[AI2613 Lecture 1] Review of Probability Theory

February 24, 2023

1 Probability Space

We start with the notion of probability space. A standard reference for the probability theory is [1].

概率论公理化：

三元组

表达概率

(随机实验)

翻牌，骰子落地？点

翻上

Definition 1 (Probability Space). *A probability space is a tuple* $(\Omega, \mathcal{F}, \mathrm{P}(\cdot))$ *satisfying the following requirements.*

- *The universe* Ω *is a set of "outcomes" (which can be either countable or uncountable).* 一系列结果 e.g. $\Omega=\{1,2,3,4,6\}$
- *The set* $\mathcal{F} \subseteq 2^\Omega$ *is a* σ*-algebra (the set of all possible "events"). Here we say* $\mathcal{F}$ *is a* σ*-algebra if* $\mathcal{F}$ *satisfies:* $\mathcal{F}$≙所有事件，事件是结果的集合； $\boxed{A \in \mathcal{F} \subseteq 2^\Omega}$ (而且)所有事件的集合为$\mathcal{F}$； $\boxed{\mathcal{F} \subseteq 2^\Omega}$ 且满足三定义
  - ① $\varnothing, \Omega \in \mathcal{F}$; 不可能事件，全事件都是事件
  - ② $\forall A \in \mathcal{F}$, *it holds* $A^c \in \mathcal{F}$; 发生A是事件 => 不发生A也是～
  - ③ *for any finite or countable sequence of sets* $A_1, \ldots, A_n, \cdots \in \mathcal{F}$, *it holds that* $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$. 有$A_1$~$A_n$这些事件 => 至少发生一个$A_i$也是事件

A是事件，$\mathcal{F}$是事件合集

e.g. $A=\{2,4,6\}$

$\mathcal{F}=\{\{1\},\{2\}\cdots\{1,2\}\cdots\{2,4,6\}\cdots\}$

- *The probability function* $\mathrm{P}(\cdot): \mathcal{F} \to [0,1]$ *satisfies* 不是所有函数都能是P，需满足以下条件。
  - $\mathrm{P}(\varnothing) = 0$, $\mathrm{P}(\Omega) = 1$;
  - $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$ *for all* $A \in \mathcal{F}$;
  - *for any finite or countable sequence of* disjoint *sets* $A_1, \ldots, A_n, \cdots \in \mathcal{F}$, *it holds that* $\mathrm{P}\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mathrm{P}(A_i)$.

P为概率函数，$\mathcal{F}$为定义域，[0,1]为值域，表现为P(A)=?，即一个事件A的概率值

$A^c := \Omega \setminus A$.

Let $\mathcal{S} \subseteq 2^\Omega$. We use $\sigma(\mathcal{S})$ to denote the minimal σ-algebra containing sets in $\mathcal{S}$. That is, for any $\mathcal{F} \subseteq 2^\Omega$, $\mathcal{F} = \sigma(\mathcal{S})$ if and only if (1) $\mathcal{F}$ is a σ-algebra; (2) $\mathcal{S} \subseteq \mathcal{F}$; (3) For any $\mathcal{F}' \subseteq \mathcal{F}$ such that $\mathcal{S} \subseteq \mathcal{F}'$, $\mathcal{F}'$ is not a σ-algebra.

The term "minimal" here is with respect to the set inclusion relation $\subseteq$.

For every $n \in \mathbb{N}$, we use $[n]$ to denote the set $\{1,2,\ldots,n\}$.

Example 1 (Tossing n fair coins). *Let* $\Omega = \{0,1\}^n$, $\mathcal{F} = 2^\Omega$ *and for every* $S \in \{0,1\}^n$, $\mathrm{P}(\{S\}) = \frac{1}{2^n}$.

建模 uniform distribution

Example 2 (Uniform Reals in $(0,1)$). *The uniform distribution on* $(0,1)$ *is defined as follows:*

- $\Omega = (0,1)$;
- $\mathcal{F}$ *is the* σ*-algebra consisiting of all Borel sets on* $(0,1)$*, namely the collection of subsets of* $(0,1)$ *obtained from open intervals by repeatedly taking* countable *unions and complements;* 见后。
- $\forall$ *interval* $I = (a,b)$, $P(I) = b - a$ *(This is the Lebesgue measure).*

The definition here, although a bit wired at the first glance, is in fact the simplest way to capture our intuition that the probability that a point is in (a,b) should be $b-a$. We cannot take $\mathcal{F} = 2^\Omega$ in Example 2 as doing so may include some *non-measurable* sets. In fact, $\mathcal{F}$ is called the *Borel algebra*, which is the smallest σ-algebra containing all open intervals. One can construct a non-Borel set in $(0,1)$ assuming the *axiom of choice*. In fact, the existence of a non-Borel set is independent of Zermelo-Fraenkel set theory without the axiom of choice. We use $\mathscr{R}$ to denote the collection of Borel sets on $\mathbb{R}$. For any $A \subseteq \mathbb{R}$, we use $\mathscr{R}(A)$ to denote $\mathscr{R} \cap 2^A$.

一个例子：　掷骰子，求建模后的 $\Omega, \mathcal{F}, \mathcal{P}$.

Ans：
$$
\begin{cases}
\Omega=\{1,2\cdots 6\} \overset{记}{=\!=} [6] \\
\mathcal{F}=2^{\Omega} \overset{记}{=\!=} 2^{[6]} \\
P(A)=\dfrac{|A|}{6}
\end{cases}
$$

☆☆☆ $|A|$ → 事件大小，可验证这个函数是满足定义的.

A → 事件

Note：$\{\Omega, \mathcal{F}, \mathcal{P}\}$:= 概率空间不唯一性：

e.g. 抛硬币：①：
$$
\begin{cases}
\Omega=\{1,0\} \quad (:=\{正,反\}) \\
\mathcal{F}=2^{\Omega} \\
P=\begin{cases} 0, & \phi \\ \frac{1}{2}, & \{1\} \\ \frac{1}{2}, & \{0\} \\ 1, & \Omega \end{cases}
\end{cases}
$$

②
$$
\begin{cases}
\Omega=(0,1) \\
\mathcal{F}=\sigma((0,\tfrac{1}{2}], (\tfrac{1}{2},1)) \\
P=\begin{cases} \frac{1}{2}, & (0,\frac{1}{2}) \\ \frac{1}{2}, & (\frac{1}{2},1) \end{cases}
\end{cases}
$$

\# $\sigma(\ )$ 意义见下

⇒ 勒贝格测度定义，

Note：最小 σ 代数 $\sigma(\ )$

$\mathcal{F}$:= σ-algebra 可以被理解为集合的集合，当然不是所有集合集合起来都满足 σ-algebra 定义，因此诞生符号 $\sigma(s_1, s_2\cdots s_n)$，代表用 $s_1\cdots s_n$ 并添加元素，生成的最小的 σ-algebra

↑元素数量

e.g. ① $\Omega=[6]$，$\sigma(\{1,2\})=\{\phi, [6], \{1,2\}, \{3,4,5,6\}\}$

② $\Omega=[6]$，$\sigma(\underset{记A}{\{1,2\}}\underset{B}{\{2,3\}})=\{\phi, \Omega, A, B, A^c, B^c, A\cup B, A^c\cup B^c, (A^c\cup B^c)^c, A^c\cap B^c, \cdots\}$

$(A^c\cup B^c)^c$ ↓ $A\cap B$

※：可以看出"交"也是要在 $\mathcal{F}$ 中的，因为 $A\cap B=(A^c\cup B^c)^c$，因此还是有很多种情况.

※ 也可以理解，$\mathcal{F}=\sigma(S_1,\cdots S_i)$ 时，只要知道 $S_1\cdots S_i$ 所有事件的概率，则定义在该 $\mathcal{F}$ 上的 $\mathcal{P}$ 就可以被推出了.（基于定义就能算出）

Note: Borel set:

在先前例中: $\Omega=(0,1)$.

$\nearrow$ (0,1)上所有开集构成的 σ(.)

$\mathcal{F}=B((0,1))=\sigma(\forall(a,b)\text{ in }(0,1))$

这之中就用到了Borel σ-set:

$\boxed{\text{Borel }\sigma\text{-set}}$: $B((x,y))\triangleq\sigma(\forall(a,b)\text{ in }(x,y))$

记 Borel sigma algebra set

$\triangleq$ $\forall(a,b)$ in (x,y), 通过交并补构成的集合, 所能这些的集合

$\downarrow$ $\boxed{\text{Borel set}}$ $\downarrow$ $\boxed{\text{Borel }\sigma\text{-set}}$

Borel set 和 σ-algebra 是类似的, 但基于拓扑空间.

Borel set 是 { 拓扑空间(a,b) 上用 σ-algebra (定义) 作出的集合, 若与制作的是(a,b)区间内所有的开子集 / (空间上的 σ-algebra)

例如 $(a,b)=(0,1)$, 则 $[(0,0.1)\cup(0.2,0.3)]$ 就是一个Borel set

Borel algebra set $B((a,b))$ 指 (a,b) 区间内

所有开集制作成的 σ-algebra 也是所有 borel set 构成的集合.

因此

* algebra 一词可以理解为集合生成的集合, 如 σ-algebra 就是满足三性质且是集合生成的集合 e.g. $\sigma(\{0,1\},\{2,3\})$

* 因为 Borel set 是空间上的 σ-algebra,

故 Borel set 有可测等好性质, 故包含所有 Borel set 的

Borel σ-algebra 就被用于了 real r.v. 的值域刻画.

即: $(\Omega',\mathcal{F}')=(\mathbb{R},B_{\mathbb{R}})$

事件

一个骰子有6种结果：$\{1,2,\cdots,6\}$，
≙一个骰子的事件结果为X.
X取值为 Ω', $\Omega'=\{1,\cdots,6\}$
or $\mathbb{R}$ ……
随机变量是概率空间的数量化

直观解释

2 Random Variables

数学定义 → 满足(映射)三大步:

Definition 2 (Measurable Space). *Consider a set Ω and a σ-algebra $\mathcal{F}$ on Ω. The tuple $(\Omega,\mathcal{F})$ is called a measurable space.*

Definition 3 (Measurable Function). *Let $(\Omega,\mathcal{F})$ and $(\Omega',\mathcal{F}')$ be two measurable spaces and $X : \boxed{\Omega} \to \Omega'$ be a function. We say X is a $\mathcal{F}$-measurable function if*

step1 将每个Ω中的ω

step2 映射到 数量值，取值范围为Ω'

$$\forall B' \in \mathcal{F}',\ X^{-1}(B') \in \mathcal{F},$$

step3 且满足 ∀事件B (或说结果ω的集合)，将每个ω按X规定的映射变化后，得到的新B'，若B'在X的值域F'中，则一定要求原事件B在F中。

$X^{-1}(B') \triangleq \{\omega \in \Omega \mid X(\omega) \in B'\}$ is the inverse of X.

For any function, we use $\sigma(X)$ to denote the minimal σ-algebra $\mathcal{F}$ such that X is $\mathcal{F}$-measurable.

Definition 4 (Random Variable). *. Let Ω' and $\mathcal{F}'$ in Definition 3 be $\mathbb{R}$ and the Borel algebra $\mathscr{B}$, then X in Definition 3 is a (real-valued) random variable.*

Borel set　其中的

We say a random variable X *discrete* if its range $\mathsf{Ran}(X)$ is countable. In other words, X can only take at most countable many distinct values. Otherwise, we say X is a *continuous* random variable.

The *measurability* of a random variable X captures the intuition that we can safely talk about *the probability of X taking some value*. Intuitively X induces a partition of Ω where two outcomes ω_1 and ω_2 are in the same partition if and only if $X(\omega_1) = X(\omega_2)$. If the partition defined by X is more "coaser" than the partition defined by a σ-algebra $\mathcal{F}$, then X is $\mathcal{F}$ measurable.

Example 3 (Measurable Functions of Tossing a Dice). *. Let $\Omega = [6]$. We have three σ-algebras on Ω: $\mathcal{F}_1 = 2^{[6]}$, $\mathcal{F}_2 = \sigma(\{1,3,5\})$ and $\mathcal{F}_3 = \sigma(\{1,2\})$. Consider three random variables $X_1, X_2, X_3 : \Omega \to \mathbb{R}$ such that $X_1 : \omega \mapsto \omega$, $X_2 : \omega \mapsto \omega \bmod 2$ and $X_3 : \omega \mapsto \mathbf{1}[\omega \le 2]$. Then all these three mappings are $\mathcal{F}_1$-measurable, only X_2 is $\mathcal{F}_2$-measurable and only X_3 is $\mathcal{F}_3$-measurable.*

※ 值得一提的是 X_1映射是 $\mathbb{Z}\to\mathbb{R}$，一种可行映射为所有R中包含1的子集，下集合都为F中的{1}。如 $\{1\}\to(0.1,\ 1.5)$. $\{1,2\}\to(0.1, 2.1)$

解释：X_2取值为 0 or 1
故可取的Borel set 类似（$\emptyset$　$\{0\}$, $\{1\}$, $\{0,1\}$）也为Ω'
而$X^{-1}(B)$为　$\emptyset$　$\{2,4,6\}$　$\{1,3,5\}$　Ω
这些恰好都有$\mathcal{F}_2$中，故$\mathcal{F}_2$-measurable.

3 Distribution

Let $(\Omega,\mathcal{F},\mathrm{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a $\mathcal{F}$-measurable random variable. Let $\mathscr{B}$ be the Borel algebra on $\mathbb{R}$. The distribution space $(\mathbb{R},\mathscr{B},\mathbf{Pr})$ induced by X is defined as

$$\forall A \in \mathscr{B}, \mathbf{Pr}[A] = \mathbf{Pr}[X \in A] \triangleq \mathbf{P}[X^{-1}(A)].$$ 解释见后

The function $F(x) := \mathbf{Pr}[X \le x] = \mathrm{P}(X^{-1}(-\infty, x))$ is called the *cumulative distribution function (cdf)* of X. CDF

If a function $f : \mathbb{R} \to \mathbb{R}$ satisfies for any $a \le b$: PDF

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a),$$

then we call $f(x)$ a *probability density function (pdf)* of X.

※ $\sigma(\underset{\text{r.v.}}{X})$ 意为使已知值域Ω'的r.v. X 能F可测的原概率空间中的F，只是一个记号，
如：$X: \omega \to \mathbf{1}[\omega\text{为素数}]$
则 $\mathcal{F} = \sigma(\{\text{所有素数}\})$.
因为 $X: \omega \to \begin{cases} 0 & \text{no prime} \\ 1 & \text{prime} \end{cases}$
0对应非素数集合
1对应素数集合即可对应上了.

Example 4 (Exponential Distribution). *If $X \sim \mathsf{Exp}(\lambda)$, or equivalently it follows exponential distribution with rate λ for $\lambda > 0$, then its pdf is*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

一个例子：$\mathcal{F} = \sigma([0,1),[1,2))$，$\Omega = \{0, 0.9, 1.6\}$.

$X: w \mapsto \mathbb{1}(w<1.5)$

X是否$\mathcal{F}$-measurable：

答：可测，$\forall B' \in \mathcal{F}'$，$X^{-1}(B') \in \mathcal{F}$.

$\Omega = \{0, 1.2, 1.6\}$.

答：不可测. $\exists B' = B([0,0],[1.2,1.2]) = \{\{0\},\{1.2\},\{0,1.2\},\phi\}$.

Note：必须注意概率的数值意义基于对底层概率空间的讨论.

即 $\forall A \in \mathcal{B}$，$Pr[A] = Pr[X \in A] \triangleq P[X^{-1}(A)]$

在r.v.的数值空间中的事件A ⟶ 的概率

$\Downarrow$

若A为单值，就是熟悉的 $Pr[X=x]$

定义为 ⟶ X映射在原概率空间上的事件的概率（由空间三元组定义）

4 Expectation and Variance

Definition 5 (Expectation)**.** *. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable.*

★★★ rv的期望是 R→R的一个函数.

- *For a discrete random variable X, its expectation is*

角度一：

$$\mathbf{E}[X] := \sum_{a \in \mathrm{Ran}(X)} a \cdot \mathbf{Pr}[X = a].$$

显然可能有多个 w, st $X(w) = a$

If Ω is at most countable, we can also write

角度二：

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} \mathrm{P}(\{\omega\})\, X(\omega).$$

- *For a continuous random variable X with pdf f, its expectation is*

$$\mathbf{E}[X] := \int_{-\infty}^{\infty} t \cdot f(t)\,\mathrm{d}t.$$

Sometimes it is more convenient to equivalently write the expectation as

$$\mathbf{E}[X] = \int_{\Omega} X(\omega)\mu(d\omega) = \int_{\Omega} X\,\mathrm{d}\mu.$$

using Lebesgue integration.

Example 5 (Expectation of Exponential Distribution)**.** *Let $X \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$, then*

$$\mathbf{E}[X] = \int_0^{\infty} t \cdot \lambda e^{-\lambda t}\,\mathrm{d}t = \frac{1}{\lambda}.$$

Definition 6 (Variance)**.** *The variance of a random variable X is*

$$\mathbf{Var}[X] := \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2.$$

Proposition 7. *Let $X_1, \ldots, X_n$ be random variables where n is a finite constant. Then*

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

5 Conditional Probability

Definition 8 (Conditional Probability)**.** *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Let $A, B \in \mathcal{F}$ be two events with $\mathrm{P}(B) > 0$. The conditional probability of A given B is*

$$\mathrm{P}(A \mid B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

首先定义事件A和B的条件概率.

This is well-defined since we know from the definition of σ-algebra that $A \cap B \in \mathcal{F}$.

In the following, we define the notion of *conditional expectation* for those *discrete* random variables.

Definition 9 (Conditional Expectation). *Let* $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space. Let* $A \in \mathcal{F}$ *be an event with* $\mathrm{P}(A) > 0$. *Let* $X : \Omega \to \mathbb{R}$ *be a* discrete *random variable. The conditional expectation of* X *conditioned on* A *is*

$$\mathbf{E}[X \mid A] := \sum_{a \in \mathsf{Ran}(X)} a \cdot \mathbf{Pr}[X = a \mid A].$$

$Pr[X=a|A] = P[\underbrace{X^{-1}(a)}_{\text{event}} | \underbrace{A}_{\text{event}}]$

仍然按X的值域求

Let $Y : \Omega \to \mathbb{R}$ *be another discrete random variable. The conditional expectation of* X *conditioned on* Y*, written as* $\mathbf{E}[X \mid Y]$*, is a random variable* $f_Y : \Omega \to \mathbb{R}$ *such that*

将A改为r.v. Y之后, E[X|Y]

$$\forall \omega \in \Omega : f_Y(\omega) = \mathbf{E}\left[X \,\middle|\, Y^{-1}(Y(\omega))\right] = \mathbf{E}[X \mid Y = Y(\omega)]. \tag{1}$$

变为了 $w \mapsto R$ 的另一个随机变量,

含义为:

在选取与Y(ω)相同的样本的情况下, 这些样本

$Y^{-1}(Y(\omega))$ 条件期望

在随机变量X的映射下的期望

$E[X|Y^{-1}(Y(\omega))]$

Proposition 10.

因为值域对应的概率空间在于Y

- $\mathbf{E}[X \mid Y]$ *is* $\sigma(Y)$*-measurable.*
- $\mathbf{E}[\mathbf{E}[X \mid Y]] = \mathbf{E}[f_Y] = \mathbf{E}[X]$.

意义上也是正确的

Proof.
- Since the value of $\mathbf{E}[X \mid Y]$ is determined by $Y(\omega)$, it is clearly $\sigma(Y)$-measurable.
- We compute $\mathbf{E}[f_Y]$ by definition.

$$\begin{aligned}
\mathbf{E}[f_Y] &= \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \mathbf{E}\left[X \,\middle|\, Y = y\right] \\
&= \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \sum_{x \in \mathsf{Ran}(X)} \mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] \cdot x \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[X = x \wedge Y = y] \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \mathbf{Pr}[X = x] \\
&= \mathbf{E}[X].
\end{aligned}$$

□

6 *Conditional Expectation for General Random Variables*

The definition of conditional expectation for continuous random variables is more subtle. For example, if $X, Y \sim N(0,1)$ are two independent random variables following standard normal distribution, then intuitively $\mathbf{E}[X \mid Y = 0]$ should be identical to $\mathbf{E}[X]$, which is zero. However, we cannot directly adopt the definition before since $\mathbf{Pr}[Y = 0] = 0$.

Definition 11. *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be the probability space. Let X be a random variable with $\mathbf{E}[|X|] < \infty$. The conditional expectation $\mathbf{E}[X \mid Y]$ is a $\sigma(Y)$-measurable random variable f_Y satisfying*

$$\forall A \in \sigma(Y), \int_A f_Y \, \mathrm{d}P = \int_A X \, \mathrm{d}P.$$

The existence and uniqueness of f_Y follow from Radon-Nikodym theorem.

7 Balls-into-Bins

Balls-into-bins is a simple random process in which a person throws m balls into n bins uniformly at random. Many interesting questions can be asked about the process.

7.1 Birthday Paradox

问题是研究m球入n bin，至少一个bin有≥2个球的概率.

Birthday paradox refers to the seemly counter-intuitive fact that some students in the class are very likely to share the same birthday. Viewing bins as dates and balls as students, the event that two students have the same birthday can be modeled as the event that some bin contains more than one ball.

Note that each ball is thrown independently. Condition on there is no collision after the $k-1$ balls are thrown, the probability that no collision occurs after throwing the k^{th} ball is $\frac{n-k+1}{n}$. Hence,

* 记X_i为第i次无重复放球的事件，

$$\begin{aligned}
\mathbf{Pr}[\text{no same birthday}] &= \prod_{k=1}^{m} \frac{n-k+1}{n} \\
&= \prod_{k=1}^{m-1} \left(1 - \frac{k}{n}\right) \\
&\le \exp\left\{-\frac{\sum_{k=1}^{m-1} k}{n}\right\} \quad (\text{by } 1 + x \le e^x) \\
&= \exp\left\{-\frac{m(m-1)}{2n}\right\}. \qquad (2)
\end{aligned}$$

⇒推导：全部无重复 $= \Pr[X_i \wedge X_{i-1} \wedge \cdots X_1]$ （由 $P(A|B) = \frac{P(A \wedge B)}{P(B)}$）

$= \Pr[X_1] \cdot \Pr[X_2 | X_1] \cdot \Pr[X_3 | X_2 \wedge X_1] \cdot \cdots \cdot \Pr[X_i | X_{i-1} \wedge \cdots \wedge X_1]$

$= \prod_{k=1}^{i} \Pr[X_k | \bigwedge_{j=1}^{k-1} X_j]$，而 $\Pr[X_k | \bigwedge_{j=1}^{k-1} X_j] = \frac{n-k+1}{n}$.

因为已经放k-1个无重复的球，再放无重复的概率显然为 $\frac{n-(k-1)}{n}$ 代入即可.

For $m = O(\sqrt{n})$, the probability can be arbitrarily close to 0.

When n is sufficiently large, Equation (2) is tight because $\frac{k}{n} \le \frac{m}{n} = O(\frac{1}{\sqrt{n}}) \to 0$ and $1 + x \le e^x$ is tight when x is small.

7.2 Coupon Collector

The coupon collector problem asks the following question: If each box of a brand of cereals contains a coupon, randomly chosen from n different types of coupons, what is the number of boxes one needs to buy to collect all n coupons? In the language of balls-into-bins, it asks how many balls one needs to throw until each of the n bins contains at least one ball.

The expectation can be easily calculated using the linearity of expectations. Let X_i be the number of balls to throw to get the i-th distinct type of coupon while exactly $i-1$ distinct types of coupons are already in had. Then the number of draws X to collect all coupons satisfies

问题是研究 n个 bin，丢多少次 n个bin全部非空的期望.

$$X = \sum_{i=1}^{n-1} X_i.$$

X_i代表i-1个框非空下，投入空框的事件.

By the linearity of expectations:

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

期望线性性

It is clear that $X_i \sim \mathrm{Gem}(\frac{n-i+1}{n})$ and therefore $\mathbf{E}[X_i] = \frac{n}{n-i+1}$. As a result,

X_i: 服从几何分布 (第一次成功所花次数)

$$\mathbf{E}[X] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot H(n),$$

代入化简即可

where $H(n)$ is the harmonic number satisfying $\lim_{n\to\infty} H(n) = \log n + \gamma$ for $\gamma = 0.577\ldots$.

γ is called the Euler constant.

8 Concentration Inequalities

In addition to the expectation, we are often interested in how a random variable deviates from certain fixed value. Concentration inequalities are inequalities of this form.

求出期望后，还要研究发生在期望附近的概率. 这里介绍了 Markov inequality 和 chebysher inequality.

8.1 Markov's Inequality

Theorem 12 (Markov's Inequality). *. For any non-negative random variable X and $a > 0$,*

$$\mathbf{Pr}[X \ge a] \le \frac{\mathbf{E}[X]}{a}.$$

Proof. Since X is non-negative, we have

$$\mathbf{E}[X] \ge a \cdot \mathbf{Pr}[X \ge a] + 0 \cdot \mathbf{Pr}[X < a].$$

This is equivalent to

$$\mathbf{Pr}[X \ge a] \le \frac{\mathbf{E}[X]}{a}.$$

□

Example 6 (Concentration for Coupon Collector). *. Recall that X is the number of balls we need. Apply Markov's inequality, for $c > 0$ we have*

$$\mathbf{Pr}[X \ge c] \le \frac{\mathbf{E}[X]}{c} = \frac{nH_n}{c}.$$

Thus, the probability that we need to draw the coupon for more than $100 \cdot nH_n$ times is less than 0.01.

8.2 Chebyshev's Inequality

A common trick to improve concentration is to consider $\mathbf{E}[f(X)]$ instead $\mathbf{E}[X]$ for some increasing function $f : \mathbb{R} \to \mathbb{R}$ since

$$\mathbf{Pr}[X \geq a] = \mathbf{Pr}[f(X) \geq f(a)].$$

Concentration inequalities give a sense that how the random variable deviate from its expectation. Then the probability we care about is actually $\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a]$ for some postive constant a. Choosing the increasing function $f(x) = x^2$, we get the following Chebyshev's inequality.

Theorem 13 (Chebyshev's Inequality)**.** . *For any random variable with bounded $\mathbf{E}[X]$ and $a \geq 0$, it holds that*

$$\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}$$

Proof. Let $Y = |X - \mathbf{E}[X]|$, then clearly $Y \geq 0$. Therefore

$$\begin{aligned}\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a] = \mathbf{Pr}[Y \geq a] = \mathbf{Pr}\left[Y^2 \geq a^2\right] &\leq \frac{\mathbf{E}\left[Y^2\right]}{a^2} \\ &= \frac{\mathbf{E}\left[(X - \mathbf{E}[X])^2\right]}{a^2} = \frac{\mathbf{Var}[X]}{a^2}.\end{aligned}$$

□

Example 7 (Coupon Collector Revisited)**.** *We apply Chebyshev's inequality to the coupon collector problem. Assuming the notation before, we have*

$$\mathbf{Pr}[X \geq nH_n + t] \leq \mathbf{Pr}[|X - \mathbf{E}[X]| \geq t] \leq \frac{\mathbf{Var}[X]}{t^2}.$$

Recall that the variable X_i indicates the number of draws to get a new coupon while there are i coupons in hands. For distinct i and j, X_i and X_j are independent. Then

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1} \mathbf{Var}[X_i].$$

For $i \in \{0, 1, \ldots, n-1\}$, $X_i \sim \mathtt{Geom}\left(\frac{n-i}{n}\right)$, so we have

$$\mathbf{Var}[X_i] = \frac{1 - \frac{n-i}{n}}{\left(\frac{n-i}{n}\right)^2} = \frac{i \cdot n}{(n-i)^2} \leq \frac{n^2}{(n-i)^2}.$$

It remains to bound $\sum_{i=0}^{n-1} \frac{1}{(n-i)^2} = \sum_{i=1}^{n} \frac{1}{i^2}$. Note that

$$\sum_{i=1}^{n} \frac{1}{i^2} \leq 1 + \int_1^\infty \frac{\mathrm{d}x}{x^2} = 2.$$

Therefore, we have $\mathbf{Var}[X] \leq 2n^2$ and $\mathbf{Pr}[X \geq nH_n + t] \leq \frac{2n^2}{t^2}$. The probability that we need to draw the coupon for more than $\sqrt{200}n + nH_n$ times is less than 0.01.

The bound obtained by Chebyshev's inequality is much tighter than that via Markov's inequality where in order to obtain the same confidence, one needs to choose $t = \Theta(n \log n)$.

References

[1] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019. 1