Probability and Statistics

- 概率
  - 随机事件
    - 确定事件与随机事件
  - 样本点和样本空间
  - 频率
    - 频率在一定程度上反映了事情发生的可能性
    - 性质
      - 0 ≤ f ≤ 1
      - 规范性：样本空间的频率为 1
      - 可列可加性
      - 稳定性（统计规律性）
  - 定义
    - 统计学定义
      - 随机事件发生频率稳定在的常数
      - 性质
        - 非负性
        - 规范性：样本空间概率为 1
        - 可列可加性
    - 公理化定义
      - 可测空间 (Ω, Σ) 中的函数 P: Σ → [0,1]，满足非负性、规范性、可列可加性，则称 P(A) 为事件 A 的概率，(Ω, Σ, P) 为概率空间
  - 计算
    - 容斥原理
    - 概率模型
      - 古典概型
      - 超几何概型
        - 概率 = 事件包含的样本点数 / 样本空间包含的样本点数
即样本空间是由有限个等可能的基本事件构成的
      - 几何概型
        - 概率 = 事件的测度 / 样本空间的测度
即样本空间无限可测，而基本事件等可能发生
    - 组合计数

n 个球	m 个箱子	无限制	每个箱子至多 1 个球（$m \ge n$）	每个箱子至少 1 个球（$n \ge m$）
不同	不同	每个球随意选箱子：m^n	箱子固定球排列：$(m)_n$	斯特林数再组合：$m!S(n,m)$
相同	不同	插板法（可空）：$\binom{n+m-1}{m-1}$	选箱子：$\binom{m}{n}$	插板法（非空）：$\binom{n-1}{m-1}$
不同	相同	斯特林数求和：$\sum_{i=1}^{m} S(n,i)$	1	斯特林数：$S(n,m)$
相同	相同	整数拆分求和：$\sum_{i=1}^{m} P(n,i)$	1	整数拆分：$P(n,m) = P(n-1,m-1) + P(n-m,m)$

  - 条件概率
  - 随机变量
    - 定义
      - 函数：随机事件取值 → 实数
    - 分布列
      - 取值-概率表
    - 数字特征
      - 期望
      - 方差
    - 概率分布函数
      - 概念
        - F(x) = P(X ≤ x)
    - 离散型随机变量
      - 分布模型

分布模型	问题描述	表示	分布列	期望	方差
01分布	伯努利试验：事件发生的概率	$X \sim \mathrm{Ber}(p)$	$P(X=0)=1-p$ $P(X=1)=p$	$E(X)=p$	$VAR(X)=p(1-p)$
二项分布	n 重伯努利试验：事件发生次数的概率	$X \sim B(n,p)$	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$E(X)=np$	$VAR(X)=np(1-p)$
泊松分布	大量实验中稀有事件发生次数的概率	$X \sim P(\lambda)$	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	$E(X)=\lambda$	$VAR(X)=\lambda$
几何分布	重复伯努利实验，第一次发生事件时实验次数的概率	$X \sim G(p)$	$P(X=k)=(1-p)^{k-1}p$	$E(X)=\frac{1}{p}$	$VAR(X)=\frac{1-p}{p^2}$
Pascal 分布	重复伯努利实验，第 r 次发生事件时实验次数的概率	$X \sim \mathrm{pascal}(r,p)$	$P(X=k)=\binom{k}{r-1}p^r(1-p)^{k-r-1}$	$E(X)=\frac{r}{p}$	$VAR(X)=\frac{r(1-p)}{p}$

      - 泊松分布与二项分布
        - 泊松定理
          - 在 n 足够大，p 足够小的时候，可以用泊松分布近似二项分布
    - 非离散型随机变量
      - 连续型随机变量
        - 概率密度函数
        - 分布模型

分布模型	问题描述	表示	分布函数	密度函数	期望	方差
均匀分布	在区间 $[a,b]$ 上取值概率相等	$X \sim U(a,b)$	$F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$	$f(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & \text{otherwise}\end{cases}$	$E(X)=\frac{a+b}{2}$	$VAR(X)=\frac{(b-a)^2}{12}$
指数分布	在 $(0,+\infty)$ 上的指数函数	$X \sim e(\lambda)$	$F(x)=\begin{cases}0 & x\le 0\\ 1-e^{-\lambda x} & x>0\end{cases}$	$f(x)=\begin{cases}0 & x\le 0\\ \lambda e^{-\lambda x} & x>0\end{cases}$	$E(X)=\frac{1}{\lambda}$	$VAR(X)=\frac{1}{\lambda^2}$
正态分布	在 $\mathbf{R}$ 上的正态分布	$X \sim \mathcal{N}(\mu,\sigma^2)$	无解析解	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(x-\mu)^2}{2\sigma^2}}$	$E(X)=\mu$	$VAR(X)=\sigma^2$

  - 多维随机变量
    - 概念
      - 多维随机变量
    - 分布与密度
    - 数字特征
      - 期望
      - 方差
    - 条件
    - 函数
      - 运算
  - 大数定律与中心极限定理
    - 大数定律
    - 中心极限定理
- 统计
  - 总体与样本
    - 概念
    - 抽样
  - 统计量
    - 原点矩
    - 中心距
    - 次序统计量
  - 正态有关分布
    - 卡方分布
    - t 分布
    - F 分布
  - 参数估计
  - 假设检验