Probability & Statistics

1.样本空间

2.事件

事件的运算与关系

- 和事件 (并集) A ∪ B = {x| x ∈ A or x ∈ B} ⇔ A与B至少有一发生
- 积事件 (交集) A ∩ B = {x| x ∈ A and x ∈ B} ⇔ A与B同时发生
- $\bigcup_{i=1}^{n} A_i$ ⇔ 该和事件发生 ⇔ $A_1, A_2, \ldots A_n$ 中至少有一发生
- $\bigcap_{i=1}^{n} A_i$ ⇔ 该积事件发生 ⇔ $A_1, A_2, \ldots A_n$ 同时发生
- 当AB = ∅, 称事件A与B不相容或互斥
- 差事件 A - B = {x ∈ A and x ∉ B}
- A的逆事件(对立事件) $A \cup \overline{A} = S, A\overline{A} = \varnothing, \overline{\overline{A}} = A$
- 引入逆事件后,**A**与**B**的差事件可表示为$A - B = A\overline{B} = A \cup B - B = A - AB$

事件的运算定律(集合的运算定律)

- 交换律 $A \cup B = B \cup A, A \cap B = B \cap A$
- 结合律 $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$
- 分配律 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- 对偶律 $\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$
- 对偶律推广 $\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}, \overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$
- **Attention:** $\overline{AB} \neq \overline{A}\overline{B}, \overline{AB} = \overline{A \cap B}, \overline{A}\overline{B} = \overline{A} \cap \overline{B}$
- $\overline{AB} = (\overline{A}\overline{B}) \cup A\overline{B} \cup \overline{A}B \Leftrightarrow A, B\text{不同时发生} = A, B\text{都不发生} \cup A\text{发生而}B\text{不发生} \cup B\text{发生而}A\text{不发生}$

Summary:

1. 至少有一发生
   - 和事件
   - $A \cup B$
   - $\bigcup_{i=1}^{n} A_i$
2. 同时发生
   - 积事件
   - $A \cap B$
   - $\bigcap_{i=1}^{n} A_i$
3. 都不发生
   - 至少有一发生的逆事件
   - $\overline{A \cup B} = \overline{A} \cap \overline{B}$
   - $\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$
4. 不同时发生 ⇔ 至少有一事件不发生 ⇔ 最多有**n-1**个事件发生
   - 同时发生的逆事件
   - $\overline{A \cap B} = \overline{A} \cup \overline{B}$
   - $\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}$

3.频率

频率的稳定值 -> p -> 概率

4. 概率

概率的统计性定义

> 当试验的次数增加时,随机事件A发生的频率的稳定值p称为概率,记为P(A) = p

概率的公理化定义

> 设随机试验对应的样本空间为S,对每个事件A定义P(A)满足三公理：

- 非负性: $P(A) \geq 0$
- 规范性: $P(S) = 1$
- 可列可加性: $A_i A_j = \varnothing, i \neq j \Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

公理化定义推导出的性质

1. $P(\varnothing) = 0$
2. $P(A) = 1 - P(\overline{A})$
3. (有限可加性) $A_1, A_2, \ldots, A_n, A_i A_j = \varnothing, i \neq j \Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
4. $A \subset B \Rightarrow P(B - A) = P(B) - P(A) \geq 0$; 一般情况: $P(B - A) = P(B) - P(AB)$
5. 概率加法公式:
   $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j \leq k < n} P(A_i A_j A_k) + \ldots + (-1)^{i-1} P(A_1 A_2 \ldots A_n))$

古典概型(等可能概型)

样本点有限且等可能

> 基本就是数数问题

1. 抽球问题(放回/不放回)(考虑顺序/不考虑顺序)
2. 分房问题
3. 随机取数问题

条件概率

$P(B) \neq P(B|A)$

> 缩小的样本空间:事件A的发生改变了样本空间,使它由原来的S缩减为新的样本空间
> $S_A = A$

> 以A的视角来看待B事件: B在A中所占的比例

定义

$$P(B|A) = \frac{P(AB)}{P(A)}, P(A) > 0$$

性质

$P(\bullet|A)$ 是概率, 因而满足三性:

- 非负性: $P(B|A) \geq 0$
- 规范性: $P(S|A) = 1$
- 可列可加性: $B_1, B_2, \ldots, B_i B_j = \varnothing, i \neq j \Rightarrow P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$

进而满足上述所有的概率性质

- $P(B|A) = 1 - P(\overline{B}|A)$
- $P(B \cup C|A) = P(B|A) + P(C|A) - P(BC|A)$
- $B \supset C \Rightarrow P(B|A) \geq P(C|A)$
- ...

乘法公式

$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$
$P(A_1A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \dots P(A_n|A_1 \dots A_{n-1})$

全概率公式

定义

$P(A) = \sum_{j=1}^{n} P(B_j) \cdot P(A|B_j)$

证明

$\because A = AS = AB_1 \cup AB_2 \cup \dots \cup AB_n,$

$AB_iAB_j = \varnothing, i \neq j$

$\therefore P(A) = \sum_{j=1}^{n} P(AB_j) \Leftrightarrow P(A) = \sum_{j=1}^{n} P(B_j) \cdot P(A|B_j)$

样本空间的变换

若设$P(B_j) = p_j, P(A|B_j) = q_j, j = 1, 2, \dots, n$

则 $P(A) = \sum_{j=1}^{n} pjq_j$

Bayes公式

定理

> 设$B_1, B_2, \dots, B_n$为S的一个划分且$P(B_i) > 0$.对$P(A) > 0$有$Bayes$公式：

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} = \frac{p_iq_i}{\sum_{j=1}^{n} p_jq_j}$$

分母即为全概率公式，分子由乘法公式得到 ; $P(B_i)$ 为先验概率, $P(B_i|A)$ 为后验概率

Note:

1. 乘法公式：
   $$P(B_i|A) \cdot P(A) = P(AB_i) = P(A|B_i) \cdot P(B_i) \Rightarrow P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)}$$
2. 全概率公式: $P(A) = \sum_{j=1}^{n} P(AB_j) \Leftrightarrow P(A) = \sum_{j=1}^{n} P(B_j) \cdot P(A|B_j)$

事件的独立性

定义

> 设A,B是两随机事件,如果$P(AB) = P(A)P(B)$,则称A,B相互独立.

特点

- A与B的对称性
- 不需要**条件概率**存在的条件,即事件的概率可以为0

$If: P(A) > 0, P(B) > 0, then: P(AB) = P(A)P(B) \Leftrightarrow P(B|A) = P(B) \Leftrightarrow P(A|B) = P(A)$

性质

A, B相互独立 $\Leftrightarrow \overline{A}, B$ 相互独立 $\Leftrightarrow A\overline{B}$ 相互独立 $\Leftrightarrow \overline{A} \cap \overline{B}$ 相互独立

推广定义

> 设$A_1, A_2, \ldots, A_n$为n个随机事件,若对$2 \le k \le n$,均有:
>
> $$P(A_{i_1}A_{i_2}\ldots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$

> 则称$A_1, A_2, \ldots, A_n$相互独立

5. 随机变量

> 中心问题: 将试验结果数量化

一维随机变量

定义

> 设随机试验的样本空间为S,e为样本点,若$X = X(e)$ 为定义在S上的实值单值函数,则称$X(e)$ 为随机变量,**简写为**X.

> 该变量实质上为一映射: 样本空间到实数上的映射

说明

- 随机变量$X(e) : S \to R$为一映射,其自变量具有随机性
- 随机事件可表示为 $A = \{e : X(e) \in I\} = \{X \in I\}, I \subset R$
- 映射的性质: 对于$i \neq j$, 必有 $\{X = i\} \cap \{X = j\} = \varnothing$

离散型随机变量

描述方式

- 分布律
- 分布函数

离散型随机变量的分布函数

> 一般地,离散型随机变量的分布函数为阶梯函数.

设离散型随机变量X的分布律为$P\{X = x_k\} = p_k, k = 1, 2, \ldots$ X的分布函数为 $F(X) = \sum_{x_k \le x} p_k$

$F(X)$ 在 $x = x_k, (k = 1, 2, \ldots)$ 处有跳跃,其跳跃值为$p_k = P\{X = x_k\} \to$ 右连续

常见的离散型随机变量

1. 退化分布
2. 0-1分布/两点分布/Bernoulli分布 $\Leftrightarrow X \sim 0-1(p) \Leftrightarrow X \sim B(1, p)$
3. 二项分布 $(Binomial)$ /n重Bernoulli试验的分布 $\Leftrightarrow X \sim B(n, p)$
4. 泊松分布 $(Poisson) \Leftrightarrow X \sim \pi(\lambda) \Leftrightarrow X \sim P(\lambda)$
5. 几何分布 $(Geometric) \Leftrightarrow X \sim Geom(p)$

二项分布

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \ldots, n.$$

泊松分布

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots, \lambda > 0$$

泰勒展开式

$e^{\lambda} = \sum_{k=0}^{+\infty} \dfrac{\lambda^k}{k!}$

作为二项分布的近似

n很大,p小时: $\lambda = np$

几何分布

$P(X=k) = p(1-p)^{k-1}, k = 1, 2, 3, \ldots.$

连续型随机变量

> 其定义是伴随着概率密度函数的定义产生的

描述方式

- 分布函数

定义

> 对于随机变量X的分布函数F(x),若存在非负的函数f(x),使对于任意实数x有:
>
> $$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

> 则称X为连续型随机变量,其中f(x)称为X的概率密度函数,**简称概率密度**,有时也写为 $f_x(x)$

概率密度函数的性质

- $f(x) \geq 0$
- $\because F(+\infty) = 1 \therefore \int_{-\infty}^{+\infty} f(x)\, dx = 1$
- $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(t)\, dt \Leftrightarrow P(X \in D) = \int_D f(x)\, dx \Rightarrow P(X=a) = 0, P(x_1 < X < x_2) = P(x_1 < X \leq x_2)$
- 在 $f(x)$ 连续点 $x, F'(x) = f(x) \Rightarrow P(x < X \leq x + \Delta x) \approx f(x) \cdot \Delta x$
- $f(x)$ 的值可以大于1
- $f(x) \xleftrightarrow[\frac{d}{dx}F(x)]{\int_{-\infty}^{x} f(t)\, dt} F(x)$

常见的连续型随机变量

1. 均匀分布 $(Uniform) \Leftrightarrow X \sim U(a,b) \Leftrightarrow X \sim Unif(a,b)$
2. 指数分布 $(Exponential) \Leftrightarrow X \sim E(\lambda) \Leftrightarrow X \sim Exp(\lambda)$
3. 正态/高斯分布 $(Normal) \Leftrightarrow X \sim N(\mu, \sigma^2)$

均匀分布

概率密度函数

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b) \\ 0, & \text{else} \end{cases}$$

分布函数

$$F(x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases}$$

性质

> 均匀分布具有等可能性,即X落入(a,b)中等长度的任意子区间上是等可能的

指数分布

概率密度函数

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}$$

分布函数

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}$$

性质

> 指数分布具有无记忆性

正态分布

概率密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty,(\sigma>0)$$

标准正态分布的概率密度函数

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}},(\sigma=1,\mu=0)$$

分布函数

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

标准正态分布的分布函数

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-t^2}{2}}\,dt,(\sigma=1,\mu=0)$$

$f(x)$ 性质

- $f(x)$ 关于$x=\mu$对称
- $f(x)$ 在$x=\mu$达到最大值 $\frac{1}{\sqrt{2\pi}\sigma}$
- $f(x)$ 的图像以Ox轴为渐近线且在 $x=\mu\pm\sigma$ 处为拐点

$\Phi(x)$ 性质

- $\Phi(-x)=1-\Phi(x)$
- $\Phi(\frac{x-\mu}{\sigma})=F(x),x\in R\Leftrightarrow Z=\frac{X-\mu}{\sigma}\sim N(0,1)$
- $P\{a<X<b\}=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})\Rightarrow 3\sigma$准则

一维随机变量的函数的分布

对于$Y=g(X),h(x)$是$g(x)$的反函数：

$$f_Y(y)=\begin{cases}f_X[h(y)]|h'(y)|, & \alpha<y<\beta;\\ 0, & \text{其他}.\end{cases}$$

一维随机变量的数字特征

一维随机变量的数学期望

一维离散型随机变量的数学期望

设离散型随机变量的分布律为: $P\{X=x_i\}=p_i$, 若如下级数绝对收敛则其为X的期望:

$$E(X)=\sum_{i=1}^{\infty} x_i p_i$$

常见离散型随机变量的期望

- $X \sim b(n,p) \to E(X)=np$
- $X \sim \pi(\lambda) \to E(X)=\lambda$
- $X \sim Ge(p) \to E(X)=\frac{1}{p}$

一维连续型随机变量的数学期望

设连续型随机变量X的概率密度函数为$f(x)$, 则当 $\Delta x(\Delta x>0)$ 充分小时,随机变量X的取值落在区间 $(x,x+\Delta x]$ 内的概率近似为 $f(x)\Delta x$, 类似于离散型随机变量的期望定义,有:

$$E(X)=\int_{-\infty}^{+\infty} x\,f(x)\,dx$$

常见连续型随机变量的期望

- $X \sim U(a,b) \to E(X)=\frac{a+b}{2}$
- $X \sim E(\lambda) \to E(X)=\frac{1}{\lambda}$
- $X \sim N(\mu,\sigma^2) \to E(X)=\mu$

一维随机变量函数的数学期望

> $g(x)$ 要为实值连续函数

一维离散型随机变量函数的数学期望

$$E(Y)=E[g(X)]=\sum_{i=1}^{\infty} g(x_i)p_i$$

一维连续型型随机变量函数的数学期望

$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)\,f(x)\,dx$$

一维随机变量的方差

$$D(X)=Var(X)=E[X-E(X)]^2$$

$$D(X)=E(X)^2-[E(X)]^2$$

常见随机变量的方差

- $X \sim b(1,p) \to D(X)=p(1-p)$
- $X \sim \pi(\lambda) \to D(X)=\lambda$
- $X \sim U(a,b) \to D(X)=\frac{(b-a)^2}{12}$
- $X \sim E(\lambda) \to D(X)=\frac{1}{\lambda^2}$
- $X \sim N(\mu,\sigma^2) \to D(X)=\sigma^2$