统计和概率

统计

- 均值 $\bar{Y} = \frac{Y_1+Y_2+...+Y_N}{N} = \frac{1}{N}\sum_{i=1}^{N}Y_i$
- 样本均值 $\bar{y} = \frac{y_1+y_2+...+y_n}{n} = \frac{1}{n}\sum_{i=1}^{n}y_i$
- 中位数：一组数据中大小在正中间的数。(如果总数为偶数，则为中间两个数的平均数)
- 众数：一组数据中出现频率最高的数。
- 总体方差(Variance)：$S^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$
- 总体标准差(Standard Variance)：$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$
- 样本标准差(Sample Variance)：$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

排列组合

- $C_n^0 = C_n^n = 1$
  $A_n^m = \frac{n!}{m!}$
  $C_n^m = C_n^{n-m} = \frac{n!}{m!(n-m)!}$
- $C_{n+1}^m = C_n^m + C_n^{m-1}$

  理解：从 n+1 个学生里选 m 个学生，分两种情况。

  1. 入选名单中不含学号 n+1 的学生，一共有C_n^m种选法
  2. 入选名单中包含学号 n+1 的学生，那么在前 n 个学生中就要少选一个，一共C_n^{m-1}种选法
- $mC_n^m = nC_{n-1}^{m-1}$

  理解：从 n 个学生里选 m 个学生，可以先选出一个学生，再从剩下的n-1个学生中选m-1个学生。发现每个学生组合都重复了m次。
- 二项式定理

  $(a+b)^n = C_n^0a^n + C_n^1a^{n-1}b + C_n^2a^{n-2}b^2 + \cdots + C_n^ra^{n-r}b^r + \cdots + C_n^nb^n \ (n \in \mathbf{N}_+)$

二点分布

- 如果随机变量X的分布列为:

X	1	0
P	p	q

  其中$0 < p < 1, q = 1-p$，则称随机变量 X 服从参数为p的二点分布。
- 期望值：$E(X) = p$
  方差：$D(X) = pq \quad (q = 1-p)$

二项分布

- 在 n 次独立重复重复试验中，每次试验事件 A 发生的概率为p，不发生的概率为$q=1-p$，离散型随机变量 X 为 A 发生的次数。称X服从参数为n,p的二项分布
- 记作$X \sim B(n,p)$
- X的分布列：$P(X=k)=C_n^k p^k q^{n-k}, \quad k=0,1,2,\ldots,n$

  期望值：$E(X)=np$

  方差：$D(X)=npq \quad (q=1-p)$

超几何分布

- 设有总数为N件的两类物品，其中一类有M件，从所有物品种任取n件$(n \leq N)$，这n件中所含这类物品件数X是一个离散型随机变量，称X服从参数为N,M,n的超几何分布

  X的分布列：$P(X=m)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \quad (0 \leqslant m \leqslant l, l\text{为}n\text{和}M\text{中较小的一个})$

  期望值：$E(X)=\frac{nM}{N}$

  方差：$D(X)=$

一堆球有N个，其中M个是红球。每次取一个球，取n次。如果每取完一个就放回去，满足二项分布。如果每去玩一个后不放回去，满足超几何分布

概率

- $P(A) \geq 0, P(\Omega)=1, P(\varnothing)=0$
- $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
  - 如果事件 A 与 B 互斥，则$P(A \cup B)=P(A)+P(B)$
  - 如果事件 A 与 B 互为对立事件，则$1=P(A \cup B)=P(A)+P(B)$
- 相互独立事件：如果 $P(AB)=P(A)P(B)$，则称 A 和 B 相互独立
- 贝叶斯定理：$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(B|A)PA}{P(B)}$
  - $P(A|B)$是已知B发生后，A的条件概率。也由于得自B的取值而被称作A的后验概率。
  - $P(A)$是A的先验概率（或边缘概率）。之所以称为"先验"是因为它不考虑任何B方面的因素。
  - $P(B|A)$是已知A发生后，B的条件概率。也由于得自A的取值而被称作B的后验概率。
  - $P(B)$是B的先验概率。
  - $\text{后验概率}=(\text{似然性}\times\text{先验概率})\div\text{标准化常量}$

正态分布

- 概率密度函数：$f(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbf{R}, \sigma>0, \mu \in \mathbf{R}$
- 通常记为 $N(\mu,\sigma^2)$
- 正态变量在区间$(\mu-\sigma, \mu+\sigma)$内，取值概率 68.3%。
- 正态变量在区间$(\mu-2\sigma, \mu+2\sigma)$内，取值概率 95.4%。
- 正态变量在区间$(\mu-3\sigma, \mu+3\sigma)$内，取值概率 99.7%。