Review of Discrete Random Variable - 2

October 8, 2023

本节内容：

- 离散型随机变量的分布列以及数字特征 (期望和方差)
- 几种常见的离散型随机变量和分布
  - 0-1 分布
  - 二项分布
  - 泊松分布
  - 几何分布
  - Pascal 分布
- 相关的经典例题

0-1 分布

0-1 分布是最经典、最简单的分布, 很多概率模型的基础。

定义 0.20 随机变量 X 的取值为 $\{0,1\}$, 其分布列

$$P(X=1)=p\,,\quad P(X=0)=1-p$$

或等价于

$$P(X=k)=p^k\,,\quad (1-p)^{1-k}$$

称 X 服从参数为 p 的 0-1 分布, 或伯努利 (Bernoulli) 分布, 记 $X\sim Ber(p)$

X	0	1
P	$1-p$	p

0-1 分布的数字特征 – 期望

若 $X \sim Ber(p)$，则 $\mathbb{E}(X) = p$.

- 若一次试验只考虑事件 A 发生或不发生两种情况, 称这样的试验为伯努利 (Bernoulli) 试验, 可以通过 0-1 分布来描述伯努利试验:

$$X = \begin{cases} 1\,, & \text{若事件 } A \text{ 发生}, \\ 0\,, & \text{否则}. \end{cases}$$

容易得到 $\mathbb{E}(X) = P(A)$, 即随机变量 X 的期望等于事件 A 发生的概率.

0-1 分布的数字特征 – 方差

若 $X \sim Ber(p)$，则 $\mathbb{VAR}(X) = p(1-p)$.

综上, 可知: 若 $X \sim Ber(p)$, 则

- $\mathbb{E}(X) = p$
- $\mathbb{VAR}(X) = p(1-p)$

由此可知 0-1 分布也可由它的数学期望唯一确定。

0-1 分布 – n 重伯努利试验

伯努利试验考虑事件 A 发生或不发生, 设 $P(A) = p \in (0,1)$.

定义 0.21 将一个伯努利试验独立重复地进行 n 次, 称这一系列重复的独立试验为 n 重伯努利试验.

一种非常重要的概率模型, 衍生出很多概率分布.

二项分布

问题: 在 n 重伯努利试验中, 用 X 表示事件 A 发生的次数, 取值 $0, 1, \ldots, n$.

解法:

- 考虑事件 $\{X = k\}$ 到底是哪 k 次发生的, 有 $\binom{n}{k}$ 种情况
- 针对一种具体的情况, 不妨设前 k 次事件 A 发生, 后 $n-k$ 次事件 A 不发生, 此种情况发生的概率为

$$\underbrace{p \times p \times \ldots \times p}_{k\text{ 个}} \times \underbrace{(1-p) \times (1-p) \times \ldots \times (1-p)}_{n-k\text{ 个}} = p^k (1-p)^{n-k}$$

- 在 n 重伯努利试验中事件 A 发生了 k 次的概率 $\binom{n}{k} p^k (1-p)^{n-k}$

二项分布

定义 0.22 设随机变量 X 的分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k\in\{0,1,2,\ldots,n\}$$

称随机变量 X 服从为参数为 n 和 p 的二项分布 (binomial distribution), 记 $X\sim B(n,p)$.

Remarks:

- 容易发现 $P(X=k)$ 是二项式 $(1-p+xp)^n$ 展开式中 x^k 项的系数.
- 当 $x=1$ 时，有

$$\sum_{k=0}^{n}P(X=k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=(1-p+1*p)^n=1$$

- 若 $n=1$, 则二项分布退化为 0-1 分布, 即 $B(1,p)=Ber(p)$.

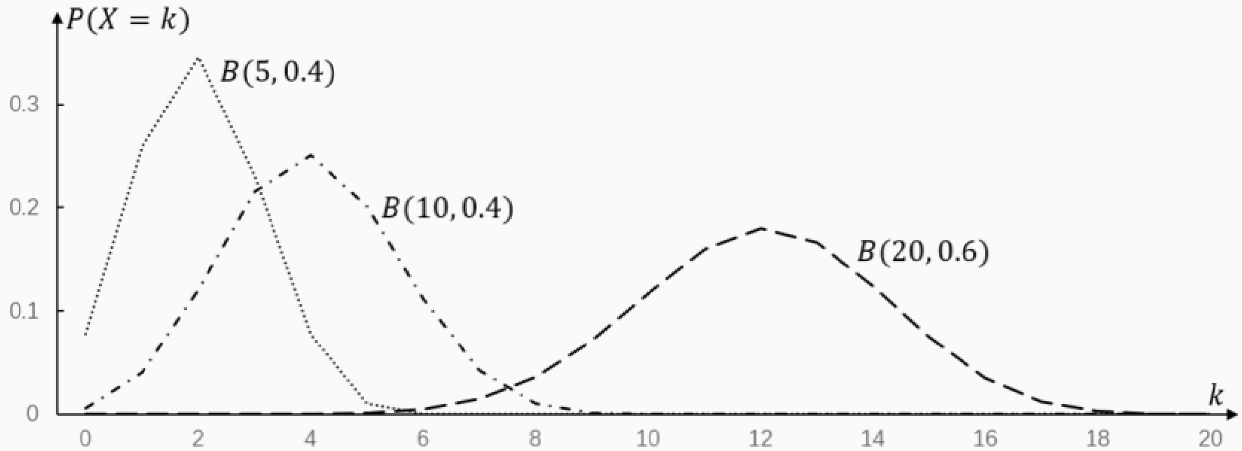
二项分布的数字特征

若 $X \sim B(n, p)$, 则

- $\mathbb{E}(X) = np$
- $\mathbb{VAR}(X) = np(1-p)$

因此，二项分布可由它的期望和方差唯一确定. 一旦知道二项分布的期望和方差, 可反解出参数 n 和 p.

二项分布的数字特征 – 分布图



若随机变量 $X \sim B(n,p)$, 则 $P(X=k)$ 从一开始单调递增, 然后一直单调递减, 一般在期望 np 附近的整数点取得最大值。

可严格证明:

$$\text{当 } k \in [0, np+p] \text{ 时}, P(X=k) \text{ 单调递增}$$

$$\text{当 } k \in [np+p, n \text{ 时}, P(X=k) \text{ 单调递减}$$

二项分布：例 0.57

例 0.57 假设有两个箱子, 每个箱子里放了 n 个球, 现在任意选取一个箱子拿走其中一球 (不放回), 重复这一过程, 求一个箱子中的球拿光而另一个箱子还剩下 r 个球的概率.

解答：例 0.57

问题: 假设有两个箱子, 每个箱子里放了 n 个球, 现在任意选取一个箱子拿走其中一球 (不放回), 重复这一过程, 求一个箱子中的球拿光而另一个箱子还剩下 r 个球的概率.

解答：注意区别

- 原题干: 一个箱子中的球拿光而另一个箱子还剩下 r 个球
- 对比题干: 当一个箱子中的球拿光时，另一个箱子恰好还剩下 r 个球

泊松分布

泊松分布用于描述大量试验中稀有事件出现次数的概率模型。

- 一个小时内公共汽车站来到的乘客数
- 一天中银行办理业务的顾客数
- 一个月内一个网站的访问量
- 一年内中国发生的地震次数

定义 0.23 若随机变量 X 的分布列为

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad (k \geq 0)$$

其中，$\lambda > 0$ 是一个常数, 称随机变量 X 服从 参数为 λ 的泊松分布, 记为 $X \sim P(\lambda)$.

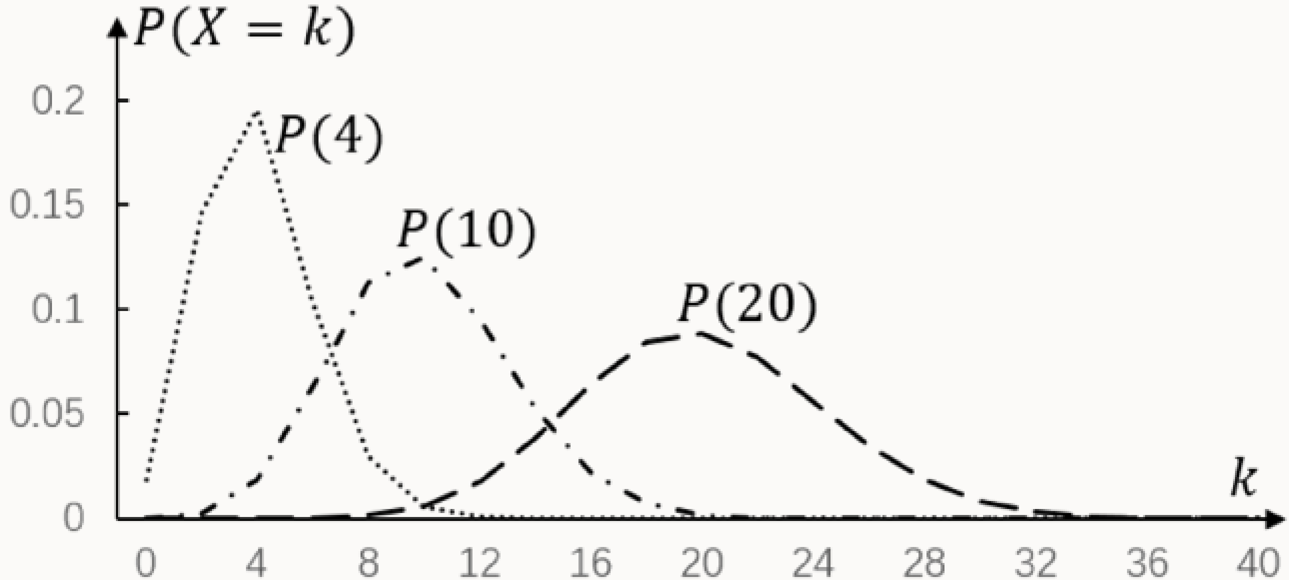
泊松分布的数字特征

若随机变量 $X \sim P(\lambda)$, 则有

$$\mathbb{E}(X) = \lambda \qquad \text{和} \qquad \mathbb{VAR}(X) = \lambda$$

泊松分布可由期望或方差唯一确定.

泊松分布的数字特征 – 分布图



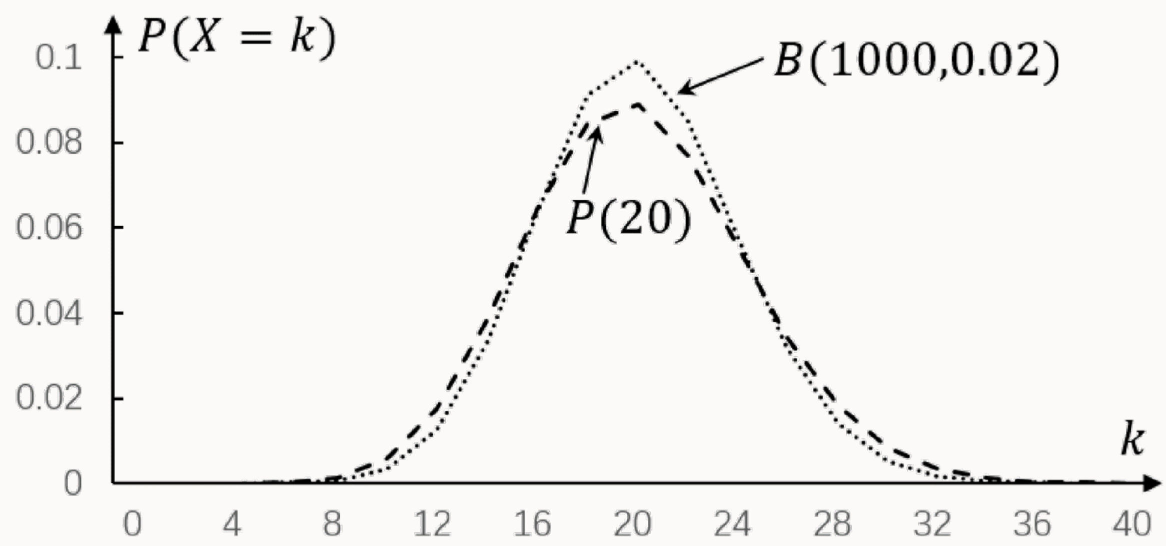
若随机变量 $X \sim P(\lambda)$, 则 $P(X = k)$ 从一开始单调递增, 然后一直单调递减, 一般在期望 λ 附近的整数点取得最大值.

泊松定理

总结来看, 泊松分布和二项分布的分布图很相似，都呈现了 “先增后减” 和 “拐点在期望附近” 的规律.

定理 0.8 对任意常数 $\lambda > 0$ 和任意正整数 n, 设 $np_n = \lambda$, 则对任意非负整数 k, 有

$$\lim_{n\to+\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$



证明：泊松定理

求证：

$$\lim_{n\to+\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

解法：

- 对齐两个分布的期望 $np_n=\lambda$
- 推导

$$\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{n!}{k!(n-k)!}p_n^k(1-p_n)^{n-k}$$

注意到每个 p_n 中包含一个 $\frac{1}{n}$, 我们需要确认与 n 直接相关的阶数

$$\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{n!}{n^k(n-k)!}\times\frac{\lambda^k}{k!}\times\left(1-\frac{\lambda}{n}\right)^{n-k}$$

第一项 $\frac{n!}{n^k(n-k)!}=1\times\frac{n-1}{n}\times\frac{n-2}{n}\times\cdots\times\frac{n-k+1}{n}\to 1$

第三项 $\left(1-\frac{\lambda}{n}\right)^{n-k}=\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}\frac{(n-k)\lambda}{\lambda}}\to \mathrm{e}^{-1\times\lambda}$

泊松定理的应用

若随机变量 $X \sim B(n,p)$, 当 n 比较大而 p 比较小时, 令 $\lambda = np$, 有

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

即利用 泊松分布近似计算二项分布.

针对彩票中奖、火山爆发、洪水泛滥、意外事故等小概率事件——试验次数 n 较多, 而试验概率 p 较小, 即试验结果较为稀疏 ——可将 n 重伯努利试验中小概率事件发生的次数近似服从泊松分布.

泊松分布：例 0.58

例 0.58 有 80 台同类型设备独立工作, 发生故障的概率是 0.01, 一台设备发生故障时只能由一人处理, 考虑方案:

I) 由四人维护, 每人单独负责 20 台

II) 由三人共同维护 80 台

方案 I) 或方案 II) 哪种更可取?

解答：例 0.58

问题：有 80 台同类型设备独立工作, 发生故障的概率是 0.01, 一台设备发生故障时只能由一人处理, 考虑方案: I) 由四人维护, 每人单独负责 20 台; II) 由三人共同维护 80 台. 请问方案 I) 或方案 II) 哪种更可取?

解答:

- 合理在于极少出现 “设备出故障不能及时维修” 的情况
- 方案一:
  - 每人单独负责 20 台时，同一时刻发生 2 台及 2 台以上的故障就不能及时维修 $P(X \geq 2; p = 0.01, n = 20)$
  - 四个人只要有一人不能及时维修，则方案出现问题 $A_1 \cup A_2 \cup A_3 \cup A_4$
- 方案二:
  - 80 台设备中同一时刻发生 4 台及 4 台以上的故障就不能及时维修 $P(X \geq 4; p = 0.01, n = 80)$
- 比较大小得知，方案二更优

泊松分布：例 0.59

例 0.59 一个公共汽车站有很多路公交车, 若一个时间段内到站乘客数

$$X \sim P(\lambda)(\lambda > 0)$$

所有到站的乘客是相互独立的、且选择 D1 路公交车的概率为 $p(p > 0)$, 求乘坐 D1 路公交车的乘客数 Y 的分布.

解答：例 0.59

问题：一个公共汽车站有很多路公交车, 若一个时间段内到站乘客数 $X \sim P(\lambda)(\lambda > 0)$. 所有到站的乘客是相互独立的、且选择 D1 路公交车的概率为 $p(p > 0)$, 求乘坐 D1 路公交车的乘客数 Y 的分布.

解答：

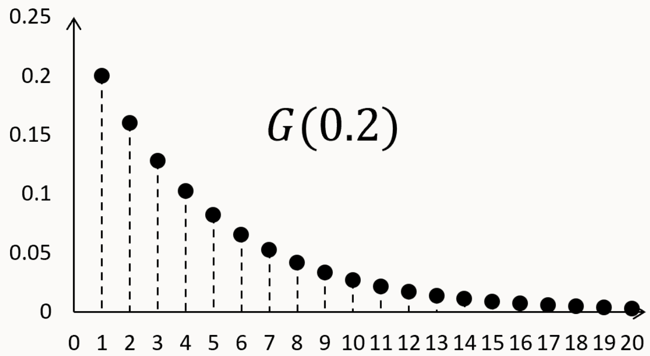
- 一个时间段内到站乘客数 $X \sim P(\lambda)(\lambda > 0)$, 那么 k 位乘客在一时间段内到站的概率
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$
- 这 k 位乘客依概率 p 选择 D1 路公交车, 那么 k 人中有 i 位乘客选择 D1 路公交车的概率为 $P(Y = i \mid X = k) = \binom{k}{i} p^k (1-p)^{k-i}$, 服从二项分布 $B(k, p)$
- 根据全概率公式
$$P(Y = i) = \sum_{k=i}^{+\infty} P(X = k) P(Y = i \mid X = k) = \frac{(p\lambda)^i e^{-p\lambda}}{i!}, \quad \Leftrightarrow \quad Y \sim P(p\lambda)$$

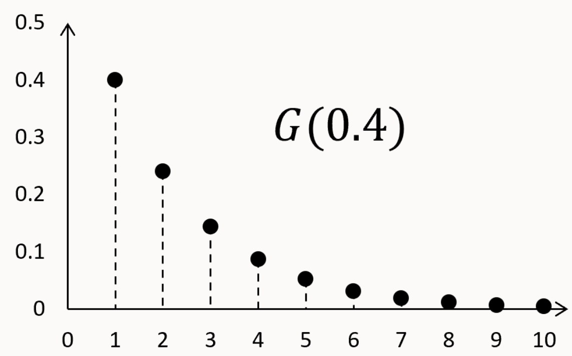
几何分布

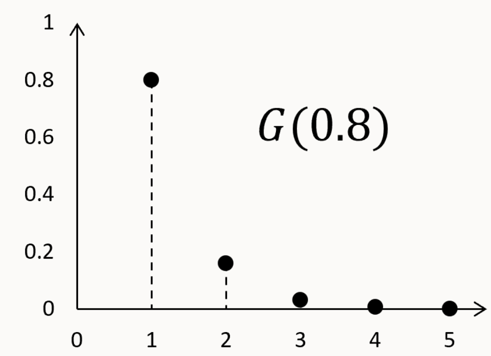
定义 0.24 在多重 Bernoulli 试验, 设事件 A 发生的概率为 p. 用随机变量 X 表示事件 A 首次发生时的试验次数, 则 X 的取值为 $1, 2, \ldots,$ 其分布列为

$$P(X = k) = (1 - p)^{k-1} p \qquad (k \geq 1)$$

称 X 服从参数为 p 的几何分布, 记为 $X \sim G(p)$.







几何分布：无记忆性 (memoryless property)

设随机变量 $X \sim G(p)$, 对任意正整数 m, n, 有

$$P(X > m + n \mid X > m) = P(X > n)$$

证明：

- 根据定义及等比数列求和计算

$$P(X > m) = \sum_{k=m+1}^{+\infty} p(1-p)^{k-1} = (1-p)^m$$

$$P(X > n) = \sum_{k=n+1}^{+\infty} p(1-p)^{k-1} = (1-p)^n$$

$$P(X > m + n) = \sum_{k=m+n+1}^{+\infty} p(1-p)^{k-1} = (1-p)^{m+n}$$

- $P(X > m + n) = P(X > m) \times P(X > n)$

几何分布：无记忆性 (memoryless property)

设随机变量 $X \sim G(p)$, 对任意正整数 m, n, 有

$$P(X > m + n \mid X > m) = P(X > n)$$

直观解释 :

- 成功的概率只与从当前起到成功的次数 n 有关，而与当前已经历的失败次数 m 无关.
- 几何分布的无记忆性：下一次是否成功与前面失败了多少次无关.

例：一赌徒在赌博时前面总是输, 总觉得下一次应该赢了.

几何分布的数字特征

若随机变量 $X \sim G(p)$, 则有

$$\mathbb{E}(X) = \frac{1}{p} \qquad \text{和} \qquad \mathbb{VAR}(X) = \frac{1-p}{p^2}$$

几何分布：例 0.60

例 0.60 古人非常重视生男孩且资源有限, 规定每个家庭可生一个男孩, 如果没男孩则可以继续生育直至有一个男孩; 若已有一个男孩, 则不再生育。不妨假设每个家庭生男孩的概率为 $p = \frac{1}{2}$, 问:

1) 一个家庭恰好有 n 个小孩的概率;

2) 一个家庭至少有 n 个小孩的概率;

3) 男女比例是否会失衡?

解答：例 0.60

问题：古人非常重视生男孩且资源有限, 规定每个家庭可生一个男孩, 如果没男孩则可以继续生育直至有一个男孩; 若已有一个男孩, 则不再生育。不妨假设每个家庭生男孩的概率为 $p = \frac{1}{2}$, 问:

1) 一个家庭恰好有 n 个小孩的概率；
2) 一个家庭至少有 n 个小孩的概率;
3) 男女比例是否会失衡?

解答：

- 认为生男孩为 "成功", 其他为 "失败"，我们关心 "成功前失败的次数", 因此这个模型符合几何分布
- $P(X = n) = p(1-p)^{n-1} = \frac{1}{2^n}$
- $P(X \geq n) = \sum_{k \geq n} p(1-p)^{n-1} = \frac{1}{2^{n-1}}$
- 每个家庭平均的孩子个数 $\mathbb{E}(X) = 1/p = 2$, 其中生男孩的期望 $E_{男} = p + p^2 + \cdots \to 1$; 生女孩的期望 $E_{女} = (1-p) + (1-p)^2 + \cdots \to 1$. 因此不会造成男女失衡.

负二项分布 (Pascal 分布)

定义 0.25 在多重 Bernoulli 试验中, 随机事件 A 发生的概率为 p. 用 X 表示事件 A 第 r 次成功时发生的试验次数, 则 X 取值 $r, r+1, r+2, \ldots$, 其分布列为

$$P(X=n)=\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}\cdot p \quad (n\geq r)$$

称 X 服从参数为 r 和 p 的负二项分布, 又称 Pascal 分布, 记为 $X\sim Pascal(r,p)$.

直观解释:

- 第 n 次成功: $\times p$
- 前 $n-1$ 次成功了 $r-1$ 次: $\times\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$

负二项分布的数字特征

若随机变量 $X \sim Pascal(r, p)$, 则有

$$\mathbb{E}(X) = \frac{r}{p} \qquad \text{和} \qquad \mathbb{VAR}(X) = \frac{r(1-p)}{p^2}$$

案例分析：德国坦克问题

在二战期间, 同盟国一直在努力确定德国坦克的生产数量, 有助于对德国战力的评估.

问题: 德国生产了 n 辆坦克, 编号分别为 $1, 2, \ldots, n$。盟军在战斗中任意击毁了 k 辆坦克, 被击毁的坦克编号为 $x_1, x_2, \ldots, x_k$, 能否通过被击毁的坦克编号来估计 n 的大小, 估计德国生产了多少辆坦克.

观察被击毁坦克编号分别为 17, 68, 94, 127, 135, 212, 估计 n.

解析：德国坦克问题

无先验的情况下，假设坦克被击毁是等可能事件.

- 该问题变为 “从 $\{1, 2, \ldots, n\}$ 中不放回地取 k 个数字”
- 进一步建模: 用被击毁坦克的最大值编号 X 来估计生产的坦克数量
- 设最大编号为 i, 其他的 $k-1$ 个编号从 $[1 : i-1]$ 中取得

$$P(X = i) = \frac{\binom{i-1}{k-1}}{\binom{n}{k}}$$

- 计算期望

$$\mathbb{E}(X) = \frac{\sum_{i=k}^{n} \binom{i-1}{k-1} i}{\binom{n}{k}} = \frac{k(n+1)}{k+1}$$

- 如今采样得到 k 辆坦克的实际编号: $x_1, x_2, \ldots, x_k$. 根据随机试验的统计性质，多

轮采样的结果应该趋近于 $\mathbb{E}(X)$. 现在仅有一次采样，令

$$\mathbb{E}(X) = \max\{x_1, x_2, \ldots, x_k\} \ .$$

- 所以

$$n \approx \max\{x_1, x_2, \ldots, x_k\}\frac{k+1}{k} - 1 \ .$$

- 真实案例

时间	统计估计	英国情报估计	德国实际产量
1940-06	169	1000	122
1941-06	244	1550	271
1942-08	327	1550	342

案例分析：集卡活动

小朋友喜欢参加各种集卡活动, 如奥特曼卡和叶罗丽卡等. 事实上很多成年人也对集卡游戏并不陌生, 例如 80 年代的葫芦娃洋画、或 90 年代的小虎队旋风卡等。

问题: 市场上有 n 种不同类型的卡片, 假设一个小朋友每次都能以等可能概率、独立地收集一张卡片, 问一个小朋友在平均情况下至少要收集多少张卡才能收集齐 n 种不同类型的卡片。

解答：集卡活动

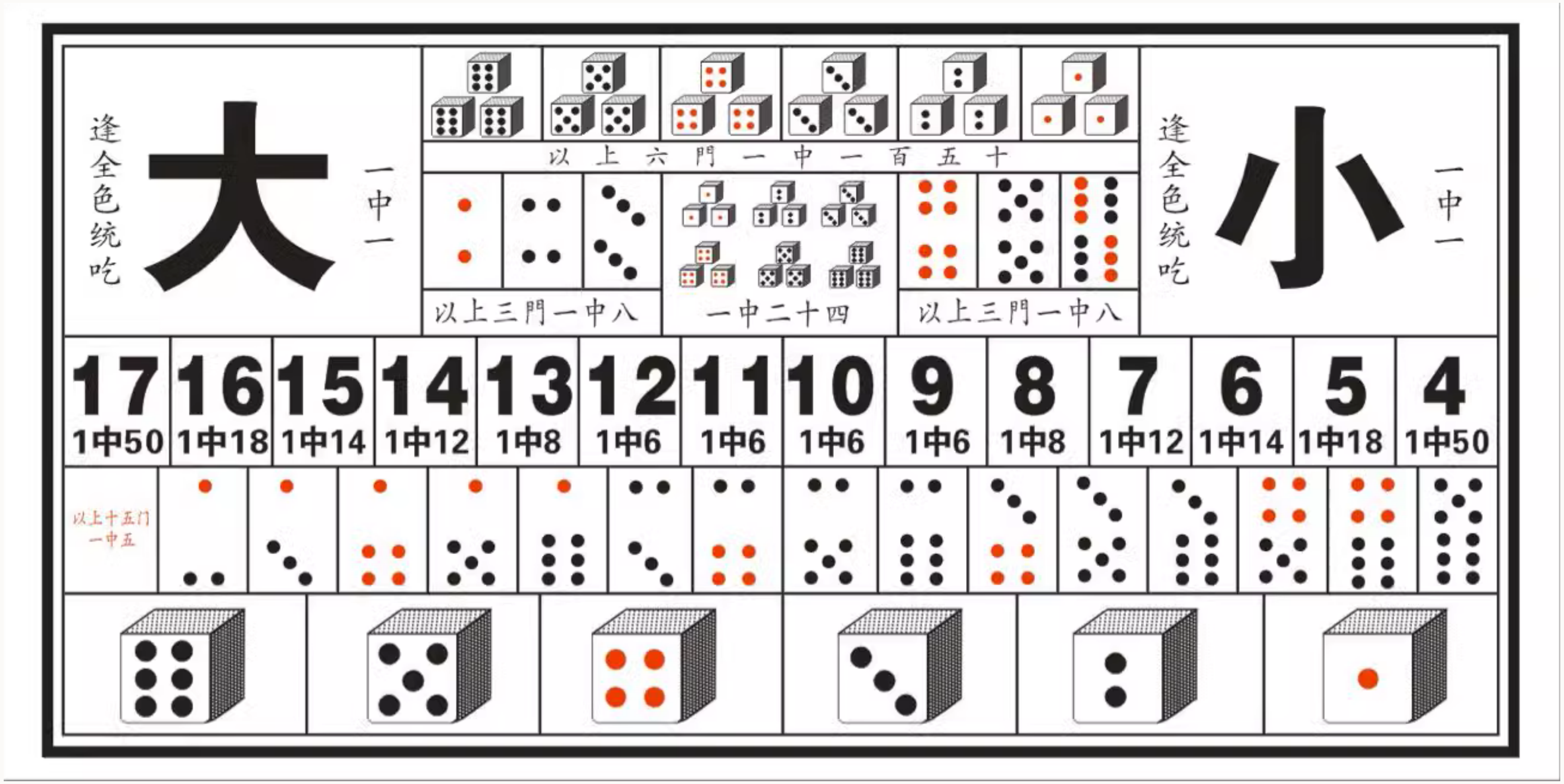
- 该问题是对收集次数建模:
  - 令 X_1 表示收集 1 种卡需要的次数: $1, p_1 = 1$
  - 令 X_2 表示已收集到 1 种卡，欲收集第 2 种卡需要的次数: $p_2 = 1 - \frac{1}{n}$
  - ...
  - 令 X_k 表示已收集到 $k-1$ 种卡，欲收集 k 种卡需要的次数: $p_k = 1 - \frac{k-1}{n}$，其中 $\frac{k-1}{n}$ 为单次抽到已有卡的概率
- X_k 服从几何分布, 因此 $\mathbb{E}(X_k) = \frac{1}{p_k} = \frac{n}{n-k+1}$
- 引理: 对任意的随机变量 X_1,X_2,...,X_n 有

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)$$

-

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \sum_{k=1}^{n} \mathbb{E}(X_k) = n \sum_{k=1}^{n} \frac{1}{k}$$

案例分析：Gambling



案例分析：Gambling

1:1				大								小			
	17 1:50	16 1:18	15 1:14	14 1:12	13 1:8	12 1:6	11 1:6		10 1:6	9 1:6	8 1:8	7 1:12	6 1:14	5 1:18	4 1:50
1:1		6		5			4		3			2		1	
1:5	1 2	1 3	1 4	1 5	1 6	2 3	2 4	2 5	2 6	3 4	3 5	3 6	4 5	4 6	5 6
1:8		1 1		2 2			3 3		4 4			5 5		6 6	
1:50								猜对三个骰子的面上数值（除去全围）							
1:150		111		222			333		444			555		666	
1:24								全围							