数理统计 25 期末

一、填空及判断题（15 分）

1. 设 $X_1, \cdots, X_n, X_{n+1} \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$，若

$$\frac{\overline{X_n} - X_{n+1}}{\sqrt{b_n \sum_{i=1}^n X_i^2 - c_n (\overline{X_n})^2}} \sim t_{n-1}$$

其中 $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$，则 $b_n =$__________，$c_n =$__________。

答案：$b_n = \frac{n+1}{n(n-1)} \quad c_n = \frac{n+1}{n-1}$

先将分子归一化为 $N(0,1)$

$$\overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \overline{X_n} - X_{n+1} \sim N\left(0, \frac{(n+1)\sigma^2}{n}\right)$$

则

$$\frac{\overline{X_n} - X_{n+1}}{\sqrt{\frac{(n+1)\sigma^2}{n}}} \sim N(0,1)$$

再处理分母

$$\frac{\sum\limits_{i=1}^n \left(X_i - \overline{X_n}\right)^2}{\sigma^2} \sim \chi^2(n-1)$$

其中

$$\begin{aligned}\sum_{i=1}^n \left(X_i - \overline{X_n}\right)^2 &= \sum_{i=1}^n X_i^2 - 2\overline{X_n}\sum_{i=1}^n X_i + n\overline{X_n}^2 \\ &= \sum_{i=1}^n X_i^2 - n\overline{X_n}^2\end{aligned}$$

则

$$\frac{\sum\limits_{i=1}^n X_i^2 - n\overline{X_n}^2}{(n-1)\sigma^2} \sim \frac{\chi^2(n-1)}{n-1}$$

也就是

$$\frac{\frac{\overline{X_n} - X_{n+1}}{\sqrt{\frac{(n+1)\sigma^2}{n}}}}{\sqrt{\frac{\sum\limits_{i=1}^n X_i^2 - n\overline{X_n}^2}{(n-1)\sigma^2}}} \sim t_{n-1}$$

其中分子分母的独立性来自正态总体的 $\overline{X_n}$ 与 S^2 独立
化简得到

$$\frac{\overline{X_n}-X_{n+1}}{\sqrt{\frac{n+1}{n(n-1)}\sum_{i=1}^{n}X_i^2-\frac{n+1}{n-1}(\overline{X_n})^2}}\sim t_{n-1}$$

就有

$$b_n=\frac{n+1}{n(n-1)}\quad c_n=\frac{n+1}{n-1}$$

2. 设 $X\sim N(\mu,2)$，求 μ 的 95% 置信区间长度小于 1 所需的最小样本量__________。
答案：$n\geq 31$
已知方差，则置信区间为

$$\left[\overline{X}-\sqrt{\frac{2}{n}}u_{\alpha/2},\overline{X}+\sqrt{\frac{2}{n}}u_{\alpha/2}\right]$$

代入数值即得

3.×

4.×

5.×

6.✓;✓;×

7. 样本量小导致有些格子计数小于 5

二、（15 分）证明来自总体分布 $U(\theta,2\theta)$ 的 n 个样本 $(X_1,...,X_n)$ 的统计量 $(X_{(1)},X_{(n)})$ 是极小充分统计量但不是完全统计量。
解：均匀总体下的充分完全统计量.
通过因子分解定理说明充分统计量：根据题意有样本联合密度

$$f(\boldsymbol{x};\theta)=\frac{\mathbb{I}_{\{\theta<x_i<2\theta,i=1,2,\dots,n\}}}{\theta^n}=\theta^{-n}\mathbb{I}_{\left\{x_{(n)}/2<\theta<x_{(1)}\right\}}.$$

因此由因子分解定理知 $T=\left(X_{(1)},X_{(n)}\right)$ 是 θ 的充分统计量.
通过定理 2.6.2 说明极小充分统计量：对任意 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，

$$\frac{f(\boldsymbol{x};\theta)}{f(\boldsymbol{y};\theta)}=\frac{\mathbb{I}_{\left\{x_{(n)}/2<\theta<x_{(1)}\right\}}}{\mathbb{I}_{\left\{y_{(n)}/2<\theta<y_{(1)}\right\}}}$$

要使得上式与 θ 无关，当且仅当

$$\left(x_{(1)},x_{(n)}\right)=\left(y_{(1)},y_{(n)}\right)$$

因此由定理 2.6.2 知 $T=\left(X_{(1)},X_{(n)}\right)$ 是 θ 的极小充分统计量.
构造函数或利用辅助统计量说明不完全性：
法一: 注意到

$$\mathbb{E}_\theta\left[X_{(1)}\right]=\theta+\frac{1}{n+1}\theta=\frac{n+2}{n+1}\theta,\quad \mathbb{E}_\theta\left[X_{(n)}\right]=\theta+\frac{n}{n+1}\theta=\frac{2n+1}{n+1}\theta,$$

上式来自 n 个独立同分布的 $\mathrm{U}(0,1)$ 的次序统计量满足

$$U_{(1)}\sim\beta(1,n)\quad U_{(n)}\sim\beta(n,1)$$

取 $g\left(x_{(1)}, x_{(n)}\right)=(2n+1)x_{(1)}-(n+2)x_{(n)}$ ，则

$$\mathbb{E}_\theta\left[g\left(X_{(1)}, X_{(n)}\right)\right]=0, \quad \forall \theta>0,$$

但函数 g 并不几乎处处为零。因此 $\left(X_{(1)}, X_{(n)}\right)$ 不是 θ 的完全统计量.
法二: 注意到 $X_i/\theta \sim \mathrm{U}(1,2)$ ，构造辅助统计量

$$T=\frac{X_{(n)}}{X_{(1)}}=\frac{X_{(n)}/\theta}{X_{(1)}/\theta},$$

于是 T 的分布与 θ 无关，是一个辅助统计量. 因此 $\left(X_{(1)}, X_{(n)}\right)$ 不是 θ 的完全统计量.

三、(15 分) 设 X 的概率分布为:

X	1	2	3
$P(X)$	2θ	3θ	$1-5\theta$

取出 20 个样本，其中有 6 个取值为 1, 有 8 个取值为 2, 有 6 个取值为 3
(1) 求 θ 的最大似然估计和矩估计，并判断该估计是否无偏；
(2) 求 θ 的一致最小方差无偏估计 (UMVUE)，并比较其方差与 C-R 下界。
解：(1) 矩估计：根据题意有

$$\mathbb{E}[X]=1\times 2\theta+2\times 3\theta+3\times(1-5\theta)=3-7\theta.$$

因此反解得 θ 的矩估计量为

$$\hat{\theta}_M(\boldsymbol{X})=\frac{3-\overline{X}}{7}$$

由 $\overline{x}=\frac{1\times 6+2\times 8+3\times 6}{20}=2$ 知 θ 的矩估计值为

$$\hat{\theta}_M(\boldsymbol{x})=1/7$$

而

$$\mathbb{E}_\theta\left[\hat{\theta}_M(\boldsymbol{X})\right]=\mathbb{E}_\theta\left(\frac{3-\overline{X}}{7}\right)=\frac{3-\mathbb{E}(\overline{X})}{7}=\theta.$$

所以矩估计 $\hat{\theta}_M(\boldsymbol{X})$ 是 θ 的无偏估计.
最大似然估计：记 $n_j=\sum_{i=1}^n \mathbb{I}_{(X_i=j)}, j=1,2,3$ 。由分布列知 θ 的似然函数为

$$L(\theta)=(2\theta)^{n_1}(3\theta)^{n_2}(1-5\theta)^{n_3}=2^{n_1}3^{n_2}\theta^{n-n_3}(1-5\theta)^{n_3}.$$

由对数似然方程

$$\frac{\partial \log L(\theta)}{\partial \theta}=\frac{n-n_3}{\theta}-\frac{5n_3}{1-5\theta}=0 \Rightarrow \theta=\frac{n-n_3}{5n}.$$

经微分检验其确为似然函数的最大值点，因此 θ 的最大似然估计为

$$\hat{\theta}_L(\boldsymbol{X})=\frac{1}{5n}\sum_{i=1}^n \mathbb{I}_{(X_i\neq 3)}$$

由样本观测值知其估计值为

$$\hat{\theta}_L(\boldsymbol{x})=0.14$$

而

$$\mathbb{E}_\theta\left[\hat{\theta}_L(\boldsymbol{X})\right]=\mathbb{E}_\theta\left(\frac{1}{5n}\sum_{i=1}^n \mathbb{I}_{(X_i\neq 3)}\right)=\frac{1}{5n}\sum_{i=1}^n(5\theta)=\theta.$$

所以最大似然估计 $\hat{\theta}_L(\boldsymbol{X})$ 也是 θ 的无偏估计.

（2）一致最小方差无偏估计：样本联合密度函数为

$$f(\boldsymbol{x};\theta)=(2\theta)^{n_1}(3\theta)^{n_2}(1-5\theta)^{n_3}=2^{n_1}3^{n_2}\theta^n\exp\left\{n_3\log\frac{1-5\theta}{\theta}\right\}$$

自然参数空间 $\Theta^*=\{\eta:-\infty<\eta<\infty\}$ 有内点，于是 $T(\boldsymbol{X})=n_3$ 是 θ 的充分完全统计量．从而由 Lehmann-Scheffé 定理知，$\hat{\theta}_L(\boldsymbol{X})$ 作为基于 n_3 的无偏估计是 θ 的 UMVUE，其方差

$$\mathrm{Var}_\theta\left[\hat{\theta}_L(\boldsymbol{X})\right]=\frac{\mathrm{Var}_\theta\left(I\left(X_1\neq 3\right)\right)}{25n}=\frac{5\theta(1-5\theta)}{25n}=\frac{\theta(1-5\theta)}{5n}.$$

而 n 个样本的 Fisher 信息量

$$I(\theta)=-\mathbb{E}_\theta\left[\frac{\partial^2\log L(\theta)}{\partial\theta^2}\right]=\mathbb{E}_\theta\left[\frac{n-n_3}{\theta^2}-\frac{25n_3}{(1-5\theta)^2}\right]=\frac{5n\theta}{\theta^2}-\frac{25n(1-5\theta)}{(1-5\theta)^2}=\frac{5n}{\theta(1-5\theta)}.$$

因而 Cramér-Rao 下界为

$$1/I(\theta)=\mathrm{Var}_\theta\left[\hat{\theta}_L(\boldsymbol{X})\right]$$

即一致最小方差无偏估计 $\hat{\theta}_L(\boldsymbol{X})$ 的方差达到 θ 的无偏估计方差的下界.

四、（12 分）设 $(X_1,...,X_m)$ 来自某总体分布；$(Y_1,...,Y_n)$ 来自另外的总体分布；设 $\overline{X}=70$，$\overline{Y}=80$，样本容量 $n=m=26$，样本方差 $S_1^2=10$，$S_2^2=12$。

（1）假设总体服从正态分布：判断两总体方差是否相同；判断两总体均值是否相同。

（2）若去掉正态性假设，检验两总体均值是否相同

解：（1）

i. 两正态总体方差齐性检验：假设检验问题为

$$H_0:\sigma_1^2=\sigma_2^2\longleftrightarrow H_1:\sigma_1^2\neq\sigma_2^2.$$

取检验统计量为

$$F(\boldsymbol{X},\boldsymbol{Y})=\frac{S_2^2}{S_1^2}\overset{H_0}{\sim}F_{n-1,m-1}$$

由 F 检验知水平 $\alpha=0.2$ 的检验拒绝域为

$$D=\left\{(\boldsymbol{X},\boldsymbol{Y}):F(\boldsymbol{X},\boldsymbol{Y})>F_{25,25}(0.1)\text{ 或}F(\boldsymbol{X},\boldsymbol{Y})<F_{25,25}(0.9)=\frac{1}{F_{25,25}(0.1)}\right\}.$$

现有

$$F(\boldsymbol{x},\boldsymbol{y})=s_2^2/s_1^2=1.2\in(1/1.68,1.68)$$

所以不能拒绝原假设，可认为方差齐性.

ii. 两正态总体均值差的检验：假设检验问题为

$$H_0:\mu_1=\mu_2\longleftrightarrow H_1:\mu_1\neq\mu_2.$$

由于我们认为方差齐性，所以可取检验统计量为

$$T(\boldsymbol{X},\boldsymbol{Y})=\frac{\overline{Y}-\overline{X}}{S_w\sqrt{1/m+1/n}}\overset{H_0}{\sim}t_{m+n-2}$$

其中 $(m+n-2)S_w^2=(m-1)S_1^2+(n-1)S_2^2$ ．由 t 检验知水平 $\alpha=0.05$ 的检验拒绝域为

$$D=\{(\boldsymbol{X},\boldsymbol{Y}):|T(\boldsymbol{X},\boldsymbol{Y})|>t_{50}(0.025)\}$$

现有

$$T(\boldsymbol{x},\boldsymbol{y})=\frac{80-70}{\sqrt{11}\times\sqrt{1/13}}=10.9>2.01$$

所以拒绝原假设，认为两班学生成绩均值有显著差异.

（2）两一般总体均值差的检验：假设检验问题仍为

$$H_0:\mu_1=\mu_2\longleftrightarrow H_1:\mu_1\neq\mu_2.$$

此时可用大样本检验，取检验统计量为

$$Z(\boldsymbol{X},\boldsymbol{Y})=\frac{\overline{Y}-\overline{X}}{\sqrt{S_1^2/m+S_2^2/n}}\xrightarrow{H_0}N(0,1),\quad \text{as } m,n\to\infty.$$

由 z 检验知水平 $\alpha=0.05$ 的检验拒绝域为

$$D=\{(\boldsymbol{X},\boldsymbol{Y}):|Z(\boldsymbol{X},\boldsymbol{Y})|>u_{0.025}\}.$$

现有

$$Z(\boldsymbol{x},\boldsymbol{y})=\frac{80-70}{\sqrt{10/26+12/26}}=10.9>1.96$$

所以拒绝原假设，认为两班学生成绩均值有显著差异.

五、（15 分）设有 n 个样本 $(X_1,...,X_n)$ 来自 $\mathrm{Beta}(2,\beta)$ 分布。检验假设：

$$H_0:\beta=1\quad\leftrightarrow\quad H_1:\beta\neq1$$

求似然比检验，Wald 检验和得分检验。

解：伽马总体下的三大检验.

似然比检验：由题意知似然函数

$$L(\beta)=\prod_{i=1}^n\beta^2x\exp\{-\beta x\}=\beta^{2n}\exp\left\{-\beta\sum_{i=1}^n x_i\right\}\prod_{i=1}^n x_i,\quad\beta>0$$

由对数似然方程

$$\frac{\partial\log L(\beta)}{\partial\beta}=\frac{2n}{\beta}-\sum_{i=1}^n x_i=0\Rightarrow\beta=\frac{2n}{\sum_{i=1}^n x_i}$$

经微分检验其确为似然函数的最大值点，因此 β 的最大似然估计为 $\hat{\beta}=2/\overline{X}$ ．于是似然比检验统计量

$$\Lambda(\boldsymbol{X})=\frac{\sup_{\beta>0}L(\beta)}{\sup_{\beta=1}L(\beta)}=\frac{L(\hat{\beta})}{L(1)}=\frac{(2/\overline{X})^{2n}\exp\{-2n\}}{\exp\{-\overline{X}\}}=(2/\overline{X})^{2n}\exp\{\overline{X}-2n\}$$

由似然比检验统计量的极限分布知

$$2\log\Lambda(\boldsymbol{X}) = 4n\log\frac{2}{\overline{X}} + 2\overline{X} - 4n \xrightarrow{H_0} \chi_1^2, \quad \text{as } n\to\infty$$

因此水平为 α 的似然比检验的拒绝域为

$$D = \left\{\boldsymbol{X} : 4n\log\frac{2}{\overline{X}} + 2\overline{X} - 4n > \chi_1^2(\alpha)\right\}.$$

Wald 检验：由似然函数计算 n 个样本的 Fisher 信息量为

$$I_n(\beta) = -\mathbb{E}_\beta\left[\frac{\partial^2\log L(\beta)}{\partial\beta^2}\right] = \frac{2n}{\beta^2}$$

所以 Wald 检验统计量为

$$W(\boldsymbol{X}) = (\hat{\beta}-1)^2 I_n(\hat{\beta}) = 2n(1-1/\hat{\beta})^2 = 2n(1-\overline{X}/2)^2 \xrightarrow{H_0} \chi_1^2, \quad \text{as } n\to\infty.$$

因此水平为 α 的 Wald 检验的拒绝域为

$$D = \left\{\boldsymbol{X} : 2n(1-\overline{X}/2)^2 > \chi_1^2(\alpha)\right\}$$

得分检验：得分函数为

$$U_n(\beta) = \frac{\partial\log L(\beta)}{\partial\beta} = \frac{2n}{\beta} - \sum_{i=1}^n x_i.$$

所以得分检验统计量为

$$S(\boldsymbol{X}) = [U_n(1)]^2 I_n^{-1}(1) = \left(2n - \sum_{i=1}^n X_i\right)^2/(2n) = n(2-\overline{X})^2/2 \xrightarrow{H_0} \chi_1^2, \quad \text{as } n\to\infty.$$

因此水平为 α 的得分检验的拒绝域为

$$D = \left\{\boldsymbol{X} : n(2-\overline{X})^2/2 > \chi_1^2(\alpha)\right\}$$

该形式与 Wald 检验的形式是一致的。

六、（10 分）男女舒张压检测数据如下：共检测男性 16 人，其中舒张压 <60 的有 4 人，>90 的有 2 人；共检测女性 21 人，其中舒张压 <60 的有 5 人，>90 的有 2 人。

是否能认为男女舒张压分布相同？

解：未合并列的齐一性检验（酌情扣分）：首先根据题目描述写出列联表如下：

舒张压	<60	$60\sim90$	>90	合计
男性人数	4	10	2	16
女性人数	5	14	2	21
合计	9	24	4	37

要检验的假设为 H_0 ：男女性舒张压分布没有显著差异．取检验统计量为

$$K(\boldsymbol{X}) = n\sum_{i=1}^r\sum_{j=1}^s\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}} \xrightarrow{H_0} \chi^2_{(r-1)(s-1)}, \quad \text{as } n\to\infty.$$

因此水平 α 的检验拒绝域为

$$D=\left\{\boldsymbol{X}: K(\boldsymbol{X})>\chi_{(r-1)(s-1)}^{2}(\alpha)\right\}$$

由样本观测值计算检验统计量值时，可以考虑等价形式

$$K(\boldsymbol{x})=n\left(\sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n_{ij}^{2}}{n_{i} . n_{. j}}-1\right)=0.1040<\chi_{2}^{2}(0.2)=3.22.$$

因此不拒绝原假设，认为男女性舒张压没有显著差异．

合并列的齐一性检验：由于有格子点计数过小，需要合并首尾两列得列联表如下：

舒张压	正常	过低或过高	合计
男性人数	10	6	16
女性人数	14	7	21
合计	24	13	37

要检验的假设为 H_0 ：男女性舒张压分布没有显著差异．取检验统计量为

$$K(\boldsymbol{X})=n \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij}-n_{i} \cdot n_{. j} / n\right)^{2}}{n_{i} \cdot n_{. j}} \xrightarrow{H_{0}} \chi_{(r-1)(s-1)}^{2}, \quad \text { as } n \rightarrow \infty .$$

因此水平 α 的检验拒绝域为

$$D=\left\{\boldsymbol{X}: K(\boldsymbol{X})>\chi_{(r-1)(s-1)}^{2}(\alpha)\right\}$$

由样本观测值计算检验统计量值时，可以考虑 2×2 列联表的等价形式

$$K(\boldsymbol{x})=\frac{n\left(n_{11} n_{22}-n_{12} n_{21}\right)^{2}}{n_{1 .} n_{.1} n_{2 .} n_{.2}}=0.0692<\chi_{1}^{2}(0.2)=1.64.$$

因此不拒绝原假设，认为男女性舒张压没有显著差异．

七、（18 分）n 个样本 $(X_1,...,X_n)$ 来自总体分布 $X\sim\Gamma\left(\alpha,\frac{1}{\theta}\right)$，其中 α 已知，先验密度为 $\pi(\theta)=\frac{1}{\theta}$ ，损失函数为

$$L^{2}(d, \theta)=\frac{1}{\theta^{2}}(d-\theta)^{2}$$

求贝叶斯解并证明其为 Minimax 解

解：后验分布的计算：样本 $\boldsymbol{X}=(X_1,\ldots,X_n)$ 的联合概率密度函数为

$$f(\boldsymbol{x} \mid \theta)=\prod_{i=1}^{n} \frac{1}{\theta^{\alpha} \Gamma(\alpha)} x_{i}^{\alpha-1} \exp \left\{-\frac{x_{i}}{\theta}\right\}=\frac{\theta^{-n \alpha}}{[\Gamma(\alpha)]^{n}} \exp \left\{-\frac{1}{\theta} \sum_{i=1}^{n} x_{i}\right\} \prod_{i=1}^{n} x_{i}^{\alpha-1}$$

因此在无信息先验 $\pi(\theta)=(1/\theta)I_{(0,\infty)}(\theta)$ 下，θ 的后验密度为

$$\pi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \theta) \pi(\theta) \propto \theta^{-n \alpha-1} \exp \left\{-\frac{1}{\theta} \sum_{i=1}^{n} x_{i}\right\}, \quad \theta>0$$

添加归一化常数后可知 θ 的后验分布为

$$\theta \mid \boldsymbol{X}=\boldsymbol{x} \sim \Gamma^{-1}\left(n \alpha, \sum_{i=1}^{n} x_{i}\right)$$

加权平方损失下的 Bayes 估计：在加权平方损失下，θ 的 Bayes 估计为

$$\hat{\theta}_B(\boldsymbol{X})=\frac{\mathbb{E}\left(\theta^{-1}\mid \boldsymbol{X}\right)}{\mathbb{E}\left(\theta^{-2}\mid \boldsymbol{X}\right)}=\frac{\frac{n\alpha}{n\overline{X}}}{\frac{n\alpha(n\alpha+1)}{(n\overline{X})^2}}=\frac{n\overline{X}}{n\alpha+1}$$

RK：在矩存在的条件下有

$$X\sim\Gamma(\alpha,\beta)\quad 有\forall n\in\mathbb{Z}\quad \mathbb{E}[X^n]=\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\beta^n}$$
$$Y\sim\Gamma^{-1}(\alpha,\beta)\quad 有\forall n\in\mathbb{Z}\quad \mathbb{E}[Y^n]=\frac{\Gamma(\alpha-n)\beta^n}{\Gamma(\alpha)}$$

事实上，Γ 分布的 n 阶矩就是 Γ^{-1} 分布的 $-n$ 阶矩
验证该 Bayes 估计为 Minimax 估计：Bayes 估计 $\hat{\theta}_B(\boldsymbol{X})$ 的风险函数，注意 $n\overline{X}=\sum\limits_{i=1}^{n}X_i\sim\Gamma\left(n\alpha,\frac{1}{\theta}\right)$

$$\begin{aligned}
R\left(\hat{\theta}_B(\boldsymbol{X}),\theta\right)&=\mathbb{E}\left[\frac{(n\overline{X}/(n\alpha+1)-\theta)^2}{\theta^2}\right]\\
&=\frac{1}{\theta^2}\mathbb{E}\left(\frac{n\overline{X}+\theta}{n\alpha+1}-\theta-\frac{\theta}{n\alpha+1}\right)^2\\
&=\frac{1}{\theta^2}\left[\mathrm{Var}\left(\frac{n\overline{X}+\theta}{n\alpha+1}\right)+\frac{\theta^2}{(n\alpha+1)^2}\right]\\
&=\frac{1}{n\alpha+1}
\end{aligned}$$

为常数，所以 θ 的 Bayes 估计 $\hat{\theta}_B(\boldsymbol{X})=n\overline{X}/(n\alpha+1)$ 是 θ 的 Minimax 估计.

八、(20 分，附加题) 将 Hardy-Weinberg 定律简化如下：n 个样本 $(X_1,...,X_n)$ 来自总体分布 X
设 X 的概率分布为：

X	1	2	3
$P(X)$	p^2	$2p(1-p)$	$(1-p)^2$

对于检验问题：

$$H_0:p\le p_0\quad\leftrightarrow\quad H_1:p>p_0$$

求 UMPT。
解：先说明样本分布族是单参数指数族：样本联合概率质量函数为

$$\begin{aligned}
f(\boldsymbol{x};p)&=\left(p^2\right)^{n_0}[2p(1-p)]^{n_1}\left[(1-p)^2\right]^{n_2}\\
&=2^{n_1}p^{2n_0+n_1}(1-p)^{n_1+2n_2}\\
&=2^{n_1}(1-p)^{2n}\exp\left\{(2n_0+n_1)\log\frac{p}{1-p}\right\}
\end{aligned}$$

这是单参数指数族，且 $Q(p)=\log\frac{p}{1-p}$ 为 p 的严格单调增函数，$T=T(\boldsymbol{X})=2n_0+n_1$.
根据推论 5.4.2 给出检验的 UMPT：注意到原检验问题等价于

$$H_0:\theta\le\theta_0\longleftrightarrow H_1:\theta>\theta_0$$

其中 $\theta_0 = \log \frac{p_0}{1-p_0}$ ．由推论 5.4.2 知，离散型要补上随机化常数，假设检验的 UMPT 可取为

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

其中常数 c 和 r 满足条件

$$c = \arg\min_{c'} \left\{ \mathrm{P}_{p_0}(T > c) \leq \alpha \right\}, \quad r = \frac{\alpha - \mathrm{P}_{p_0}(T > c)}{\mathrm{P}_{p_0}(T = c)}.$$

确定检验函数中的常数 c 和 r ：记

$$p_t = \sum_{1 \leq i,j \leq n, i+j \leq n, 2i+j=t} \frac{n!}{i!j!(n-i-j)!} 2^i p_0^{2i+j} \left(1-p_0\right)^{j+2(n-i-j)},$$

则

$$\mathrm{P}_{p_0}(T = c) = p_c, \quad \mathrm{P}_{p_0}(T > c) = \sum_{t=c+1}^{n} p_t.$$