最优化学习笔记附录——B3 概率基础

BD S

2021 年 7 月 14 日

目录

1 概率空间

概率空间由一个三元组 (Ω, F, P) 组成，其中 Ω 是样本空间，F 是事件域，以及在 0-1 之间的概率函数 P。

2 随机变量

为了更好的描述随机现象量化的过程。随机变量 X 是定义在样本空间 Ω 上的实值函数，即 $X : \Omega \to \mathbb{R}$。随机变量满足要求可测性，对于任意的 $t \in \mathbb{R}$，集合 $\{\omega|X(\omega) > t\}$ 总是属于 F。

给定一个样本空间中的概率函数 P，随机变量 X 大于某个值 x 的概率为 $P(\omega|\omega(X) > x)$，简写为 $P(X > x)$，最后的这句话是重点，结合到了所涉及到的知识点。再具体一点，一

个例子。样本空间设为 $\Omega = \{1, 2, \ldots, 8\}$，随机变量的定义为

$$X(\Omega) = \begin{cases} 0, \omega\ is\ even \\ 1, \omega\ is\ odd \end{cases} \tag{1}$$

就是奇数偶数的时候取值不一样。在这个例子中，我们感兴趣的就是使得 X 具有可测性的事件域中最小的那个，即 $F = \{\emptyset, \{1, 3, 5, 7\}, \{2, 4, 6, 8\}, \Omega\}$。个人理解就是根据要求来设定这个随机变量函数 X，然后诱导出对应的事件域 F。

2.1 离散型随机变量和连续型随机变量

首先讲一下分布函数。可以知道，$P(a < X \leq b) = P(x \leq b) - P(x \leq a)$。定义一个随机变量的分布函数为 $F(x) = P(X \leq x), x \in \mathbb{R}$，此时也称 X 服从分布 $F(x)$，记为 $X \sim F(x)$，这是一个单调有界并且右连续的函数。

再讲一下概率密度函数。对于一个连续型变量 X，如果存在函数 $f(x)$，使得

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

有性质

$$\int_{-\infty}^{+\infty} f(t)dt = 1$$

对于连续型随机变量，每个点的概率密度都是 0，$P(X = x) = 0, \forall x \in \mathbb{R}$。对于离散型随机变量，可列个点的概率密度为 $p(x) = P(X = x)$

2.2 数学期望

对于离散型的随机变量 X，设其取值的集合为 $x_1, x_2, \ldots$，分布律为 $P(x)$，若级数 $\sum\limits_{i=1}^{\infty} x_i P_i(x_i)$ 是绝对收敛的，则称其为 X 的数学期望。

2.3 方差

定义

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

同时也是

$$Var(X) = E(X^2) - E(X)^2$$

2.4 联合分布

上面讲的都是单个变量的情况，还有多个变量的情况。比如对于两个随机变量 X，Y，定义联合分布函数为

$$F(a,b) = P(X \leq a, Y \leq b), -\infty < a, b < +\infty$$

通过联合分布也可以得到随机变量 X 的分布函数

$$F_X(a) = P(X \leq a) = P(X \leq a, Y < +\infty) = \lim_{b \to +\infty} F(a,b)$$

对于这种只含部分分量的分布函数称为边缘分布函数。

以上是分布函数，下面是分布律。分布律定义为

$$p(x,y) = P(X = x, Y = y)$$

边缘分布律也可以类似计算

$$P_X(x) = \sum_{y:p(x,y)>0} p(x,y)$$

接下来是联合密度函数

$$P(X \in A, Y \in B) = \int_B \int_A f(x,y)dxdy$$

那么这个 $f(x,y)$ 就说联合密度函数。边缘密度函数也类似，$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)$

一个概念，相互独立。也就是

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$$

$$F(a,b) = F_X(a)F_Y(b)$$

对于离散情况

$$p(x,y) = p_X(x)p_Y(y)$$

对于连续情况

$$f(x,y) = f_X(x)f_Y(y)$$

对于两个变量不独立的情况，引入协方差。定义为

$$Cov(X,Y) = E((X - E(X)(Y - E(Y))) = E(XY) - E(X)E(Y)$$

如果协方差是 0，就说明相互独立。

本章概念，分布函数 F、分布律 p、边缘分布函数、边缘分布律（这四个差不多）联合密度函数 f、边缘密度函数。密度函数积分就是分布律，分布律积分就是分布函数。都可以由概率函数 P 引出。

3 条件期望

假设现在有随机向量 $(X, Y_1, Y_2, \ldots, Y_n)$，并且已经观测到了 $(Y_1, Y_2, \ldots, Y_n)$，那么就要用 $(Y_1, Y_2, \ldots, Y_n)$ 来预测 X。

3.1 L^2 空间

设概率空间 (Ω, F, P)，定义 L^2 空间为

$$L^2(\Omega, F, P) = \{X|E(X^2) < +\infty\}$$

是二阶矩有限的随机变量构成的集合。

考虑 L^2 空间，可以引入内积的概念。定义 X 与 Y 的内积 $\langle X, Y\rangle = E[XY]$ 引入范数和距离

$$||X||_{L^2} = \sqrt{E[X^2]}$$

$$d(X, Y) = \sqrt{\langle E[(X-Y)^2]\rangle}$$

3.2 L^2 空间上随机变量的条件期望