概率论 **Spring 2024**

第 2 节: 随机变量 期望

Lecturer: 尹一通 *Scribes:* 张桄玮

§1 随机变量

随机变量一开始希望刻画那些“取值看似是随机”的变量. 但是, 在更加严格地定义它们之前, 我们就无法定义什么叫相同的随机变量. 比如 X, Y 是两个掷骰子的结果, 请问 X^2, XY 是相同的随机变量吗? $2X, X+Y$ 呢? 在给出定义之前, 我们不好回答.

随机变量还可能是

- 连续抛硬币, 直到正面朝上为止的次数;
- 从 M 个白球和 $N-M$ 个黑球中 (有/无放回) 取出 n 个球的白球数目;
- n 个顶点, 任意两点之间以概率 p 产生一条边的随机图的最小染色数;
- $[0,1]$ 中随机取一个数它的值.

我们先从最简单的例子看一看.

例子 1.1 (掷骰子)**.** 投掷一枚骰子, 定义 $X \in \Omega$ 为掷出来的结果, $Y \in \{0,1\}$ 表示它的奇偶性.

我们有如下的观察:

Ω中的样本	X的值	Y的值
1 点	1	1
2 点	2	0
3 点	3	1
4 点	4	0
5 点	5	1
6 点	6	0

可以发现, 随机变量只不过是把样本空间里面的元素映射到了实数.

好比说投掷两枚骰子, 样本空间是 $\Omega \times \Omega$, 如果这随机变量是两次骰子的和, 那就是把这空间的每一个元素和一个实数相对应.

刚刚我们从样本空间 Ω 来理解了这映射. 那么这映射应当满足一些约束条件. 具体地, 需要与事件集合 Σ 以及概率律 $\Pr$ 有些联系. 所以我们给出下面的定义.

定义 1.1 (随机变量)**.** 给出一概率空间 $(\Omega, \Sigma, \Pr)$, 其上的**随机变量 (random variable, r.v.)** 是一个函数 $X : \Omega \to \mathbb{R}$, 满足 $\forall x \in \mathbb{R}, \{w \in \Omega : X(\omega) \leq x\} \in \Sigma$.

这定义实际上说的是, 任给我一个实数值, 样本空间中那些满足 $X(\omega) \leq x$ 的 ω 构成的集合要在事件集合中. 比如, 我们的 Σ-可测的事件集天然满足这一性质.

当定义看上去很难懂的时候，想办法把它读出来.

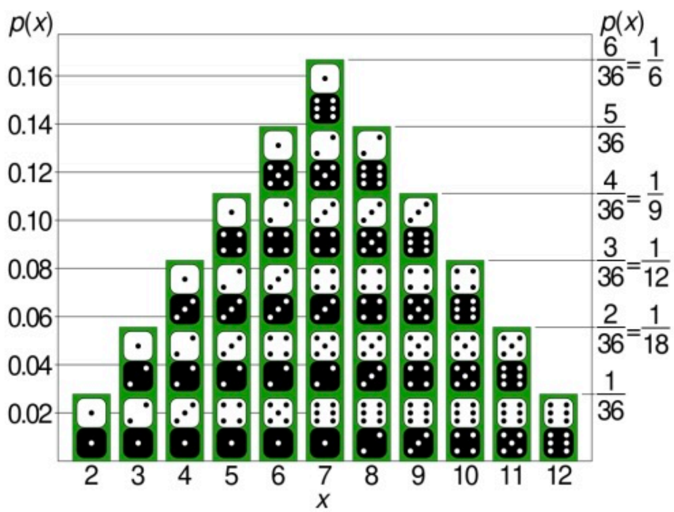
后续, 为了方便起见, 我们引入一些简化记号:

- $X \leq x (x \in \mathbb{R})$ 表示事件集合 $\{\omega \in \Omega : X(\omega) \leq x\}$;
- $X > x (x \in \mathbb{R})$ 表示事件集合 $\{\omega \in \Omega : X(\omega) > x\}$;
- $X \in S (S \subset \mathbb{R}$, 且是由有限个交、并生成的) 表示事件 $\{\omega \in \Omega : X(\omega) \in S\}$.

由于实数的复杂性, 我们才特意规定了上面的 S 是由有限个交、并生成出来的集合. (不然你可能构造出千奇百怪的东西). 对于离散的随机变量 $X : \Omega \to \mathbb{Z}$ 而言, $X \in S$ 中的限制条件就变成了 $S \subset Z$ 了.

1.1. 随机变量的分布. 既然随机变量是一个从样本空间到实数上的映射, 自然需要一个方法来直观地表述这一映射. 由于随机变量的值域通常是可以观测的, 在下面看到表示随机变量的过程中, 通常拿这一值域当做自变量. 因变量就是对应值域的可能的 Ω 中元素的集合.

令 X 为两个独立的掷骰子得到的点数之和. 于是我们可以画出这样的图像.



这些集合如果这样画上去还是有些太麻烦了. 我们干脆把这些集合 (根据定义保证会在概率空间的事件集里面) 塞到概率律 Pr 中, 这样我们又得到了一个实数. 于是, 便可以用中学学过的处理 $f : \mathbb{R} \to \mathbb{R}$ 的手段来直观表示 (也就是上图右侧的数字).

上述考量自然地揭示了分布的想法. 在离散的情形下, 我们可以使用概率质量函数定义分布.

定义 1.2 (概率质量函数)**.** 随机变量 $X : \Omega \to \mathbb{R}$ 被叫做离散的, 如果 $X(\Omega)$ 是可数的.

对于一离散的随机变量 X, 其**概率质量函数 (probability mass function, pmf)**$p_X : \mathbb{R} \to [0, 1]$ 定义做

$$P_X(x) = \Pr(X = x).$$

但是, 在连续的情形, 单单盯着一个点谈论概率往往是没有意义的 (因为任何一个点发生的概率都是 0, 而 $0 \cdot \infty$ 是未定式, 不会违反加起来为 1 的限制). 因此, 我们通常使用累积分布 (看 $x \leq X$ 的概率) 来描述一个分布.

定义 1.3 (累积分布函数)**.** 随机变量 X 的**累计分布函数 (cumulative distribution function)**(或简单叫做分布函数) 是一个映射: $F_X : \mathbb{R} \to [0, 1]$, 对应法则为

$$F_X(x) := \Pr(X \leq x).$$

关于 X 的一切概率都可以从 F_X 中得到. 因此, 一旦能够确定累积分布函数 $F_X(x)$ 在后续中, 我们实际上便不再需要概率空间.

现在就可以讨论什么叫两个随机变量相同了.

定义 1.4. 称两个随机变量 X, Y **相同分布 (identically distributed)**. 如果两个变量具有相同分布, 那么 $F_X = F_Y$.

从上面的定义来看, 对于离散的情况, 随机变量 X 的 CDF 就是 $F_X(y) := \sum_{x \leq y} p_X(x)$. 如果我们能把 CDF 表示成某个可积的积分的形式: $F_X(y) = \Pr(X \leq y) = \int_{-\infty}^{y} f_X(x)dx$, 我们就称它为连续随机变量.

实际上这积分并不一定 Riemann 可积, 有可能是 Lebesgue 可积

上面的叙述, 实际上暗示了有些随机变量既不是离散的, 也不是连续的. 但是这里先不谈这些.

累积分布函数还满足一些 (显而易见的) 性质.

性质 1.1 (累积分布函数的性质)**.** 累积分布函数满足如下的性质:

1. 单调性: $\forall x, y \in \mathbb{R}$, 如果 $x \leq y$, 那么 $F_X(x) \leq F_Y(y)$.
2. 有界性: $\lim_{n \to -\infty} F_X(x) = 0$, 而且 $\lim_{x \to \infty} F_X(x) = 1$.

可以说, 这两条分布的性质直接继承了概率律函数 Pr 的属性.

1.2. 独立性. 我们上次在介绍概率空间的时候说了事件的独立性. 描述随机变量号称可以 "不用再考虑概率空间" 的分布函数当然也要提一提.

定义 1.5 (随机变量的独立性)**.** 对于随机变量 $X_1, X_2, ..., X_n$, 我们说它们 (互相) 独立, 当且仅当任意的 $x_1, x_2, ..., x_n$, 有

$$p_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \Pr(X_1 \leq x_1 \cap \cdots \cap X_n \leq x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

当然上述定义并不好用 (!) 对于接下来我们要考虑的离散型随机变量, 定义就简化成了

- 两个离散随机变量 X, Y 是独立的, 当且仅当对于任何的两个数值 x, y, 有 $X = x, Y = y$ 两个事件也是独立的.
- 如果有一组离散随机变量 $X_1, X_2, ..., X_n$, 称它们是随机的, 当且仅当对于任意的 $x_1, x_2, ..., x_n$ 而言, 事件 $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ 是独立的. 换言之, 就是 PMF 可以直接乘起来. 即

$$p_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

1.3. 随机向量. 有了一个随机变量, 可否考虑一系列随机变量? 即, 我们有 $X_1, X_2, X_3, ...$, 他们都是随机变量. 为了方便起见, 干脆把它们放在一起作为一个整体来考虑. 叫做随机向量.

定义 1.6 (随机向量)**.** 给定一个概率空间 $(\Omega, \Sigma, \Pr)$, 某**随机向量 (random vector)**$\boldsymbol{X}$ 记作 $\boldsymbol{X} := (X_1, \ldots, X_n)$, 其中每一个元素 X_i 都是定义在概率空间 $(\Omega, \Sigma, \Pr)$ 上的随机变量.

例子 1.2. 比如我们有两个随机变量, X, Y. X 可以取 x_1, x_2, x_3, x_4; Y 可以取 y_1, y_2, y_3. 那么 (X, Y) 就是一个随机向量.

如需寻求一个直观的表示, 对于离散的随机变量而言, 当然可以枚举每一种可能的组合. 并且定义联合质量函数.

定义 1.7 (联合质量函数)**.** 对于离散的随机变量而言, 其**联合质量函数 (joint mass function)** 被定义作

$$p_X(x_1, \ldots, x_n) = \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n)$$

例子 1.3. 对于上例子, 若把全部的可能组合写出来, 实际上可以画出一张表. 比如

$Y \backslash X$	x_1	x_2	x_3	x_4
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$
y_3	$\frac{9}{32}$	0	0	0

同样对于非离散的情形. 仍然可以通过联合累计密度函数表示.

定义 1.8 (联合累积密度函数)**.** 一个随机向量的**联合累积密度函数 (joint CDF)** 是 $F_X : \mathbb{R}^n \to [0,1]$, 其对应法则为

$$F_X(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1 \cap \cdots \cap X_n \leq x_n)$$

如果有选择性地忽略某一变量, 即按照行列相加, 会得到边缘分布. 如下所示.

$Y \backslash X$	x_1	x_2	x_3	x_4	$p_y(y) \downarrow$
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{8}{32}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{15}{32}$
y_3	$\frac{9}{32}$	0	0	0	$\frac{9}{32}$
$p_X(x) \rightarrow$	$\frac{16}{32}$	$\frac{8}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{32}{32}$

这样便得到了一个只含有 X(或 Y) 的随机变量的质量函数, 将二维化为了一维. 像这样的分布叫做边缘分布, 因为其总是在表格的边上.

定义 1.9 (边缘分布)**.** 对于一个随机向量 $(X_1, \ldots, X_n)$, 其边缘分布由

$$p_{X_i}(x_i) := \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} p_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n)$$

定义.

§2 离散随机变量

接下来我们主要考虑整数值的随机变量 $X : \Omega \to \mathbb{Z}$. 根据先前的定义, 其 PMF $p_X : \mathbb{Z} \to [0,1]$ 由 $p_X(k) = \Pr(X = k)$ 给出. 我们可以把这样的表示解读做

- 柱状图: p_X 描绘了概率分布的直方图.
- 向量: 如果 $R :=$ 随机变量 X 的值域, 那么 $p_X \in [0,1]^R$ 可以看做一个向量, 满足 $||p_X(x)||_1 = 1$.

一个随机变量的函数也是一个随机变量. 就像我们创造复合函数一样. 譬如我们有 $Y = f(X)$, 那么

$$p_Y(y) = \sum_{x:f(x)=y} p_X(x)$$

2.1. 若干离散随机变量及其分布. 下面来看若干离散随机变量及其分布.

a) Bernoulli 随机变量 Bernoulli 试验只可能产生两个结果. 这个实验的结果只可能是是 "成功" 或者 "失败". 倘若将 "成功" 记作 1, "失败" 记作 0, 并用随机变量表示之, 便得到了 Bernouli 随机变量.

定义 2.1 (Bernouli 随机变量)**.** Bernouli 随机变量 X 是在 $\{0,1\}$ 中取值的随机变量, 满足

$$p_X(k) = \Pr(X = k) = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \end{cases}$$

其中 $p \in [0,1]$ 是参数.

Bernouli 随机变量通常用于一个事件发生与否的**指示器 (indicator)**. 例如

$$X = I(A) = \begin{cases} 1 & \text{如果} A \text{ 发生} \\ 0 & \text{否则} \end{cases} \qquad \text{是一个参数为 } \Pr(A) \text{ 的 Bernouli r.v.} \Pr(A)$$

b) 几何分布 不断地投一枚硬币, 第一次抛出正面的时候, 我们抛出了几次? 这也是一个随机变量. 其分布类似几何分布.

定义 2.2 (几何分布随机变量)**.** 随机变量 X 是第一次独立同分布的 Bernoulli 实验成功的时候, 做实验的次数. 其取值范围是 $\{1,2,3,...,\}$, 其分布是

$$p_X(k) = \Pr(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

其中 p 是一次 Bernouli 实验成功的概率, 是一个参数.

若一随机变量服从这样的分布, 简单记作 $X \sim \text{Geo}(p)$.

需要注意的是, 几何分布没有记忆性. 也就是说, 无论从何时开始, 都和从第一次开始的分布无异.

性质 2.1. 几何随机变量 $X \sim \text{Geo}(p)$ 不具有记忆性. 即对 $k \geq 1, n \geq 0$,

$$\Pr(X = k + n \mid X > n) = \Pr(X = k).$$

Proof. 只要回到条件概率的定义证明即可.

$$\begin{aligned} \Pr(X = k + n \mid X > n) &= \frac{\Pr(X = k + n)}{\Pr(X > n)} = \frac{(1-p)^{n+k-1}p}{\sum_{k=n}^{\infty}(1-p)^k p} \\ &= \frac{(1-p)^{k-1}p}{\sum_{k=0}^{\infty}(1-p)^k p} = (1-p)^{k-1}p \end{aligned}$$

□

顺带一提, 几何分布是值域为 $\{1,2,...\}$ 的离散随机变量中唯一一个没有记忆性的分布.

c) 二项分布 二项分布是 n 次抛硬币中, 抛出正面的个数的分布.

定义 2.3 (二项随机变量)**.** 在 n 次独立的参数为 p 的 Bernouli 实验中, 成功的次数称为**二项随机变量 (Binomial r.v.)**. 其分布称为**二项分布 (binomial distribution)**, 若一随机变量 X 服从二项分布, 则可简记为 $X \sim B(n,p)$(或 $X \sim \text{Bin}(n,p)$).

二项分布的取值范围在 $\{0,1,...,n\}$, 并且有

$$p_X(k) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^k, \quad k = 0,1,\dots,n$$

d) Poisson 分布 在实际生活中, 通常面对大部分的情况有二项分布的 $n \to \infty$, 但是 $np = \lambda$ 是一个常数. 通过计算, 我们知道

$$p_{\text{Bin}(n,\lambda/n)}(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n}{n}\frac{n-1}{n}\dots\frac{n-k+1}{n}\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \approx \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

这就是 Poisson 分布.

定义 2.4 (Possion 分布)**.** 某一 Possion 随机变量 X 取值范围为 $\{0,1,2...\}$, 其服从的分布为

$$p_X(k) = \Pr(X=k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0,1,2,\dots$$

可以验证, 这是一个良定义 (没有与定义中描述的相冲突) 的分布. 因为它满足 $\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = 1$. 如果某个随机变量服从 Poisson 分布, 可以简单记作 $X \sim \text{Pois}(\lambda)$.

另外指出, 独立的 Possion 随机变量的和也是 Possion 随机变量. 这是由于二项分布的类比: $X \sim \text{Bin}(n_1,p), Y \sim \text{Bin}(n_2,p) \Longrightarrow X+Y \sim \text{Bin}(n_1+n_2,p)$

我们来证明相互独立的 X,Y, $X \sim \text{Pois}(\lambda_1), Y \sim \text{Pois}(\lambda_2) \Longrightarrow X+Y \sim \text{Pois}(\lambda_1+\lambda_2)$.

Proof.

$$\begin{aligned}
p_{X+Y}(k) &= \Pr(X+Y=k) = \sum_{i=0}^{k}\Pr(X=i \cap Y=k-i) = \sum_{i=0}^{k}p_X(i)p_Y(k-i) \\
&= \sum_{i=0}^{k}\frac{\mathrm{e}^{-\lambda_1}\lambda_1^i}{i!}\frac{\mathrm{e}^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^i\lambda_2^{k-i} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}
\end{aligned}$$

□

e) 多项式分布 这是对二项分布的推广. 假设有 n 个球扔进了 m 个桶里面, 每个球扔进哪个桶里面是随机的, 满足第 i 个桶接收到这个球的概率是 $p_i(p_1+p_2+...+p_m=1)$. 我们用一组随机变量 $(X_1,X_2,...,X_n)$ 表示第 i 个桶恰好收到的 X_i 个球. 那么 $(X_1,X_2,...,X_m)$ 从 $(k_1,k_2,...,k_m) \in \{0,1,...,n\}^m$ 中取值, 且 $k_1+k_2+...+k_m=n$. 且其分布 $p_{(X_1,\dots,X_m)}(k_1,\dots,k_m)$ 应该满足什么样的规律呢?

我们首先从 n 个球里面选出 k_1 个 (每个的概率为 p_1), 然后从 $n-k_1$ 当中选出 k_2 个 (每个的概率是 p_2), ..., 就得到了如下的式子:

$$\underbrace{\binom{n}{k_1}p_1^{k_1}}_{\text{第一个桶放入}k_1\text{个球的概率}}\underbrace{\binom{n-k_1}{k_2}p_2^{k_2}}_{\text{第二个桶放入}k_2\text{个球的概率}}\binom{n-k_1-k_2}{k_3}p_3^{k_3}\cdots\binom{n-k_1-k_2-\cdots-k_{m-1}}{k_m}p_m^{k_m}$$

$$\xlongequal{\text{组合数定义}}\frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)}\cdots\frac{(n-k_1-\cdots-k_{m-1})!}{k_m!(n-k_1-k_2-\cdots-k_{m-1}-k_m)!}p_1^{k_1}\cdots p_m^{k^m}$$

$$\xlongequal{\text{约分}}\frac{n!}{k_1!\cancel{(n-k_1)!}}\frac{\cancel{(n-k_1)!}}{k_2!\cancel{(n-k_1-k_2)}}\cdots\frac{\cancel{(n-k_1-\cdots-k_{m-1})!}}{k_m!\underbrace{\cancel{(n-k_1-k_2-\cdots-k_{m-1}-k_m)!}}_{\text{等于 }0}}p_1^{k_1}\cdots p_m^{k^m}$$

$$=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}\cdots p_m^{k^m}$$

所以我们得到:

$$p_{(X_1,\ldots,X_m)}(k_1,\ldots,k_m)=\Pr\left(\bigcap_{i=1}^m(X_i=k_i)\right)=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}$$

实际上, 刚刚推导的正是多重组合数. 对于多重组合数, 有时候也可以记 $\binom{n}{k_1,k_2,k_3,\cdots,k_m}$ 作 "把 $k_1+\cdots+k_m$ 个球放进 m 个桶里面, 第一个桶里面有 k_1 个球, 第二个桶里有 k_2 个球, ..., 第 m 个桶里有 k_m 个球的方案数. "

其计算公式是

$$\binom{n}{k_1,k_2,k_3,\cdots,k_m}=\frac{n!}{k_1!k_2!\cdots k_m!}.$$

如果考察这分布的边缘分布, 任意的 $1\le i\le m$, 边缘分布 X_i 服从 $\mathrm{Bin}(n,p_i)$.

f) Possion 高维分布 将多项分布推向极限, 如果 $(Y_1,...,Y_m)$ 中的每个随机变量都服从独立的 Possion 分布 $(Y_i\sim\mathrm{Pois}(\lambda_i))$, 并且 $\lambda_i=np_i$, 那么实际上高维的 Possion 分布和多项分布是同分布的.

性质 2.2. $\boldsymbol{X}=(X_1,...,X_n)$ 遵循参数为 $m,n,p_1+p_2+...+p_m$ 的多维分布. $\boldsymbol{Y}=(Y_1,Y_2,...,Y_n)$ 满足独立的 $\forall i,Y_i\sim\mathrm{Pois}(\lambda_i),\lambda_i=np_i$. 如果 $\sum_{i=1}^m Y_i=n$, 那么 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 同分布.

Proof.

$$\Pr\left[(Y_1,\ldots,Y_m)=(k_1,\ldots,k_m)\mid Y_1+\cdots+Y_m=n\right]=\left(\prod_{i=1}^m\frac{\mathrm{e}^{-np_i(np_i)^{k_i}}}{k_i!}\right)/\left(\frac{\mathrm{e}^{-n}n^n}{n!}\right)$$

$$=\frac{n!}{k_1!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m}=\Pr\left[(X_1,\ldots,X_m)=(k_1,\ldots,k_m)\right]$$

□

2.2. 构造随机变量的方法. 事实上, 从上面可以看出, 一般有两种方法构造随机变量. 第一种是对已有的随机变量做一个函数映射. 如已经知道了 $X_1, X_2, ..., X_n$, 可以通过 $Y = f(X_1, X_2, \cdots, X_n)$ 构造这一新的随机变量 Y. 如二项分布随机变量就是 n 个独立的 Bernoulli 随机变量求和得到的.

另一个方法是为考虑随机变量序列 $X_1, X_2, ..., X_T$ 的**停止时间 (stopping time)T**, 把 T 当做随机变量. 如几何分布中考虑的 "Bernoulli 试验的第一次成功的时候, 做实验的次数. "

定义 2.5 (随机变量的停止时间)**.** 对于随机变量序列 $X_1, X_2, ...$, 我们说这组随机变量的**停止时间 (stopping time)** 是 T(也是一个随机变量), 当且仅当对于所有的 $t \geq 1, T = t$ 仅由 $X_1, X_2, ..., X_t$ 决定.

上述定义表示我们不再考虑停止时间 X_t 之后的那些元素造成的影响. 自然有在某一随机变量之后停止之意.

a) 独立随机变量和的分布 如上例, 独立随机变量的和的分布非常常见. 我们可以为它导出一个通用的公式. 比如 X, Y 是两个独立的离散随机变量, 我们需要得到 $X + Y$ 的分布.

只要求出 $X + Y = 1$ 的概率, $X + Y = 2$ 的概率, ..., 我们就可以得到这一随机变量的分布. 我们若要求出 $X + Y = z$ 的概率, 只需要使用全概率公式, 先固定住 $X = x$, 然后使用独立性, 最后把所有可能的 x 加起来.

$$\begin{aligned} p_{X+Y}(z) = \Pr(X + Y = z) &= \sum_x \Pr(X = x \cap Y = z - x) \\ &= \sum_x p_X(x) p_Y(z - x) = \sum_y p_X(z - y) p_Y(y) \end{aligned}$$

如果我们把这个函数看做一个整体, 而不是看做输入某个固定的值之后如何计算, 我们便说: "p_{X+Y} 的 PMF 就是 p_X 的 PMF 和 p_Y 的 PMF 做了一个**卷积 (convolution)**". 记作

$$p_{X+Y} = p_X * p_Y$$

例子 2.1. 对于二项分布而言是若干个 i.i.d. Bernouli 随机变量的和. 假设一次成功概率为 p, 那么根据上述的公式照样可以推出二项分布:

$$\begin{gathered} p_{X_1+\cdots+X_n}(k) = p \cdot p_{X_1+\cdots+X_{n-1}}(k-1) + (1-p) \cdot p_{X_1+\cdots+X_{n-1}}(k) \\ = \binom{n-1}{k-1} p^k (1-p)^{n-k} + \binom{n-1}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{gathered}$$

b) 另一个停止时间随机变量的例子 我们来考虑负二项分布. 它考虑的是成功概率为 p 的 iid. Bernouli 实验中成功 r 次这一段时间中失败的次数. 先记为随机变量 X.

定义 2.6 (负二项分布)**.** 负二项随机变量 X 在 $\{0, 1, 2, ...\}$ 取值, 其分布为

$$p_X(k) = \Pr(X = k) = \binom{k+r-1}{k} (1-p)^k p^r = (-1)^k \binom{-r}{k} (1-p)^k p^r, k = 0, 1, 2, ...$$

其中 r, p 是参数.

§3 期望及其计算

回顾中学定义的期望, 用随机变量的语言重述如下.

定义 3.1 (期望)**.** 一个离散型随机变量的**期望 (expectation)** 定义做

$$\mathbb{E}[X]=\sum_x x p_X(x)$$

其中, p_X 表示 X 的 pmf, 求和指标 x 取遍 $p_X(x)>0$ 的 x.

需要注意的是, 期望不总是存在. 如对于正整数 k, 定义 $p_X\left(2^k\right)=2^{-k}$. 其期望并不是一个有限数!

下面将介绍三种常见的期望计算方法. 如果我们发现期望很难算, 那么我们可以采取估计的方式. 在后续的内容中, 将说明计算的期望有什么含义.

3.1. 直接计算. 根据期望的定义, 我们可以直接计算.

例子 3.1 (指示器变量的期望)**.** 对于一个成功概率为 p 的 Bernouli 随机变量, 其期望为

$$\mathbb{E}[X]=0\cdot(1-p)+1\cdot p=p$$

如果指示某事件 A 发生与否的随机变量 $I(A)$, $I(A)$ 取得 1 表示事件 A 发生, 否则表示事件没有发生. 那么

$$\mathbb{E}[X]=0\cdot\Pr(A^c)+1\cdot\Pr(A).$$

这例子看上去简单, 但是后续一些性质可以把复杂的任务拆成如此简单的求期望的内容.

例子 3.2 (Possion 随机变量)**.** 如果 $X\sim\operatorname{Pois}(\lambda)$, 那么

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k\geq 0} k\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k\geq 1}\frac{\mathrm{e}^{-\lambda}\lambda^k}{(k-1)!} \\
&= \sum_{k\geq 0}\frac{\mathrm{e}^{-\lambda}\lambda^{k+1}}{k!}=\lambda\sum_{k\geq 0}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \\
&= \lambda
\end{aligned}$$

既然随机变量经过某个函数之后也是随机变量, 那么这个新的随机变量的期望是什么? 下面的命题展示了随机变量的函数的期望怎么求.

性质 3.1. 对于 $f:\mathbb{R}\to\mathbb{R}$, X 和 $\boldsymbol{X}=\{X_1,X_2,...,X_n\}$, 我们有

- $\mathbb{E}[f(X)]=\sum_x f(x)p_X(x)$
- $\mathbb{E}\left[f\left(X_1,\ldots,X_n\right)\right]=\sum_{(x_1,\ldots,x_n)} f\left(x_1,\ldots,x_n\right)p_X\left(x_1,\ldots,x_n\right)$

Proof. 令 $Y=f(X_1,X_2,...,X_n)$, 那么

$$\begin{aligned}\mathbb{E}\left[f\left(X_1,\ldots,X_n\right)\right]&=\sum_y y\Pr(Y=y)=\sum_y y\sum_{(x_1,\ldots,x_n)\in f^{-1}(y)}\Pr\left(\left(X_1,\ldots,X_1\right)=\left(x_1,\ldots,x_n\right)\right)\\&=\sum_{(x_1,\ldots,x_n)}f\left(x_1,\ldots,x_n\right)\Pr\left(\left(X_1,\ldots,X_1\right)=\left(x_1,\ldots,x_n\right)\right)\\&=\sum_{(x_1,\ldots,x_n)}f\left(x_1,\ldots,x_n\right)p_X\left(x_1,\ldots,x_n\right)\end{aligned}$$

□

3.2. 期望的线性性. 下面的定理叙述的是, 无论两个随机变量是否独立, 期望的和等于和的期望.

定理 3.2 (期望的线性性)**.** 对于任意的 $a,b\in R$, 以及随机变量 X,Y 有

- $\mathbb{E}[aX+b]=a\mathbb{E}[X]+b$;
- $\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$

Proof.

$$\mathbb{E}[aX+b]=\sum_x(ax+b)p_X(x)=a\sum_x xp_X(x)+b\sum_x p_X(x)=a\mathbb{E}[X]+b$$

$$\begin{aligned}\mathbb{E}[X+Y]&=\sum_{x,y}(x+y)\Pr((X,Y)=(x,y))\\&\overset{x,y}{=}\sum_x x\sum_y\Pr((X,Y)=(x,y))+\sum_y y\sum_x\Pr((X,Y)=(x,y))\\&=\sum_x x\Pr(X=x)+\sum_y y\Pr(Y=y)=\mathbb{E}[X]+\mathbb{E}[Y]\end{aligned}$$

□

这可以推广到线性函数 f 和一组随机变量 $X_1,...,X_n$, 同样有 $\mathbb{E}\left[f\left(X_1,\ldots,X_n\right)\right]=f\left(\mathbb{E}\left[X_1\right],\ldots,\mathbb{E}\left[X_n\right]\right)$, 而且不用关心它们这些随机变量的相关性.

例子 3.3 (二项分布的期望)**.** 上文提到, 要从定义计算二项随机变量 $X\sim\text{Bin}(n,p)$ 的期望, 必须计算

$$\mathbb{E}[X]=\sum_{k=0}^n k\binom{n}{k}p^k(1-p)^{n-k}.$$

但是, 由于二项分布可以被看做一系列随机变量的和 $X=X_1+\cdots+X_n$, 其中每一个 X_i 是 iid. Bernouli 随机变量. 于是根据期望的线性性有:

$$\mathbb{E}[X]=\mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_n\right]=np$$

例子 3.4 (几何分布的期望)**.** 要从定义计算几何分布随机变量 $X\sim\text{Geo}(p)$, 必须计算

$$\mathbb{E}[X]=\sum_{k\geq1}k(1-p)^{k-1}p.$$

但是, 可以用另一种方法看这问题: 令 $X=\sum_{k>1} I_k$, 其中 $I_k\in\{0,1\}$ 表示头 $k-1$ 次试验是否都失败了, 那么

$$\mathbb{E}[X]=\sum_{k\geq 1}\mathbb{E}\left[I_k\right]=\sum_{k\geq 1}(1-p)^{k-1}=\frac{1}{p}$$

例子 3.5 (负二项分布的期望)**.** 按照定义, 对于服从参数为 r,p 的负二项分布的随机变量 X, 应该计算

$$\mathbb{E}[X]=\sum_{k\geq 1}k\binom{k+r-1}{k}(1-p)^k p^r$$

但是, X 可以看做由若干个 iid. 几何随机变量组成. 即 $X_1,X_2,...,X_r$ 个参数为 p 的 iid. 几何随机变量. 那么 $X=(X_1-1)+\cdots+(X_r-1)$. 因此根据期望的线性性, 有

$$\mathbb{E}[X]=\mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_r\right]-r=r(1-p)/p$$

例子 3.6 (超几何分布的期望)**.** 一个盒子中总共有 N 个球. 其中有 M 个是红球, $N-M$ 个是蓝球. 从中无放回地抽取 n 个球. 假设抽取红球的个数记为随机变量 X, 那么我们说 X 服从参数为 N,M,n 的超几何分布.

对于一个服从参数为 N,M,n 的超几何分布的随机变量 X, 按照定义, 我们应该计算

$$\mathbb{E}[X]=\sum_{k=0}^{n}k\binom{M}{k}\binom{N-M}{n-k}/\binom{N}{n}$$

但是, 定义指示变量 $X_i\in\{0,1\}$ 表示第 i 个红球被抽出来了. 那么, $X=X_1+\cdots+X_M$. 每一个红球被抽出来的概率是 $\binom{N-1}{n-1}/\binom{N}{n}=\frac{n}{N}$. 因此使用期望的线性性, 有

$$\mathbb{E}[X]=\mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_M\right]=\frac{nM}{N}$$

例子 3.7. 一只猴子在胡乱敲键盘 (假设敲击每个字母的概率是均等的). 请问它敲击 m 次键盘之后, 期望出现了多少次 “hamlet”?

更加具体地, 假设我们有一个字母表 Q, 其大小 $|Q|=q$. 定义 $s:=(s_1,\ldots,s_n)\in Q^n$ 表示字母表中由 n 个字母组成的串的集合. 对于我们要匹配的字符串是 $\pi\in Q^k$, 定义随机变量 X 表示 π 在 s 中作为子串出现的次数. 要计算这个问题, 我们可以设置指示变量 $I_i\in\{0,1\}$, 表示 $\pi=(s_i,s_{i+1},\ldots,s_{i+k-1})$. 那么, 我们的随机变量就可以变为 $X=\sum_{i=1}^{n-k+1}I_i$.

根据期望的线性性,

$$\mathbb{E}[X]=\sum_{i=1}^{n-k+1}\mathbb{E}\left[I_i\right]=(n-k+1)q^{-k}$$

可以发现, 出现的次数仅仅与要匹配的字符串的长度以及敲了几次键盘有关.

如果这只猴子不断地敲击键盘, 直到最后的几个字符是 “hamlet” 停止. X 是截止停止时间的时候敲击键盘的次数. 那么这样, 这期望的次数就要和这个串的性质决定了.

例子 3.8 (邮票收集者)**.** 现在去集邮. 假设现在有 n 种邮票, 每次去集邮的时候他会随机等概率给你一张. 在集齐的时候去集邮的次数作为随机变量 X. 请问, 你期望要去多

少次, 才能把所有的邮票集齐?

实际上, 这问题如果用球和盒子的比喻, 可以说作: “向 n 个篮子里面不断一个一个地扔球, 直到所有的篮子都填满, 扔的球的个数作为随机变量 X. ”

可以设随机变量 X_i 表示当现在有且仅有$i-1$ 个非空的篮子的时候扔的球的数目. 实际上, X_i 是参数为 $p_i = 1 - \frac{i-1}{n}$ 的几何随机变量, 且 $X = \sum_{i=1}^{n} X_i$.

那么根据期望的线性性, 有

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n\sum_{i=1}^{n} \frac{1}{i} = nH(n)$$

其中 $H(n)$ 是调和级数.

由于期望是概率的加权平均和, 我们可以枚举每一个事件把它拆做一个一个小事情.

定理 3.3. 对于非负取值在 $\{0, 1, 2, ...\}$ 的随机变量 X, 有

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \Pr[X > k]$$

成立.

Proof. 这是一个双射. 可以交换求和记号证明.

$$\mathbb{E}[X] = \sum_{x\geq 0} x\Pr[X=x] = \sum_{x\geq 0}\sum_{k=0}^{x-1} \Pr[X=x] = \sum_{k\geq 0}\sum_{x>k} \Pr[X=x] = \sum_{k\geq 0} \Pr[X>k]$$

□

Proof. 也可以使用期望的线性性. 定义 $I_k \in \{0, 1\}$ 表示随机变量 $X > k$ 是否成立. 那么, $X = \sum_{k\geq 0} I_k$. 根据期望的线性性, 有

$$\mathbb{E}[X] = \sum_{k\geq 0} \mathbb{E}\left[I_k\right] = \sum_{k\geq 0} \Pr[X > k]$$

□

面对复杂的事件, 难免使用容斥原理. 假设我们有指示事件 A 发生与否的变量 $I(A) \in \{0, 1\}$, 那么可以知道:

- $I\left(A^c\right) = 1 - I(A)$;
- $I(A \cap B) = I(A) \cdot I(B)$.

我们化并为交的容斥原理就可以使用了. 在上一节中说到的证明方法同样适用于这里.

定理 3.4 (容斥原理)**.** 对于 n 个事件 $A_1, A_2, \ldots, A_n$, 有

$$I\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\varnothing \neq S \subseteq \{1,\ldots,n\}} (-1)^{|S|-1} I\left(\bigcap_{i\in S} A_i\right)$$

期望的线性性帮助我们省去了很多很繁杂甚至不可能的计算. 但是, 在面对无穷多个随机变量的时候, 期望的线性性也有其限制和条件.

假设有无穷多个随机变量 $X_1, X_2, \ldots$, 满足级数 $\sum_{i=1}^{\infty} \mathbb{E}\left[|X_i|\right]$ 绝对收敛的时候, 式子 $\mathbb{E}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[X_i\right]$ 才会成立.

一个更加有趣的情况是随机个随机变量的和, 即, 假设 N 是一个非负整值随机变量, 以及有随机变量 $X_1, X_2, \ldots, X_N$, 那么, $\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}[N]\mathbb{E}\left[X_1\right]$ 成立吗?

3.3. 一些不等式.

a) Jensen 不等式 对于一般的非线性函数 f, 以及随机变量 X, 通常不满足 $\mathbb{E}[f(X)] = f(\mathbb{E}[X])$. 但是, 如果我们知道了 f 是凸函数, 根据 Jensen 不等式, 有

$$\begin{aligned} f\text{是凸函数} &\iff f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \\ &\iff \mathbb{E}[f(X)] \ge f(\mathbb{E}[X]) \end{aligned}$$

这可以使用 Taylor 展开以及中值定理证明.

b) 期望的单调性 对于随机变量 X, Y, 以及 $c \in \mathbb{R}$,

- 如果 $X \le Y$ a.s.(almost surely, 即 $\Pr(X < Y) = 1$), 那么 $\mathbb{E}[X] \le \mathbb{E}[Y]$.
- 如果 $X \le c(X \ge c)$ a.s, 那么 $\mathbb{E}[X] \le c(\mathbb{E}[X] \ge c)$.
- $\mathbb{E}[|X|] \ge |\mathbb{E}[X]| \ge 0$.

Proof.

$$\begin{aligned} \mathbb{E}[X] &= \sum_x x \Pr(X = x) = \sum_x x \sum_y \Pr((X,Y) = (x,y)) \\ &= \sum_x x \sum_{y \ge x} \Pr((X,Y) = (x,y)) = \sum_y \sum_{x \le y} x \Pr((X,Y) = (x,y)) \\ &\le \sum_y \sum_{x \le y} y \Pr((X,Y) = (x,y)) \le \sum_y y \Pr(Y = y) = \mathbb{E}[Y] \end{aligned}$$

□

c) 平均原理 平均原理说的是肯定有元素大于等于均值的元素. 即 $\Pr(X \ge \mathbb{E}[X]) > 0$. 不然均值就无法维持.

§4 条件分布与条件期望

在取条件的时候, 实际上是更换我们讨论的样本空间. 因此, 当然可以在这个新的空间上面再次考察其分布. 比如得到其 pmf(概率质量函数).

定义 4.1 (条件分布)**.** 离散型随机变量 X 在 A 发生的条件下的概率密度函数 (pmf) 记作 $p_{X|A} : \mathbb{Z} \to [0,1]$. 其对应法则为

$$p_{X|A}(x) = \Pr(X = x \mid A)$$

需要注意的是, $(X|A)$ 也是一个随机变量. 其分布完全由 pmf $p_{X|A}$ 表示. 既然如此, 它自然可以用于期望的计算. 如 $\mathbb{E}[X \mid A]=\sum_x x \Pr(X=x \mid A)$. 以及满足期望的各种性质 (如最常用的, 线性性).

定义 4.2 (条件期望)**.** 离散型随机变量 X 在事件A 发生的条件下的**条件期望 (conditional expectation)** 定义做

$$\mathbb{E}[X \mid A]=\sum_x x \Pr(X=x \mid A)$$

为了满足概率空间的定义, 还需满足

- $\Pr(A)>0$;
- 级数 $\sum_x x \Pr(X=x \mid A)$ 绝对收敛.

定义 4.3 (条件期望)**.** 对于两个随机变量X, Y, 条件期望 $\mathbb{E}[X \mid Y]$ 是一个随机变量 $f(Y)$, 其分布为当 $Y=y$ 的时候

$$f(y)=\mathbb{E}[X \mid Y=y]$$

这样的定义自然可以推广到多于 2 个变量的情况.

为什么要引入条件期望? 许多时候我们希望对样本空间做划分, 在那些划分过后的样本空间中, 往往就好求解期望了. 如果把每一个小块的期望值加起来, 结果上就等于全空间的期望了. 这就是下一个定理阐述的全期望定理.

定理 4.1 (全期望定理)**.** 假设随机变量 X 的期望存在, $B_1, B_2, \ldots, B_n$ 是样本空间 Ω 的一个划分, 满足 $\Pr\left(B_i\right)>0, \forall i$. 那么有

$$\mathbb{E}[X]=\sum_{i=1}^n \mathbb{E}\left[X \mid B_i\right] \Pr\left(B_i\right)$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x \Pr(X=x)=\sum_x x \sum_{i=1}^n \Pr\left(X=x \mid B_i\right) \Pr\left(B_i\right) \\
&= \sum_{i=1}^n \Pr\left(B_i\right) \sum_x x \Pr\left(X=x \mid B_i\right)=\sum_{i=1}^n \mathbb{E}\left[X \mid B_i\right] \Pr\left(B_i\right)
\end{aligned}$$

□

有了这样的定理就会得出一个有趣的结论: $\mathbb{E}[\mathbb{E}[X \mid Y]]=\mathbb{E}[X]$. 这是因为

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \textstyle\sum_y \mathbb{E}[X \mid Y=y] \quad \Pr(Y=y) \\
&= \mathbb{E}[X]
\end{aligned}$$

4.1. 快速排序的期望运行时间. 我们在《算法导论》的课程上了解了如下的随机快速排序算法. 使用自然语言大概可以看做算法 1.

我们声称: 每一次从元素中独立地随机选取基准元素, 那么对于任意的输入, 快速排序比较的期望次数为 $2n \lg n+O(n)$.

Algorithm 1 随机快速排序算法

输入: 待排序的数组 $S=[x_1,x_2,\cdots,x_n]$
输出: 排序后的数组 S.

- 如果 S 只有一个或者零个元素, 返回 S. 否则继续.
- 随机选择 S 中的元素 s 作为基准元素.
- 把 S 分为两个小的列表 S_1,S_2. 其中, 任何一个 S_1 中的元素比 s 要小, 任何一个 S_2 中的元素比 s 要大.
- 对 S_1,S_2 进行快速排序.
- 返回列表 $[S_1,x,S_2]$.

Proof. 设 $y_1,y_2,\cdots,y_n$ 是输入值 $x_1,x_2,\cdots,x_n$ 按照升序排列的结果. 我们定义 $X_{ij}(i<j)$ 是一个随机变量. 如果在算法执行的某一时刻 y_i 和 y_j 发生了比较, X_{ij} 取值为 1, 否则为 0. 那么比较的总次数满足

$$X=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}X_{ij}$$

根据期望的线性性, $\mathbb{E}X=\mathbb{E}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}X_{ij}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}X_{ij}$.

由于 X_{ij} 只能取 0 和 1, 是指示变量, $\mathbb{E}X_{ij}$ 是 X_{ij} 等于 1 的概率.

什么时候 y_i 和 y_j 会发生比较呢? 我们发现 y_i 和 y_j 发生比较, 当且仅当 y_i 或 y_j 是从集合 $Y_{ij}=\{y_i,y_{i+1},\cdots,y_{j-1},y_j\}$ 中选取的一个基准元素. 否则, 他们会被分在不同的子列表中, 因而不会比较.

由于我们的基准元素是独立选取的, 因此 y_i 和 y_j 是从 Y_{ij} 中选取的一个基准元素的概率, 也就是 X_{ij} 取 1 的概率, 是 $2/(j-i+1)$. 也就是

$$\begin{aligned}
\mathbb{E}X&=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{2}{j-i+1}\\
&\xlongequal{k:=j-i+1}\sum_{i=1}^{n-1}\sum_{k=2}^{n-i+1}\frac{2}{k}=\sum_{k=2}^{n}\sum_{i=1}^{n+1-k}\frac{2}{k}\\
&=\sum_{k=2}^{n}(n+1-k)\frac{2}{k}=\left((n+1)\sum_{k=2}^{n}\frac{2}{k}\right)-2(n-1)\\
&=(2n+2)\sum_{k=1}^{n}\frac{1}{k}-4n.
\end{aligned}$$

□

4.2. 随机个随机变量的和的期望. 假设某种生物在生命的最后自动繁殖, 且每一代的存活时间相同. 每一个个体会繁殖的个数的均值为 μ. 现在请问你 n 代之后存在的生物量均值有多少.

上述问题可以转换为一列随机变量 $X_0,X_1,X_2,\ldots$ 由

$$\begin{cases}X_0=1\\ X_{n+1}=\sum_{j=1}^{X_n}\xi_j^{(n)}\end{cases}$$

定义. 其中 $\xi_j^{(n)} \in \mathbb{Z}_{\geq 0}$ 是 iid. rv. 服从均值 $\mu = \mathbb{E}\left[\xi_j^{(n)}\right]$.

这问题奇怪的地方在于, 随机变量的个数是随机的. 但是没关系, 我们可以把当前的随机变量 X_n 依据 X_{n-1} 的值划分成若干份: 即在 $X_{n-1} = k$ 的条件下的期望. 最后把它们加起来.

$$\mathbb{E}\left[X_n \mid X_{n-1} = k\right] = \mathbb{E}\left[\sum_{j=1}^{k} \xi_j^{(n-1)} \mid X_{n-1} = k\right] = k\mu \Longrightarrow \mathbb{E}\left[X_n \mid X_{n-1}\right] = X_{n-1}\mu$$

那么根据期望的期望等于它本身, 就得到了

$$\mathbb{E}\left[X_n\right] = \mathbb{E}\left[\mathbb{E}\left[X_n \mid X_{n-1}\right]\right] = \mathbb{E}\left[X_{n-1}\mu\right] = \mathbb{E}\left[X_{n-1}\right] \cdot \mu = \mu^n$$

这表明, 在第 n 代后代数量期望有 μ^n 个.