北京大学信息科学技术学院试题
信息学中的概率统计：期末考试

考试科目：______________________ 姓名：______________ 学号：__________________
考试时间：___________年_______月_______日 任课教师：__________________________

注意事项：

1. 考试时间为 120 分钟（上午 8:30 至 10:30）。
2. 共 5 个大题，总分为 100 分。试卷单面打印，共**两页**。**请不要擅自拆掉装订！**
3. 请合理分配时间，并将所有答案**写在答题纸**上，试题纸或草稿纸上作答无效。
4. 考试结束后将回收答题纸，试卷和草稿纸。

题目一（共 20 分）

考虑一元线性回归问题。数据 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ 服从关系 $y_i = \beta x_i + \epsilon_i$，其中 x_i<u>不全相同</u>，误差（噪声）ϵ_i 满足 $E(\epsilon_i) = 0$，$\mathrm{Var}(\epsilon_i) = \sigma^2$ 且相互独立。这里 β 和 σ^2 均为未知参数。注意，本题中我们假设线性相关关系中的截距项 $\alpha = 0$。定义

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2。$$

考虑 β 的最小二乘估计量 $\hat{\beta} = \mathrm{argmin}_\beta Q(\beta)$，也即最小化 $Q(\beta)$ 的参数 β。

1. （10 分）写出正规方程，并给出 $\hat{\beta}$ 的表达式。
2. （10 分）给出估计量 $\hat{\beta}$ 的偏差和均方误差。

题目二（共 20 分）

对于正整数 $n \geq 1$，令 $X_1, X_2, \ldots, X_n$ 为独立同分布的随机变量，且 $X_i \sim \mathrm{Exp}(\lambda)$，即参数为 λ 的指数分布。令 $Y = \sum_{i=1}^{n} X_i$，本题中，我们将对 $a > n/\lambda$ 给出 $P(Y \geq a)$ 的上界。

1. （5 分）计算 $E(Y)$ 和 $\mathrm{Var}(Y)$。
2. （5 分）证明 $P(Y \geq a) \leq \frac{n}{(\lambda a - n)^2}$。
3. （5 分）计算 Y 的矩生成函数 $M_Y(t) = E\left(e^{tY}\right)$。
4. （5 分）使用 Chernoff bound，证明 $P(Y \geq a) \leq (\lambda a/n)^n e^{-(\lambda a - n)}$。

题目三（共 20 分）

随机变量 $X \sim U(0,1)$，$Y \sim U(0,1)$，且 X 与 Y 相互独立。令 $U = X + Y$，$V = 2X$。

1.（5 分）使用卷积公式，给出 U 的边际密度函数。
2.（5 分）给出 (U, V) 的联合密度函数。
3.（5 分）对于任意 $0 < u < 2$，计算给定 $U = u$ 条件下 V 的条件密度函数。
4.（5 分）计算 $\mathrm{Cov}(U, V)$ 以及 $\mathrm{Corr}(U, V)$。

题目四（共 20 分）

令总体 X 服从概率密度函数如下的连续分布，其中 $\theta > 0$ 为未知参数，

$$f(x) = \begin{cases} \frac{2x}{\theta} \cdot e^{-x^2/\theta} & x > 0 \\ 0 & x \le 0 \end{cases}。$$

给定简单随机样本 $X_1, X_2, \ldots, X_n$，回答以下问题。
提示：利用结论（无需证明）$2X^2/\theta \sim \chi^2(2)$，也即 $2X^2/\theta$ 服从自由度为 2 的卡方分布。

1.（5 分）计算 X 的二阶矩 $E(X^2)$，并将二阶矩 $E(X^2)$ 替换为样本二阶矩 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$，给出 θ 的矩估计量 $\hat{\theta}_1$。判断 $\hat{\theta}_1$ 是否为无偏估计量以及渐进无偏估计量。
2.（5 分）给出 θ 的最大似然估计量 $\hat{\theta}_2$。
3.（5 分）计算 $\hat{\theta}_2$ 的均方误差，并判断 $\hat{\theta}_2$ 是否为一致估计量。
4.（5 分）对于 $0 < \alpha < 1$，构造统计量 $\hat{\theta}_L$ 和 $\hat{\theta}_R$ 使得 $P(\hat{\theta}_L \le \theta \le \hat{\theta}_R) = 1 - \alpha$，也即构造 θ 的置信水平为 $1 - \alpha$ 的置信区间。最终结果应依赖于 χ^2 分布的分布函数的逆函数 F^{-1}。

题目五（共 20 分）

对于正整数 $n \ge 2$，令 $X_1, X_2, \ldots, X_n$ 为独立同分布的随机变量，且 X_i 服从标准正态分布 $N(0,1)$。令 $Y_n = \max\{X_1, X_2, \ldots, X_n\}$。

1.（5 分）证明
$$E(Y_n) = \int_0^{\infty} P(Y_n > x)dx - \int_{-\infty}^{0} P(Y_n < x)dx。$$
2.（5 分）证明存在常数 $C_1 > 0$，$P(Y_n \ge C_1 \cdot \sqrt{\ln n}) \le \frac{1}{n}$。
3.（5 分）证明存在常数 $C_2 > 0$，对于充分大的 n，$E(Y_n) \le C_2 \cdot \sqrt{\ln n}$。
4.（5 分）证明存在常数 $C_3 > 0$，对于充分大的 n，$E(Y_n) \ge C_3 \cdot \sqrt{\ln n}$。