信息学中的概率统计：作业八

截止日期：2024 年 1 月 3 日（周五）期末考试前。

第一题

给定 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，其中 $y_i = \alpha + \beta x_i + \epsilon_i$，$\epsilon_i$ 相互独立，且 ϵ_i 服从拉普拉斯分布，其概率密度函数（参考作业三第五题）满足对于任意实数 $x \in \mathbb{R}$，

$$f(x) = \frac{1}{2b} e^{-|x|/b},$$

这里 α, β 和 $b > 0$ 为未知参数。证明 α 和 β 的最大似然估计量为

$$\mathrm{argmin}_{\alpha,\beta} \sum_{i=1}^{n} |y_i - (\alpha + \beta x_i)|\text{。}$$

第二题

给定 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，令 $\hat{\alpha}$ 和 $\hat{\beta}$ 为最小二乘估计量，$\hat{y}_i = \hat{\beta} x_i + \hat{\alpha}$ 为 x_i 的预测值，$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$，$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$。证明

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\overline{y} - \hat{y}_i)^2\text{。}$$

提示：利用正规方程，并证明

$$\hat{y}_i = \overline{y} + \hat{\beta}(x_i - \overline{x})\text{。}$$

第三题

给定 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，其中 $y_i = \alpha + \beta x_i + \epsilon_i$，$\epsilon_i$ 相互独立且 $\epsilon_i \sim N(0, \sigma^2)$。沿用第二题中的记号，并令 $s^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2$，$s_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$。

(1) 令

$$q_1 = [1/\sqrt{n}, 1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T \in \mathbb{R}^n,$$

$$q_2 = \left[\frac{x_1 - \overline{x}}{\sqrt{s_{xx}}}, \frac{x_2 - \overline{x}}{\sqrt{s_{xx}}}, \ldots, \frac{x_n - \overline{x}}{\sqrt{s_{xx}}} \right]^T \in \mathbb{R}^n\text{。}$$

证明存在 $q_3, q_4, \ldots, q_n \in \mathbb{R}^n$，使得 $q_1, q_2, q_3, q_4, \ldots, q_n$ 为 $\mathbb{R}^n$ 中的一组标准正交基。

(2) 将 y 视作 $\mathbb{R}^n$ 中的向量。对于 $1 \le i \le n$，令 $z_i = q_i^T y$，也即 $z = Qy \in \mathbb{R}^n$，其中 $Q \in \mathbb{R}^{n \times n}$ 的第 i 行为 $q_i \in \mathbb{R}^n$。给出 n 维随机变量 z 服从的分布。提示：计算随机向量 y 的数学期望，并验证其与 $q_3, q_4, \ldots, q_n$ 的正交性。

(3) 证明 $z_1 = \sqrt{n}\overline{y}$，$z_2 = \sqrt{s_{xx}}\hat{\beta}$。

(4) 利用第二题中提示和结论，证明 $\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2=z_2^2$ 及 $(n-2)s^2=\sum(y_i-\hat{y}_i)^2=\sum_{i=3}^{n}z_i^2$。

(5) 给出 $(n-2)s^2/\sigma^2$ 服从的分布，并证明 s^2 与 $\hat{\alpha}$ 和 $\hat{\beta}$ 均相互独立。

(6) 当 $\beta=0$，给出统计量 $t=\frac{\hat{\beta}}{\sqrt{s^2}/\sqrt{s_{xx}}}$ 服从的分布。

(7) 若 σ^2 未知，考虑假设检验问题，原假设 $H_0:\beta=0$，备择假设 $H_1:\beta\neq 0$。拒绝域为

$$W=\{((x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n))\mid |t|\geq c\},$$

其中 c 为待定常数。若显著性水平为 α，给出 c 的取值。