1 - p^{n - \sum_{i=1}^{n} X_i}) \\ &= (\frac{p}{1-p})^{\sum_{1=1}^{n}} \cdot (1-p)^n \\ &= (1-p)^n exp\{ln\frac{p}{1-p} \cdot (\sum_{i=1}^{n} X_i)\} \end{aligned}$$

由 5.4.2 可知, $Q = ln\frac{p}{1-p}$, $Q' = \frac{1}{p(1-p)} > 0$, 是关于 P 的严格单增函数, 故是关于$T(x) = \sum_{i=1}^{n} x_i$单增的 MLR, 有水平 $\alpha$ 的 UMPT:

$$\Phi(T(x)) = \begin{cases} 1, & \text{if } T(x) < c_1 \text{ or } T(x) > c_2; \\ \delta_i, & \text{if } T(x) = c_i, i = 1, 2; \\ 0 & \text{if } c_1 < T(x) < c_2. \end{cases}$$

且 $E_{H_0}\Phi(x) = \alpha$, $E_{H_0}[T(x)\Phi(T(x)) = \alpha E_{\theta_0}T(x)$.
因为 $T(x) = \sum_{i=1}^{n} X_i \sim B(n, p)$, 故 $E_{H_0}T(x) = np_0$

$$\begin{aligned} E_{H_0}\Phi(T(x)) &= P_{H_0}(T(x) < c_1) + P_{H_0}(T(x) > c_2) + \delta_1 P_{H_0}(T(x) = c_1) + \delta_2 P_{H_0}(T(x) = c_2) \\ &= \alpha \end{aligned}$$

$$\begin{aligned} E_{H_0}[T(x)\Phi(T(x))] &= \sum_{k<c_1} kC_n^k p_0^k (1-p_0)^{n-k} + \sum_{c_2<k\le n} kC_n^k p_0^k (1-p_0)^{n-k} \\ &\quad + \delta_1 c_1 C_n^{c_1} p_0^{c_1} (1-p_0)^{n-c_1} + \delta_2 c_2 C_n^{c_2} p_0^{c_2} (1-p_0)^{n-c_2} \\ &= np_0\alpha \end{aligned}$$

故满足上面两式的 $c_1, c_2$ 即为所求.

14. 设$X_1, \ldots, X_n$是$n$个来自正态总体的相互独立的样本，且

$$EX_i = \begin{cases} \mu_1, & i = 1, \cdots, m, \\ \mu_0, & i = m+1, \cdots, n, \end{cases} \qquad \mathrm{Var}X_i = \sigma^2,\ i = 1, \cdots, n.$$

则关于假设
(1) $H_0 : \mu_1 \le \mu_0 \longleftrightarrow H_1 : \mu_1 \ge \mu_0$;
(2) $H_0 : \mu_1 = \mu_0 \longleftrightarrow H_1 : \mu_1 \ne \mu_0$,
存在水平$\alpha$的UMPU检验，且分别为

$$\phi_1(x) = \begin{cases} 1, & T(x) < c_1, \\ 0, & \text{其它}, \end{cases} \qquad \phi_2(x) = \begin{cases} 1, & |T(x)| > c_2, \\ 0, & \text{其它}, \end{cases}$$

其中

$$T(x) = \frac{\sqrt{n-m}(x_1 - \mu_0)}{\sqrt{\sum_{i=m+1}^{n} x_i^2}},$$

且当$\mu_1 = \mu_0$时，$T(X) \sim t(n-m)$.

15. 设甲乙各有资本$M, N$元($M, N$均为正整数)，且两人赌博时没有平局. 另设每一局若甲胜，则乙给甲元；若乙胜，则甲给乙一元，且设每局中甲胜的概率

为$p(0 < p < 1)$. 问：如果一局一局地赌下去，甲输光的概率是多少？平均几局后有一方输光？

解：令$X_n$表示赌了n局后甲手中的赌金，则$\{X_n, n \geq 0\}$是一个齐次马尔科夫链( MC )，状态空间$E = \{0, 1, \cdots, c\}$,状态0，c 为吸收态，其一步转移概率为

$$p_{ij} = \begin{cases} p & j = i + 1 \\ q & j = i - 1 \\ 0 & \text{其他} \end{cases} \qquad 0 < i < c, p_{00} = p_{cc} = 1$$

设$P_i$表示甲从状态i出发到破产的概率，显然有$P_0 = 1, P_c = 0$，对任意$0 \leq i \leq c - 1$，由全概率公式可得，

$$P_i = pP_{i+1} + qP_{i-1} \Leftrightarrow P_{i+1} - P_i = \beta(P_i - P_{i-1}), \beta = \frac{1-p}{p}$$

即有$P_i = \{^{\frac{\beta^i - \beta^c}{1-\beta}(1-P_1) \quad p \neq q}_{(c-i)(1-P_1) \quad p=q=1/2}$,

$$P_i = \{^{\frac{c-i}{c} \qquad p=q}_{\frac{\beta^i - \beta^c}{1-\beta}(1-P_1) \quad p \neq q}$$

$$P_a = \{^{\frac{N}{M+N} \qquad p=q}_{\frac{p^N(1-p)^M - (1-p)^{M+N}}{p^{M+N} - (1-p)^{M+N}} \quad p \neq q}$$

.

16. 设$X_1, \ldots, X_n$为来自密度函数为

$$f(x, \sigma) = \begin{cases} \frac{\sigma^3}{\sqrt{2\pi}} \sqrt{x} \exp\{-\sigma^2 x/2\}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

的总体的iid样本，其中$\sigma > 0$为未知参数. 对于给定的$0 < \sigma_0 < \sigma_1$，试求假设

$$H_0 : \sigma = \sigma_0 \longleftrightarrow H_1 : \sigma = \sigma_1$$

的序贯概率比检验.

解：总体的概率分布 $f(x, \sigma) = \frac{\sigma^{3n}}{(2\pi)^{\frac{n}{2}}} (\prod_{i=1}^n x^{\frac{1}{2}}) exp\{\frac{\sigma^2}{2} (\sum_{i=1}^n x_i)\}$

$$0 < \sigma_0 < \sigma_1, \ \lambda(x) = \frac{f(x,\sigma_1)}{f(x,\sigma_0)} = (\frac{\sigma_1}{\sigma_2})^{3n} exp\{\frac{\sigma_0^2 - \sigma_1^2}{2} \sum_{i=1}^n x_i\}$$

$$ln\lambda(x) = 3nln\frac{\sigma_1}{\sigma_0} - \frac{\sigma_1^2 - \sigma_0^2}{2} \sum_{i=1}^n x_i$$

记 $T_n=\sum_{i=1}^n x_i$, 且 $c=\frac{3ln\frac{\sigma_1}{\sigma_0}}{\frac{\sigma_1^2-\sigma_0^2}{2}}$, $\alpha_1=\frac{-lnB}{\frac{\sigma_1^2-\sigma_0^2}{2}}$, $\alpha_2=\frac{-lnA}{\frac{\sigma_1^2-\sigma_0^2}{2}}$.
$c>0$, $d_1>d_2$, 故对给定的 A, B, 有:

$$\lambda_n \geq B \Longleftrightarrow T_n \geq R_n = cn+d_1$$

$$\lambda_n \leq A \Longleftrightarrow T_n \leq A_n = cn+d_2$$

综上, 对该假设的 SPRT 的停时为:

$$T^X = min\{n : T_n \leq A_n \text{ 或 } T_n \geq R_n, n \geq 1\}.$$

17. 对于给定的$\alpha\in(0,1/2)$, 设常数$c$满足：$\frac{\alpha}{2(1-\alpha)}<c<\alpha$. 令$\Theta=\{-1\}\cup[0,1$, 设样本$X$的分布函数$p(x,\theta)$为：

| $X$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| $\theta=-1$ | $\frac{\alpha}{2}$ | $\frac{1-\alpha}{2}$ | $\alpha$ | $\frac{1-\alpha}{2}$ | $\frac{\alpha}{2}$ |
| $\theta\neq-1$ | $c\theta$ | $\frac{1-c}{1-\alpha}(\frac{1}{2}-\alpha)$ | $\frac{1-c}{1-\alpha}\alpha$ | $\frac{1-c}{1-\alpha}(\frac{1}{2}-\alpha)$ | $c(1-\theta)$ |

(1)求假设$H_0:\theta=-1\longleftrightarrow H_1:\theta\neq-1$的水平为$\alpha$的似然比检验;
(2) 证明当且仅当$X=0$时拒绝原假设的检验有水平$\alpha$，且无论$\theta$取什么值，它都具有更大的功效 (本例说明似然比检验未必是功效最大的).

18. 令$X_1,\cdots,X_m$ 和$Y_1,\cdots,Y_n$ 分别是来自正态总体$N(\mu_x,\sigma_x^2)$和 $N(\mu_y,\sigma_y^2)$ 的IID随机变量. 假设$X_i$和$Y_j$是相互独立的.
(1) 当$\sigma_x=\sigma_y=1$ 时，求

$$H_0:\mu_x\leq\mu_y \leftrightarrow H_1:\mu_x>\mu_y$$

水平为$\alpha$的UMP检验;
(2) 当$\mu_x$ 和$\mu_y$ 已知时，求

$$H_0:\sigma_x\leq\sigma_y \leftrightarrow H_1:\sigma_x>\sigma_y$$

水平为$\alpha$的UMP检验.
解: (1)$\bar{X}\sim(u_x,\frac{\delta_x^2}{m})$,$\bar{Y}\sim(u_y,\frac{\delta_y^2}{n})$

$$\bar{X}-\bar{Y}\sim(u_x-u_y,\frac{\delta_x^2}{m}+\frac{\delta_y^2}{n})$$

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim N(u,1)$$

则原假设化为: $H_0: u \le 0 \longleftrightarrow H_1: u > 0$.
$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}+\frac{1}{n}}}$ 的 PDF 为:

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u)^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}e^{-\frac{x^2}{2}}e^{xu}$$

可知是关于 x 的非降的 MCR 分布族, 故存在水平为 $\alpha$ 的 UMP 检验函数:

$$\Phi(x) = \begin{cases} 1, & \text{if } x \ge c; \\ 0 & \text{if } x < c. \end{cases}, \text{且 } E_{\theta_0}\Phi(x) = \alpha.$$

即 $P_{u=0}\{x \ge c\} = \alpha, c = u_\alpha$

$$\Longrightarrow \Phi(x) = \begin{cases} 1, & \text{if } \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}+\frac{1}{n}}} \ge u_\alpha; \\ 0 & \text{if } \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}+\frac{1}{n}}} < u_\alpha;. \end{cases}$$

(2) $u_x, u_y$ 已知且相等,

$$S_m^2 = \frac{(x_1-u_x)^2 + (x_2-u_x)^2 + \cdots + (x_m-u_x)^2}{\delta_x^2}$$

$$S_n^2 = \frac{(y_1-u_y)^2 + (y_2-u_y)^2 + \cdots + (y_n-u_y)^2}{\delta_y^2}$$

$$F = \frac{S_m^2/m}{S_n^2/n} \sim F(m,n)$$

所以原假设化为: $H_0: \frac{\delta_y^2}{\delta_x^2} \ge 1 \longleftrightarrow H_1: \frac{\delta_y^2}{\delta_x^2} < 1$.
设 $S^* = \frac{\delta_y^2}{\delta_x^2} \cdot \frac{S_m^2/m}{S_n^2/n}$, $S^*$ 的 PDF 为:

$$f(S^*, \frac{\delta_y^2}{\delta_x^2}) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \frac{m}{n} \cdot \frac{\delta_y^2}{\delta_x^2} \cdot (\frac{m}{n}\frac{\delta_y^2 x}{\delta_x^2})^{\frac{m}{2}-1} \cdot (1 + \frac{x\delta_y^2}{\delta_x^2})^{-\frac{m+n}{2}}$$

是关于 $x$ 的 MLR 非降分布族, 所以存在 $\alpha$ 的 UMP 检验

$$\Phi(S^*) = \begin{cases} 1, & \text{if } S^* \le c; \\ 0 & \text{if } S^* > c. \end{cases} \text{ 且 } E_{\theta_0}\Phi(S^*) = \alpha$$

即 $P_{\frac{\delta_y^2}{\delta_x^2=1}}\{S^* \le c\} = \alpha \Longrightarrow c = F_\alpha(m,n)$.

19. 设$X_1,\ldots,X_n$来自均匀分布$U(\theta,\theta+1),\theta\in\mathbb{R}$. 假设$n\geq 2$.
(1)求$X_{(1)}$和$X_{(n)}$ 的联合分布;
(2) 对于检验

$$H_0:\theta\leq 0\leftrightarrow H_1:\theta>0$$

存在水平$\alpha$的UMP检验，且

$$T(X_{(1)},X_{(n)})=\begin{cases}0, & X_{(1)}<1-\alpha^{1/n},X_{(n)}<1;\\ 1, & \text{否则}.\end{cases}$$

解：(1)$(X_{(1)},X_{(n)})$的联合密度函数为，

$$f_\theta(x,y)=n(n-1)(y-x)^{n?2}I_{(\theta,y)}(x)I_{(x,\theta+1)}(y)$$

(2)检验T的势函数为$\beta_T(\theta)=\int T(x,y)f_\theta(x,y)dxdy$，即可得，

$$\beta_T(\theta)=\begin{cases}0 & \theta<-\alpha^{1/n}\\ (\theta+\alpha^{1/n})^n & -\alpha^{1/n}\leq\theta\leq 0\\ 1+\alpha-(1-\theta)^n & 0<\theta\leq 1-\alpha^{1/n}\\ 1 & \theta>1-\alpha^{1/n}\end{cases}$$

对任意的$\theta_1\in(0,1-\alpha^{1/n}]$,由N-P引理，对于假设$H_0:\theta=0,H_1:\theta=\theta_1$，水平为$\alpha$的UMP 检验为

$$T_1=\begin{cases}1 & X_{(n)}>1\\ \alpha/(1-\theta_1)^n & \theta_1<X_{(1)}<X_{(n)}<1\\ 0 & \text{否则}.\end{cases}$$

$T_1$在$\theta_1$的势函数为$\beta_{T_1}(\theta_1)=1-(1-\theta_1)^n+\alpha$，跟T 在$\theta_1$ 处的势函数相同。当$\theta>1-\alpha^{1/n}$，T的势为1。所以，T是假设为$H_0:\theta\leq 0,H_1:\theta>0$的水平为$\alpha$ 的UMP 检验。

20. 假设$F$和$G$为$\mathbb{R}$上已知的累积分布函数，$X$是来自分布$\theta F(x)+(1-\theta)G(x)$的一个观测，这里$\theta\in[0,1]$是未知的.
(1)求关于检验

$$H_0:\theta\leq\theta_0\leftrightarrow H_1:\theta>\theta_0$$

的水平为$\alpha$的UMP检验;
(2) 试证对于检验$H_0:\theta\leq\theta_1$或$\theta\geq\theta_2\leftrightarrow H_1:\theta_1<\theta<\theta_2$，$T(X)\equiv\alpha$是一个水平

为$\alpha$的UMP检验.
解：(1) 假设f(x)和g(x)分别是对应于F(x)+G(x)的测度v的F(x)和G(x)的导函数，则X的密度函数为$\theta f(x) + (1-\theta)g(x)$. 对于 $0 \le \theta_1 < \theta_2 \le 1$,

$$\frac{\theta_2 f(x) + (1-\theta_2)g(x)}{\theta_1 f(x) + (1-\theta_1)g(x)} = \frac{\theta_2 \frac{f(x)}{g(x)} + (1-\theta_2)}{\theta_1 \frac{f(x)}{g(x)} + (1-\theta_1)}$$

关于$Y(x) = f(x)/g(x)$非降。则X的分布族是关于统计量$Y(x) = f(x)/g(x)$的MRL分布族，则UMP 检验，

$$T = \begin{cases} 1 & Y(X) > c \\ \gamma & Y(X) = c \\ 0 & Y(X) < c \end{cases}$$

c和$\gamma$可以由在$\theta = \theta_0$下,$E[T(X) = \alpha$确定。
(2)对于任意检验T，它的势为

$$\begin{aligned} \beta_T(\theta) &= \int T(x)[\theta f(x) + (1-\theta)g(x)dv \\ &= \int T(x)[f(x) - g(x)d\Theta + \int T(x)g(x)du \end{aligned}$$

上述势函数是$\theta$ 在$[0,1$的线性函数。如果T的水平为$\alpha$，则对任意$\theta \in [0,1, \beta_T(\theta) \le \alpha$ 。则对于检验T若其势函数等于$\alpha$，则此检验为水平为$\alpha$的UMP。

21. 在上题中，如果检验为

$$H_0 : \theta \in [\theta_1, \theta_2 \leftrightarrow H_1 : \theta \notin [\theta_1, \theta_2,$$

其中$0 < \theta_1 \le \theta_2 < 1$.
(1) 证明不存在UMP检验;
(2) 求一个水平为$\alpha$ 的UMPU检验.
解：(1)令$\beta T(\theta)$是检验T的势函数，对任意势函数不是常数的水平为$\alpha$的检验T，有$\beta T(0), \beta T(1)$均严格小于$\alpha$。不失一般性，假设$\beta T(\theta) < \alpha$。在$\theta = 0$，即在备选假设下，T的势函数要小于$T_\star \equiv \alpha$。所以，对任意势函数不是常数的水平为$\alpha$的检验T 都不是UMP。由20 题可知，对于假设$H_0 : \theta \le \theta_0, H_1 : \theta > \theta_1$ 的水平为$\alpha$ 的检验在$\theta = 1$的势都大于$\alpha$。所以，$T_\star \equiv \alpha$ 不是UMP。即不存在UMP。
(2)对于势函数不是常数的水平为$\alpha$的检验T，因为$\beta T(0), \beta T(1)$均严格小于$\alpha$，T一定不是无偏的。所以，只有势函数是常数的检验才是无偏检验。即$T_\star \equiv \alpha$是UMPU。

22. 假设$X_i = \beta_0 + \beta_1 t_i + \varepsilon_i$, 其中$t_i$是一些不完全相同的固定的常数， $\varepsilon_i$是来自正态总体$N(0, \sigma^2)$的IID随机变量，并且$\beta_0, \beta_1, \sigma^2$ 是未知参数. 对于下列检验求一个水平

为$\alpha$ 的UMPU检验.
(1) $H_0 : \beta_0 \le \theta_0 \leftrightarrow H_1 : \beta_0 > \theta_0$;
(2) $H_0 : \beta_0 = \theta_0 \leftrightarrow H_1 : \beta_0 \ne \theta_0$;
(3) $H_0 : \beta_1 \le \theta_0 \leftrightarrow H_1 : \beta_1 > \theta_0$;
(4) $H_0 : \beta_1 = \theta_0 \leftrightarrow H_1 : \beta_1 \ne \theta_0$.
解：注意到$(X_1, \cdots, X_n)$ 服从简单的线性回归模型，令$D = n\sum_{i=1}^n t_i^2 - (\sum_{i=1}^n)^2$, 则

$$\hat{\beta}_0 = \frac{1}{D}(\sum_{i=1}^n t_i^2 \sum_{i=1}^n X_i - \sum_{i=1}^n t_i \sum_{i=1}^n t_i X_i)$$

$$\hat{\beta}_1 = \frac{1}{D}(n\sum_{i=1}^n t_i X_i - \sum_{i=1}^n t_i \sum_{i=1}^n X_i)$$

是最小二乘估计，令$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{\beta}_0\hat{\beta}_1 t_i)^2$，则$\hat{\beta}_0$ 服从均值为$\beta_0$，方差为$\sigma^2 \sum_{i=1}^n t_i^2/D$的正态分布，$\hat{\beta}_1$ 服从均值为$\beta_1$，方差为$\sigma^2 n/D$的正态分布，$(n?2)\hat{\sigma}^2/\sigma^2$ 服从$\chi^2_{n-2}$，且$(\hat{\beta}_0, \hat{\beta}_1)$和$\sigma^2$ 相互独立。
(1)水平为UMPU的检验为$T = \{^0_1 \quad {}^{t_0 > t_{n-2,\alpha}}_{\text{其他}}$，其中$t_0 = \frac{\sqrt{D}(\hat{\beta}_0 - \theta_0)}{\hat{\sigma}\sqrt{\sum_{i=1}^n t_i^2}}$;
(2)水平为UMPU的检验为$T = \{^0_1 \quad {}^{|t_0| > t_{n-2,\alpha}}_{\text{其他}}$;
(3)水平为UMPU的检验为$T = \{^0_1 \quad {}^{t_1 > t_{n-2,\alpha}}_{\text{其他}}$，其中$t_1 = \frac{\sqrt{D}(\hat{\beta}_1 - \theta_0)}{\sqrt{n}\hat{\sigma}}$;
(4)水平为UMPU的检验为$T = \{^0_1 \quad {}^{|t_1| > t_{n-2,\alpha}}_{\text{其他}}$。

# 6 习题六

1. 请检验下面的20个数据是否来自$N(0, 1)$:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $-2.4$ | $-2.1$ | $-1.2$ | $-0.7$ | $-0.4$ | $-0.3$ | $-0.2$ | $-0.1$ | $-0.05$ | $-0.02$ |
| 0.01 | 0.04 | 0.15 | 0.2 | 0.4 | 0.8 | 1.4 | 2.2 | 2.8 | 3.1 |

解: 由于所检验分布完全已知, 并且为连续型分布, 我们可以采用 6.4 节 Kolmogorov 检验; 类似例 6.4.1, 计算得 $D_n = 0.1446 < 0.2645 = D_{0.1}(20)$, 因而在显著性水平 0.1 下, 我们不能拒绝该数据来自正态分布的假设.

2. 在$\pi$的前800位小数的数字中，0至9十个数字分别出现了74，92，83，79，80，73，77，75，76和91次. 试在水平0.05下检验这十个数字出现的可能性相等.
解: 我们要检验 $H_0 : p_i = 1/10, i = 0, 1, \ldots, 9$. 采用 6.2 节 $\chi^2$ 检验, 检验统计量为$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_{i0})^2}{np_{i0}} = 5.125$, 其中 $r = 10, p_{i0} = 1/10$; 从而我们可以计算检验的 $p$ 值为 0.823 , 则我们没有理由拒绝这 10 个数字等可能出现的假设.

3. 在著名的Rutherford-Chadwik-Ellis实验中，观察了一放射性物质每1/8分钟内放射的$\alpha$粒子数，他共观察了2612次，结果如下：

| 粒子数： | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 频数： | 57 | 203 | 383 | 525 | 532 | 408 | 273 | 139 | 49 | 27 | 10 | 6 |

请问在水平0.1下，上述数据是否与Poisson分布相符？

解: 采用 $\chi^2$ 检验. 首先, 我们简单取样本均值来估计泊松分布的均值 $\lambda$, 得到 $\hat{\lambda} = \bar{X} = 3.875957$. 然后, 我们由泊松分布 $P(\hat{\lambda})$ 计算理论概率 $p_{i0}, i = 1, \ldots, 12$. 从而得到检验统计量 $\chi^2 = 10.42$, $p$ 值为 0.40, 注意到这里自由度取 10. 因而我们可以认为数据是来自泊松分布的.

4. 设总体分布关于原点对称，且$X_1, \ldots, X_n$为来自此总体的iid样本，$X_{(1)} \leq \cdots \leq X_{(n)}$ 为次序统计量，并记

$$EX_{(i)} = m_i, \;\; \mathrm{Cov}(X_{(i)}, X_{(j)}) = v_{ij}, \; i, j = 1, \cdots, n.$$

请证明：

(1)$m_i = -m_{n+1-i}$，且$\sum_{i=1}^n m_i = 0$. (2) $v_{ij} = v_{n+1-i, n+1-j}$，且$V = (v_{ij})$不仅关于主对角线对称，而且还关于副对角线对称.

证: 假设 $F$ 绝对连续且定义域为 R.

(1) 易知, $X_{(i)}$ 的密度函数为

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1}(1-F(x))^{n-i} f(x),$$

则

$$m_i = \mathrm{E}X_{(i)} = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x F(x)^{i-1}(1-F(x))^{n-i} f(x) \mathrm{d}x.$$

注意到 $F$ 关于原点对称, 我们有 $F(-x) = 1 - F(x)$ 且 $f(-x) = f(x)$. 从而,

$$\begin{aligned}
m_{n+1-i} &= \frac{n!}{(n-i)!(i-1)!} \int_{-\infty}^{\infty} x F(x)^{n-i}(1-F(x))^{i-1} f(x) \mathrm{d}x \\
&= \frac{n!}{(n-i)!(i-1)!} \int_{-\infty}^{\infty} x (1-F(-x))^{n-i}(F(-x))^{i-1} f(-x) \mathrm{d}x \\
&= \frac{n!}{(n-i)!(i-1)!} - \int_{-\infty}^{\infty} (-y)(1-F(y))^{n-i}(F(y))^{i-1} f(y) \mathrm{d}(-y) \\
&= -\frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} y (F(y))^{i-1}(1-F(y))^{n-i} f(y) \mathrm{d}y \\
&= -m_i.
\end{aligned}$$

进一步, $\sum m_i = 0$.
(2) 只需知道 $X_{(i)}, X_{(j)}$ 的联合密度函数为

$$f_{(X_{(i)},X_{(j)})}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times$$
$$F(x)^{i-1}(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j}f(x)f(y), x<y,$$

并注意到 $F$ 关于原点对称即可, 其余步骤类似(1).

7. 为研究慢性气管炎与吸烟量的关系，现调查了272人，其结果如下：

| 是否患病 | 吸烟量(支/日) | | | 和 |
|---|---|---|---|---|
| | 0–9 | 10–19 | $\geq 20$ | |
| 是 | 22 | 98 | 25 | 145 |
| 否 | 22 | 89 | 16 | 127 |
| 和 | 44 | 187 | 41 | 272 |

请问慢性气管炎与每日的吸烟量有关吗？($\alpha = 0.05$)
解: 采用 6.3 节列联表检验. 检验统计量 $\chi^2 = 1.2229 < 5.9915 = \chi^2_{0.05}(2)$, 则我们认为慢性气管炎与每日吸烟量无关.

8. 消费者转会为了解消费者对市场上五种矿泉水的偏好，现随机地调查了1000人，得到如下数据：

| 品牌 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 喜欢的人数 | 210 | 212 | 170 | 185 | 223 |

请问调查结果是否说明消费者对这五种品牌的矿泉水存在着不同的偏好？($\alpha = 0.05$)
解: 与第 2 题类似. 检验统计量 $\chi^2 = 9.49$, $p$ 值=0.04995; 则我们认为消费者对五种品牌存在某种偏好.

9. 假设$F$为$\mathbb{R}$上的连续分布函数，记 $D_n^+(F) = \sup_{x\in\mathbb{R}}[F_n(x) - F(x)$， $D_n^-(F) = \sup_{x\in\mathbb{R}}[F(x) - F_n(x)]$. 证明：对于固定的$n$，

$$P(D_n^+(F) \leq t) = \begin{cases} 0, & t \leq 0, \\ n!\prod_{i=1}^n \int_{\max(0,\frac{n-i+1}{n}-t)}^{u_{n-i+2}} du_1 \dots du_n, & 0 < t < 1, \\ 1, & t \geq 1, \end{cases}$$

且$D_n^-(F)$与$D_n^+(F)$分布相同.
证: 由 6.4 节知: $D_n^+(F) = \max_{1\leq i\leq n}\{\frac{i}{n} - F(X_{(i)})\}$. 假设 $X \sim F(\cdot)$, 则 $F(X) \sim U(0,1)$.

记 $U_i = F(X_{(i)})$, 则 $D_n^+(F) = \max_{1\le i\le n}\{\frac{i}{n} - U_i\}$. 注意到 $U_1, \dots, U_n$ 的联合密度函数为

$$g(u_1, \dots, u_n) = n!, \quad \text{如果} \quad u_1 < \dots < u_n,$$

简单的积分即可得到所证.

# 7 习题七

1. 说明为什么复合抽样法能够产生我们想要的随机数.
解: 注意到条件概率公式 $p_X(x) = p_{X|Y}(x|y)p_Y(y)$ 即可.

2. 假设$x_1, \dots, x_n$是一组来自未知连续分布函数$F$的简单随机样本，现在我们想要检验

$$H_0: \text{ 分布}F\text{的中位数为零} \quad versus \quad H_1: \text{ 分布}F\text{的中位数大于零}.$$

考虑使用著名的非参数符号检验统计量$T_n = \sum_{i=1}^n I(X_i > 0)$，显然当 $T_n$ 很大时拒绝．这是一个与分布无关的检验（也就是其零分布不依赖于$F$ ）．请描述使用Monte Carlo模拟方法来求取该检验的$p$值的具体步骤.
解: 注意到该检验与分布无关, 我们可以取 $F$ 为任意一个连续型分布, 比如说, 标准正态分布. 进一步, 考虑 $N$ 次独立的模拟, 每次均从标准正态分布中独立地抽取 $n$ 个样本, 据此计算检验统计量 $T$ 的值; 这样 $N$ 次模拟我们产生 $N$ 个检验统计量 $T_1, \dots, T_N$. 另一方面, 我们由 $n$ 个样本 $X_1, \dots, X_n$ 计算检验统计量的值, 并记为 $T_{\text{obs}}$. 则我们所求的 $p$ 值可以近似为 $\frac{1}{N}\sum_{i=1}^N \mathbf{1}_{\{T_i > T_{\text{obs}}\}}$.

3. 考虑例7.2.2 中的指数加权移动平均方法，现固定 $\lambda = 0.1$，试用模拟方法求 $B$ 使得在可控状态时（即数据始终来自标准正态分布）$E(N)$ 达到500(使用二分法时可控制$|E(N) - 500| \le 3$ 即可，模拟次数推荐至少 2000.
解: 0.6435.

4. 编写使用逆变换法生成密度函数为$\lambda e^{-\lambda x}(x > 0)$ 的指数分布随机数的程序; 自己给定某一$\lambda$值，生成100 个独立的该随机数，使用Kolmogorov-Smirnov 进行拟合优度检验，判断生成的样本是否服从参数为$\lambda$的指数分布
解: 假设 $U \sim U(0,1)$, 则 $-\frac{\log U}{\lambda} \sim E(\lambda)$.

5. 如何使用逆变换法生成几何分布和Cauchy分布的随机变量.
解: (1) 假设 $X \sim G(p)$, 即: $P(X = k) = (1-p)^{k-1}p, k = 1, 2, \dots$. 易得 $F(k) = 1 - (1-p)^k$. 由逆变换法, 我们可以生成 $U \sim U(0,1)$, 然后根据 $\log(U)/\log(1-p)$ 生成几何分布.

(2) 标准 Cauchy 分布的 CDF 为 $F(x)=\frac{1}{2}+\arctan(x)/\pi$. 由逆变换法, 我们可以生成 $U\sim U(0,1)$, 然后根据 $\tan(\pi(U-\frac{1}{2}))$ 生成标准 Cauchy 分布.

8. 假设$X\sim \text{Binomial}(100,0.25)$. 请编写程序使用重要抽样法计算尾部概率$P(X\geq 75)$.

解: 首先我们可以计算 $\mathrm{P}(X\geq 75)=1.56e^{-26}$; 也就是说平均模拟 $6.4e^{25}$ 次, 我们才能观测到 1 个大于 75 的 $X$. 采用重要抽样法, 记 $h(X)=\mathbf{1}_{\{X\geq 75\}}$, $X\sim f(x)$ (离散情形下指概率质量函数). 考虑 $X$ 的矩母函数 $M(t)=\mathrm{E}_f e^{tX}=(pe^t+1-p)^n$. 引入 $f_t(x)=\frac{e^{tX}f(x)}{M(t)}$, 则

$$\begin{aligned}\theta=\mathrm{E}h(x) &= \int h(x)f(x)\mathrm{d}x\\ &= \int h(x)e^{-tx}M(t)\mathrm{d}x\\ &= \mathrm{E}_{f_t}h(x)e^{-tx}M(t).\end{aligned}$$

所以 $\theta$ 的一个无偏估计为:

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\geq 75\}}e^{-tX_i}(pe^t+1-p)^n,$$

其中 $X_1,\ldots,X_n\overset{\text{iid}}{\sim}f_t(x)$, 事实上 $f_t(x)\sim\text{Binomial}(n,p_t)$, 这里 $p_t=\frac{pe^t}{pe^t+1-p}$. 取 $t=2.1972$, 模拟 5000 次, 最终我们得到 $\hat{\theta}=1.5e^{-25}$.

10. 请计算使用对偶变量抽样法估计积分$\int_0^1 e^x dx$所降低的方差大小(相对于样本均值法). 解: 样本均值法: $\hat{\theta}_1=\frac{1}{2n}\sum_{i=1}^{2n}e^{U_i}$, $U_i\overset{\text{iid}}{\sim}U(0,1)$, 且易知 $\mathrm{Var}\hat{\theta}_1=0.121/n$.
对偶抽样法: $\hat{\theta}_2=\frac{1}{2n}\sum_{i=1}^{n}(e^{U_i}+e^{1-U_i})$, $U_i\overset{\text{iid}}{\sim}U(0,1)$. 且易知 $\mathrm{Cov}(e^U,e^{1-U})=-0.234$, $\mathrm{Var}e^U=0.242$,
从而 $\mathrm{Var}\hat{\theta}_2=0.004$.
则相比于样本均值法, 对偶抽样法估计该积分的方差降低了 96.7%.

# 8 第八章

1. 假设总体均值为$\mu$，我们感兴趣的是估计$\mu^2$，现考虑直接使用$\bar{X}^2$ 作为估计，请问如何使用bootstrap方法进行偏差修正?

解：对$X_1,\ldots,X_n$进行$B$次有放回抽样，得到的样本记为

$$\tilde{X}_1^{(i)},\ldots,\tilde{X}_n^{(i)},i=1,\ldots,B,$$

得到的样本均值的平方记为$\tilde{X}^{(1)2},\ldots,\tilde{X}^{(B)2}$ 则偏差修正后的boostrap估计值为$2\bar{X}^2-\frac{\sum_{i=1}^{B}\tilde{X}^{(i)2}}{B}$.

2. 假设我们有一组来自于形状参数为$\alpha$、尺度参数为1的gamma分布的简单随机样本$Y_1,\ldots,Y_n$，
(1)请描述如何使用参数bootsrap方法获得关于$\sqrt{\alpha}$的下95%单侧置信区间;
(2) 我们如何构造一个置信水平为0.05的Monte Carlo检验$H_0:\sqrt{\alpha}\le 10$ vs $H_1\sqrt{\alpha}>10$.
(3) 上面两问的解案之间有什么关系？
解:(1)样本均值$\bar{Y}$是参数$\alpha$的一个好的估计，记$\hat{\alpha}=\bar{Y}$,我们从$\Gamma(\hat{\alpha},1)$中抽取$B$次i.i.d 样本$Y_1^{(i)},\ldots,Y_n^{(i)},i=1,\ldots,B$ 则得到$\alpha$的相合估计为

$$\sqrt{\bar{Y}^{(1)}},\ldots,\sqrt{\bar{Y}^{(B)}}.$$

则95%的置信区间为$(0,\sqrt{\bar{Y}^{(B)}}_{([0.95B)}$,其中$\sqrt{\bar{Y}^{(B)}}_{([0.95B)}$ 表示第[0.95$B$小的统计量。
(2).从$\Gamma(100,1)$中抽取i.i.d样本$Y_1^{(i)},\ldots,Y_n^{(i)},i=1,\ldots,B$, 得到样本均值的平方根为$\sqrt{\bar{Y}^{(1)}},\ldots,\sqrt{\bar{Y}^{(B)}}$, 则该检验的$p$值为$\frac{\#\{i,\sqrt{\bar{Y}^{(i)}}\ge\sqrt{\bar{Y}}\}}{B}$. 若该$p$值小于0.05则我们拒绝原假设，否则接受。
(3)如果10落入(1)中的置信区间，我们就接受$H_0$，否则拒绝。 3假设我们观测到来自某一连续分布的成对数据$\{X_i,Y_i\}_{i=1}^{n}$. 我们感兴趣的是$H_0$：在$X\le 100$的条件下$Y$的中位数等于50. 请考虑如何使用经验似然方法构造检验统计量.
解：首先记$X$小于等于100的个数为$n_1$,再记这$n_1$个$X$对应的$Y$大于50的有$m_1$ 个，小于五十的有$m_2$个，则记对应于这些$X$的取值概率分别为$p_1,\ldots,p_{m_1}$,
$p_{m_1+1},\ldots,p_{m_1+m_2}$,及$p_{m_1+m_2+1},\ldots,p_{n_1}$，另一些$X$ 的取值概率为$p_{n_1+1},\ldots,p_n$则经验似然为

$$R(\mu_{\text{med}})=\max\{\prod np_i|0\le p_i\le 1,\sum_{i=1}^{m_1}p_i=\sum_{i=1}^{m_2}p_{i+m_1}\}$$

4. 继续考虑习题8.3中的问题. 假设$n=100$，且共有70个$X_i\le 100$，而这70个小于等于100的$X_i$所对应的$Y_i$中共有 30个小于50而另外40个大于50. 请问在$H_0$下极大经验似然比的值是多少？而那30个大于100的$X_i$所对应的样本在最大化的分布下获得的权重$(\widehat{p}_i)$是多少？
解:由第3题知，此时的经验似然为

$$R(\mu_{med})=\max\{\prod_{i=1}^{100}p_i|0\le p_i\le 1,\sum_{i=1}^{40}p_i=\sum_{i=41}^{70}p_i\}$$

记$\sum_{i=1}^{40}X_i=x$,则$\sum_{i=41}^{70}X_i=x$ , $\sum_{i=71}^{100}x_i=1-2x$,则易知$x^{70}(1-2x)^{30}$的最大值在$\hat{x}=\frac{7}{20}$处取得。则又知给定$x$的条件下，使经验似然达到最大值的 $p_i$为$\hat{p}_i=\frac{7}{800},i=$

$1,\ldots,40$,$\hat{p_i}=\frac{7}{600}$,$i=41,\ldots,70$,$\hat{p_i}=\frac{1}{100}$. 即那三十个大于100的$X_i$所对应的样本在最大化分布下的权重为$\frac{1}{100}$.

5. 假设$X_1,\ldots,X_n$是来自某一未知连续分布的简单随机样本. 请问如何构造经验似然比检验$H_0:\mu=\mu_0$ 且 $\sigma^2=\sigma_0^2$，其中$\mu$和$\sigma^2$分别代表其期望和方差.
解：经验似然为

$$R(\mu,\sigma^2)=\max\{\prod np_i|0\le p_i\le 1,i=1,\ldots,n,\sum_{i=1}^n p_iX_i=\mu_0,\sum_{i=1}^n p_i(X_i-\sum_{i=1}^n X_ip_i)^2=\sigma_0^2\}$$

极限分布为$\chi^2(2)$,则$R(\mu,\sigma^2)>\chi^2_{1-\alpha}(2)$则我们拒绝原假设，否则接受原假设。