

概率统计微课第 1 课：概率论基础

管枫

七月在线

May, 2017

主要内容

- 积分学
 - 理解积分：无穷求和，体积
 - 微积分基本定理：牛顿-莱布尼茨公式
- 概率空间
 - 随机变量与概率：概率密度函数的积分
 - 随机变量与概率：概率空间
 - 条件概率与贝叶斯公式
 - 常见分布与共轭分布

- 本节课常用数学记号

\mathbb{R}^n 实坐标空间

$f(x)$ 函数

$\int_a^b f(x)dx$ 函数积分

X 随机变量

$E(X)$ 随机变量的期望

$Var(x)$ 随机变量的方差

$\mu_n(x)$ 随机变量的 n 阶矩

积分学: 理解积分: 无穷求和, 体积

Definition (单变量函数黎曼积分)

令 $f(x)$ 为开区间 (a, b) 上的一个连续函数, 对于任何一个正整数 n 定义, $x_i = a + \frac{i(b-a)}{n}$ 求和式:

$$S_n(f) = \sum_{i=0}^{n-1} f(x_i)(x_{i+1} - x_i)$$

如果极限 $\lim_{n \rightarrow \infty} S_n(f)$ 存在, 那么函数 $f(x)$ 在这个区间上的黎曼积分为

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} S_n(f)$$

积分学: 理解积分: 无穷求和, 体积

理解积分

- 代数意义: 无穷求和
- 几何意义: 函数与 X 轴之间的有向面积

此处课堂画图举例说明

积分学: 微积分基本定理: 牛顿-莱布尼茨公式

Theorem (牛顿-莱布尼茨公式)

如果 $f(x)$ 是定义在闭区间 $[a, b]$ 上的可微函数, 那么就有

$$\int_a^b f'(t)dt = f(b) - f(a)$$

不定积分表示为

$$\int f'(t)dt = f(x) + C$$

牛顿-莱布尼茨公式展示了微分与积分的基本关系: 在一定程度上微分与积分互为逆运算.

积分学: 微积分基本定理: 牛顿-莱布尼茨公式

Example

函数 $\ln(x)$ 的不定积分

令 $f(x) = x \ln(x) - x$, 则 $f'(x) = 1 \cdot \ln(x) + x \cdot \frac{1}{x} - 1 = \ln(x)$.
根据牛顿-莱布尼茨公式我们得到

$$\int \ln(t) dt = \int f'(t) dt = x \ln(x) - x + C$$

积分学: 微积分基本定理: 牛顿-莱布尼茨公式

Example

函数 $\frac{1}{x^\alpha} = x^{-\alpha}$ 的不定积分

我们知道幂函数的导数公式 $\frac{d}{dx}x^n = n \cdot x^{n-1}$.

反向思维如果 $n - 1 = -\alpha$, 那么 $n = -\alpha + 1$. 所以令

$f(x) = x^{-\alpha+1}$, 那么 $f'(x) = (-\alpha + 1) \cdot x^{-\alpha}$.

于是当 $\alpha \neq 1$ 时,

$$\int \frac{1}{x^\alpha} dx = \frac{1}{(1 - \alpha)x^{\alpha-1}} + C$$

当 $\alpha = 1$ 的时候呢, 我们熟知 $\frac{d}{dx} \ln(x) = 1/x$. 所以

$$\int \frac{1}{x} dx = \ln(x) + C$$

积分学: 多变量积分

多变量函数的积分

如果一个函数 $f(x, y)$ 有多个变量, 那么在矩形 $[a, b] \times [c, d]$ 上的多重积分可以看成是每一个变量的依次积分

$$\int_c^d \int_a^b f(x, y) dx dy$$

- 如果积分区域形状不规则, 可以用一个矩形把积分区域包起来, 并令函数在积分区域外边等于 0.
- 二重积分的几何意义是积分函数与 $X - Y$ 坐标平面之间部分的有向体积.

小结 (积分学)

- 积分的代数意义是无穷求和，几何意义是带符号的体积
- 微分和积分在一定程度上互为逆运算
- 熟悉微分公式有助于计算积分
- 多重积分可以理解成是依次进行的单重积分

随机变量与概率：概率密度函数的积分

离散随机变量

假设随机变量 X 的取值域为 $\Omega = \{x_i\}_{i=1}^{\infty}$ ，那么对于任何一个 x_i ，事件 $X = x_i$ 的概率记为 $P(x_i)$ 。

对于 Ω 的任何一个子集 $S = \{x_{k_i}\}_{i=1}^{\infty}$ ，事件 $X \in S$ 的概率为

$$P(S) = \sum_{i=1}^{\infty} P(x_i)$$

对于离散随机变量，概率为概率函数的求和。

随机变量与概率：概率密度函数的积分

连续随机变量

假设随机变量 X 的取值域为 \mathbb{R} , 那么对于几乎所有 $x \in \mathbb{R}$, 事件 $X = x$ 的概率 $P(X = x)$ 都等于 0. 所以我们转而定义概率密度函数 $f: \mathbb{R} \rightarrow [0, \infty)$. 对于任何区间 (a, b) , 事件 $X \in (a, b)$ 的概率为

$$P((a, b)) = \int_a^b f(x) dx$$

- 对于连续型随机变量, 概率为概率密度函数的积分.
- 不论是离散还是连续型随机变量, 概率函数和概率密度函数的定义域即为这个随机变量的值域.
- 作为一个特殊的概率函数, 分布函数定义为
$$\Phi(x) = P(X < x).$$

我们在此课中只考虑几乎处处连续的概率密度函数, 我们不考虑离散, 连续混合型的随机变量

随机变量与概率: 概率空间

概率空间是概率论问题的一种数学表达方式

概率空间

概率空间由一个集合 Ω , 这个集合的若干子集组成的 σ 代数 \mathcal{F} , 以及一个概率测度 $P: \mathcal{F} \rightarrow \mathbb{R}$ 组成。

直观上可以这样理解概率空间的组成部分:

- Ω 代表全部可能出现的结果。
- \mathcal{F} 中每一个元素 A 都是 Ω 的一个子集合, 代表一类结果, 称为事件。
- $P: \mathcal{F} \rightarrow \mathbb{R}$ 作为概率函数代表每一个事件应对应的概率。
- 通常把事件 A 的概率写成 $\int_A dP(\omega)$.

随机变量与概率: 概率空间

概率空间

当事件可以由随机变量的取值唯一描述时, 可以使用随机变量的取值空间来定义概率空间。此时取值的区间 $I_{ab} = (a, b)$ 就构成基本的事件, 而密度函数的积分 $\int_a^b \rho(x)dx$, 则定义了事件 (a, b) 的概率 $P(I_{ab})$.

比如当问题研究的是滴滴司机何时到达目的地这样的随机变量时。

随机变量与概率: 概率空间

概率空间

当样本本身具有其自然的定义而不能由随机变量的取值唯一描述时, 可以使用全部样本来构成概率空间。此时样本的分类则构成了事件, 而每一种类型的样本占总体的比例就构成了这个事件的概率。

比如在垃圾邮件分类问题中, 可以将全部可能的邮件全体作为 Ω , 而邮件的每一种类型都对应一个事件, 比如垃圾邮件的全体 S . 而任选一个邮件, 这个邮件是垃圾邮件的概率 $P(S) = |S|/|\Omega|$ 为垃圾邮件占全部邮件的比例. 垃圾邮件分类要解决的问题就是, 在某些特定条件下估计一个邮件是垃圾邮件的概率。

随机变量与概率： 如何理解概率

事件的概率

- 整个概率空间是一个事件，这个事件一定发生所以全空间的概率为 1
- 事件是随机变量值域的子集 S
- 事件的概率则表示 S 里面概率之和或概率密度之积分.

随机变量的基本特征

随机变量常用的基本特征: 期望, 方差, 协方差, 相关系数。假设 X, Y 是概率空间 Ω 上的两个随机变量那么:

期望, 方差, 协方差, 相关系数的积分定义

- 期望 $E(X) = \int_{\Omega} X(\omega) dP(\omega)$
- 方差 $\text{Var}(X) = E(X^2) - E(X)^2$
- 协方差 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- 相关系数 $\text{Corr}(X, Y) = \text{Cov}(X, Y) / (\text{Var}(X)\text{Var}(Y))^{1/2}$

思考题: 为什么如此定义的方差一定非负?

事件的条件概率

- 条件也是事件，也可表示为随机变量值域的子集: A
- 条件概率里面的事件，又是这个条件的子集: $S \cap A \subset A$
- 事件的条件概率则表示 $S \cap A$ 在 A 里面所占的比例. 故而
$$P(S|A) = \frac{P(S \cap A)}{P(A)}$$

概率其实就是集合的大小比例，而概率函数或者概率密度函数可以理解为比较大小时候的权重

随机变量与概率：贝叶斯公式

贝叶斯公式

如果 A, B 是两个事件，那么条件概率满足公式

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

利用前面的定义我们知道，事件 A, B 同时发生的概率为 $P(A \cap B)$ ，一方面

$$P(A \cap B) = P(B|A)P(A)$$

另一方面对称的有

$$P(A \cap B) = P(A|B)P(B)$$

所以 $P(B|A)P(A) = P(A|B)P(B)$ ，两边同时除以 $P(B)$ 就得到了贝叶斯公式。

事件的独立性

如果两个事件 A, B 满足 $P(A) = P(A|B)$ 那么就说这两个事件是独立的。利用贝叶斯公式，这个定义等价于

独立事件

$$P(A \cap B) = P(A)P(B)$$

随机变量的独立性

两个随机变量独立，直观理解就是说这两个变量的取值彼此之间没有影响。具体来说：

独立变量

如果 X, Y 是两个随机变量而且对于任意的 x, y 都有 $P(X \leq x) = P(X \leq x | Y \leq y)$ 则说这两个变量是独立的。

思考题：是否存在相关系数为零，但是互相不独立的随机变量。

随机变量与概率：共轭分布

常见的概率分布基本上都有参数，比如正态分布有 (μ, σ) 两个参数，泊松分布有一个参数 λ . 对于一个具体的问题而言，关于这些参数有两种不同的看法

- 利用经验得到一个关于参数的先验分布.(Bayesian)
- 不对参数先验分布做任何假设，只利用当前观测的数据来对参数进行估计.(Frequentist)

本课对这两种观点相关的方法都进行介绍，本章介绍的先验分布，以及共轭分布将来会出现在贝叶斯学习中，而下一章的极大似然估计方法就更像是 Frequentists 的做法

先验分布, 似然函数, 后验分布

- 参数先验分布为 $p(\theta)$
- 似然函数为 $p(x|\theta)$
- 观测值为 X

贝叶斯的思想是根据观测值来调整参数的先验分布从而得到参数的后验分布. 参数后验分布为

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{\int_{\theta'} p(X|\theta')p(\theta')d\theta'}$$

共轭分布

如果参数的后验分布与先验分布属于同一类分布，那么我们说这种先验分布为共轭分布 (Conjugate prior). 比如

- 似然函数为正态分布时, 如果 σ 已知, 关于 μ 的正太分布是共轭分布
- 似然函数为正态分布时, 如果 μ 已知, 关于 σ 的反 Gamma 分布是共轭分布

具体共轭分布列表可以参考

https://en.wikipedia.org/wiki/Conjugate_prior

共轭分布的好处在于，先验与后验分布属于一个大类，这样计算和理解上都比较方便.

小结 (随机变量与概率)

- 概率可以理解为事件所代表的集合在全概率空间中的比例
- 对于概率分布参数的先验分布有不同的观点
- 如果参数先验分布与后验分布属于同一类，则叫做共轭分布

谢谢大家!