

概统复习讲座

初等概率论

概率基础

- 样本空间与概率公理化定义（一般不考，但是要熟悉定义）
- 事件运算与概率空间性质（熟悉即可，较为直觉）
- 条件概率与Bayes公式、全概率公式（常考，但简单，熟记）
- 独立性与条件独立

随机变量（一维）

- 随机变量类型：离散型、连续型、混合型
- 随机变量的刻画方式：CDF（泛用）；PDF（连续）；PMF（离散）
 - CDF三性质：单调不减、双侧极限、右连续
- 离散随机变量
 - 期望与方差（标准差）
 - PMF
- 常见离散分布
 - Bernoulli分布与二项分布： $X \sim B(n, p)$
 - 可加性
 - $E(X) = np, Var(X) = np(1 - p)$
 - 现实意义
 - Poisson分布： $X \sim P(\lambda)$
 - 可加性
 - $E(X) = Var(X) = \lambda$
 - 与Bernoulli的关系
- 连续随机变量
 - PDF（是否具有唯一性？大小的意义？）
 - 离散与方差
- 常见连续分布

- 均匀分布: $X \sim U(a, b)$
 - $E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a)^2}{12}$
- 正态分布: $X \sim N(\mu, \sigma^2)$
 - 可加性
 - 标准正态分布
 - 应用意义
- 指数分布: $X \sim Exp(\lambda)$
 - $E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$
 - 无记忆性
 - Weibull分布
- 卡方分布: $X \sim \chi^2(n)$
 - 与正态分布的关系
 - $E(X) = n, Var(X) = 2n$
 - 卡方检验
- Cauchy分布
 - 方差与期望?
 - 与正态分布的关系
- t分布
 - 与正态分布的关系
 - t检验 (对比正态检验)
- F分布
 - 与正态分布的关系
 - F检验
- 随机变量的函数
 - PMF/PDF求解
 - 线性函数的期望与方差

联合分布 (多维)

- 随机向量与CDF
- 联合PMF与PDF
- 联合分布

- 多元正态分布
- 多项分布
- 均匀分布
- 边际分布与条件分布
 - 计算公式
 - 常见条件分布
- 独立性
 - 独立同分布
 - 常见判断
- 随机向量函数
 - 密度函数变化法
 - 卷积法

随机变量的数字特征

- 期望与方差
- 分位数（唯一性）
- 众数
- 协方差与相关系数
 - 随机变量线性组合的方差
 - 协方差矩阵
 - 不相关vs独立
- 矩
 - 中心矩与标准矩
 - 偏度与峰度
 - 矩母函数(MGF)
 - 母函数，特征函数
 - 与各阶矩间关系
 - 证明同分布
 - 常见分布矩母函数
 - 条件期望
 - 全期望公式

- 均方最优预测

概率不等式与各类定理

- 概率不等式（很少考到）
 - Markov
 - Chebyshev
 - Chernoff
- 大数定律
 - Khinchin弱大数定律
 - 应用条件
 - 相关推广
 - Kolmogorov强大数定律
 - 应用条件
 - 概率的频率解释
 - Monte Carlo方法
 - 几乎必然收敛、依概率收敛、依分布收敛
- 中心极限定理
 - 二项分布的正态近似
 - 正态分布的优越性

统计推断

统计基础

- 样本期望、样本方差
- 统计量

参数估计

- 矩估计
 - 原理依据：大数定律
 - 尽可能使用低阶矩
- 极大似然估计
 - 似然函数与对数似然函数
 - 不变性

- Fisher信息量（常考！）
- Bayes估计
 - 先验知识与后验分布
 - 估计方式：3种
 - MLE与Bayes估计
- 优良性准则
 - 小样本
 - 无偏性
 - MSE拆分：Bias-Variance trade-off
 - MVUE
 - 大样本
 - 渐近无偏性
 - 相合性
 - 渐进正态性
- 置信区间
 - 双侧区间与单侧区间
 - 统计量选取
 - 渐进置信区间
 - MLE置信区间

假设检验

- 检验统计量
- 假设检验与置信区间
- 一、二类错误
 - α & β
 - 功效
 - 拒绝与接受
- P值：定义与意义
- 常见假设检验
 - t检验：未知方差，小样本（Paired t-test!）
 - 正态检验：已知方差/大样本/异方差

- F检验：比较方差
- 卡方检验：拟合优度检验与独立性检验
- 以上均需注意单边/双边!
- 似然比检验与Bayes检验
 - 似然比卡方检验

线性回归

- 有监督学习
- 基本假设
 - Gaussian-Markov
 - 同方差
- 简单线性回归
 - 最小二乘参数估计
 - sigma估计
 - beta标准误估计
 - 显著性检验 (t检验)
 - 预测
 - 预测区间与置信区间

样题

1. 参数求解

假设 X 和 Y 都是正的随机变量，其联合概率密度函数是 $f(x, y) = \frac{c}{y} \exp\left\{-\left(y + \frac{x}{y}\right)\right\}$, $0 < x, y < \infty$, 这里 c

是常数.

- (1) 确定常数 c 的值.
- (2) 计算协方差 $\text{Cov}(X, Y)$.
- (3) 计算条件期望 $E[e^{-X} | Y = 1]$.

2. 随机变量的函数

设随机样本 X_i ($i=1, \dots, n$) 来自总体 $U(0,1)$, 假设 $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$.

- (1) 分别求 $X_{(1)}$, $X_{(n)}$ 的概率密度函数.
- (2) 求 $X_{(1)}$, $X_{(n)}$ 的联合概率密度函数.
- (3) 求 $E(X_{(n)} + X_{(1)})$.
- (4) 求 $X_{(n)} - X_{(1)}$ 的概率密度函数.

3. 极限定理应用

某校有1600名学生将参加通识讲座, 共有5个讲座同时开讲, 假定每位同学是随机地选择一个讲座去听, 而且同学之间的选择是相互独立的, 如果某间阶梯教室想以不低于95%的概率保证在这间教室听讲座的学生都有座位, 那么这间阶梯教室至少需要设有多少座位? (需给出求解过程)

4. 状态转移问题

一个小孩在吃彩虹糖, 共有n种颜色, 每次等概率吃到其中一种颜色的糖果, 求吃遍所有颜色糖果所需要次数的期望。

5. 配对问题

有n个人, 每人有一顶帽子。现将所有帽子放到一起, 再随机分配给每人一顶, 记拿到自己帽子的人数为X, 求X的期望与方差。

6. 置信区间求解

某个工厂生产螺丝, 公差iid服从某一正态分布, 相关参数未知。现有其生产的n个零件 x_1, x_2, \dots, x_n , 现在希望给出某一新生产螺丝公差的95%显著区间。

7. 极大似然估计

(22 分) 设随机样本 X_1, \dots, X_n 来自具有概率密度函数为 $f(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{其它} \end{cases}$ 的分布, 其中 $\theta > 0$ 是未知参数.

知参数.

- (1) 求 θ 的极大似然估计量 θ^* .
- (2) 判断 $\frac{1}{\theta^*}$ 是否是 $\frac{1}{\theta}$ 的无偏估计并给出相应理由.
- (3) 当 $n=4$ 时, 一组样本观测值为 $\frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{5}$, 利用 Fisher 信息量和这组观测值给出 θ^* 的标准误差的估计.
- (4) 利用 θ^* 和 (3) 中样本观测值给出 θ 的满足 95% 置信要求的一个具体置信区间.
- (5) 考虑假设 $H_0: \theta=1$ v.s. $H_1: \theta \neq 1$, 利用 (3) 中样本观测值, 在检验水平 $\alpha=0.05$ 下进行似然比检验.