Prerequisite for L2

谢润烁

2023/10/10

> 由于大家在高中数学和离散数学中已经学习过基础的概率论，故概率论的很多基础内容在之后的讨论班就不展开了。这份handout会将这些内容整理出来（包括常用的符号、定义、定理、术语），以供大家回顾和参考用。此外，我可能会插入一些我自己的注解，其中有些想法并不成熟，大家不必过分较真，因为我写注解只是想给大家提出些想法，看能不能启发大家想到一些有意思的东西。

1 Notation

Notation	Meaning
A	a set
Ω	sample space (the universe)
$A \setminus B$	$\{x : x \in A \land x \notin B\}$
A^c	the compliment of A, *i.e.*, $\Omega \setminus A$
Σ	the set of events
$\mathcal{F}$	a collection of set / a set family
$\Pr$	the probability measure

2 Definition

The occurrence or non-occurrence of a random event depends upon the chain of circumstances involved, which is called an **experiment** or **trial**; the result of an experiment is called its **outcome**.

> 正是因为随机事件可以不断重复（试验是可重复的），概率和频率才能统一起来。不过，<u>要是处理一些不能重复的随机现象，那概率还有什么意义吗？毕竟你都无法得到频率了</u>。

Definition (Sample Space):

The set of all possible outcomes of an experiment is called the **sample space** and is denoted by Ω.

> 找出样本空间是用概率论建模问题的第一步——如果一上来就开始算概率，很容易算着算着就迷失在一堆数字里面，最后算出来一个不知道对不对的数字。

Definition (σ-field/σ-algebra):

A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if it satisfies the following conditions:

1. $\varnothing \in \mathcal{F}$

2. If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=0}^{\infty} A_i \in \mathcal{F}$
3. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

> 集合的集合，一般称为“集族”，英文为family或者collection，通常使用$\mathcal{F}$来表示。
>
> σ代数了解一下即可，我们不会怎么涉及到，但是还是可以稍微了解一下为什么要引入测度论来作为概率论的基础。
>
> 此外，条件2中的...可能不是很严谨，这个地方应该强调是有countable（包括finite和infinite）个集合。如果是我来写，我可能会用指标集(Index set)来表述：
> If $I \subseteq \mathbb{N}$ and for any $i \in I, A_i \in \mathcal{F}$, then $\bigcup_{i \in I} A_i \in \mathcal{F}$.

We think of ***events*** as subsets of the sample space Ω. We require the set of events Σ is a σ-field.

Events A and B are called **disjoint** if their intersection is the empty set $\varnothing$;

- $\varnothing$ is called the **impossible event**.
- The set Ω called the **certain event**.

Note

The power set of Ω, which is written 2^{Ω}, is obviously a σ-field. However, when Ω is infinite, its power set is too large a collection for probabilities to be assigned reasonably to all its members.

> 当Ω为$\mathbb{R}$时，$(\mathbb{R}, \mathcal{B}, \Pr)$是一个良定义的概率空间：
>
> - $\mathcal{B}$称为Borel σ-field，是包含$\mathcal{F}$的最小的σ-field
> - $\mathcal{F}$是$\mathbb{R}$上所有开区间的集合
>
> 此时，对于任意$B \in \mathcal{B}$，$\Pr(B) = \int_B f(x)dx$是Lebesgue可积的（不一定Riemaan可积）。

Definition (Probability Measure):

A **probability measure** Pr on (Ω, Σ) is a function $\Pr : \Sigma \to [0, 1]$ satisfying

1. $P(\varnothing) = 0, P(\Omega) = 1$.
2. if $A_1, A_2, \dots$ is a collection of disjoint members of Σ, in that $A_i \cap A_j = \varnothing$ for all pairs i, j satisfying $i \neq j$, then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

An event A is called **null** if $\Pr(A) = 0$. If $\Pr(A) = 1$, we say that A occurs **almost surely**.

> 注意：null event不一定是impossible event $\varnothing$，almost surely的事件也不一定是certain event。为什么？
>
> 为什么这里要管Pr叫做“测度”呢？因为满足第二条规则的函数就是一个测度(measure)，而加上$\Pr(\Omega)=1$这条规则就变成“概率测度”了。当然，你管它叫概率函数(probability function)也没问题，Probability and Computing一书没引入测度的语言，它就将其称为probability function。

Definition (Probability Space):

The triple $(\Omega, \Sigma, \Pr)$, comprising a set Ω, a σ-field Σ of subsets of Ω, and a probability measure $\Pr$ on (Ω, Σ), is called a **probability space**.

> 咱们做第一次讨论班的时候提到，不存在一个合法的概率测度Pr使得：
>
> $$\forall m, n \in \mathbb{N}. \Pr(\omega=m)=\Pr(\omega=n)$$
>
> 因此，“从均匀分布的自然数抽取一个数字的概率”是没有良定义的。但是大家在离散数学的作业中做过这样一道题：
>
> > 试构造适当的概率模型证明：从正整数中随机取2个数，它们互素的概率为$\frac{6}{\pi^2}$。
>
> 大家认为这道题是似乎是违反了概率论的公理体系，而知乎上也确实有人提问了这个问题：面对如「求两个随机自然数a，b互质的概率？」这类涉及「算术密度」的问题，如何使概率定义符合公理化体系？ - 知乎。
>
> 但是，既然题目说的是“构造适当的概率模型”，那我们不一定需要假设$\{i\} \in \Sigma, i \in \mathbb{N}$。也就是说，抽到某个数字的事件不需要是可测的——**注意，概率测度Pr不一定要对Ω中的元素可测，我们只要对Σ中的元素可测即可。**
>
> 因此，我们可以这么定义概率空间中的Σ和Pr：
>
> - A_i：抽到的第一个数字为p_i的倍数的概率
> - B_i：抽到的第二个数字为p_i的倍数的概率
> - $\Pr(A_i)=\Pr(B_i)=1/p_i$
> - $\Sigma=\bigcup_{i=1}^{\infty}(A_i \cup B_i)$
>
> 这个时候，我们就可以在有良定义的情况下求解了。我们要抽到的两个数字互素，对于每一个素数，最多只有一个数是这个素数的倍数，也就是：
>
> $$\bigcap (A_i \cap B_i)^c$$
>
> 而由于A_i和B_i是相互独立的，我们有$\Pr(A_i \cap B_i)=1/p_i^2$。因此我们可以得到：
>
> $$\Pr\left(\bigcap_{i=1}^{\infty}(A_i \cap B_i)^c\right)=\prod_{i=1}^{\infty}\Pr((A_i \cap B_i)^c)=\prod_{i=1}^{\infty}(1-\frac{1}{p_i^2})$$

> 注意到这个地方之所以能把$\bigcap_{i=1}^{\infty}$提出来变成$\prod$，是因为借助了事件的独立性（请见下文）。此处的独立性证明在有限情况下是可以证明的（无限情况下可能不行，那我们就先算有限情况下的概率，然后对概率表达式取极限）。
>
> 得到这个结果之后，根据Euler's product formula：
>
> $$\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{i=1}^{\infty}\frac{1}{1-p_i^{-s}}$$
>
> 于是乎：
>
> $$\Pr\left(\bigcap_{i=1}^{\infty}(A_i\cap B_i)^c\right)=\prod_{i=1}^{\infty}(1-\frac{1}{p_i^2})=\prod_{i=1}^{\infty}(1-p_i^{-2})=\frac{1}{\sum_{n=1}^{\infty}\frac{1}{n^2}}=\frac{6}{\pi^2}$$
>
> (本解答参考了https://www.zhihu.com/question/23376401/answer/24390803和钟开莱的 Elementary Probability Theory With Stochastic Processes and an Introduction to Mathematical Finance的2.5节Arithmetical density)

Definition (Conditional Probability of Event)

If $\Pr(B)>0$ then the **conditional probability** that A occurs given that B occurs is defined to

$$\Pr(A\mid B)=\frac{\Pr(A\cap B)}{\Pr(B)}$$

pronounced 'the probability of A given B', or sometimes 'the probability of A conditioned (or conditional) on B'.

> 你可以把求$\Pr(A\mid B)$的过程看成是把样本空间中不属于B的样本点全部砍掉，得到一个全新的样本空间，然后再来计算新样本空间里面属于A的样本点在新的样本空间中的概率。

Definition (Independence of Events)

Events A and B are called **independent** if

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

More generally, a family $\{A_i : i\in I\}$ is called **independent** if

$$\Pr\left(\bigcap_{i\in J}A_i\right)=\prod_{i\in J}\Pr(A_i)$$

for any *finite* subsets J of I.

A family $\{A_i : i\in I\}$ is called **k-wise independent** if

$$\Pr\left(\bigcap_{i\in J}A_i\right)=\prod_{i\in J}\Pr(A_i)$$

for any *finite* subsets J of I such that $|J| = k$. **2-wise independent** is also called **pairwise independent**.

> 注意，这里目前讨论的是**事件**的独立性和条件概率，后面学到**随机变量**的时候会有随机变量的独立性和条件概率。<u>不妨猜猜随机变量和事件是什么关系？</u>

Definition (Random Variable)

A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \in \Sigma\}$ for each $x \in \mathbb{R}$. Such a function is said to be **Σ-measurable**.

> 在我看来，随机变量可以称得上是概率论研究的最重要的一个对象了。如果概率论没有随机变量这个概念，那么它可能不需要单独成立一门学科，而是只需要些朴素的组合数学技巧就够了。也许大家可以带着这个问题去学习之后的内容：<u>随机变量到底强大在什么地方？</u>[1]

> 值得注意的是，尽管随机变量被定义成一个从样本空间Ω到实数集$\mathbb{R}$的函数，但是在研究随机变量的时候，我们从来不关心其样本空间是什么。我们甚至都没有给它假设一个样本空间，我们仅关心其分布(distribution)。<u>问题来了，怎么去描述一个随机变量的分布呢？</u>

3 Theorem

Basic Lemma derived from Axioms

1. $\Pr(A^c) = 1 - \Pr(A)$
2. If $A \subseteq B$ then $\Pr(B) \geq \Pr(A)$
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
4. (**Inclusion-Exclusion Principle / Poincaré's Formula**)

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{I \subseteq \{1,...,n\}} (-1)^{|I|-1} \Pr\left(\bigcap_{i \in I} A_i\right).$$

> 这里面可能除了容斥原理，其它结论都比较显然。集合论中的容斥原理是这样的（如果要把$||$去掉，只需要把运算$+, -$变成集合的运算$\cup, \setminus$即可）：
>
> $$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{I \subseteq \{1,...,n\}} (-1)^{|I|-1} \left|\bigcap_{i \in I} A_i\right|.$$
>
> <u>不妨思考一下两者是什么关系</u>。此外，<u>你可以回忆一下容斥原理是怎么证明的，以及能不能有更简单的证明方法</u>[2]。

> 与容斥原理配套的还有一个**Bonferroni's inequality**，不过目前我好像没见过这个不等式的应用，所以就不放在这里了。

Theorem (Law of Total Probability)

let $B_1, B_2, \ldots, B_n$ be a partition of Ω such that $\Pr(B_i) > 0$ for all i. Then

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

> 这些定理的证明都很trivial，但是都用得挺多。

Theorem (Boole’s Inequalities / Union Bounds)

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i)$$

> 这是一个用得很多的定理，比方说用概率法证明离散对象性质时就很常用。一般比较少叫Boole’s Inequalities，都是叫Union Bounds。

Theorem (Bayes's Formula)

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \Pr(A)}{\Pr(B)}$$

> 有时候分母的$\Pr(B)$会用全概率公式展开。

---

1. 我自己对这个问题的回答是来自于我对定量和定性的理解。当处理的对象不能映射到数集（比方说$\mathbb{N}$，$\mathbb{R}$）上时，你只能做一些比较定性的分析（“是”与“否”，或者说处理的对象只能映射到一个布尔变量$\{\textbf{True}, \textbf{False}\}$？）。定性分析在我看来很像决策树——If XXX then XXX else XXX（是不是有点像人们所说的“非黑即白”的简单思维？当然了，定性分析也没那么菜，“非黑即白”是只有一层的决策树，多层的决策树还是挺有用的）。而决策树这个模型在真实世界的复杂度面前显得无比简单，所以我认为定性分析仅限于解决简单的任务，解决困难的任务需要足够好的测量和足够好的数学工具/模型，也就是定量分析。 ↩

2. 这么说当然是有的……有一个用**指示函数(Indicator)**来证明的方法，这东西还能用来证明对称差运算(symmetric difference, $A \Delta B = (A \setminus B) \cup (B \setminus A)$)具有结合性(associative) ↩