北京大学数学科学学院期末试题

2020–2021 年第一学期

考试科目__数理统计__　考试时间 2020 年 1 月 11 日

姓名____________　学号____________

本试题共 8 道大题，满分 100 分。

1. (12 分) 设 $X_1, \ldots, X_n$ 是来自参数为 μ 的正态分布 $N(\mu, \sigma_0^2)$ 的简单随机样本，σ_0^2 已知。

(1) 求 μ 的最大似然估计；

(2) 求 μ 的矩估计；

(3) 求 Fisher 信息量 $I(\mu)$；

(4) 求 μ 的无偏估计的方差下界；

(5) (1) 中的最大似然估计是否是 μ 的最小方差无偏估计 (需说明理由)？

(6) 试找出 μ^2 的一个无偏估计。

1. (12 points) Suppose $X_1, \ldots, X_n$ are i.i.d random samples from normal distribution $N(\mu, \sigma_0^2)$, with σ_0^2 known.

(1) Find the MLE of μ;

(2) Find the moment estimate of μ;

(3) Calculate the Fisher information $I(\mu)$ of μ;

(4) Find the variance lower bound of unbiased estimator for μ;

(5) Is the MLE in (1) the UMVUE of μ (Please give your reason)?

(6) Try to find an unbiased estimator for μ^2.

2. (12 分) 若随机变量 X 的分布密度可取下面的 $f_0(x)$ 或 $f_1(x)$：

$$f_0(x)=\begin{cases}1 & \text{当}0\le x\le 1\\ 0 & \text{其他}\end{cases};\quad f_1(x)=\begin{cases}\frac{C_1}{1+x^2} & \text{当}0\le x\le 1\\ 0 & \text{其他}\end{cases}$$

其中 C_1 是未知常数。基于 X 的一个观测值，对检验问题 $H_0: f(x)=f_0(x)\leftrightarrow H_1: f(x)=f_1(x)$, 利用 N-P 引理求检验水平为 $\alpha=0.1$ 的 UMP 检验 ϕ，并求其第二类错误的概率。

2. (12 points) Suppose random variable X has density $f_0(x)$ or $f_1(x)$:

$$f_0(x)=\begin{cases}1 & \text{when } 0\le x\le 1\\ 0 & \text{otherwise}\end{cases};\quad f_1(x)=\begin{cases}\frac{C_1}{1+x^2} & \text{when } 0\le x\le 1\\ 0 & \text{otherwise}\end{cases}$$

where C_1 is unknown constant. Based on an observation of X, for the hypothesis testing $H_0: f(x)=f_0(x)\leftrightarrow H_1: f(x)=f_1(x)$, use N-P lemma to get a UMP test ϕ of significance level $\alpha=0.1$, and calculate the probability of Type 2 error.

3. (13 分) 设 $X_1,\ldots,X_n$ 和 $Y_1,\ldots,Y_m$ 是分别来自的参数为 λ_1 和 λ_2 的指数分布的简单随机样本，λ_1 和 λ_2 未知, 即

$$p(X_i=x)=\lambda_1 e^{-\lambda_1 x},\quad p(Y_j=y)=\lambda_2 e^{-\lambda_2 y}$$

试求 λ_1/λ_2 的置信水平为 $1-\alpha$ 的置信区间。

3. (13 points) Suppose $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ are ramdom samples from exponential distribution with parameter λ_1 and λ_2 respectively, where λ_1 and λ_2 are unknown, that is:

$$p(X_i=x)=\lambda_1 e^{-\lambda_1 x},\quad p(Y_j=y)=\lambda_2 e^{-\lambda_2 y}$$

Give the $1-\alpha$ confidence interval for λ_1/λ_2

4. (13 分) 设 X_i, $i=1,2,...,n$ 为参数为 λ 的泊松分布的独立同分布随机样本。λ 的先验分布为 $Gamma(\alpha,\beta)$，即 $\pi_{\alpha,\beta}(\lambda)=\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$。

(1) 求 λ 的后验分布

(2) 在平方误差下，求 λ 的贝叶斯估计量。

4. (13 points) Suppose X_i, $i=1,2,...,n$ are i.i.d samples from a Poisson distribution

with parameter λ. The prior of λ is $\pi(\lambda) \sim Gamma(\alpha, \beta)$, that is $\pi_{\alpha,\beta}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$.

(1) Find the posterior distribution of λ.

(2) Find the Bayes estimator of λ using squared error loss.

5. (16 分) 设 ANOVA 模型 $Y_{ij} = \mu+\tau_i+\epsilon_{ij}$, $i = 1, 2, ..., I$, $j = 1, ..., n_i$, 其中 $\sum_{i=1}^I \tau_i = 0$, 且 ϵ_{ij} 独立, $E(\epsilon_{ij}) = 0$, $Var(\epsilon_{ij}) = \sigma^2$。定义

$$SS_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..})^2;$$

$$SS_W = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2;$$

$$SS_B = \sum_{i=1}^{I} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2.$$

其中 $\overline{Y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $\overline{Y}_{..} = \frac{1}{\sum_{i=1}^I n_i}\sum_{i=1}^I\sum_{j=1}^{n_i} Y_{ij}$.

(1) 证明 $SS_{TOT} = SS_W + SS_B$。

(2) 若 $n_i = J$, 证明 $E(SS_W) = I(J-1)\sigma^2$ 以及 $E(SS_B) = J\sum_{i=1}^I \tau_i^2 + (I-1)\sigma^2$

5. (16 points) Suppose an ANOVA model: $Y_{ij} = \mu+\tau_i+\epsilon_{ij}$, $i = 1, 2, ..., I$ and $j = 1, ..., n_i$, where $\sum_{i=1}^I \tau_i = 0$. The ϵ_{ij} are independent and $E(\epsilon_{ij}) = 0$, $Var(\epsilon_{ij}) = \sigma^2$. Define

$$SS_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..})^2;$$

$$SS_W = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2;$$

$$SS_B = \sum_{i=1}^{I} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2.$$

where $\overline{Y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ and $\overline{Y}_{..} = \frac{1}{\sum_{i=1}^I n_i}\sum_{i=1}^I\sum_{j=1}^{n_i} Y_{ij}$.

(1) Show that $SS_{TOT} = SS_W + SS_B$.

(2) If $n_i = J$ for $i = 1, 2, ..., I$. Prove that $E(SS_W) = I(J-1)\sigma^2$ and $E(SS_B) = J\sum_{i=1}^I \tau_i^2 + (I-1)\sigma^2$

6. (16 分) 设参数 $\theta \in [0,1]$, $X|\theta \sim p(x,\theta)$, 数据为 $X_1,\dots,X_n$，损失函数为 $L(\theta,a)=(\theta-a)^2$。

(1) 试证明，存在决策 δ_1，对 θ 的任意先验分布，其贝叶斯风险 $\rho(\delta_1)\le\frac{1}{4}$;

(2) 若 $X|\theta\sim B(1,\theta)$，即参数为 θ 的两点分布，先验分布 $\pi(\theta)$ 为 Beta 分布 $Be(\alpha,\beta)$，即

$$\pi(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

求贝叶斯决策 δ^*。

6. (16 points) Suppose parameter $\theta\in[0,1]$, $X|\theta\sim p(x,\theta)$, data are $X_1,\dots,X_n$, and the loss function is $L(\theta,a)=(\theta-a)^2$.

(1) Prove that there exists a decision rule δ_1 such that for any prior of θ, the Bayes risk $\rho(\delta_1)\le\frac{1}{4}$;

(2) If $X|\theta\sim B(1,\theta)$, that is Binomial distribution with parameter θ, prior $\pi(\theta)$ is Beta distribution $Be(\alpha,\beta)$:

$$\pi(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

Find the Bayes decision rule δ^*.

7. (18 分) 考虑测量误差的问题，假设有下述数据生成机制：

$$Y_i=\beta_0+\beta_1\xi_i+\varepsilon_i,\varepsilon_i\sim N(0,\sigma_\varepsilon^2);$$
$$X_i=\xi_i+\delta_i,\delta_i\sim N(0,\sigma_\delta^2)$$

其中 ε_i 和 δ_i 独立。在方程中，ξ_i 为常数，$\mathbb{E}X_i=\xi_i$。并且假设 $\sigma_\delta^2=\lambda\sigma_\varepsilon^2$，其中 λ 已知。

(1) 用极大似然方法求出 β_0 和 β_1 的估计，分别记作 $\hat\beta_0(\lambda)$ 和 $\hat\beta_1(\lambda)$。

(2) 证明 $\lim_{\lambda\to0}\hat\beta_1(\lambda)=S_{xy}/S_{xx}$，即 y 对 x 回归的斜率。

(3) 证明 $\lim_{\lambda\to+\infty}\hat\beta_1(\lambda)=S_{yy}/S_{xy}$，即 x 对 y 回归的斜率的倒数。

其中 $S_{uv}=\sum_{i=1}^n(u_i-\bar u)(v_i-\bar v)$。

7. (18 points) Consider a measurement error problem, suppose we have the following data generating process:

$$Y_i=\beta_0+\beta_1\xi_i+\varepsilon_i,\varepsilon_i\sim N(0,\sigma_\varepsilon^2);$$
$$X_i=\xi_i+\delta_i,\delta_i\sim N(0,\sigma_\delta^2)$$

where ε_i and δ_i are independent. In the equation, ξ_i are constants and $\mathbb{E}X_i = \xi_i$. Also, suppose $\sigma_\delta^2 = \lambda\sigma_\varepsilon^2$, where λ is known.

(1) Using MLE method to get the estimates for β_0 and β_1, denoted as $\hat{\beta}_0(\lambda)$ and $\hat{\beta}_1(\lambda)$ respectively.

(2) Show that $\lim_{\lambda\to 0}\hat{\beta}_1(\lambda) = S_{xy}/S_{xx}$, the slope of the ordinary regression of y on x.

(3) Show that $\lim_{\lambda\to +\infty}\hat{\beta}_1(\lambda) = S_{yy}/S_{xy}$, the reciprocal of the slope of the ordinary regression of x on y.

where $S_{uv} = \sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})$.