北京师范大学 2022 - 2023 学年第二学期期中考试试卷（A 卷-答案）

课程名称：　数理统计　　　　任课老师姓名：＿＿＿＿＿＿

卷面总分：＿＿＿ 分　考试时长:8：00-10：45　考试类别：　闭卷 ☒　开卷 ☐　其他 ☐

院（系）：＿＿＿＿＿＿　专 业：＿＿＿＿＿＿　年 级：＿＿＿＿＿＿

姓 名：＿＿＿＿＿＿　学 号：＿＿＿＿＿＿

题号	一	二	三	四	五	六	七	八	总分
得分									

阅卷老师（签字）：＿＿＿＿＿＿

Exercise 1 刀切统计量.

设 $X_1, \cdots, X_n$ 是总体 X 的简单随机样本，X 服从指数分布，参数 λ 未知。

1. 请计算 λ 的极大似然估计 $T(X_1, \cdots, X_n)$.
2. 请计算极大似然估计 $T(X_1, \cdots, X_n)$ 的偏差 Bias. 这是一个无偏估计吗?
3. 对于 $I \subset \{1, 2, \cdots, n\}$，设 $\mathbf{X}_I = (X_i)_{i \in I}$ 为对应的部分样本. 类似的，可以得到估计量 $T(\mathbf{X}_I)$. 刀切统计量定义为
$$T_J(\mathbf{X}_{I_n}) = nT(\mathbf{X}_{I_n}) - \frac{n-1}{n} \sum_{i=1}^{n} T(\mathbf{X}_{I_n \setminus i})$$
其中 $I_n = \{1, 2, \cdots, n\}$, $I_n \setminus i = \{1, 2, \cdots, n\} \setminus \{i\}$. 请计算刀切统计量的偏差 Bias. 这是一个无偏估计吗?

Exercise 2. 一致最小方差无偏估计 UMVUE.

对于参数分布族 $\mathcal{F} = \{f(x, \theta); \theta \in \mathbf{\Theta}\}$, 设 $g(\theta)$ 为参数的可测函数, 设 $X_1, \cdots, X_n$ 为样本. 我们引入

$$\begin{aligned} U_g =& \{T = T(X_1, \cdots, X_n) | \mathbf{E}_\theta T = g(\theta), \mathbf{E}_\theta T^2 < \infty, \forall \theta \in \mathbf{\Theta}\} \\ U_0 =& \{T = T(X_1, \cdots, X_n) | \mathbf{E}_\theta T = 0, \mathbf{E}_\theta T^2 < \infty, \forall \theta \in \mathbf{\Theta}\} \end{aligned}$$

其中 $T(X_1, \cdots, X_n)$ 表示 $X_1, \cdots, X_n$ 的可测函数。我们称 $T_\star$ 是 $g(\theta)$ 在 U_g 中的一个 UMVUE，如果 $T_\star \in U_g$ 且
$$Var(T_\star) = \min_{T \in U_g} Var(T).$$
那么我们有如下定理：估计量 $T_\star$ 是 $g(\theta)$ 在 U_g 中的 UMVUE 的充要条件是

$$\mathbf{E}_\theta[T_0 T_\star] = 0, \forall \theta \in \mathbf{\Theta}, \forall T_0 \in U_0.$$

1. 承认充分性，请证明必要性.（提示：考虑 $T_a = T_\star - aT_0$）
2. 如果我们有充分统计量 T_s，且 $g(\theta)$ 的某个无偏估计 T_* 满足 $T_* = h(T_s)$，其中 h 是可测函数，那么可以引入

$$U_0' = \{T' \Big| \exists T \in U_0 \text{ s.t. } T' = \mathbf{E}_\theta[T|T_s]\}.$$

   请证明: 如果

$$\mathbf{E}_\theta[T'T_*] = 0, \forall \theta \in \mathbf{\Theta}, \forall T' \in U_0',$$

   那么 T_* 是 $g(\theta)$ 在 U_g 中的 UMVUE. (提示：可以使用前文中的充要条件)
3. 设 $X_1, \cdots, X_n$ 是来自 $U(0,\theta)$ 的简单随机样本，其顺序统计量为 $X_{(1)} < \cdots < X_{(n)}$. 参数 θ 未知。
   (a) 请证明 $T_s = X_{(n)}$ 为充分统计量
   (b) 请验证 $T_* = \frac{n+1}{n} T_s$ 是 θ 的无偏估计。
   (c) 请验证 T_* 是 θ 的 UMVUE。

Exercise 3. 极大似然估计. C-R 正则族

1. 设 $X_1, \cdots, X_n$ 为总体 X 的简单随机样本，其中 X 的密度函数为

$$f(x,\theta) = \frac{1}{2\theta} e^{-|x|/\theta}, x \in \mathbb{R}, \theta > 0.$$

   (a) 请计算 θ 的极大似然估计. 这是无偏估计吗?
   (b) 请给出 θ 的矩估计. 这是无偏估计吗?
   (c) 验证此分布族是 C-R 正则分布族.
   (d) 请计算 θ 的无偏估计方差的 C-R 下界.
   (e) 请给出 θ 的 UMVUE.
2. 设 $X_1, \cdots, X_n$ 为总体 X 的简单随机样本，其中 $X \sim N(\theta, \theta^2)$, 参数 $\theta > 0$. 请计算 θ 的极大似然估计 $\hat{\theta}_n$。并且证明 $\hat{\theta}_n$ 是 θ 的强相合估计。
3. 设 $X_1, \cdots, X_n$ 为总体 X 的简单随机样本，其中 $X \sim IG(\mu, \lambda)$(Inverse Gaussian distribution) 的密度函数为

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi} x^{-3}} \exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\}, x > 0.$$

   其中 $\mu > 0, \lambda > 0$. 请计算 μ 和 λ 的极大似然估计.

Solution. 根据样本联合分布密度得到似然函数

$$L(\lambda,\mu)=(2\pi)^{-n/2}\lambda^{n/2}(\prod_{i=1}^{n}x_i)^{-3/2}\exp\left\{-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}\right\}$$

其对数似然函数为

$$\ell(\lambda,\mu)=\frac{n}{2}\ln\lambda-\frac{3}{2}\sum_{i=1}^{n}\ln x_i+\frac{n\lambda}{\mu}-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}x_i-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}+cste.$$

求偏导可得

$$\begin{cases}\frac{\partial\ell(\lambda,\mu)}{\partial\lambda} &= \frac{n}{2\lambda}+\frac{n}{\mu}-\frac{\sum_{i=1}^{n}x_i}{2\mu^2}-\frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i}=0\\ \frac{\partial\ell(\lambda,\mu)}{\partial\mu} &= -\frac{n\lambda}{\mu^2}+\frac{\lambda}{\mu^3}\sum_{i=1}^{n}x_i=0\end{cases}$$

于是得到 MLE

$$\hat{\mu}=\overline{X}_n,\quad \hat{\lambda}=\frac{n}{\sum_{i=1}^{n}(\frac{1}{X_i}-\frac{1}{\overline{X}_n})}.$$

Exercise 4. 假设检验.

假设 X 大学的同学们近期进行了一次体检，教授 A 拿到了匿名的检验结果，包含身高、体重、血压等等信息。教授 A 希望找一位数学系的同学对数据进行一些假设检验。

1. 教授 A 将身高和体重的数据列为 $(h_i,w_i)_{1\le i\le n}$. 请你设计一假设检验，检验身高和体重是否独立。
2. 请你设计一假设检验，检验身高与体重的相关性。
3. 参加此次体检的同学共计 998 人，请你设计一假设检验，检验体重的分布是否满足正态分布.
4. (*) 一般而言，人们认为视力与身高和体重都独立，体检得到的视力结果为 $\{v_i,1\le i\le n\}$. 请设计一假设检验，检验视力是否与身高和体重独立。

Exercise 5. 设 $X_1,\cdots,X_n$ 是独立的随机变量，且 $X_i\sim N(0,\sigma_i^2)$. 记

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

1. 计算 Y 的分布.
2. 证明 $Z\sim\chi^2(n-1)$.
3. 如果 $\sigma_i^2=\sigma^2,\forall i=1,2,\cdots,n$. 记 $\overline{X}_n$ 为样本均值，S_n^2 为样本方差，

$$\tau=\frac{X_1-\overline{X}_n}{S_n}\sqrt{\frac{n}{n-1}}$$

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

Solution.

- Let us consider the joint law of (Y, Z). For convenience, set $b_j = 1/\sigma_j^2$ and $B = \sum_{j=1}^n b_j$. For any continuous and bounded function h and g,

$$\mathbf{E}[h(Y)g(Z)] = \int_{\mathbb{R}^n} h(\frac{\sum_{j=1}^n b_j x_j}{B}) g(\sum_{i=1}^n b_i (x_i - \frac{\sum_{j=1}^n b_j x_j}{B})^2) \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} e^{-\sum_{i=1}^n \frac{x_i^2}{2\sigma_i^2}} dx_1 \cdots dx_n.$$

We use the following change of variables:

$$z_i = x_i - \frac{\sum_{j=1}^n b_j x_j}{B}, \forall 1 \le i \le n-1 \text{ and } z_n = \frac{\sum_{j=1}^n b_j x_j}{B}.$$

Observe that

$$\sum_{i<n} b_i z_i + b_n(x_n - z_n) = \sum_{i \le n} b_i x_n - z_n B = 0.$$

So, inversely, one has

$$x_i = z_i + z_n, \forall 1 \le i \le n-1 \text{ and } x_n = z_n - \frac{\sum_{i<n} b_i z_i}{b_n}.$$

Consequently,

$$\mathbf{E}[h(Y)g(Z)] = \int_{\mathbb{R}^n} h(\frac{\sum_{j=1}^n b_j x_j}{B}) g(\sum_{i=1}^n b_i (x_i - \frac{\sum_{j=1}^n b_j x_j}{B})^2) \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} e^{-\sum_{i=1}^n \frac{x_i^2}{2\sigma_i^2}} dx_1 \cdots dx_n$$

$$= \int_{\mathbb{R}^n} h(z_n) g(\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2) \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}[\sum_{i<n} b_i (z_i+z_n)^2 + b_n (z_n - \frac{\sum_{i<n} b_i z_i}{b_n})^2]} \frac{B}{b_n} dz_1 \cdots dz_n$$

$$= \int_{\mathbb{R}^n} h(z_n) g(\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2) \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}[\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2 + B z_n^2]} \frac{B}{b_n} dz_1 \cdots dz_n$$

$$= \int_{\mathbb{R}} h(z_n) \frac{\sqrt{B} e^{-\frac{1}{2} B z_n^2}}{\sqrt{2\pi}} dz_n$$

$$\times \int_{\mathbb{R}^{n-1}} g(\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2) \frac{\sqrt{B}}{\sqrt{b_n}} \prod_{i<n} \sqrt{\frac{b_i}{2\pi}} e^{-\frac{1}{2}[\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2]} dz_1 \cdots dz_{n-1}$$

This shows that Y and Z are independent and that $Y \sim N(0, \frac{1}{B})$. Take $u_i = b_i z_i$ for all $1 \le i \le n-1$, we get that

$$\int_{\mathbb{R}^{n-1}} g(\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2) \frac{\sqrt{B}}{\sqrt{b_n}} \prod_{i<n} \sqrt{\frac{b_i}{2\pi}} e^{-\frac{1}{2}[\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2]} dz_1 \cdots dz_{n-1}$$

$$= \int_{\mathbb{R}^{n-1}} g(\sum_{i<n} \frac{u_i^2}{b_i} + \frac{1}{b_n}(\sum_{i<n} u_i)^2) \frac{\sqrt{B}}{\sqrt{b_n}} \prod_{i<n} \frac{1}{\sqrt{2\pi b_i}} e^{-\frac{1}{2}[\sum_{i<n} \frac{u_i^2}{b_i} + \frac{1}{b_n}(\sum_{i<n} u_i)^2]} du_1 \cdots du_{n-1}$$

where

$$\sum_{i<n} \frac{u_i^2}{b_i} + \frac{1}{b_n}(\sum_{i<n} u_i)^2 = (u_1, \cdots, u_{n-1}) \begin{pmatrix} \frac{1}{b_1} + \frac{1}{b_n} & \frac{1}{b_n} & \cdots & \frac{1}{b_n} \\ \frac{1}{b_n} & \frac{1}{b_2} + \frac{1}{b_n} & \cdots & \frac{1}{b_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{b_n} & \frac{1}{b_n} & \cdots & \frac{1}{b_{n-1}} + \frac{1}{b_n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{pmatrix}.$$

Note that

$$\Lambda = \begin{pmatrix} \frac{1}{b_1}+\frac{1}{b_n} & \frac{1}{b_n} & \cdots & \frac{1}{b_n} \\ \frac{1}{b_n} & \frac{1}{b_2}+\frac{1}{b_n} & \cdots & \frac{1}{b_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{b_n} & \frac{1}{b_n} & \cdots & \frac{1}{b_{n-1}}+\frac{1}{b_n} \end{pmatrix} = U^T \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} \end{pmatrix} U \text{ with } U^T U = I_{n-1}.$$

Moreover,

$$\det(\Lambda) = \frac{B}{\prod_{i\le n} b_i} = \prod_{i<n} \lambda_i.$$

Therefore, by taking $v = Uv$

$$\begin{aligned}
&\int_{\mathbb{R}^{n-1}} g(\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2)\frac{\sqrt{B}}{\sqrt{b_n}}\prod_{i<n}\sqrt{\frac{b_i}{2\pi}}e^{-\frac{1}{2}[\sum_{i<n} b_i z_i^2 + \frac{1}{b_n}(\sum_{i<n} b_i z_i)^2]}dz_1\cdots dz_{n-1} \\
&= \int_{\mathbb{R}^{n-1}} g(\sum_{i<n}\lambda_i v_i^2)\frac{1}{(2\pi)^{\frac{n-1}{2}}\sqrt{\det(\Lambda^{-1})}}e^{-\frac{1}{2}\sum_{i<n}\lambda_i v_i^2}dv_1\cdots dv_{n-1} \\
&= \int_{\mathbb{R}^{n-1}} g(\sum_{i<n} z_i^2)\frac{1}{(2\pi)^{\frac{n-1}{2}}}e^{-\frac{1}{2}\sum_{i<n} z_i^2}dz_1\cdots dz_{n-1}.
\end{aligned}$$

This shows that $Z \sim \chi^2(n-1)$.

- Without loss of generality, we could assume that $\sigma^2 = 1$. Take $W = (X_1 - \overline{X}_n)\sqrt{\frac{n}{n-1}}$ and $V = \sum_{i\le n}(X_i - \overline{X}_n)^2 - W^2$. Observe that

$$T = \frac{W\sqrt{n-2}}{\sqrt{V}}$$

In order to prove that $T \sim t(n-2)$, it suffices to show that $W \sim N(0,1)$ and $V \sim \chi^2(n-2)$ and that W and V are independent. For any continuous and bounded functions h and g,

$$\begin{aligned}
&\mathbf{E}[h(W)g(V)] \\
&= \int_{\mathbb{R}^n} h(\sqrt{\frac{n}{n-1}}(x_1 - \frac{1}{n}\sum_{j=1}^n x_j))g(\sum_{i\le n}(x_i - \frac{1}{n}\sum_{j\le n} x_j)^2 - \frac{n}{n-1}(x_1 - \frac{1}{n}\sum_{j\le n} x_j)^2)\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \\
&= \int_{\mathbb{R}^{n-1}} h(-\sqrt{\frac{n}{n-1}}y_1)g(y^T\Lambda y - \frac{n}{n-1}y_1^2)\frac{1}{(2\pi)^{\frac{n-1}{2}}\sqrt{\det(\Lambda^{-1})}}e^{-\frac{1}{2}y^T\Lambda y}dy_1\cdots dy_{n-1}
\end{aligned}$$

Note that

$$y^T\Lambda y - \frac{n}{n-1}y_1^2 = y^T A_1 y$$

with A_1 semi-definitely positive with rank $n-2$. So,

$$y^T\Lambda y = y^T A_1 y + y^T A_2 y$$

with A_2 semi-definitely positive with rank 1. Thus,

$$\begin{aligned}&\mathbf{E}[h(W)g(V)]\\&=\int_{\mathbb{R}} h(-v_1)\frac{e^{-\frac{1}{2}v_1^2}}{\sqrt{2\pi}}dv_1\int_{\mathbb{R}^{n-2}} g(v^Tv)\frac{e^{-\frac{1}{2}v^Tv}}{(2\pi)^{\frac{n-2}{2}}}dv_1\cdots dv_{n-2}\end{aligned}$$

This is exactly what we need.

Exercise 6. 设随机变量 $U\sim\chi^2(n)$, $V\sim\chi^2(m)$, 且 U 和 V 独立。令

$$X=\frac{U/n}{V/m},\quad Y=\frac{1}{X},\qquad Z=\frac{nX}{m+nX}$$

1. 请计算 Y 的分布.
2. 请计算 Z 的分布.

Solution.

- $Y\sim F(n,m)$.
- For any bounded continuous function h,

$$\begin{aligned}\mathbf{E}[h(Z)]=&\mathbf{E}[h(\frac{\frac{U}{V/m}}{m+\frac{U}{V/m}})]=\mathbf{E}[h(\frac{U}{V+U})]\\&=\int_{(0,\infty)^2} h(\frac{u}{u+v})\frac{u^{n/2-1}e^{-u/2}}{2^{n/2}\Gamma(n/2)}\frac{v^{m/2-1}e^{-v/2}}{2^{m/2}\Gamma(m/2)}dudv\end{aligned}$$

by change of variables

$$z=\frac{u}{u+v},t=u+v,$$

we get that

$$\begin{aligned}\mathbf{E}[h(Z)]&=\int_{t>0,0<z<1} h(z)t^{\frac{m+n}{2}-2}z^{n/2-1}(1-z)^{m/2-1}e^{-\frac{1}{2}t}\frac{tdtdz}{2^{\frac{m+n}{2}}\Gamma(n/2)\Gamma(m/2)}\\&=\int_0^1 h(z)z^{n/2-1}(1-z)^{m/2-1}dz\int_0^\infty t^{\frac{m+n}{2}-1}e^{-\frac{1}{2}t}\frac{dt}{2^{\frac{m+n}{2}}\Gamma(n/2)\Gamma(m/2)}\\&=\int_0^1 h(z)z^{n/2-1}(1-z)^{m/2-1}\frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)}dz\end{aligned}$$

So the density of Z is

$$f(z)=z^{n/2-1}(1-z)^{m/2-1}\frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)}1_{z\in(0,1)}.$$

备注:

- $\chi^2(k)$ 的密度函数为

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, x > 0.$$

- $t(k)$ 的密度函数为

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k}\Gamma(\frac{k}{2})}(1 + \frac{t^2}{k})^{-\frac{k+1}{2}}, t \in \mathbb{R}.$$

- $F(r, s)$ 的密度为

$$f(w) = \frac{(r/s)^{r/2}\Gamma(\frac{r+s}{2}) w^{r/2-1}}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2})[1 + \frac{rw}{s}]^{\frac{r+s}{2}}}, w > 0.$$

- Gamma 函数 $\Gamma(\alpha)$ 为

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du, \forall \alpha > 0.$$