随机过程

讲师：吴明燕
笔记由 Dafu Zhu 编写
基于 2025 春季厦大数院《随机过程》

最后修改：2025/03/05

目录

成绩：平时（作业 + 考勤）+ 期中论文 + 期末

概率论准备知识

概率论中，随机变量的本质是可测函数。

$$X : \Omega \to S$$

S 的 σ-代数记为 $\mathcal{S}$，是个 Borel σ-代数（由开集/闭集生成）
Q: 为什么要给 Ω 一个 σ-代数？
A: 样本空间是抽象的，给它 σ-代数赋予它结构，相当于对信息进行重整/提取
概率测度的本质是集函数，

$$\text{集合} \to \text{函数}$$

将信息具象化，

$$\begin{aligned} \mathbb{P} :& \mathcal{F} \to [0, 1] \\ & A \to \mathbb{P}(A) \end{aligned}$$

随机过程：一族随机变量 $\{X_t\}_{t\in\mathbb{T}}$
其中 $\mathbb{T}$ 为指标集，$X_t : \Omega \to S$

Example 1

$\mathbb{T} = \mathbb{N}_0$: 时间离散；$\mathbb{T} = [0, T]$: 时间连续

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S}, \mu_X)$$

思考：什么是随机过程的分布 $\{\mu_t\}_{t\in\mathbb{T}}$？

0.1 事件概率

0.1.1 事件域

Definition 1 (样本空间、事件)

样本点、样本空间、事件和事件的运算：

- 样本点 ω：一次试验的结果
- 样本空间 Ω：全体样本点
- 事件：Ω 的子集
- 事件的运算：集合的运算，即交并补 $(A \cap B, A \cup B, A^c)$

Definition 2

若 $A \cap B = \varnothing$，则称 A 与 B 不相交，更一般地，若 $A_i \cap A_j = \varnothing (i \neq j)$，则称 $\{A_i\}_{i \geq 1}$ 互不相交

Definition 3

称 $\mathcal{F} \subset 2^\Omega = \{A | A \subset \Omega\}$ 是一个 σ 代数/事件域（其中 2^Ω 表示所有 Ω 的子集构成的集合，是一个集类）若

1. $\Omega \subset \mathcal{F}$
2. (对补封闭) $A \in \mathcal{F} \to A^c \in \mathcal{F}$
3. (对可列并封闭) $A_n \in \mathcal{F}, n \geq 1 \Rightarrow \cup_{n \geq 1} A_n = \cup_{n=1}^{\infty} A_n \in \mathcal{F}$

σ 代数是满足以上特定条件的集类，是由 Ω 的子集构成的集合

注：σ 代数对有限交/有限并/可列交封闭

现在给出了一个定义，我们会想“为什么定义会这样给呢”，现在要举一些例子说明“定义有意义”

Example 2

最小的 σ 代数：$\{\varnothing, \Omega\}$

最大的 σ 代数：2^Ω

以上这两个例子一个太小、一个太大，似乎没意义，所以叫它们“平凡的”

Example 3

$A \subset \Omega, \sigma(\{A\}) = \sigma(A) = \{A, A^c, \Omega, \varnothing\} = \sigma(A^c)$

这是由 A 生成的 σ 代数

Definition 4 (划分/分割)

称 $\Pi_\Omega := \{\Lambda_n, n \geq 1\}$ 是 Ω 的一个分划，若 $\Omega = \sum_{n \geq 1} \Lambda_n$

1. $\Omega = \cup_{n \geq 1} \Lambda_n$
2. $\{\Lambda_n\}_{n \geq 1}$ 互不相交

Example 4

$\Omega = \sum_{n\geq 1} \Lambda_n, \Pi_\Omega := \{\Lambda_n\}_{n\geq 1}$

$$\sigma(\Pi_\Omega) = \left\{\sum_{k\in J} \Lambda_k, J \subset \mathbb{N}\right\}$$

Problem 1 (作业 1-1)

证明：

1. $\sigma(\Pi_\Omega)$ 是一个 σ 代数
2. $\sigma(\Pi_\Omega)$ 是包含集类 Π_Ω 的最小 σ 代数

$(S, \mathcal{S}) = (S, 2^S)$: S 可列时，取 2^S 为 σ 代数

$(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$: S 为实数集时，取博雷尔集 $\mathcal{B}(\mathbb{R})$ 为 σ 代数

0.1.2 概率测度

Definition 5 (概率测度)

$(\Omega, \mathcal{F})$ 称 $\mathbb{P} : \mathcal{F} \to [0, 1]$ 是概率测度

1. 非负性
2. 归一性
3. 可列可加性 *

Property 1

$\mathbb{P}$ 满足有限可加性（可列可加一定有限可加，如果既不是可列可加、也不是有限可加，则不可测）

Corollary 1

1. $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$
2. 若 $A \subset B$，则 $\mathbb{B} = \mathbb{A} + \mathbb{P}(BA^c) \geq \mathbb{P}(A)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Remark 1. 引用知乎上三维之外的大白话解释可列可加性：

首先，在我们总是习惯于处理有限相加，而很少遇到无限相加的情况。从测度论内容理解，有限相加与事实（数学的）不符，比如 $(0, 1)$ 区间有不可数个点，每个点的测度（理解为直径吧）是 0，按照习惯想法（有限相加），直径的加和（总宽度）应该为 0，显然，$(0, 1)$ 区间的宽度不可能是 0；

如果规定为“只要是无穷多个点相加，其宽度就不再是 0”的话，还是存在矛盾，我们知道，区间 $(0, 1)$ 上的有理数是是无穷多个的（而且是可列的），那么其宽度就应该为 1，可是无理数还是不可数的呢——理解为无理数是有理数的无穷大量或有理数是无理数的无穷小量，那么无理数的宽度是多少呢？即使还是 1，显然 $(0, 1)$ 区间的宽度不可能是 2 吧!?

于是，勒贝格说道：在测量长度、面积、体积时，我们采用可列可加性，即可列个点相加，规定其宽度（测度）为 0，如果点的个数超过了可列个（这时必是连续统的），那么，就不满足了——即这些点的总宽度就不是 0 了，而是具有了非 0 的宽度（正测度），当然，具有测度的这些点是紧挨在一起的，否则不一定有测度，比如康托大师制造的三分集就很诡异。
到这里，可列可加性事实上讲完了，再啰嗦一下次可列可加性。这是因为不论作为集合，还是概率上的事件（也是集合），一般是存在公共元素的，因此，一般情形下，当然满足次可列可加性的性质了，可列可加性只有在集合之间的距离大于 0 或事件之间完全独立的情形下，才会满足。

Property 2 (次可列可加性)
$A_n \subset \mathcal{F}, n \geq 1$

$$\mathbb{P}(\bigcup_{n\geq 1} A_n) \leqslant \sum_{n\geq 1} \mathbb{P}(A_n)$$

证明：$\cup_{n\geq 1} A_n = \sum_{n\geq 1} B_n$，其中 $B_1 = A_1, B_2 = A_2 \cap (A_1)^c, \cdots, B_n = A_n \cap A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c$
$B_n \subset A_n$，由可列可加性和推论1(2)

Problem 2 (作业 1-2)
证明 $\cup_{n\geq 1} A_n = \sum_{n\geq 1} B_n$

证明：

1. 先证 $\bigcup_{n\geqslant 1} A_n \subseteq \sum_{n\geqslant 1} B_n$。
   假设 $x \in \bigcup_{n\geqslant 1} A_n$，
   若 $x \in A_1$，则 $x \in B_1$，
   若 $x \in A_2$ 且 $x \notin A_1$，则 $x \in B_2$
   $\cdots$
   若 $x \in A_n$ 且 $x \notin A_1, x \notin A_2, \ldots, x \notin A_{n-1}$，则 $x \in B_n$
   $\forall x \in \bigcup_{n\geq 1} A_n$，都有 $x \in \bigcup_{n\geq 1} B_n$
   $\because B_i \cap B_j = \varnothing, i \neq j$，$\therefore \bigcup_{n\geqslant 1} B_n = \sum_{n\geqslant 1} B_n$，$x \in \sum_{n\geqslant 1} B_n$。
2. 再证 $\sum_{n\geqslant 1} B_n \subseteq \bigcup_{n\geqslant 1} A_n$
   假设 $x \in \sum_{n\geqslant 1} B_n$，则 $\exists n_0 \in \mathbb{N}^+$，使得 $x \in B_{n_0}$，
   由 B 的定义

$$B_{n_0} = A_{n_0} \cap \left(\bigcap_{k=1}^{n_0-1} A_k^c \right)$$

   $\therefore x \in A_{n_0} \subseteq \bigcup_{n\geqslant 1} A_n$
   $\therefore \bigcup_{n\geqslant 1} A_n = \sum_{n\geqslant 1} B_n$ □

Property 3 (连续性)
(1) $A_n \uparrow$ 单调上升，即 $A_n \subset A_{n+1}$，$\lim_{n\to\infty} A_n = \cup_{n\geq 1} A_n$，则 $\mathbb{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$
(2) $B_n \downarrow$ 单调下降，即 $B_n \supset B_{n+1}$，$\lim_{n\to\infty} B_n = \cap_{n\geq 1} B_n$，则 $\mathbb{P}(\lim_{n\to\infty} B_n) = \lim_{n\to\infty} \mathbb{P}(B_n)$

证明：(1) $\cup_{n\geq 1} A_n = A_1 + A_2 \setminus A_1 + A_3 \setminus A_2 + \cdots$

$$\begin{aligned}
\mathbb{P}(\bigcup_{n\geq 1} A_n) &= \mathbb{P}(A_1) + \sum_{n\geq 1} \mathbb{P}(A_{n+1} \setminus A_n) \\
&= \mathbb{P}(A_1) + \lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{P}(A_{n+1} \setminus A_n) \\
&= \mathbb{P}(A_1) + \lim_{m\to\infty} \sum_{n=1}^{m} [\mathbb{P}(A_{n+1}) - \mathbb{P}(A_n)] \\
&= \mathbb{P}(A_1) + \lim_{m\to\infty} [\mathbb{P}(A_{m+1}) - \mathbb{P}(A_1)] \\
&= \lim_{m\to\infty} \mathbb{P}(A_{m+1}) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n) \quad \square
\end{aligned}$$

(2) $B_n \downarrow B \Rightarrow \forall n, B_{n+1} \subseteq B_n \Rightarrow \forall B_n^c \subseteq B_{n+1}^c$

$$\begin{aligned}
\mathbb{P}(B) = \mathbb{P}(\cap_{n\geq 1} B_n) &= 1 - \mathbb{P}((\cap_{n\geq 1} B_n)^c) \\
&= 1 - \mathbb{P}(\cup_{n\geq 1} B_n^c) \\
&= 1 - \mathbb{P}(B_1^c \cup (\cup_{n\geq 2} (B_n^c \setminus B_{n-1}^c))) \\
&= 1 - \mathbb{P}(B_1^c) - \sum_{n\geq 2} (\mathbb{P}(B_n^c) - \mathbb{P}(B_{n-1}^c)) \\
&= 1 - \mathbb{P}(B_1^c) - \lim_{m\to\infty} \sum_{n=2}^{m} (\mathbb{P}(B_n^c) - \mathbb{P}(B_{n-1}^c)) \\
&= 1 - \mathbb{P}(B_1^c) - \lim_{m\to\infty} (\mathbb{P}(B_m^c) - \mathbb{P}(B_1^c)) \\
&= 1 - \mathbb{P}(B_1^c) - \lim_{n\to\infty} \mathbb{P}(B_n^c) + \mathbb{P}(B_1^c) \\
&= 1 - \lim_{n\to\infty} \mathbb{P}(B_n^c) \\
&= \lim_{n\to\infty} \mathbb{P}(B_n) \quad \square
\end{aligned}$$

第二个等式用到 De Morgan's Law

0.2 独立性

Definition 6 (事件间的独立性)

$(\Omega,\mathcal{F},\mathbb{P}), A,B\in\mathcal{F}$，称 A 与 B 独立，若 $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$，记为 $A\perp\!\!\!\perp B$

Definition 7 (事件间的相互独立)

$\{A_n\}_{n\geqslant 1}\subset\mathcal{F}$，称其相互独立，若 $\forall J\subset\mathbb{N}, \#J\geq 2$

$$\mathbb{P}(\bigcap_{k\in J}A_k)=\prod_{k\in J}\mathbb{P}(A_k)$$

Property 4

$A\perp\!\!\!\perp B\Rightarrow A\perp\!\!\!\perp B^c, A^c\perp\!\!\!\perp B, A^c\perp\!\!\!\perp B^c$

Definition 8 (σ 代数间的独立性)

$(\Omega,\mathcal{F}_1,\mathbb{P}),(\Omega,\mathcal{F}_2,\mathbb{P})$ 称 $\mathcal{F}_1$ 与 $\mathcal{F}_2$ 独立，若 $\forall A_1\in\mathcal{F}_1, A_2\in\mathcal{F}_2$，有 $A_1\perp\!\!\!\perp A_2$，记为 $\mathcal{F}_1\perp\!\!\!\perp\mathcal{F}_2$

Definition 9 (σ 代数间相互独立)

$(\Omega,\mathcal{F}_k,\mathbb{P})(k\geq 1)$ 称 $\{\mathcal{F}_k\}_{k\geqslant 1}$ 相互独立，若 $\forall J\subset\mathbb{N},\#J\geq 2,\forall A_k\in\mathcal{F}_k(k\in J)$，有

$$\mathbb{P}(\bigcap_{k\in J}A_k)=\prod_{k\in J}P(A_k)$$

Property 5

$\{\mathcal{F}_k\}_{k\geqslant 1}$ 相互独立 $\Leftrightarrow\forall A_k\in\mathcal{F}_k,\mathbb{P}(\cap_{k\geq 1}A_k)=\prod_{k=1}^{\infty}\mathbb{P}(A_k)$

证明：$\Rightarrow$ 显然，J 取 $\mathbb{N}$ 即可，$\mathbb{N}\subset\mathbb{N}$

$\Leftarrow$ 注意到右侧 $\forall A_k\in\mathcal{F}$ 对于左侧条件 $\forall A_k\in\mathcal{F}(k\in J)$ 更加一般，所以证 $\Leftarrow$ 的过程也是从一般到特殊。从 $\cap_{k\geq 1}A_k\to\cap_{k\in J}A_k$ 即从 $k\in\mathbb{N}\to k\in J$。思路是把 $k\in\mathbb{N}$ 分成 $k\in J$ 和 $k\in J^c$，在 $k\in J^c$ 上取 $A_k=\Omega$，再利用性质 $\Omega\perp\!\!\!\perp A$。

对于 $\forall J\subseteq\mathbb{N}$

$$\bigcap_{k\geq 1}A_k=\left(\bigcap_{k\in J}A_k\right)\cap\left(\bigcap_{k\in J^c}\Omega\right)$$

$$\begin{aligned}\mathbb{P}(\bigcap_{k\geq 1}A_k)&=\mathbb{P}\left((\bigcap_{k\in J}A_k)\cap(\bigcap_{k\in J^c}\Omega)\right)\\&=\mathbb{P}(\bigcap_{k\in J}A_k)\mathbb{P}(\bigcap_{k\in J^c}\Omega)\qquad[\Omega\perp\!\!\!\perp A_k]\\&=\mathbb{P}(\bigcap_{k\in J}A_k)\end{aligned}$$

$$\prod_{k\geq 1}\mathbb{P}(A_k)=\prod_{k\in J}\mathbb{P}(A_k)\cdot\prod_{k\in J^c}\mathbb{P}(\Omega)=\Pi_{k\in J}\mathbb{P}(A_k)$$

又因为 $\mathbb{P}(\cap_{k\geq 1}A_k)=\prod_{k=1}^{\infty}\mathbb{P}(A_k)$

$$\mathbb{P}(\bigcap_{k\in J}A_k)=\prod_{k\in J}\mathbb{P}(A_k)\quad\square$$

Definition 10 (离散随机变量)

令取值空间 $S=\{x_k\}_{k\geqslant 1}$ (x_k 互不相同)，$\Omega=\sum_{k\geq 1}\Lambda_k$ (划分)，则称

$$X(\omega)=\sum_{k\geq 1}x_k\mathbb{I}_{\Lambda_k}(\omega),\omega\in\Omega$$

为离散随机变量。其中

$$\mathbb{I}_{\Lambda_k}(\omega)=\begin{cases}1 & \text{if } \omega\in\Lambda_k\\ 0 & \text{if } \omega\notin\Lambda_k\end{cases}$$

这个定义的核心思想是：

- 对于每个样本点 $\omega\in\Omega$，$X(\omega)$ 的取值是 x_k，当且仅当 $\omega\in\Lambda_k$
- 因此，X 的取值由样本点 ω 所在的划分 Λ_k 决定

由于随机变量是个可测函数

$$X:(\Omega,?)\to(S,2^S)$$

那么 X 生成的 σ 代数表示为 $\sigma(X):=X^{-1}(2^S)=\{X^{-1}(A)|A\in 2^S\}$

Property 6

$\sigma(X):=X^{-1}(2^S)$，则

1. $\sigma(X)=\sigma(\Pi_\Omega)$ 故称 $\sigma(X)$ 为由 X 生成的 σ 代数。其中 $\Pi_\Omega=\{\Lambda_k,k\geq 1\},\Lambda_k=\{X=x_k\}$
2. $X:(\Omega,\sigma(X))\to(S,2^S)$. 这个记号的解释是 $\forall A\in 2^S,X^{-1}(A)=\{\omega\in\Omega|X(\omega)\in A\}\in\sigma(X)$

证明：要证 $\sigma(X)=\sigma(\Pi_\Omega)$，即证两个集合互相包含

$\sigma(\Pi_X)=\{\sum_{k\in J}\Lambda_k|J\subseteq\mathbb{N}\}$ 由划分生成，$\sigma(X)=X^{-1}(2^S)$ 由 X 生成

下证 $\sigma(X)\subseteq\sigma(\Pi_X)$

$$
\begin{aligned}
\forall A \in 2^S, X^{-1}(A) &= \{\omega | X(\omega) \in A\} \\
&= \sum_{x_k \in A} \{\omega \in \Omega | X(\omega) = x_k\} \\
&= \sum_{x_k \in A} \{X = x_k\} \\
&= \sum_{x_k \in A} \Lambda_k \in \sigma(\Pi_X)
\end{aligned}
$$

第二个等式用到离散 r.v. 定义10

下证 $\sigma(\Pi_X) \subseteq \sigma(X)$

$$
\begin{aligned}
J \subseteq \mathbb{N}, \quad \sum_{k \in J} \Lambda_k &= \sum_{k \in J} \{\omega | X(\omega) = x_k\} \\
&= \{\omega | X(\omega) \in \{x_k, k \in J\}\} \\
&= X^{-1}(\{x_k, k \in J\}) \in \sigma(X)
\end{aligned}
$$

最后一个等式中 $\{x_k, k \in J\} \in 2^S$ □

Example 5

$X = \mathbb{I}_A$ 由划分的定义 $\Pi_X = \{\Lambda_k\}_{k \geqslant 1}, \Lambda_k = \{X = x_k\}$，知道划分将全集分成两部分

$$
\begin{aligned}
\Pi_X &= \{\{X = 1\}, \{X = 0\}\} \\
&= \{\{\omega \in \Omega | X(\omega) = 1\}, \{\omega \in \Omega | X(\omega) = 0\}\} \\
&= \{A, A^c\}
\end{aligned}
$$

$\sigma(\Pi_A) = \{\varnothing, A, A^c, \Omega\} = \sigma(A) = \sigma(A^c)$

其中 $\sigma(\Pi_A)$ 由划分生成，$\sigma(A)$ 由 A 生成，两者相等

另外，$\sigma(X) = \sigma(\mathbb{I}_A) = \sigma(\Pi_X) = \{\varnothing, A, A^c, \Omega\} = \sigma(A) \Rightarrow \sigma(\mathbb{I}_A) = \sigma(A)$

Definition 11 (离散随机变量间的独立性)

$X : \Omega \to S_1, Y : \Omega \to S_2$ 为两离散随机变量，称 $X \perp\!\!\!\perp Y$，若 $\sigma(X) \perp\!\!\!\perp \sigma(Y)$[定义8]，即 $X^{-1}(2^{S_1}) \perp\!\!\!\perp X^{-1}(2^{S_2})$

即 $\forall E_1 \subseteq S_1, E_2 \subseteq S_2$，有 $\mathbb{P}(X \in E_1, Y \in E_2) = \mathbb{P}(X \in E_1)\mathbb{P}(Y \in E_2)$

S_1, S_2 分别为 X, Y 的取值空间，$E_1 \subseteq S_1$ 为 X 的一个取值，$X \in E_1 := \{\omega \in \Omega | X(\omega) \in E_1\}$，$E_2$ 同理

Theorem 1

$X \perp\!\!\!\perp Y \Leftrightarrow \forall x \in S_X, y \in S_Y$ 有$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$

证明：$\Rightarrow$ 一般到特殊，取 $E_1 = \{x\}, E_2 = \{y\}$，由 $\{x\} \in S_X, \{y\} \in S_Y$ 易证

$\Leftarrow$

$$
\begin{aligned}
\mathbb{P}(X \in E_1, Y \in E_2) &= \mathbb{P}(\bigcup_{x \in E_1} \{X = x\} \cap \{Y \in E_2\}) \\
&= \sum_{x \in E_1} \mathbb{P}(\{X = x\} \cap \sum_{y \in E_2} \{Y = y\}) \\
&= \sum_{x \in E_1} \sum_{y \in E_2} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in E_1} (\sum_{y \in E_2} \mathbb{P}(X = x)\mathbb{P}(Y = y)) \\
&= \sum_{x \in E_1} \mathbb{P}(X = x)\mathbb{P}(Y \in E_2) \\
&= \mathbb{P}(X \in E_1)\mathbb{P}(Y \in E_2)
\end{aligned}
$$

第一个等式中，$\{X = x\} \cap \{Y \in E_2\}$ 看作一整个集合 $\subseteq \{X = x\}$，因为离散、每个 x 不相交，所以这是个不交并，由练习2，可以改写成加法形式。
第四个等式由条件 $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ 成立。 □

Theorem 2

$X \perp\!\!\!\perp Y \Leftrightarrow \forall x \in S_X, y \in S_Y, \mathbb{P}(X \leqslant x, Y \leqslant y) = \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y)$

用定理1证明

$\Rightarrow$ 已知 $X \perp\!\!\!\perp Y$，由定义11，$\forall E_1 \subseteq S_1, E_2 \subseteq S_2$，有 $\mathbb{P}(X \in E_1, Y \in E_2) = \mathbb{P}(X \in E_1)\mathbb{P}(Y \in E_2)$。取 $E_1 = \{\omega | X(\omega) \leqslant x\}, E_2 = \{\omega | Y(\omega) \leqslant y\}$

$\Leftarrow$

$$
\begin{aligned}
\mathbb{P}(X = x, Y = y) &= \mathbb{P}(X \leqslant x, Y \leqslant y) - \mathbb{P}(X \leqslant x^-, Y \leqslant y) - \mathbb{P}(X \leqslant x, Y \leqslant y^-) + \mathbb{P}(X \leqslant x^-, Y \leqslant y^-) \\
&= \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y) - \mathbb{P}(X \leqslant x^-)\mathbb{P}(Y \leqslant y) - \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y^-) + \mathbb{P}(X \leqslant x^-)\mathbb{P}(Y \leqslant y^-) \\
&= [\mathbb{P}(X \leqslant x) - \mathbb{P}(X \leqslant x^-)][\mathbb{P}(Y \leqslant y) - \mathbb{P}(Y \leqslant y^-)] \\
&= \mathbb{P}(X = x)\mathbb{P}(Y = y)
\end{aligned}
$$

其中 x^-, y^- 为小于 x, y 的最大值，由于离散，$\{X \leqslant x\} - \{X \leqslant x^-\} = \{X = x\}, \{Y \leqslant y\} - \{Y \leqslant y^-\} = \{Y = y\}$

Definition 12

称一列离散随机变量 $\{X_n\}_{n \geqslant 1}$ 相互独立，若 $\sigma(X_n), n \geq 1$ 相互独立

Theorem 3

$\{A_n\}_{n \geqslant 1}$ 事件列下列等价

1. $\{A_n\}_{n \geqslant 1}$ 相互独立
2. $\sigma(A_n), n \geq 1$ 相互独立
3. $\mathbb{I}_{A_n}, n \geq 1$ 相互独立

证明：

1. 由例题5，$\sigma(\mathbb{I}_{A_n})=\sigma(A_n)$，所以 (2) $\Leftrightarrow$ (3)
2. 下证 (2) $\to$ (1)，一般到特殊，$A_n\subseteq\sigma(A_n)$
3. 下证 (1) $\to$ (2)，$\sigma(A_n)=\{A_n,A_n^c,\varnothing,\Omega\}$，$\varnothing\perp\!\!\!\perp A_n,\Omega\perp\!\!\!\perp A_n$，由性质4，$\varnothing\perp\!\!\!\perp A_n^c,\Omega\perp\!\!\!\perp A_n^c$

由定理5，$\forall A_k\in\sigma(A_n),\mathbb{P}(\cap_{k\geq1}A_k)=\prod_{k=1}^{\infty}\mathbb{P}(A_k)$

由于条件 (1)，上面等式成立 $\Rightarrow$ 满足 σ 代数相互独立的定义 □

0.3 条件概率与条件独立

Definition 13 (条件概率)

$B\in\mathcal{F},\mathbb{P}(B)>0$ 定义

$$\mathbb{P}(A|B)=\frac{\mathbb{P}(AB)}{\mathbb{P}(B)}=:\mathbb{P}_B(A)\quad\forall A\in\mathcal{F}$$

Theorem 4 (乘法公式)

$\mathbb{P}(AB)=\mathbb{P}(A|B)\mathbb{P}(B)$,

$$\mathbb{P}(\bigcap_{k=1}^{n}A_k)=\mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)\cdots\mathbb{P}(A_n|\bigcap_{k=1}^{n-1}A_k)$$

Theorem 5 (全概公式)

(1) $\Omega=\sum_{k\geq1}\Lambda_k$ 划分 $\mathbb{P}(\Lambda_k)>0$，则 $\forall A\in\mathcal{F}$,

$$\mathbb{P}(A)=\sum_{k\geq1}\mathbb{P}(A|\Lambda_k)\mathbb{P}(\Lambda_k)$$

(2) * 一般地，$\{B_n\}_{n\geqslant1}$ 互不相交，$\mathbb{P}(B)>0,\mathbb{P}(\sum_{n\geq1}B_n)=1$，则 $\forall A\in\mathcal{F}$

$$\mathbb{P}(A)=\sum_{n\geq1}\mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

注：$\mathbb{P}(\cdot)=1$ 不一定是全集，但概率测度是 1。同样，$\mathbb{P}(\cdot)=0$ 不一定是 $\varnothing$，而是叫零测集

证明：

(1) 由 $A=A\cap\Omega=A\cap(\sum_{k\geq1}\Lambda_k)=\sum_{k\geq1}(A\cap\Lambda_k)$，$A$ 被划分成若干不相交的集合 $A\cap\Lambda_k$，根据可列可加性，得到

$$\mathbb{P}(A)=\sum_{k\geq1}\mathbb{P}(A\cap\Lambda_k)=\sum_{k\geq1}\mathbb{P}(A|\Lambda_k)\mathbb{P}(\Lambda_k)$$

(2) $\Omega=(\sum_{n\geq1}B_n)+(\sum_{n\geq1}B_n)^c=\sum_{n\geq0}B_n$，其中 $B_0=(\sum_{n\geq1}B_n)^c$

$\mathbb{P}(B_0)=1-\mathbb{P}(\sum_{n\geq1}B_n)=0\to0\leqslant\mathbb{P}(AB_0)\leqslant\mathbb{P}(B_0)=0$

左边不等号成立是因为概率测度非负，右边不等号成立是因为 $AB_0\subseteq B_0$，所以 $\mathbb{P}(AB_0)=0$

$$\begin{aligned}\mathbb{P}(A) &= \sum_{n\geq 0}\mathbb{P}(AB_n) \quad [\text{可列可加性}] \\ &= \sum_{n\geq 1}\mathbb{P}(AB_n) \quad [\mathbb{P}(AB_0)=0] \\ &= \sum_{n\geq 1}\mathbb{P}(A|B_n)\mathbb{P}(B_n) \quad [\text{全概公式}] \quad \square\end{aligned}$$

Theorem 6

$\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$

$$A \perp\!\!\!\perp B \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$$

$\mathbb{P}(A|B)$ 见定义13

Theorem 7

$\mathbb{P}_B : \mathcal{F} \to [0,1]$ 也是 $(\Omega, \mathcal{F})$ 上的概率测度 [定义5]

Property 7

$\mathbb{P}(C) > 0, \mathbb{P}(B) > 0$，则

$$\mathbb{P}_B(\cdot|C) = \mathbb{P}(\cdot|BC) = \mathbb{P}_{BC}(\cdot)$$

$\mathbb{P}_B(\cdot|C)$ 见定义13

Definition 14

称 C 条件发生下，A 与 B 独立，若

$$\mathbb{P}_C(AB) = \mathbb{P}_C(A)\mathbb{P}_C(B)$$

记为 $A \perp\!\!\!\perp_C B$（条件独立）

Theorem 8

$\mathbb{P}(C) > 0, \mathbb{P}(BC) > 0$ 则 $A \perp\!\!\!\perp_C B \Leftrightarrow \mathbb{P}_C(A|B) = \mathbb{P}_C(A)$

证明：由 $A \perp\!\!\!\perp_C B$，$\mathbb{P}_C(AB) = \mathbb{P}_C(A)\mathbb{P}_C(B)$

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

0.4 期望与条件期望

0.4.1 离散随机变量的期望

Definition 15 (X 的期望)

$X:\Omega\to S$

$$\mathbb{E}(X)=\sum_{x\in S}x\mathbb{P}(X=x)=\mathbb{E}^{\mathbb{P}}(X)$$

当此求和绝对收敛

注：$\mathbb{E}^{\mathbb{P}}(X)$ 强调这是在概率测度 $\mathbb{P}$ 下的期望

Definition 16 ($g(X)$ 的期望)

$g:\mathbb{R}\to\mathbb{R}$

$$\mathbb{E}g(X)=\sum_{x\in S}g(x)\mathbb{P}(X=x)$$

当此求和绝对收敛

关于“求和绝对收敛”的讨论：

Example 6

$\mathbb{E}(\mathbb{I}_A)=\mathbb{P}(A),A\in\mathcal{F}$

Example 7

X 是离散随机变量，由定义10，$X=\sum_{x\in S}x\mathbb{I}_{A_x}$，其中 $A_x:=\{X=x\}$。B 是任意的，求 $\mathbb{E}(\mathbb{I}_BX)$

Remark 2. 对于 $A_x:=\{X=x\}$ 应这样理解，A_x 是样本空间 Ω 的一个子集，包含了所有使得 $X(\omega)=x$ 的样本点 ω。

根据离散随机变量的定义，$X(\omega)=x_k$ 当且仅当 $\omega\in\Lambda_k$。因此对于每个 $x_k\in S$，有

$$A_{x_k}=\{X=x_k\}=\{\omega\in\Omega|X(\omega)=x_k\}=\Lambda_k$$

所以 $A_x=\{A_{x_k}\}_{k\geq1}$ 就是离散随机变量的划分

对于 $X=\sum_{x\in S}x\mathbb{I}_{A_x}$ 可以这样理解。对于每个 $x\in S$，$\mathbb{I}_{A_x}(\omega)$ 是事件 $A_x=\{X=x\}$ 的指示函数

$$\mathbb{I}_{A_x}(\omega)=\begin{cases}1 & \textit{if } X(\omega)=x\\ 0 & \textit{if } X(\omega)\neq x\end{cases}$$

Solution. 要先求 $\mathbb{E}(|\mathbb{I}_BX|)<\infty$ 说明期望存在

对 $\forall\omega\in B$

$$\begin{aligned}\mathbb{I}_BX(\omega)&=\mathbb{I}_B(\omega)\sum_{x\in S}(x\cdot\mathbb{I}_{A_x}(\omega))\\&=\sum_{x\in S}x\mathbb{I}_{A_x\cap B}(\omega)\end{aligned}$$

其中 $\mathbb{I}_{A_x\cap B}$ 也可记为 $\mathbb{I}_{A_xB}$
$\{A_xB, x\in S\}\cup\{B^c\}$ 构成了样本空间 Ω 的一个划分。因为 A_x 本身是对 Ω 的一个划分，其与 B 的交是对 B 的划分。并上 B^c，则满足划分的定义*4*
对于 $\omega\in\Omega$，由划分

$$\mathbb{I}_BX(\omega)=0\cdot\mathbb{I}_{B^c}(\omega)+\sum_{x\in S}x\mathbb{I}_{A_x\cap B}$$

$$\therefore\mathbb{E}|\mathbb{I}_BX|=\sum_{x\in S}|x|\mathbb{P}(A_xB)\leqslant\sum_{x\in S}|x|\mathbb{P}(A_x)=\mathbb{E}|x|<\infty$$

最后一个等号参考期望的定义*15*

$$\mathbb{E}(\mathbb{I}_BX)=\sum_{x\in S}x\mathbb{P}(A_xB)=\sum_{x\in S}x\mathbb{P}(\{X=x\}\cap B)$$

Theorem 9
$\mathbb{E}(aX+bY)=a\mathbb{E}X+b\mathbb{E}Y$

离散随机变量有两种表达形式，如定义10和练习7所示

$$X=\sum_{x\in S}x\mathbb{I}_{\{X=x\}}=\sum_{k\geq1}x_k\mathbb{I}_{\Lambda_k}$$

$$\sum_{x\in S}x\mathbb{P}(X=x)=\sum_{k\geq1}\mathbb{P}(X=x_k)$$

只有在“求和绝对收敛”（见定义15）的条件下，等式才成立

Remark 3.

1. $\sum_{x\in S}$ *(1)* 级数的重排 *(2)* 可和族
2. X 是离散随机变量，$g:\mathbb{R}\to\mathbb{R}$，则

$$g(X)=\sum_{x\in S}g(x)\mathbb{I}_{X=x}$$

   是一个离散随机变量，且 $\sigma(g(X))\subseteq\sigma(X)$。下面说明这个结论
   当 $x_1\neq x_2$ 时可能 $g(x_1)=g(x_2)$，因此

$$\Pi_X=\{\{X=x\}|x\in S\}\neq\Pi_{g(X)}$$

   其实 $\Pi_{g(X)}\subseteq\sigma(\Pi_X)$，因为对于 $x_1\neq x_2$ 但 $g(x_1)=g(x_2)$ 的情况，比如在 Π_X 上 x_1,x_2 对应的样本空间是 Ω_1,Ω_2，但在 $\Pi_{g(X)}$ 上是 $\Omega_1\cup\Omega_2$。这一项在 Π_X 里有，因为 σ 代数对可列并封闭。但 Ω_1,Ω_2 分别在 $\Pi_{g(X)}$ 上没有。把 σ 代数理解成信息，则 $g(X)=y$ 提供的信息是比直接提供 x 的值要少的（在 $g(\cdot)$ 已知的情况下）。
3. $X\perp\!\!\!\perp Y,\quad g,h:\mathbb{R}\to\mathbb{R}$，则 $g(X)\perp\!\!\!\perp h(Y)$。因为 $\sigma(X)\perp\!\!\!\perp\sigma(Y)$，而 $\sigma(g(X))\subseteq\sigma(X),\sigma(h(Y))\subseteq\sigma(Y)$
   如果 X,Y 是连续随机变量，则对 g,h 有其他要求。特殊地，结论 *3* 对 g,h 连续时成立。

Theorem 10

(1) $X \perp\!\!\!\perp Y, \mathbb{E}|X| < \infty, \mathbb{E}|Y| < \infty$，则 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

(2) $X_1, X_2, \cdots, X_n$ 相互独立，则 $\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}X_1 \cdots \mathbb{E}X_n$

(3) $X \perp\!\!\!\perp Y, g, h : \mathbb{R} \to \mathbb{R}, \mathbb{E}|g(X)| < \infty, \mathbb{E}|h(Y)| < \infty$

$$\Rightarrow g(X) \perp\!\!\!\perp h(Y), \mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$$

Theorem 11

若 $X \geq 0$ 取整数值，则 $\mathbb{E}(X) = \sum_{k \geq 1} \mathbb{P}(X \geq k)$

证明：

0.4.2 条件期望

1° 关于“给定集合”的条件期望

Definition 17

$(\Omega, \mathcal{F}, \mathbb{P}), X : \Omega \to S, A \in \mathcal{F}, \mathbb{P}(A) > 0, \mathbb{E}|X| < \infty$，定义 X 关于 A 的条件期望

$$\begin{aligned} \mathbb{E}(X|A) &:= \sum_{x \in S} \mathbb{P}(X = x|A) \\ &= \sum_{x \in S} x\mathbb{P}_A(X = x) \\ &= E^{\mathbb{P}_A}(X) \end{aligned}$$

Property 8 (线性性)

$\mathbb{E}(aX + bY|A) = a\mathbb{E}(X|A) + b\mathbb{E}(Y|A)$

证明：（用期望的性质）

Example 8

$\mathbb{E}(\mathbb{I}_B|A) = 1 \cdot \mathbb{P}(B|A) + 0 \cdot \mathbb{P}(B^c|A) = \mathbb{P}(B|A)$

Example 9

$B \perp\!\!\!\perp A \Rightarrow \mathbb{E}(\mathbb{I}_B|A) = \mathbb{E}(\mathbb{I}_B)$

Property 9

$\mathbb{E}|X| < \infty, \mathbb{P}(A) > 0, X \perp\!\!\!\perp \mathbb{I}_A \Rightarrow \mathbb{E}(X|A) = \mathbb{E}(X)$

证明：

$\because X \perp\!\!\!\perp \mathbb{I}_A, \therefore \{X=x\} \perp\!\!\!\perp A$

$$\mathbb{E}(X|A) = \sum_{x\in S} x\mathbb{P}(X=x|A) = \sum_{x\in S} x\mathbb{P}(X=x) = \mathbb{E}(X)$$

其中

$$\sum_{x\in S} x\mathbb{P}(X=x|A) = \sum_{x\in S} x\frac{\mathbb{P}(\{X=x\}\cap A)}{P(A)} = \mathbb{E}(X\mathbb{I}_A)/\mathbb{P}(A)$$

最后一个等号由例题7

至此没有用到独立性，可以得到以下推论

Corollary 2

$\mathbb{E}(X|A) = \mathbb{E}(X\mathbb{I}_A)/\mathbb{P}(A)$

Problem 3 (作业 2-1)

Y 在 A 上取常数 c，证明：$\mathbb{E}(XY|A) = c\mathbb{E}(X|A)$

2° 关于“给定划分生成的 σ 代数”的条件期望

Definition 18

设 $\Pi = \{\Lambda_k\}_{k\geqslant 1}$ 是 Ω 的划分，X 为离散随机变量，$\mathbb{E}|X| < \infty$，定义

$$\mathbb{E}(X|\sigma(\Pi))(\omega) := \mathbb{E}(X|\Lambda_k)$$

当 $\omega \in \Lambda_k$，即

$$\mathbb{E}(X|\sigma(\Pi)) = \sum_{k\geq 1} \mathbb{I}_{\Lambda_k}\mathbb{E}(X|\Lambda_k)$$

期望的本质是积分，现在因为数分里的积分不够用了，我们要定义新积分，希望它也能保留原先的好性质

Property 10 (线性性)

$\mathbb{E}(aX+bY|\sigma(\Pi)) = a\mathbb{E}(X|\sigma(\Pi)) + b\mathbb{E}(Y|\sigma(\Pi))$

证明：$\omega \in \Lambda_k$, $LHS = \mathbb{E}(aX+bY|\Lambda_k) = a\mathbb{E}(X|\Lambda_k) + b\mathbb{E}(Y|\Lambda_k)$

第二个等号由性质8成立。

Example 10

$$\begin{aligned}
\mathbb{E}(X|\{\varnothing,\Omega\}) &= \mathbb{E}(X|\sigma(\Omega)) \\
&= \mathbb{I}_\Omega \mathbb{E}(X|\Omega) \qquad [\text{定义}(18)] \\
&= \sum_{x\in S} x\mathbb{P}(X=x|\Omega) \qquad [\text{定义}(17), \Omega \perp\!\!\!\perp X] \\
&= \sum_{x\in S} x\mathbb{P}(X=x) \\
&= \mathbb{E}(X)
\end{aligned}$$

独立可以理解为：什么信息也没提供

Example 11

$$\begin{aligned}
\mathbb{E}(\mathbb{I}_B|\sigma(A)) &= \mathbb{E}(\mathbb{I}_B|\{A, A^c, \Omega, \varnothing\}) \\
&= \mathbb{E}(\mathbb{I}_B|\sigma(A, A^c)) \\
&= \mathbb{I}_A\mathbb{E}(\mathbb{I}_B|A) + \mathbb{I}_{A^c}\mathbb{E}(\mathbb{I}_B|A^c)
\end{aligned}$$

更进一步，若 $A \perp\!\!\!\perp B$，由 $\sigma(B) \perp\!\!\!\perp \sigma(A) \to \sigma(\mathbb{I}_B) \perp\!\!\!\perp \sigma(\mathbb{I}_A) \Rightarrow \mathbb{E}(\mathbb{I}_B|\sigma(A)) = \mathbb{E}(\mathbb{I}_B)$

可以把这个结果推广：

Property 11

$\sigma(X) \perp\!\!\!\perp \sigma(\Pi)$，则 $\mathbb{E}(X|\sigma(\Pi)) = \mathbb{E}(X)$

证明：$\Pi_X = \{\{X=x\}|x\in S\}$，默认 x 不相同

$\sigma(X) = \sigma(\Pi_X) = \{\{X=x\}|x\in S\}$

不妨设 $\Pi = \{\Lambda_k, k\geq 1\}$

则 $\sigma(X) \perp\!\!\!\perp \sigma(\Pi) \Rightarrow \forall x\in S, k\geq 1, \{X=x\} \perp\!\!\!\perp \Lambda_k$

$$\begin{aligned}
\mathbb{E}(X|\sigma(\Pi)) &= \sum_{k\geq 1} \mathbb{I}_{\Lambda_k}\mathbb{E}(X|\Lambda_k) \\
&= \sum_{k\geq 1} \mathbb{I}_{\Lambda_k} \sum_{x\in S} x\mathbb{P}(X=x|\Lambda_k) \\
&= \sum_{k\geq 1} \mathbb{I}_{\Lambda_k} \sum_{x\in S} x\mathbb{P}(X=x) \\
&= \sum_{k\geq 1} \mathbb{I}_{\Lambda_k}\mathbb{E}(X) \\
&= \mathbb{I}_\Omega\mathbb{E}(X) \\
&= \mathbb{E}(X)
\end{aligned}$$

Example 12

$\mathbb{E}(X|\sigma(X)) = X$

$\sigma(X)$ 作为条件相当于知道了与 X 相关的所有信息，即提取已知量

证明：$\sigma(X) = \sigma(\Pi_X)$，其中 $\mathbb{I}_X = \{\{X = x\}|x \in S\}$

$$\begin{aligned}
\mathbb{E}(X|\sigma(X)) &= \sum_{x\in S} \mathbb{I}_{\{X=x\}}\mathbb{E}(X|X=x) \\
&= \sum_{x\in S} \mathbb{I}_{\{X=x\}}\mathbb{E}(X\mathbb{I}_{\{X=x\}})/\mathbb{P}(X=x) \quad [\text{推论(2)}] \\
&= \sum_{x\in S} \mathbb{I}_{\{X=x\}} \cdot \frac{x\cdot \mathbb{P}(X=x) + 0\cdot \mathbb{P}(X\neq x)}{\mathbb{P}(X=x)} \\
&= \sum_{x\in S} x\mathbb{I}_{\{X=x\}} = X \quad \square
\end{aligned}$$

Property 12 (提取已知量)

设 $\Pi = \{\Lambda_k, k \geq 1\}$ 为 Ω 的划分，$\mathbb{E}|X| < \infty, \mathbb{E}|XY| < \infty$，则当 $\sigma(X) \subseteq \sigma(\Pi)$ 时，有

1. $\mathbb{E}(X|\sigma(\Pi)) = X$
2. $\mathbb{E}(XY|\sigma(\Pi)) = X\mathbb{E}(Y|\sigma(\Pi))$

特别地，取 $X = \mathbb{I}_A, A \in \sigma(\Pi)$，则

1. $\mathbb{E}(\mathbb{I}_A|\sigma(\Pi)) = \mathbb{I}_A$
2. $\mathbb{E}(\mathbb{I}_A Y|\sigma(\Pi)) = \mathbb{I}_A\mathbb{E}(Y|\sigma(\Pi))$

证明：只需证 (2)，因为从 (2)→(1) 即 $Y = \mathbb{I}_\Omega$

$X = \sum_{x\in S} x\mathbb{I}_{A_x}$，其中 $A_x := \{X = x\}$

(Step 1) $\sigma(X) = \{\sum_{x\in S'_X} A_x | S'_X \subseteq S_X\}$

$\sigma(X) = \{\sum_{k\in J} \Lambda_k | J \subseteq \mathbb{N}\}$

已知：$\sigma(X) \subseteq \sigma(\Pi) \Rightarrow \exists$ 一族 $\{x_k\}_{k\geqslant 1}$（可能有相同元素），使得 $X = \sum_{k\geq 1} x_k\mathbb{I}_{\Lambda_k}$，其中 $\cup_{k\geq 1}\{x_k\} = S_x$ (S_x 为取值空间)

注：Π 是 $\Pi_X = \{A_x|x \in S\}$ 的加细划分

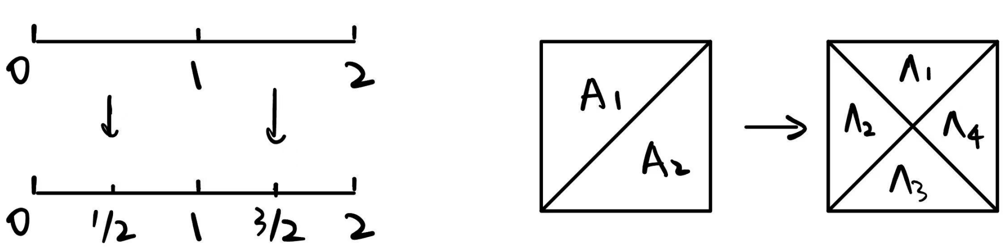

图 1: 加细划分

(Step 2) 对于 $\omega \in \Lambda_j, \forall j \geq 1$

$$\begin{aligned}
\mathbb{E}(XY|\sigma(\Pi))(\omega) &= \mathbb{E}(\sum_{k\geq 1} x_k \mathbb{I}_{\Lambda_k} Y|\sigma(\Pi))(\omega) \qquad [X = \sum_{k\geq 1} x_k \mathbb{I}_{\Lambda_k}] \\
&= \mathbb{E}(\sum_{k\geq 1} x_k \mathbb{I}_{\Lambda_k} Y|\Lambda_j) \qquad [\sigma(\Pi)\text{定义}] \\
&= \mathbb{E}(\sum_{k\geq 1} x_k \mathbb{I}_{\Lambda_k} Y \mathbb{I}_{\Lambda_j})/\mathbb{P}(\Lambda_j) \qquad [\text{推论}(2)] \\
&= \mathbb{E}(Y x_j \mathbb{I}_{\Lambda_j})/\mathbb{P}(\Lambda_j) \qquad [\mathbb{I}_{\Lambda_k}\mathbb{I}_{\Lambda_j}\text{当}\Lambda_k \neq \Lambda_j\text{时} = 0] \\
&= x_j \mathbb{E}(Y \mathbb{I}_{\Lambda_j})/\mathbb{P}(\Lambda_j) \\
&= x_j \mathbb{E}(Y|\mathbb{I}_{\Lambda_j}) \\
&= X(\omega)\mathbb{E}(Y|\mathbb{I}_{\Lambda_j})
\end{aligned}$$

$$\Rightarrow \mathbb{E}(XY|\sigma(\Pi)) = X \sum_{j\geq 1} \mathbb{I}_{\Lambda_j} \mathbb{E}(Y|\Lambda_j) = X\mathbb{E}(Y|\sigma(\Pi))$$

数学上有种现象叫“法国人的伎俩”，即把定理当定义用。严格地讲，这么做有时会出现存在性和唯一性不满足的问题。下面介绍一个常被当做定义用的定理：

Theorem 12

$\Pi = \{\Lambda_k, k \geq 1\}$ 为 Ω 的划分，$\mathbb{E}|X| < \infty$。记 $Y := \mathbb{E}(X|\sigma(\Pi)) = \sum_{k\geq 1} \mathbb{I}_{\Lambda_k} \mathbb{E}(X|\Lambda_k)$，则

1. Y 仍是一个离散随机变量，且 $\mathbb{E}|Y| \geq \mathbb{E}|X| < \infty$
2. $\sigma(Y) \subseteq \sigma(\Pi)$ (记作 $Y \in \sigma(\Pi)$，即 Y 的所有信息都在 $\sigma(\Pi)$ 里)
3. $\forall A \in \sigma(\Pi)$，有 $\mathbb{E}(Y\mathbb{I}_A) = \mathbb{E}(X\mathbb{I}_A)$

证明：(1)$E|X| = \sum_{x\in S_x} |x|\mathbb{P}(X = x) < \infty$

$$\mathbb{E}|Y| = \sum_{k\geq 1} |\mathbb{E}(X|\Lambda_k)|\mathbb{P}(\Lambda_k) \geq \sum_{k\geq 1}\sum_{x\in S} |x|\mathbb{P}(\{X = x\} \cap \Lambda_k)$$

逻辑上，现在第一个等号不成立，但之后 $< \infty$ 一写出来，之前的所有等号立刻成立，此处只为书写简便

$$\mathbb{E}|X| = \sum_{x\in S_x} |x|\mathbb{P}(X = x) = \sum_{x\in S} |x| \sum_{k\geq 1} \mathbb{P}(\Lambda_k \cap \{X = x\})$$

我们知道 $\sum_{x\in S} |x| \sum_{k\geq 1} \mathbb{P}(\Lambda_k \cap \{X = x\})$ 绝对收敛，若求和次序交换后的 $\sum_{k\geq 1}\sum_{x\in S} |x|\mathbb{P}(\{X = x\} \cap \Lambda_k)$ 也绝对收敛，则 $\mathbb{E}|Y| < \infty$ 得证。有一个引理可以保证绝对收敛：

Lemma 1 (菲赫金哥尔茨《微积分学教程》(2).P280. 推论)

从 273-280

Corollary 3 (来自定理12(1)) 1.（重期望公式）$\mathbb{E}|\mathbb{E}(X|\sigma(\Pi))| = \mathbb{E}|X|, \mathbb{E}(\mathbb{E}(X|\sigma(\Pi))) = \mathbb{E}(X)$

2. $|\mathbb{E}(X|\Lambda_k)| \leqslant \mathbb{E}(|X| \mid \Lambda_k), |\mathbb{E}(X|\sigma(\Pi))| \leqslant \mathbb{E}(|X| \mid \sigma(\Pi))$

(2) 由定义，$Y=\sum_{k\geq 1} y_k \mathbb{I}_{\Lambda_k}$，其中 $y_k := \mathbb{E}(X|\Lambda_k)$
记 $S_Y = \cup_{k\geq 1}\{y_k\}$，注意到，可能 $\exists i \neq j$，但 $y_i = y_j$
故 $J_y = \{k|y_k = y\}(y \in S_Y)$ 中个数可能大于 1

$$Y = \sum_{y\in S_Y} y\mathbb{I}_{\sum_{k\in J_y}\Lambda_k}$$

$$\{Y=y\} = \sum_{k\in J_y}\Lambda_k \in \sigma(\Pi)$$

$$\sigma(Y) \subseteq \sigma(\Pi) \quad \square$$

(3) $\mathbb{E}(Y\mathbb{I}_A) = \mathbb{E}(\mathbb{I}_A\mathbb{E}(X|\sigma(\Pi)))$

$$\begin{aligned}\mathbb{E}(Y\mathbb{I}_A) &= \mathbb{E}(\mathbb{I}_A\mathbb{E}(X|\sigma(\Pi))) \\ &= \mathbb{E}(\mathbb{E}(X\mathbb{I}_A|\sigma(\Pi))) \qquad [A\in\sigma(\Pi), \text{性质}(12)] \\ &= \mathbb{E}(X\mathbb{I}_A) \qquad [\text{重期望-推论}(3)]\end{aligned}$$

3° 关于离散随机变量的条件期望

Definition 19
概率空间 $(\Omega,\mathcal{F},\mathbb{P})$，$X,Y$ 为离散随机变量，$\mathbb{E}|X|<\infty$。定义 $\mathbb{E}(X|Y)=\mathbb{E}(X|\sigma(Y))=\mathbb{E}(X|\sigma(\Pi_Y))$，称为 X 关于 Y 的条件期望

注：$\omega=\{Y=y\}\in\Pi_Y$ 或 $Y(\omega)=y$，$\mathbb{E}(X|Y)(\omega)=\mathbb{E}(X|Y=y)$

Example 13
$\mathbb{E}(X|\Pi_\Omega)=\mathbb{E}(X|\sigma(\Omega))=\mathbb{E}(X)$

Example 14
$\mathbb{I}_A \perp\!\!\!\perp \mathbb{I}_B \Rightarrow \mathbb{E}(\mathbb{I}_A|\mathbb{I}_B) = [\text{Exa}(11)]\mathbb{E}(\mathbb{I}_A)$

Example 15
$\mathbb{E}(X|X)=\mathbb{E}(X|\sigma(X))=X[\text{Exa } 12]$

Property 13

假设以下期望、条件期望都有意义

1. $\mathbb{E}(aX+bY|Z)=a\mathbb{E}(X|Z)+b\mathbb{E}(Y|Z)$
2. $X \perp\!\!\!\perp Y \Rightarrow \mathbb{E}(X|Y)=\mathbb{E}(X)$
3. $\sigma(X)\subseteq\sigma(Z)\Rightarrow\mathbb{E}(XY|Z)=X\mathbb{E}(Y|Z)$
4. $\mathbb{E}(\mathbb{E}(X|Z))=\mathbb{E}(X)$
5. $|\mathbb{E}(X|Z)|\leqslant\mathbb{E}(|X|\mid Z)$

4° 关于多个离散随机变量的条件期望

$\mathbb{E}(Y|X_1,\cdots,X_n)$

1. 由 $X_1,\cdots,X_n$ 生成的 σ 代数 $\sigma(X_1,\cdots,X_n)$
2. $:=\mathbb{E}(Y|\sigma(X_1,\cdots,X_n))$

怎样生成 σ 代数可以包含 $X_1,\cdots,X_n$ 尽可能多的信息？

直觉是 $\bigcup_{k=1}^{\infty}\sigma(X_k)$，然而它不一定是 σ 代数，因为它对可列并不封闭。

每个 $\sigma(X_k)$ 是一个 σ 代数，因此它对可列并封闭。

然而,$\bigcup_{k=1}^{\infty}\sigma(X_k)$ 只是将每个 $\sigma(X_k)$ 中的集合简单地并在一起,并没有保证这些集合的可列并仍然在 $\bigcup_{k=1}^{\infty}\sigma(X_k)$ 中。

例如，假设 $X_k\in\sigma(X_k)$，那么 X_k 在 $\bigcup_{k=1}^{\infty}\sigma(X_k)$ 中，但 $\bigcup_{k=1}^{\infty}X_k$ 可能不在 $\bigcup_{k=1}^{\infty}\sigma(X_k)$ 中，因为它可能不属于任何一个单独的 $\sigma(X_k)$。问题出在 $\bigcup_{k=1}^{\infty}\sigma(X_k)$ 缺少 $\{\sigma(X_k)\}_{k\geq 1}$ 交互的部分

怎样把 $\bigcup_{k=1}^{\infty}\sigma(X_k)$ 变成 σ 代数？

Definition 20 (多个离散随机变量的条件期望)

定义由离散随机变量 $X_1,\cdots,X_n$ 生成的 σ 代数

$$\begin{aligned}\sigma(X_1,\cdots,X_n)&:=(X_1,\cdots,X_n)^{-1}(2^{S_1}\times\cdots\times 2^{S_n})\\&:=\{\underbrace{(X_1,\cdots,X_n)^{-1}(A_1\times\cdots\times A_n)}_{\text{柱集}}|A_1\times\cdots\times A_n\subseteq\underbrace{S_1\times\cdots\times S_n}_{\text{乘积空间}}\}\\&=\{\bigcap_{k=1}^{\infty}X_k^{-1}(A_k)|A_k\in 2^{S_k},1\leqslant k\leqslant n\}\end{aligned}$$

Theorem 13

令 $x_k = \sum_{i\geq 1} x_{k,i}\mathbb{I}_{\Lambda_{k,i}}, 1 \leqslant k \leqslant n$，为离散随机变量，对每一个 k，$\Pi_k := \{\Lambda_{k,i} | i \geq 1\}$ 为 Ω 的划分，定义

$$\Pi_{(X_1,\cdots,X_n)} := \{\Lambda_{1,i_1} \cap \cdots \cap \Lambda_{n,i_n} | i_k \geq 1, 1 \leqslant k \leqslant n\}$$

则

1. $\Pi_{(X_1,\cdots,X_n)}$ 是 Ω 的划分，且

$$\sigma(\Pi_{(X_1,cdots,X_n)}) = \left\{ \sum_{\substack{(i_1,\cdots,i_n) \\ \in J_1\times\cdots\times J_n}} (\Lambda_{1,i_1} \cap \cdots \cap \Lambda_{1,i_n}) | J_k \subseteq \mathbb{N}, 1 \leqslant k \leqslant n \right\}$$

2. $\sigma(X_1,\cdots,X_n) = \sigma(\Pi_{(X_1,\cdots,X_n)})$（即定义20是有意义的，well-defined，make sense，良定义）

Problem 4 (作业 2-2)

证明定理13在 $n=2$ 时成立

Definition 21

$\mathbb{E}|Z| < \infty$ 定义

$$\mathbb{E}(Z|X_1,\cdots,X_n) = \mathbb{E}(Z|\sigma(X_1,\cdots,X_n)) := \mathbb{E}(Z|\sigma(\Pi_{(X_1,\cdots,X_n)}))$$

Definition 22

$\Omega, \mathcal{F}, \mathbb{P}, Y : \Omega \to S_Y, X_1 : \Omega \to S_1, X_2 : \Omega \to S_2$ 为离散随机变量，称 Y 和 (X_1, X_2) 独立，若 $\sigma(Y) \perp\!\!\!\perp \sigma(X_1, X_2)$. $[\sigma(Y) = Y^{-1}(2^{S_Y}), \sigma(X_1,X_2) = (X_1,X_2)^{-1}(2^{S_1} \times 2^{S_2})]$
即 $\forall A \subseteq S_Y, B \subseteq 2^{S_1} \times 2^{S_2}, B = B_1 \times B_2$，有

$$\mathbb{P}(Y \in A, (X_1,X_2) \in B) = \mathbb{P}(Y \in A)\mathbb{P}((X_1,X_2) \in B)$$

其中 $\mathbb{P}((X_1,X_2) \in B) = \mathbb{P}(X_1 \in B_1, X_2 \in B_2)$

Problem 5 (作业 2-3)

证明：

$$\begin{aligned} Y \perp\!\!\!\perp (X_1,X_2) \Leftrightarrow & \forall y \in S_Y, x_1 \in S_1, x_2 \in S_2 \\ & \text{有}\mathbb{P}(Y=y,(X_1,X_2)=(x_1,x_2)) \\ & = \mathbb{P}(Y=y)\mathbb{P}((X_1,X_2)=(x_1,x_2)) \end{aligned}$$

有了上述定义，可以推广：

1. $(Y_1,\cdots,Y_n) \perp\!\!\!\perp (X_1,\cdots,X_n)$
2. $Y \perp\!\!\!\perp_A (X_1,\cdots,X_n) (A \in \mathcal{F}, \mathbb{P}(A) > 0)$

Property 14

$Y \perp\!\!\!\perp (X_1, X_2) \Rightarrow Y \perp\!\!\!\perp X_1, Y \perp\!\!\!\perp X_2$

证明：在定义22中取 $B_2 = \Omega$

$$\begin{aligned}
\mathbb{P}(Y \in A, X_1 \in B_1) &= \mathbb{P}(Y \in A, X_1 \in B_1, X_2 \in S_2) \\
&= \mathbb{P}(Y \in A)\mathbb{P}(X_1 \in B_1, X_2 \in S_2) \qquad [Y \perp\!\!\!\perp (X_1, X_2)] \\
&= \mathbb{P}(Y \in A)\mathbb{P}(X_1 \in B_1)
\end{aligned}$$

注：看到 $\Rightarrow$ 要自然地问，反过来 $\Leftarrow$ 成立吗？做数学要多问自己一些问题，即便没有答案

Corollary 4

$(Y_1, \cdots, Y_n) \perp\!\!\!\perp (X_1, \cdots, X_n) \Rightarrow Y_k \perp\!\!\!\perp X_j, 1 \leqslant k \leqslant m, 1 \leqslant j \leqslant n$

1 马氏链

2 泊松过程