

中期Project候选题目

2009年11月17日

注意事项

Project作业应以研究论文的形式提交. 包括: 标题, 作者, 摘要, 引言(背景介绍, 回答这个Project的问题有什么意义, 已有的研究方法等, 本Project拟研究的问题), 方法(理论模型, 理论结果, 模拟方法介绍等); 模拟研究(应用模拟数据分析等), 总结, 参考文献, 附录(模拟研究中的代码等). 原则上使用L^AT_EX软件书写, 最后提交pdf文档.

时间: 11月9日至11月30日

一些题目:(选择其一或者自己寻找题目)

1 使用Monte Carlo模拟方法研究 t 检验对总体假设的稳健性. (查找理论研究文献), 通过选择一些非正态分布, 如 χ_1^2 , $U(0, 2)$ $Exp(1)$, 正态混合等等总体, 研究假设检验 $H_0: \mu = 1 \leftrightarrow H_1: \mu \neq 1$ 的一型错误率和名义错误率之间的关系.

2 对二项分布 $B(n, p)$, 在 n 较小时 p 的置信区间已有几种不同的构造方法, 如Wald置信区间, Adjusted Wald置信区间, "Score"(渐近的)置信区间, Clopper-Pearson置信区间等等. 查找文献了解这些区间的形式和想法. 使用Monte Carlo方法研究, 在不同的样本量 n 大小情形下, 各种方法的优劣(即覆盖率和名义置信水平之间的关系).

3 构造不同的概率模型, 用其估计圆周率. 并对每种模型, 选用几种不同的Monte Carlo方法以及方差减少技术, 研究它们的差异.

4 R中的函数`cor.test`应用Pearson相关系数 ρ , Spearman秩相关系数 ρ_s , 或者Kendall系数 τ 这三种统计量来检验两组数据之间的相关性. 查找文献了解这三种统计量的结构, 使用Monte Carlo方法研究(1) (X, Y) 的联合分布为二元正态分布时, 证实基于Pearson相关系数统计量的检验功效最高. (2) 寻求一个二元分布, 使得基于Spearman秩相关系数 ρ_s , 或者Kendall系数 τ 的非参数检验, 至少有一个功效比基于Pearson相关系数统计量的检验功效高.

5 比较Bootstrap和Jackknife方法在统计推断中的应用.

6 使用Monte Carlo方法研究标准正态分布, χ_n^2 , ($n = 5, 20$)下总体均值的几种Bootstrap置信区间: 标准正态Bootstrap置信区间, Basic Bootstrap置信区间, Percentile Bootstrap置信区间等的经验覆盖率. 找出这些置信区间分别在左边和右边错失均值的比例.(左边错失: 即置信区间的下界比真值要大; 右边错失: 即置信区间的上界比真值要小). 与总体分布偏度的程度有何关联?

7 在应用Metropolis-Hastings算法中, 以目标分布(1) $0.5N(1, 1) + 0.5N(-1, 0.5^2)$ (2) t_5 (3) Gumbel分布 $f(x) = e^{-x-\gamma} \exp(-e^{-x-\gamma})$, $\gamma = -\text{Euler's constant}$ 为例, 研究提议分布为 $q(y|x_{t-1}) = N(a, b)$ (独立抽样) 与 $q(y|x_{t-1}) = N(x_{t-1}, d)$ (随机游动Metropolis算法) 时, 超参数 a, b, d 的值对MCMC抽样性质的影响. 比较准则比如是目标分布期望和方差等数字特征的均方误差等等.