第 0 讲 概率回顾, Markov 链

概率空间

全概率公式

独立性

Markov 链

条件期望

- 关于随机事件的条件期望
- 关于随机变量的条件期望
- 条件期望的性质

参考资料

- 李贤平, 概率论基础.
- R. W. Yeung, Information Theory and Network Coding, 2.1.
- T. M. Cover and J. A. Thomas, Elements of Information Theory, 2.8.

定义 1.1

一个试验每个可能的结果称为 样本点，所有样本点构成的集合 Ω 称为是这个试验的样本空间. 样本空间 Ω 的子集 A 称为是一个 事件. 如果样本点 ω 包含在事件 A 中，则记为 $\omega \in A$，否则记为 $\omega \notin A$. 对于事件 A，如果一次试验的结果 $\omega \in A$，则称事件 A 发生，否则称事件 A 不发生.

定义 1.2

Ω 的子集族 $\mathcal{F}$ 称为一个 σ-域，如果其满足下列条件：

(i) $\emptyset \in \mathcal{F}$;

(ii) 如果 $A_1, A_2, \ldots \in \mathcal{F}$ 则 $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$;

(iii) 如果 $A \in \mathcal{F}$, 则 $A^c \in \mathcal{F}$.

定义 1.3

$(\Omega, \mathcal{F})$ 上的一个 **概率测度** P 是一个函数 $P : \mathcal{F} \to [0,1]$，满足:

(i) $P(\Omega) = 1$;

(ii) 如果 $A_1, A_2, \ldots$ 是 $\mathcal{F}$ 中的一列不交的元素，即对任意 $i \neq j$，$A_i \cap A_j = \emptyset$，那么

$$P(\cup_{j=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

由集合 Ω，Ω 的子集构成的 σ-域 $\mathcal{F}$，以及 $(\Omega, \mathcal{F})$ 上的一个 **概率测度** P 构成的一个三元组 $(\Omega, \mathcal{F}, P)$ 称为是一个**概率空间**. σ-域 $\mathcal{F}$ 称为 **事件域**.

定义 1.4

可测空间 $(\Omega, \mathcal{F})$ 上的函数 $X : \Omega \to \mathbb{R}$ 称为是 (空间 $(\Omega, \mathcal{F})$ 上的) 一个随机变量，如果对任意 $x \in \mathbb{R}$，有

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

这样的函数称为是 $\mathcal{F}$-可测的.

- 设 X 是以 $\mathcal{X}$ 为字母表的一个随机变量.
- 我们用 $\{p_X(x), x \in \mathcal{X}\}$ 表示 X 的分布, 其中 $p_X(x) = \Pr\{X = x\}$.
- 当不会引起混淆时，我们将 $p_X(x)$ 简记为 $p(x)$.

定义 1.5

设 X 是以 $\mathcal{X}$ 为字母表的一个随机变量, 分布为 $\{p(x), x \in \mathcal{X}\}$. X 的**支集 (支撑集, 支撑)** S_X 定义为 $\{x \in \mathcal{X} : p(x) > 0\}$.

如果 $S_X = \mathcal{X}$, 则称 X 是**严格正的**.

定义 2.1

设 A 和 B 是两个事件. 如果 $P(B)>0$，则在给定事件 B 发生的条件下事件 A 的条件概率为

$$P(A|B)=\frac{P(A\cap B)}{P(B)}. \tag{2.1}$$

我们称 $P(A)$ 为 A 的先验概率，而 $P(A|B)$ 为 A 的后验概率.

样本空间 Ω 中的事件族 $A_1, A_2, \ldots, A_n$ 称为是 Ω 的一个分割或分划，如果

$$A_i \cap A_j = \emptyset \ (i \neq j \text{ 时}), \quad \text{且} \quad \cup_{i=1}^n A_i = \Omega.$$

我们也称 $A_1, A_2, \ldots, A_n$ 是一个完备事件组.

定理 2.2: 全概率公式

设 A 和 B 是样本空间 Ω 中的两个事件，其中 $0 < \mathbb{P}(A) < 1$，则

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

一般地，设事件 $A_1, A_2, \ldots, A_n$ 为 Ω 的一个分划，满足对任意 i 有 $P(A_i) > 0$. 则

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

注记 2.3

实际上全概率公式的条件可以减弱为 $B \subset \cup_{i=1}^n A_i$. 实际上我们补充定义 $A_{n+1} = (\cup_{i=1}^n A_i)^c$, 注意到 $P(B|A_{n+1}) = 0$, 从而可得所需结论.

定理 2.4: Bayes 公式

设 A, B 是两个事件，且 $P(A), P(B) > 0$，则

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

利用 Bayes 公式和全概率公式，我们有如下结论.

定理 2.5

设 $A_1, \ldots, A_n$ 为样本空间 Ω 的一个分划, 满足对任意 i 均有 $\mathbb{P}(A_i) > 0$. 则对任意满足 $P(B) > 0$ 的事件 B，有

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}(B|A_j)P(A_j)}.$$

定义 3.1

我们称随机变量 X 和 Y 是独立的, 记做 $X \perp Y$, 如果对任意 $x \in \mathcal{X}, y \in \mathcal{Y}$, 有

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

定义 3.2: 相互独立

设 $n \geq 3$, 随机变量 $X_1, X_2, \cdots, X_n$ 是 **相互独立的**, 如果对于任意 $x_1, x_2, \cdots, x_n$, 有

$$p(x_1, x_2, \cdots, x_n) = p(x_1)p(x_2)\cdots p(x_n).$$

定义 3.3: 两两独立

设 $n \geq 3$, 随机变量 $X_1, X_2, \cdots, X_n$ 是 **两两独立的**, 如果对任意 $1 \leq i < j \leq n$, 有 X_i 和 X_j 相互独立.

定义 3.4: 条件独立

对于随机变量 X, Y 以及 Z, 称 X 和 Z 关于 Y **条件独立**, 记为 $X \perp Z | Y$, 如果对任意 x, y and z,

$$p(x, y, z) p(y) = p(x, y) p(y, z).$$

或等价地,

$$p(x, y, z) = \begin{cases} \dfrac{p(x, y) p(y, z)}{p(y)} = p(x, y) p(z|y) & \text{如果 } p > 0, \\ 0 & \text{其他}. \end{cases} \tag{3.1}$$

命题 3.5

对于随机变量 X, Y, 和 Z, $X \perp Z | Y$ 当且仅当对任意 x, y 以及 z, 其中 $p(y) > 0$, 有

$$p(x, y, z) = a(x, y) b(y, z).$$

证明.
“仅当” 部分直接由 (3.1) 中关于条件独立的定义得出, 所以接下来我们证明 “当 “部分. 假设对于任意满足 $p(y) > 0$ 的 x, y, z 有

$$p(x,y,z) = a(x,y)b(y,z).$$

则对于上述 x, y, z, 我们有

$$p(x,y) = \sum_z p(x,y,z) = \sum_z a(x,y)b(y,z) = a(x,y)\sum_z b(y,z)$$

以及

$$p(y,z) = \sum_x p(x,y,z) = \sum_x a(x,y)b(y,z) = b(y,z)\sum_x a(x,y).$$

证明.
又

$$p(y)=\sum_z p(y,z)=\Big(\sum_x a(x,y)\Big)\Big(\sum_z b(y,z)\Big)>0,$$

我们有

$$\frac{p(x,y)p(y,z)}{p(y)}=\frac{\big(a(x,y)\sum_z b(y,z)\big)\big(b(y,z)\sum_x a(x,y)\big)}{\big(\sum_x a(x,y)\big)\big(\sum_z b(y,z)\big)}=a(x,y)b(y,z)=p(x,y,z).$$

对于满足 $p(y)=0$ 的 x, y, z, 注意到

$$0\le p(x,y,z)\le p(y)=0,$$

我们可知

$$p(x,y,z)=0.$$

综上由 (3.1) 我们有 $X\perp Z|Y$, 从而得证. $\square$

概率空间

全概率公式

独立性

Markov 链

条件期望

- 关于随机事件的条件期望
- 关于随机变量的条件期望
- 条件期望的性质

定义 4.1: Markov 链

设 $X_1, X_2, \cdots, X_n$ 是一列随机变量, 其中 $n \geq 3$, 称 $X_1 \to X_2 \to \cdots \to X_n$, **如果**对任意 $x_1, x_2, \cdots, x_n$, 有

$$p(x_1, \cdots, x_n)p(x_2)p(x_3)\cdots p(x_{n-1}) = p(x_1, x_2)p(x_2, x_3)\cdots p(x_{n-1}, x_n),$$

或等价地,

$$\begin{aligned} & p(x_1, x_2, \cdots, x_n) \\ = & \begin{cases} p(x_1, x_2)p(x_3|x_2)\cdots p(x_n|x_{n-1}) & \text{如果 } p(x_2), p(x_3), \cdots, p(x_{n-1}) > 0 \\ 0 & \text{其他.} \end{cases} \end{aligned}$$

安德烈·安德烈耶维奇·马尔可夫 (Andrey Andreyevich Markov, 1856 年 6 月 14 日—1922 年 7 月 20 日) 是一位俄国数学家. 他在随机过程领域做出重要工作.
马尔可夫于 1856 年 6 月 14 日出生于俄国梁赞市. 1874 年，马尔科夫考入圣彼得堡大学数学系. 1884 年以题为《关于代数连分数的某些应用》的论文通过博士答辩, 博士导师是切比雪夫. 1893 年成为圣彼得堡大学正教授. 1922 年 7 月 20 日在彼得格勒去世.

马尔可夫对数学最大的贡献是在概率论领域做出的. 他沿着切比雪夫开创的方向, 改进和完善了大数定律和中心极限定理. 他的另一项重要工作是分析普希金的作品《叶甫盖尼·奥涅金》中辅音和元音的统计分布, 建立了一个如今被称为马尔可夫链的模型.

命题 4.2

$X_1 \to X_2 \to \cdots X_n$ 构成一个 Markov 链当且仅当 $X_n \to X_{n-1} \to \cdots \to X_1$ 构成一个 Markov 链.

命题 4.3

$X_1 \to X_2 \to \cdots X_n$ 构成一个 Markov 链当且仅当

$$
\begin{aligned}
&X_1 \to X_2 \to X_3 \\
&(X_1, X_2) \to X_3 \to X_4 \\
&\qquad\qquad \vdots \\
&(X_1, X_2, \cdots, X_{n-2}) \to X_{n-1} \to X_n
\end{aligned}
$$

构成 Markov 链.

命题 4.4

$X_1 \to X_2 \to \cdots X_n$ 构成一个 Markov 链当且仅当任意满足 $p(x_2), p(x_3), \cdots, p(x_{n-1}) > 0$ 的 $x_1, x_2, \cdots, x_n$, 有

$$p(x_1, x_2, \cdots, x_n) = f_1(x_1, x_2) f_2(x_2, x_3) \cdots f_{n-1}(x_{n-1}, x_n).$$

命题 4.5: Markov 子链

设 $\mathcal{N}_n = \{1, 2, \cdots, n\}$, $X_1 \to X_2 \to \cdots \to X_n$ 是一个 Markov 链. 对任意 $\alpha \subseteq \mathcal{N}_n$, 我们将 $(X_i, i \in \alpha)$ 简记为 X_α. 对于任意两两不交的 $\alpha_1, \alpha_2, \cdots, \alpha_m \subseteq \mathcal{N}_n$, 其中

$$k_1 < k_2 < \cdots < k_m,$$

我们有

$$X_{\alpha_1} \to X_{\alpha_2} \to \cdots X_{\alpha_m}$$

构成一个 Mrkov 链. 也就是说, $X_1 \to X_2 \to \cdots \to X_n$ 的一个子链也是一个 Markov 链.

命题 4.6

设 X_1, X_2, X_3 以及 X_4 是随机变量, 满足 $p(x_1, x_2, x_3, x_4)$ 是严格正的. 那么我们有

$$\left.\begin{array}{l} X_1 \perp X_4 | (X_2, X_3) \\ X_1 \perp X_3 | (X_2, X_4) \end{array}\right\} \Rightarrow X_1 \perp (X_3, X_4) | X_2.$$

证明.

由 $X_1 \perp X_4|(X_2, X_3)$, 我们有

$$p(x_1, x_2, x_3, x_4) = \frac{p(x_1, x_2, x_3)p(x_2, x_3, x_4)}{p(x_2, x_3)}. \tag{4.1}$$

另一方面, 由 $X_1 \perp X_3|(X_2, X_4)$, 我们可以得到

$$p(x_1, x_2, x_3, x_4) = \frac{p(x_1, x_2, x_4)p(x_2, x_3, x_4)}{p(x_2, x_4)}. \tag{4.2}$$

证明.
对比上面两式，我们有

$$p(x_1, x_2, x_3) = \frac{p(x_2, x_3)p(x_2, x_3, x_4)}{p(x_2, x_4)}.$$

从而

$$p(x_1, x_2) = \frac{p(x_2)p(x_1, x_2, x_4)}{p(x_2, x_4)},$$

也就是说，

$$\frac{p(x_1, x_2, x_4)}{p(x_2, x_4)} = \frac{p(x_1, x_2)}{p(x_2)}.$$

证明.

于是对任意 x_1, x_2, x_3, x_4, 有

$$p(x_1, x_2, x_3, x_4) = \frac{p(x_1, x_2, x_4)p(x_2, x_3, x_4)}{p(x_2, x_4)}$$
$$= \frac{p(x_1, x_2)p(x_2, x_3, x_4)}{p(x_3)},$$

也即 $X_1 \perp (X_3, X_4) | X_2$. $\square$

思考

在上述定理中, 如果存在 x_1, x_2, x_3, x_4 使得 $p(x_1, x_2, x_3, x_4) = 0$, 结论是否依旧成立?

思考

在上述定理中, 如果存在 x_1, x_2, x_3, x_4 使得 $p(x_1, x_2, x_3, x_4) = 0$, 结论是否依旧成立?

例 4.8

设 Y, Z 是两个独立的随机变量. 记 $X_1 = Y$, $X_2 = Z$, $X_3 = X_4 = (Y, Z)$. 则 $X_1 \perp X_4 | (X_2, X_3)$ 且 $X_1 \perp X_3 | (X_2, X_4)$, 但 $X_1 \not\perp (X_3, X_4) | X_2$.

概率空间
全概率公式
独立性
Markov 链
条件期望
关于随机事件的条件期望
关于随机变量的条件期望
条件期望的性质

定义 5.1: 关于随机事件的条件期望

设 A 是一个概率为正的随机事件. 如果 Y 是一个离散型随机变量，那么 Y 在给定事件 A 时的条件期望定义为

$$E[Y|A] = \sum yP(Y = y|A),$$

其中求和是在 Y 的支撑集上进行. 如果 Y 是一个连续型随机变量，具有概率密度函数 f，则

$$E[Y|A] = \int yf(y|A)dy,$$

其中条件概率密度函数 $f(y|A)$ 定义为条件分布函数 $F(y|A) = P(Y \leq y|A)$ 的导数，由 Bayes 公式我们有

$$f(y|A) = \frac{P(A|Y = y)f(y)}{P(A)}.$$

定理 5.2

设 $A_1, \cdots, A_n$ 是样本空间 Ω 的一个划分，且对任意 i 有 $P(A_i) > 0$. 设 Y 是同一样本空间的一个随机变量. 则

$$E[Y] = \sum_{i=1}^{n} E[Y|A_i]P(A_i).$$

设 B 是样本空间 Ω 中的一个事件，取 $Y = I_B$ 应用上面定理我们可以得到全概率公式.

关于随机变量的条件期望

设 X 是一个随机变量，Y 是一个离散型随机变量，我们定义

$$E[Y|X=x]=\sum_{y} yP(Y=y|X=x).$$

类似地，如果 Y 是连续的，我们定义

$$E[Y|X=x]=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy.$$

定义 5.3

设 $g(x) = E[Y|X=x]$. 则给定 X 时 Y 的条件期望，记为 $E[Y|X]$，定义为随机变量 $g(X)$.

设 X 和 Y 是两个随机变量.

命题 5.4

如果 X 和 Y 是独立的，那么

$$E[Y|X] = E[Y].$$

命题 5.5

设 h 是一个 Borel 可测函数，那么

$$E[h(X)Y|X] = h(X)E[Y|X].$$

例 5.6

设 $Z \sim \mathcal{N}(0,1)$，$Y = Z^2$. 求 $E(Y|Z)$.

例 5.7

设 $Z \sim \mathcal{N}(0,1)$, $Y = Z^2$. 求 $E(Y|Z)$.

解.

我们有 $E[Y|Z] = E[Z^2|Z] = Z^2$. $\square$

命题 5.8

设 c 是一个常数，我们有

$$E[Y_1+Y_2|X]=E[Y_1|X]+E[Y_2|X],\ E[cY|X]=cE[X|Y].$$

例 5.9

设 $X_1, \cdots, X_n$ 是独立同分布的，$S_n = X_1 + X_2 + \cdots + X_n$. 求 $E[X_1|S_n]$.

解.
由对称性，

$$E[X_1|S_n] = E[X_2|S_n] = \cdots = E[X_n|S_n],$$

由线性性，我们有

$$E[X_1|S_n] + \cdots + E[X_n|S_n] = E[S_n|S_n] = S_n.$$

于是

$$E[X_1|S_n] = S_n/n = \bar{X}_n. \qquad \square$$

命题 5.10

$$E[E[Y|X]] = E[Y].$$

定理 5.11

设 X, Y 是同一样本空间上的两个随机变量，$E[Y^2] < \infty$. 设 h 是一个 Borel 可测函数，满足 $E[h^2(X)] < \infty$. 那么随机变量 $Y - E[Y|X]$ 与 $h(X)$ 是不相关的. 也就是说，

$$E[(Y - E[Y|X])h(X)] = 0.$$

证明.

由 Cauchy-Schwarz 不等式,

$$E[Yh(X)] \leq \sqrt{E[Y^2]E[h^2(X)]} < \infty.$$

于是我们有

$$\begin{aligned} E[(Y - E[Y|X])h(X)] &= E[h(X)Y] - E[h(X)E[Y|X]] \\ &= E[h(X)Y] - E[E[h(X)Y|X]] = 0, \end{aligned}$$

从而得证. $\square$