北京工业大学2023-2024学年第一学期
统计计算试题A卷

一、（15分） 已知随机变量密度函数为

$$f(x)=\frac{1}{\pi\{1+(x-\mu)^2\}},$$

有样本$(1.82, 0.21, 1.32, -0.29, 1.81)$，用牛顿法估计参数$\mu$。

二、（20分）

根据下列要求产生t分布随机数。

1. 若独立随机变量$X_i \sim N(0,1), i=1,\ldots,n$，$Y \sim N(0,1)$，则$\frac{Y}{\sum_{i=1}^n X_i^2/n} \sim t(n)$。

(1)根据上述结论产生2000个来自分布$t(10)$的独立随机数。

(2)画出这2000个随机数的密度直方图。

(3)画出$t(10)$的密度曲线图比较随机数的产生效果。

2. 已知$t(10)$分布在0处密度达到最大值。

(1)给出$t(10)$密度函数的最大值M。

(2)根据均匀分布$U(-10,10)$和M，用舍选法产生2000个$t(10)$分布的随机数。

(3)画出随机数的密度直方图和出$t(10)$的密度曲线图。

三、（25分）

设$X_1,\ldots,X_n$是来自指数分布$Exp(\lambda)$的独立样本，检验原假设和备择假设$H_0:\lambda=5$和$H_1:\lambda>5$。

1.取样本量$n=10$，检验统计量$T=n/\sum_{i=1}^n X_i$。用统计模拟的方法给出2000个样本得到的检验统计量的抽样分布，画出密度直方图。

2.根据上述的模拟结果给出检验水平为$\alpha=0.05$时检验拒绝域$\{T\leq c\}$的临界值c。

3.当$\lambda=6$时，给出2000个样本得到的检验统计量的抽样分布，画出密度直方图，并根据c给出检验功效。

四、（20分）

设$X_1, \ldots, X_n$是来自指数分布$Exp(\lambda)$的独立样本，用蒙特卡洛和Bootstrap法估计λ的偏差和方差。

1.给定$\lambda = 5$。取样本量$n = 10$，用蒙特卡洛方法估计估计量$\hat{\lambda}_1 = n/\sum_{i=1}^n X_n$和$\hat{\lambda}_2 = \sqrt{(n-1)/\sum_{i=1}^n (X_i - \bar{X})^2}$的偏差和方差。

2.给定随机数种子$set.seed(12)$，产生样本量为$n = 10$的来自指数分布$Exp(5)$的样本。根据该样本用Bootstrap方法估计$\hat{\lambda}_1$和$\hat{\lambda}_2$的偏差和方差。

五、（20分）

考虑混合泊松分布$p(x) = \sum_{j=1}^2 \alpha_j p_j(x; \lambda_j)$，其中$p_j(x; \lambda_j)$是参数为$\lambda_j$的泊松分布的分布列。

1.产生$m = 2000$个来自混合泊松分布$p(x)$的随机数，其中$\alpha_1 = 0.3, \alpha_2 = 0.7, \lambda_1 = 4, \lambda_2 = 8$。计算这组随机数的均值和方差。

2.给定样本量为$n = 10$的样本$(11, 3, 3, 9, 2, 10, 7, 7, 8, 5)$，用EM算法估计参数$\alpha_1, \alpha_2, \lambda_1, \lambda_2$。(给定初值$\alpha_1^0 = 0.4, \alpha_2^0 = 0.6, \lambda_1^0 = 2, \lambda_2^0 = 4$。)

提示：若随机变量$Y_i = 1$表示第i个样本点来自第1个泊松分布，$Y_i = 2$表示第i个样本点来自第2个泊松分布，则根据完全数据(X_i, Y_i)，对于$j = 1, 2$，参数的极大似然估计为

$$\hat{\alpha}_j = \frac{\sum_{i=1}^n \mathbf{1}\{y_i = j\}}{n},$$
$$\hat{\lambda}_j = \frac{\sum_{i=1}^n x_i \mathbf{1}\{y_i = j\}}{\sum_{i=1}^n \mathbf{1}\{y_i = j\}}.$$

隐变量$Y_i = j$的后验概率为

$$\Pr\{Y_i = j | x, \alpha_1, \alpha_2, \lambda_1, \lambda_2\} = \frac{\alpha_j p_j(x_i; \lambda_j)}{\alpha_1 p_1(x_i; \lambda_1) + \alpha_2 p_2(x_i; \lambda_2)}.$$