期末复习

多元随机变量

知识点

略，详见相关Slides

题目

设 $X_1, \ldots, X_n$ 为分别来自来自指数分布 $\mathrm{Exp}(\theta_1), \mathrm{Exp}(\theta_2), \ldots, \mathrm{Exp}(\theta_n)$ 的独立随机变量，求

$$Y = \max\{\frac{a_1}{X_1}, \frac{a_2}{X_2}, \ldots, \frac{a_n}{X_n}\}, \quad \text{其中 } a_1, \ldots, a_n > 0$$

的密度函数

大数定理 & Chernoff

知识点

- 尾不等式
  - Chernoff-Hoeffding不等式：若$X = \sum_{i=1}^{n} X_i$，X_i相互独立且 $a \leq X_i \leq b$
    - $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
    - $P(X \leq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
- 大数定律
  - 马尔可夫大数定律
    - **马尔可夫大数定律**：若$\frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \to 0$，则$\{X_n\}$服从大数定律
  - 辛钦大数定律

    辛钦大数定律：$\{X_n\}$独立同分布，且数学期望$\mu = E(X_i)$存在，则$\{X_n\}$服从大数定律

题目

重要性抽样在计算分布某些特征时非常有用。设随机变量 X 的分布为 f，但从该分布 f 直接模拟比较困难。于是，我们从另一分布 g（易于模拟）生成独立同分布样本 $Y_1, Y_2, \ldots, Y_m$，对任意给定函数 h，计算下式的平均：

$\frac{1}{m} \sum_{i=1}^{m} \frac{f(Y_i)}{g(Y_i)} h(Y_i).$

我们假设 f 和 g 拥有相同的支撑集，且 $\mathrm{Var}[h(X)] < \infty$

1. 证明

   $\mathbb{E}\left[\frac{1}{m} \sum_{i=1}^{m} \frac{f(Y_i)}{g(Y_i)} h(Y_i)\right] = \mathbb{E}_f[h(X)].$
2. 假设 $\frac{f(Y_i)}{g(Y_i)} h(Y_i) \in [a, b]$，用Chernoff不等式证明

   $\frac{1}{m} \sum_{i=1}^{m} \frac{f(Y_i)}{g(Y_i)} h(Y_i) \xrightarrow{p} \mathbb{E}_f[h(X)].$

3. 证明 $\frac{1}{m}\sum_{i=1}^{m}\frac{f(Y_i)}{g(Y_i)}h(Y_i)\xrightarrow{p}\mathbb{E}_f[h(X)]$.

估计

知识点

最大似然

矩估计

无偏估计量、渐进无偏估计量、一致估计量

题目

1. 设 $X_1,\dots,X_n$ 为来自对数正态分布 $\mathrm{Lognormal}(\theta,\sigma^2)$ 的独立同分布随机变量，求 θ 的最大似然估计量
2. 设 $X_1,\dots,X_n$ 为来自反伽马分布的独立同分布随机变量，其密度可写为
$$f(x\mid\theta)=\theta\,x^{-2}\exp(-\theta/x),x>0$$
求 θ 的最大似然估计量
3. 设 $X_1,\dots,X_n$ 为来自一个单参数指数族的独立同分布随机变量，其密度可写为
$$f(x|\theta)=\exp\{\theta h(x)-H(\theta)\}g(x),$$
其中 $h=H'$ 且 h 为单调递增函数。求 θ 的最大似然估计量

假设检验

知识点

第一步：建立假设

- **原假设** H_0：不应轻易拒绝的假设
  - 例：新的策略没有效果；网络设备没有异常
- **备择假设**H_1：与原假设 H_0对立的假设
- $H_0{:}\,\theta\le 110, H_1{:}\,\theta>110$

第二步：选择统计量，给出拒绝域的形式

将样本取值$x_1,x_2,\dots,x_n$分为两个区域W和$\overline{W}$

- 若$(x_1,x_2,\dots,x_n)\in W$，拒绝原假设 H_0
- 若$(x_1,x_2,\dots,x_n)\in\overline{W}$，接受原假设，也即拒绝备择假设$H_1$

W：**拒绝域**，$\overline{W}$：**接受域**

通常根据统计量的取值的设计拒绝域

第三步：选择显著性水平α，给出拒绝域

$H_0: \theta \in \Theta_0, H_1: \theta \in \Theta_1$
若检验对于任意$\theta \in \Theta_0$，都有$\alpha(\theta) \le \alpha$，称该检验**显著性水平**为$\alpha$

- 也即，控制第一类错误的概率不超过α，再尽量减少第二类错误的概率
- **Neyman-Pearson原则**
- 一般取 $\alpha = 0.05$

题目

I. 均匀分布假设检验

线性回归

知识点

线性相关关系

正规方程-最小二乘估计

均方误/方差/协方差

题目

给定 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，其中 $y_i = \alpha + \beta\eta_i + \epsilon_i$， ϵ_i 相互独立且 $\epsilon_i \sim N(0, \sigma_\epsilon^2)$，而 $x_i = \eta_i + \delta_i$ ， δ_i 相互独立且 $\delta_i \sim N(0, \sigma_\delta^2)$，其中 η_i 为未知常量，

a. 求 α, β 的最大似然估计（提示，需要先估计 η_i）