

Part III – Ch03: 离散型随机变量

试验结果数值化

- 有些随机试验的结果本身就是数值
 - 抛一枚骰子的点数: $1, 2, \dots, 6$
 - 国家一年出生的婴儿数: $1, 2, \dots, n$
- 有些试验结果看起来与数值无关, 但可以用数值来表示
 - 抛硬币: 用 1 表示正面朝上, 用 0 表示正面朝下
 - 针对一般事件 A , 可以定义

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

- 事件驱动的 (event-driven), 数值化意味着可计算

随机变量

试验结果用数值表示, 引入变量来表示随机事件——随机变量.

定义 0.15 设 Ω 是一个样本空间, 如果对每个基本事件 $\omega \in \Omega$, 都对应于一个实数 $X(\omega)$, 称这样的单射实值函数 $X : \Omega \rightarrow \mathbb{R}$ 为 **随机变量 (random variable)**, 一般简写为 X .

随机变量与 (确定性) 普通函数存在着本质的不同:

- $X(\omega)$ 随样本点 ω 的不同而取不同的值, 具有一定的随机性
- 各试验结果的出现具有一定的概率, X 的取值具有统计规律性

From 随机事件 **To** 随机变量:

- 通过随机变量来描述随机现象或随机事件, 可以利用各种数学分析工具, 通过对随机变量的研究来揭示随机现象.

随机变量

根据取值类型, 可以将随机变量进行分类:

- 离散型随机变量: X 的取值是有限的、无限可列的

- 抛一枚骰子的点数: $1, 2, \dots, 6$ —— 有限的

X	1	2	3	4	5	6
-----	---	---	---	---	---	---

- 国家一年出生的婴儿数: $1, 2, \dots, n$ —— 有限的或者无限可列的

X	1	2	...	n	...
-----	---	---	-----	-----	-----

- 非离散型随机变量: X 的取值是无限不可列的

提纲: 离散型随机变量

- 分布列
- 数字特征
 - 期望
 - 函数期望 (Jensen 不等式)
 - 方差
- 常见的离散型随机变量和分布
 - 0-1 分布: 伯努利试验、分布列、数字特征
 - 二项分布: 分布列、数字特征
 - 泊松分布: 分布列、数字特征
 - 几何分布: 分布列、无记忆性、数字特征
 - 负二项分布: 分布列、数字特征

Ch03: 离散型随机变量

Discrete Random Variable and Its Variants

September 4, 2025

分布列

分布列包含随机变量的取值和概率, 可以完全刻画其概率属性

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

- 分布列: 对于 X 所有的取值 x_k , 其概率为 $p_k = P(X = x_k)$
- 性质:
 - 对于任意的 k 都有 $p_k \geq 0$
 - $\sum_k p_k = 1$

分布列: 例 0.65

例 0.65 设随机变量 X 的分布列为

$$P(X = k) = \frac{c}{4^k} \quad \text{且} \quad k \geq 0 .$$

求概率 $P(X = 1)$.

解答: 例 0.65

题目: 设随机变量 X 的分布列为

$$P(X = k) = \frac{c}{4^k} \quad \text{且} \quad k \geq 0.$$

求概率 $P(X = 1)$.

解答:

- 根据概率的完备性有

$$1 = \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{c}{4^k} = \frac{4}{3}c,$$

可得 $c = 3/4$, 从而得到 $P(X = 1) = 3/16$.

分布列: 例 0.66

例 0.66 给定常数 $\lambda > 0$, 随机变量 X 的分布列为

$$P(X = k) = \frac{c\lambda^k}{k!} \quad \text{且} \quad k \geq 0.$$

- 求概率 $P(X = 1)$;
- 求概率 $P(X > 2)$.

解答: 例 0.66

题目: 给定常数 $\lambda > 0$, 随机变量 X 的分布列为

$$P(X = k) = \frac{c\lambda^k}{k!} \quad \text{且} \quad k \geq 0.$$

求概率 $P(X = 1)$ 和 $P(X > 2)$.

解答:

- 根据概率的完备性有

$$1 = \sum_{k=0}^{\infty} P(X = k) = c \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = c \cdot e^{\lambda},$$

可得 $c = e^{-\lambda}$.

- 进一步得到 $P(X = 1) = \lambda e^{-\lambda}$, 及

$$P(X > 2) = 1 - P(X \leq 2) = 1 - p_0 - p_1 - p_2 = 1 - e^{-\lambda}(1 + \lambda + \lambda^2/2)$$

数字特征 or 统计量

分布列 能一目了然的看出离散型随机变量得取值及相应的概率.

- 例 0.65 及例 0.66

针对一个具体的问题, 完全求解出概率分布列可能不是一件容易的事.

- 统计某个区域的工资水平, 包括不同职业不同人群得收入——不容易完全求解
- 在例 0.66中, 求解 $P(X > 2)$ 不需要依赖于 $P(X = 3)$, $P(X = 4)$, ...
——不需要完全求解

通过一些指标来描述离散随机变量的总体特征, 或称为统计量.

- 期望、方差、相关系数、矩等.

统计量一：期望

定义 0.16 设随机变量 X 的分布列为 $P(X = x_k) = p_k (k \geq 1)$. 若级数 $\sum_k p_k x_k$ 绝对收敛, 则称级数 $\sum_k x_k p_k$ 为随机变量 X 的期望, 记为 $\mathbb{E}(X)$

$$\mathbb{E}(X) = \sum_k x_k p_k$$

Remarks:

- 期望是常量不是变量.
- 期望反映随机变量 X 的平均值.
- 期望的性质:
 1. 设 $c \in \mathbb{R}$, 若随机变量 $X \equiv c$, 则 $\mathbb{E}(X) = c$, 不考虑其分布列.
 2. 线性: 对随机变量 X 及常数 $a, b \in \mathbb{R}$, $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- 除非特别说明, 通常直接利用定义计算期望, 不考虑绝对收敛性.

期望：例 0.67

例 0.67 随意抛掷一枚骰子, X 为掷出的点数. 试求 $\mathbb{E}(X)$.

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

解答：例 0.67

题目：随意抛掷一枚骰子, X 为掷出的点数. 试求 $\mathbb{E}(X)$.

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

解答：

- 根据期望的定义有

$$\mathbb{E}(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2}.$$

期望：例 0.68

例 0.68 设随机变量 X 的分布列为

$$P\left(X = \frac{(-2)^k}{k}\right) = \frac{1}{2^k}, \quad k \in \mathbb{N}^+,$$

求期望 $\mathbb{E}(X)$.

解答: 例 0.68

题目: 设随机变量 X 的分布列为 $P\left(X = \frac{(-2)^k}{k}\right) = \frac{1}{2^k}$, $k \in \mathbb{N}^+$, 求期望 $\mathbb{E}(X)$.

解答:

- 根据期望的定义有

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} P\left(X = \frac{(-2)^k}{k}\right) \frac{(-2)^k}{k} = \sum_{k=1}^{+\infty} \frac{(-1)^k}{k} = -\ln 2.$$

- 但是

$$\sum_{k=1}^{+\infty} P\left(X = \frac{(-2)^k}{k}\right) \left|\frac{(-2)^k}{k}\right| = \sum_{k=1}^{+\infty} \frac{1}{k} \rightarrow +\infty.$$

该级数并非绝对收敛, 其级数和可能随着求和顺序的改变而改变, 级数和并非唯一的数值, 故该随机变量的期望 $\mathbb{E}(X)$ 不存在.

期望: 例 0.69

例 0.69 有 n 把钥匙只有一把能打开门. 随机选取一把试开门, 若打不开则除去, 求打开门次数的平均数.

解答: 例 0.69

题目: 有 n 把钥匙只有一把能打开门. 随机选取一把试开门, 若打不开则除去, 求开门次数的平均数.

解答:

- 设随机变量 X 表示尝试开门的次数, 其分布列为

$$P(X = k) = \frac{\binom{n-1}{k-1}}{\binom{n}{k-1}} \cdot \frac{1}{n - k + 1} = \frac{1}{n} \quad k \in [n].$$

- 根据期望的定义有

$$\mathbb{E}(X) = \sum_{k=1}^n \frac{\binom{k}{n}}{n} = \frac{(1+n)n}{2n} = \frac{n+1}{2}.$$

期望: 例 0.70

例 0.70 有 4 个盒子编号分别为 1, 2, 3, 4. 将 3 个不同的球随机放入 4 个盒子, 用 X 表示有球盒子的最小号码, 求 $\mathbb{E}(X)$.

解答: 例 0.70

题目: 有 4 个盒子编号分别为 1, 2, 3, 4. 将 3 个不同的球随机放入 4 个盒子, 用 X 表示有球盒子的最小号码, 求 $\mathbb{E}(X)$.

解答:

• 分布列:

- $X = 1$ 时, 分 1 号盒子中有 1 个球、2 个球、3 个球三种情况讨论 $C_3^1 3^2 + C_3^2 C_3^1 + 1$
- $X = 2$ 时, 分 2 号盒子中有 1 个球、2 个球、3 个球三种情况讨论 $C_3^1 2^2 + C_3^2 C_2^1 + 1$
- $X = 3$ 时, 分 3 号盒子中有 1 个球、2 个球、3 个球三种情况讨论 $C_3^1 1^2 + C_3^2 C_1^1 + 1$
- $X = 4$ 时, 化简为子问题 “三个球都放在编号为 4 的盒子里”

X	1	2	3	4
P	$\frac{37}{4^3}$	$\frac{19}{4^3}$	$\frac{7}{4^3}$	$\frac{1}{4^3}$

• 期望: $\mathbb{E}(X) = \frac{100}{64} = \frac{25}{16}$

期望: 例 0.71

例 0.71 一个口袋中有 5 个球, 在这 5 个球上分别标有数字 1, 2, 3, 4, 5. 从袋中不放回任取 3 个球, 每个球被取到的可能性相同, 求取到的球上标明的最大数字 X 的分布列及其期望.

解答：例 0.71

题目：一个口袋中有 5 个球，在这 5 个球上分别标有数字 1, 2, 3, 4, 5. 从袋中不放回任取 3 个球，每个球被取到的可能性相同，求取到的球上标明的最大数字 X 的分布列及其期望.

X	3	4	5
P	$\frac{1}{C_5^3}$	$\frac{C_3^2}{C_5^3}$	$\frac{C_4^2}{C_5^3}$

$$\mathbb{E}(X) = 3 \cdot \frac{1}{C_5^3} + 4 \cdot \frac{1}{C_5^3} + 5 \cdot \frac{1}{C_5^3} = \frac{9}{2}.$$

统计量二：函数期望

定理 0.6 设离散型随机变量 X 的分布列是 $P(X = x_k) = p_k > 0 (k \geq 1)$. 对任意的实值函数 $g : \mathbb{R} \rightarrow \mathbb{R}$. 若级数 $\sum_k g(x_k)p_k$ 绝对收敛, 则有

$$\mathbb{E}[g(X)] = \sum_k g(x_k)p_k$$

Remarks:

- 求 $Y = g(X)$ 的期望, 无需知道 Y 的分布列. 使用 X 的分布列即可计算 $\mathbb{E}(Y)$.
- 定理 0.6 及期望的线性性.

计算：函数期望

设离散型随机变量 X 的分布列是 $P(X = x_k) = p_k > 0 (k \geq 1)$, 以及 $g_i : \mathbb{R} \rightarrow \mathbb{R}$ 是实值函数 ($i \in [n]$). 若级数 $\sum_k g_i(x_k)p_k$ 绝对收敛, 则对任意常数 $c_1, c_2, \dots, c_n$, 有

$$\mathbb{E} \left[\sum_{i=1}^n c_i g_i(X) \right] = \sum_i c_i \mathbb{E}[g_i(x_k)] = \sum_i \sum_k c_i g_i(x_k) p_k$$

举例: $\mathbb{E}[X^2 + X + \sin X + 4] = ?$

函数期望：例 0.72

例 0.72 随意抛掷一枚骰子, X 为掷出的点数. 试求

- 期望 $\mathbb{E}(X)$
- 函数期望 $\mathbb{E}[g(X)]$, 其中 $g(x) = x^2$
- 函数期望 $\mathbb{E}[f(X)]$, 其中 $f(x) = x^2 + 3x + 1$

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

解答：例 0.72

根据题干求解新的分布列

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$g(X) = X^2$	1	4	9	16	25	36
$3X$	3	6	9	12	15	18
$f(X) = X^2 + 3X + 1$	5	11	19	29	41	55

或者

$$\mathbb{E}[f(X)] = \mathbb{E}[X^2 + 3X + 1] = \mathbb{E}[X^2] + 3\mathbb{E}[X] + \mathbb{E}[1]$$

凸函数和凹函数

定义 0.17 (凸函数) 设函数 $g : [a, b] \rightarrow \mathbb{R}$, 对任意 $x_1, x_2 \in [a, b]$ 和 $\lambda \in [0, 1]$. 若

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2),$$

则称 g 是定义在 $[a, b]$ 的凸函数 或者下凸函数.

定义 0.18 (凹函数) 设函数 $g : [a, b] \rightarrow \mathbb{R}$, 对任意 $x_1, x_2 \in [a, b]$ 和 $\lambda \in [0, 1]$. 若

$$g(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda g(x_1) + (1 - \lambda)g(x_2),$$

则称 g 是定义在 $[a, b]$ 的凹函数 或者上凸函数.

Jensen 不等式

定理 0.7 对随机变量 $X \in [a, b]$ 和若实值函数 $f : [a, b] \rightarrow \mathbb{R}$ 是凸函数, 实值函数 $g : [a, b] \rightarrow \mathbb{R}$ 是凹函数, 则有

$$f[\mathbb{E}(X)] \leq \mathbb{E}[f(X)] \quad \text{和} \quad g[\mathbb{E}(X)] \geq \mathbb{E}[g(X)]$$

推论:

$$(\mathbb{E}(X))^2 \leq \mathbb{E}(X^2)$$

$$e^{\mathbb{E}(X)} \leq \mathbb{E}[e^X]$$

$$\ln \mathbb{E}(X) \geq \mathbb{E}[\ln X]$$

统计量三：方差

定义 0.19 设随机变量 X 的分布列为 $P(X = x_k) = p_k > 0 (k \geq 1)$. 若期望 $\mathbb{E}(X)$ 和

$$\mathbb{E} \left[(X - \mathbb{E}(X))^2 \right] = \sum_{k \geq 1} p_k \left(x_k - \sum_k p_k x_k \right)^2,$$

存在, 则称 $\mathbb{E} (X - \mathbb{E}(X))^2$ 为随机变量 X 的**方差**, 记为 $\text{VAR}(X)$. 称

$$\sigma(X) = \sqrt{\text{VAR}(X)} \quad \text{或} \quad \mathbb{D}(X) = \sqrt{\text{VAR}(X)}$$

为**标准差**.

Remarks:

- 方差是常量而不是变量.
- 方差反映随机变量 X 与其期望 $\mathbb{E}(X)$ 的偏离程度.

离散型随机变量方差的性质

1. $\text{VAR}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$
2. 设 $c \in \mathbb{R}$ 是常数, 若随机变量 $X \equiv c$, 则 $\text{VAR}(X) = 0$
3. 对随机变量 X 和常数 $a, b \in \mathbb{R}$, 有 $\text{VAR}(aX + b) = a^2 \text{VAR}(X)$
4. 方差不具有线性性:

$$\text{VAR}(f(X) + g(X)) \neq \text{VAR}(f(X)) + \text{VAR}(g(X)), \text{ for } f, g \in \mathcal{C}$$

5. 对随机变量 X 和常数 a , 有 $\text{VAR}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 \leq \mathbb{E}(X - a)^2$
6. Bhatia-Davis 不等式: 对随机变量 $X \in [a, b]$, 有

$$\text{VAR}(X) \leq (b - \mathbb{E}(X))(\mathbb{E}(X) - a) \leq \frac{(b - a)^2}{4}$$

方差: 例 0.73

例 0.73 某人有一笔资金, 可投入两个项目: 住房地产和商铺地产, 其收益都与市场状态相关. 若把未来市场划分为好、中、差三个等级, 其发生的概率分别为 0.2, 0.7, 0.1. 通过调查, 该投资者认为投资于住房地产的收益 X 和投资商铺地产的收益 Y 的分布列如下表所示, 请问投资者该如何投资?

X	11	3	-3
Y	11	3	-3
P	0.2	0.7	0.1

方差: 例 0.73

问题:

X	11	3	-3
Y	11	3	-3
P	0.2	0.7	0.1

解答:

- 这个问题要综合考察两个投资项目收益的期望和方差
- 先考察期望

$$\mathbb{E}(X) = 11 \times 0.2 + 3 \times 0.7 + (-3) \times 0.1 = 4$$

$$\mathbb{E}(Y) = 6 \times 0.2 + 4 \times 0.7 + (-1) \times 0.1 = 3.9$$

- 再来考察方差

$$\mathbb{VAR}(X) = (11 - 4)^2 \times 0.2 + (3 - 4)^2 \times 0.7 + (-3 - 4)^2 \times 0.1 = 15.4$$

$$\mathbb{VAR}(Y) = (6 - 3.9)^2 \times 0.2 + (4 - 3.9)^2 \times 0.7 + (-1 - 3.9)^2 \times 0.1 = 3.29$$

- 如何抉择呢?

方差: 例 0.74

例 0.74 假设有 10 只同种电器元件, 其中有两只不合格品. 装配仪器的时候, 从这批原件中任取一只, 如果是不合格的产品, 则扔掉重新任取一只; 如果仍是不合格的产品, 则继续扔掉再任取一支. 求取得合格品之前, 已取到的不合格品数的方差.

方差: 例 0.74

题干: 假设有 10 只同种电器元件, 其中有两只不合格品. 装配仪器的时候, 从这批原件中任取一只, 如果是不合格的产品, 则扔掉重新任取一只; 如果仍是不合格的产品, 则继续扔掉再任取一支. 求取得合格品之前, 已取到的不合格品数的方差.

解答:

- 已知 $P(X = 0) = 4/5, P(X = 1) = 8/45, P(X = 2) = 1/45$
- 则 $\mathbb{E}(X) = 2/9, \mathbb{E}(X^2) = 4/15$
- 所以方差 $\text{VAR} = 88/405 \approx 0.2173$

几种常见的离散型随机变量和分布

类型	知识点
0-1 分布	伯努利试验、分布列、数字特征
二项分布	分布列、数字特征
泊松分布	分布列、数字特征
几何分布	分布列、无记忆性、数字特征
Pascal 分布	分布列、数字特征
相关的经典例题	

伯努利试验

定义 0.20 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**.

伯努利试验可以用 0 和 1 标记来实现试验结果数值化, 即

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

拓展:

- 伯努利试验是离散变量的概统学的基础
- 在人工智能领域有相关应用, 比如 masked modeling, spiking computing, ...

0-1 分布

定义 0.21 随机变量 X 的取值为 $\{0, 1\}$, 其分布列为

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

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

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

$$\mathbb{E}(X) = 1 \cdot P(A) + 0 \cdot (1 - P(A)) = P(A)$$

即随机变量 X 的期望等于事件 A 发生的概率.

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

由此可知 0-1 分布也可由它的数学期望 p 唯一确定.

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 0.22 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 $0, 1, \dots, n$.

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

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

- 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 0.23 设随机变量 X 的分布列为

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

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

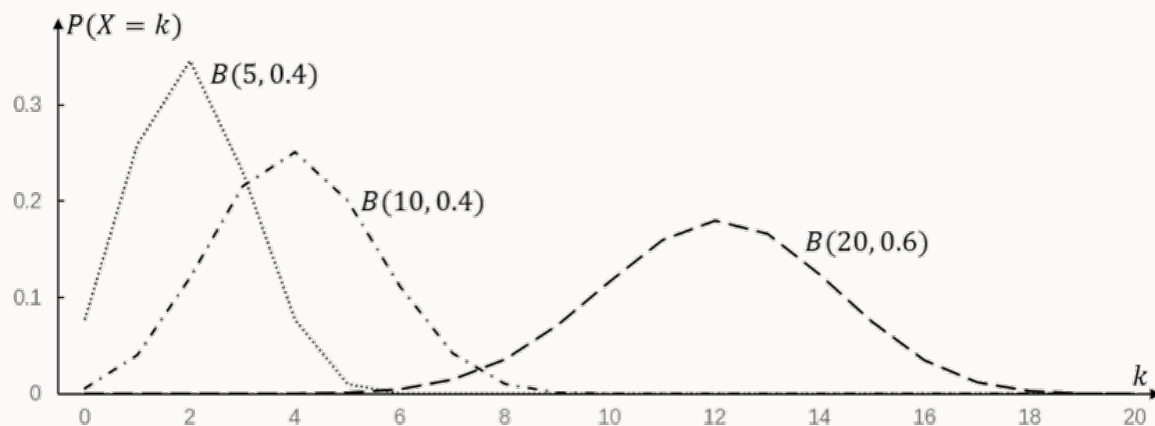
$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} 1^k = (1 - p + 1 \cdot p)^n = 1$$

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

二项分布的期望和方差

若 $X \sim B(n, p)$, 则有 $\mathbb{E}(X) = np$ 和 $\mathbb{V}\mathbb{A}\mathbb{R}(X) = np(1 - p)$.

- 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 .
- 若随机变量 $X \sim B(n, p)$, **可严格证明**: (1) 当 $k \in [0, np + p]$ 时, $P(X = k)$ 单调递增; (2) 当 $k \in [np + p, n]$ 时, $P(X = k)$ 单调递减; (3) 一般在期望 np 附近的整数点取得最大值. (留作作业)



二项分布: 例 0.75

例 0.75 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 0.75

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

$$\binom{2n-r}{n} p^n (1-p)^{n-r} + \binom{2n-r}{n-r} p^{n-r} (1-p)^n$$

- 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

$$\binom{2n-r-1}{n-1} p^{n-1} (1-p)^{n-r} \cdot p + \binom{2n-r-1}{n-1} p^{n-r} (1-p)^{n-1} \cdot (1-p)$$

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 0.24 若随机变量 X 的分布列为

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

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 $X \sim P(\lambda)$ 或者 $X \sim \text{Poisson}(\lambda)$.

- 若随机变量 $X \sim P(\lambda)$, 则有 $\mathbb{E}(X) = \lambda$ 和 $\mathbb{V}\text{AR}(X) = \lambda$. 因此, 泊松分布可由期望或方差唯一确定.

泊松分布: 例 0.76

例 0.76 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 0.76

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 $P(Y = i | X = k) = \binom{k}{i} p^i (1-p)^{k-i}$, 服从二项分布 $B(k, p)$
- 根据全概率公式

$$P(Y = i) = \sum_{k=i}^{+\infty} P(X = k) P(Y = i | X = k) = \frac{(p\lambda)^i e^{-p\lambda}}{i!}, \quad \Leftrightarrow \quad Y \sim P(p\lambda)$$

泊松分布: 例 0.77

例 0.77 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?

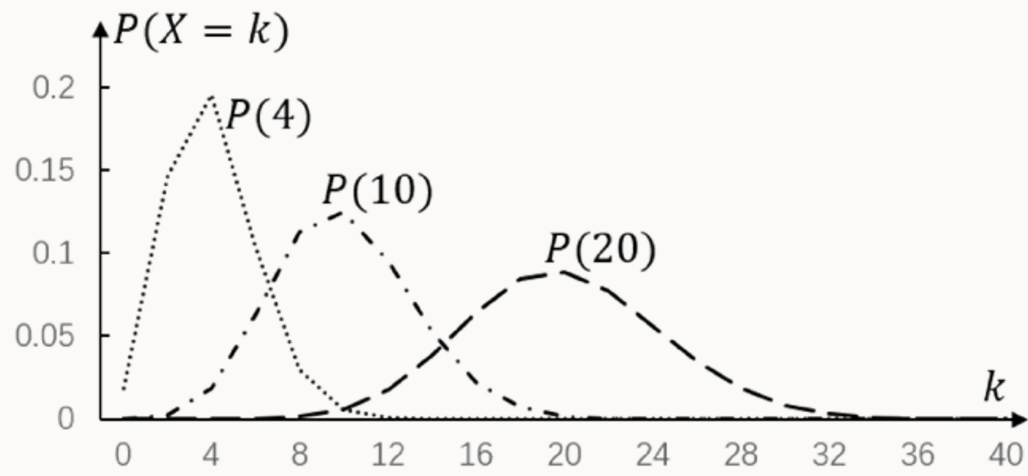
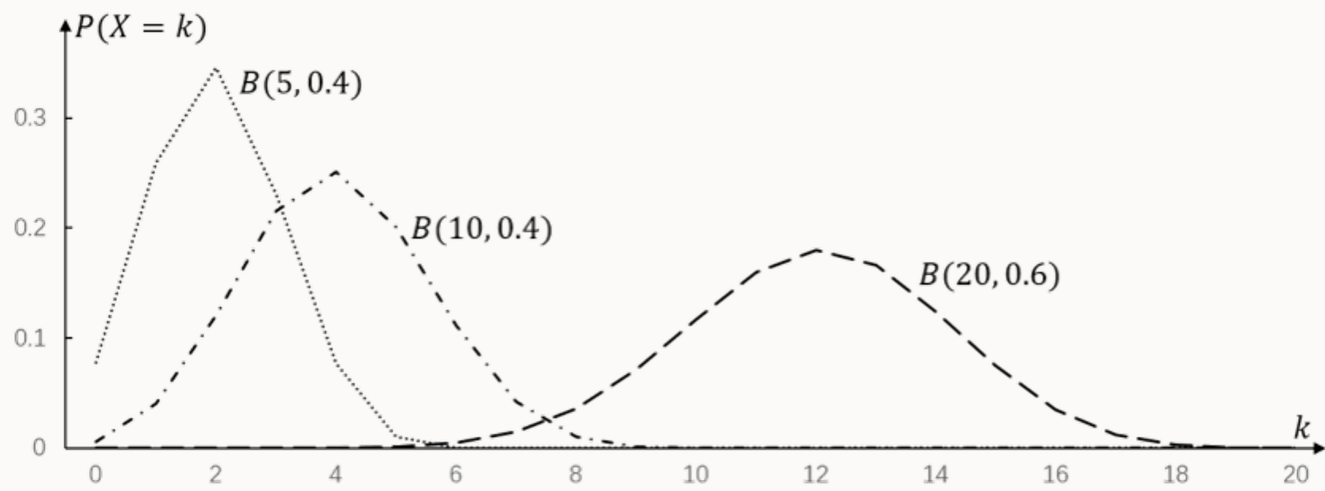
解答：例 0.77

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

回顾二项分布和泊松分布的分布图

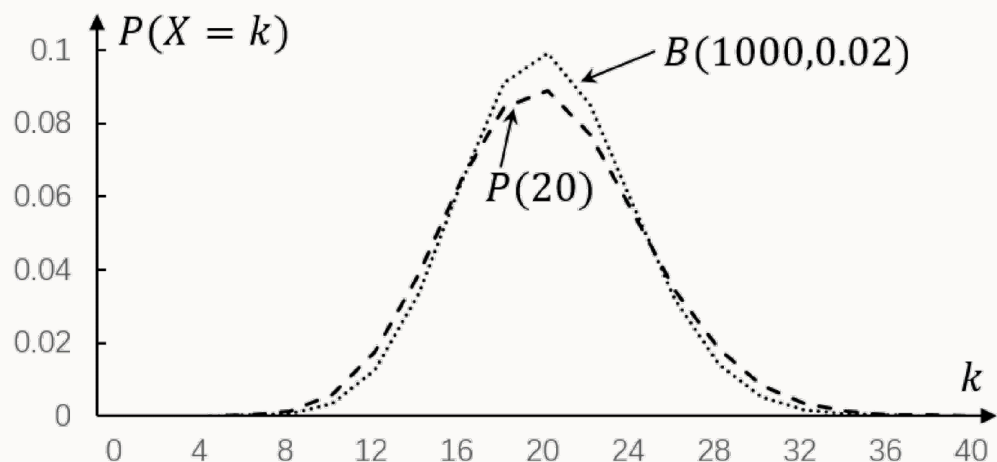


泊松定理

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

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



证明: 泊松定理

求证:

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

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 $\frac{n!}{n^k(n-k)!} = 1 \times \frac{n-1}{n} \times \frac{n-2}{n} \times \dots \times \frac{n-k+1}{n} \rightarrow 1$

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

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.
- 通常来讲, 满足 $n \geq 20$ 且 $p \leq 0.05$ 的条件下, 近似计算效果比较好,
比如例 0.77 中 $n = 80$ 且 $p = 0.01$.
- 主要: 这只是近似计算, 为了方便计算, 并不是精确计算.

泊松定理: 例 0.78

例 0.78 假设有一个工厂生产大量零件, 某种零件在生产过程中出现次品的概率为 $p = 0.02$, 该工厂每天生产 $n = 500$ 个零件. 我们希望计算每天出现 $k = 5$ 个次品的概率.

解答: 例 0.78

问题: 假设有一个工厂生产大量零件, 某种零件在生产过程中出现次品的概率为 $p = 0.02$, 该工厂每天生产 $n = 500$ 个零件. 我们希望计算每天出现 $k = 5$ 个次品的概率.

解答:

- [使用二项分布计算] 用 X 表示每天次品的数量, 即 $X \sim B(500, 0.02)$, 我们需要计算 $P(X = 5)$:

$$P(X = 5) = \binom{500}{5} 0.02^5 0.98^{495}$$

难以计算.

- [使用泊松分布计算] 根据泊松分布的近似条件, 首先计算 $\lambda = np = 10$. 因此可以假设 $Y \sim \text{Poisson}(10)$, 并计算得到 $P(Y = 5)$:

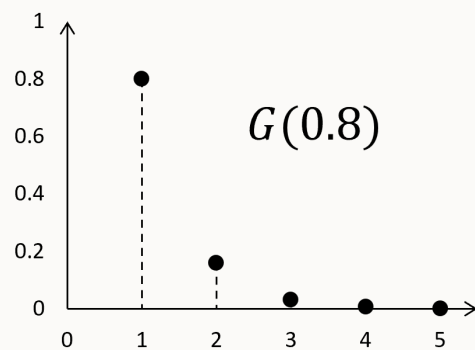
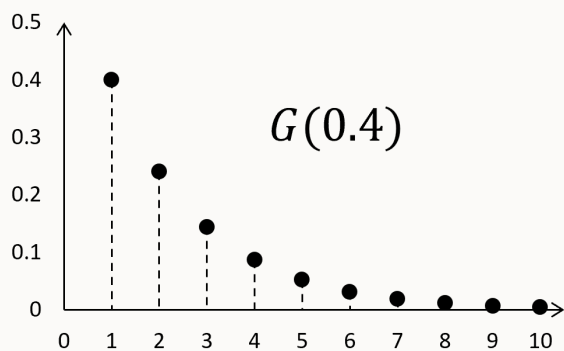
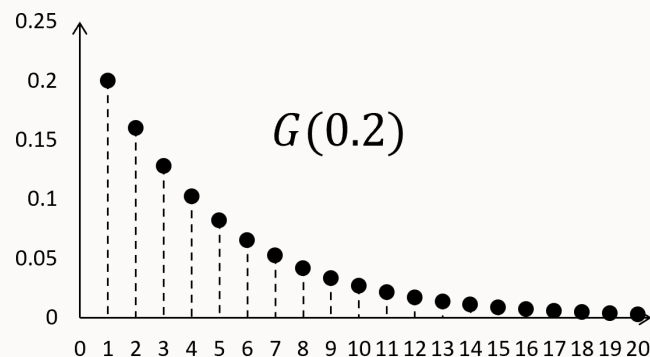
$$P(Y = 5) = \frac{10^5 e^{-10}}{5!} \approx 0.0378$$

几何分布

定义 0.25 在多重伯努利试验中设事件 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, 则其分布列为

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

称 X 服从参数为 p 的几何分布, 记为 $X \sim G(p)$.



- 若随机变量 $X \sim G(p)$, 则有 $\mathbb{E}(X) = \frac{1}{p}$ 和 $\mathbb{V}\text{AR}(X) = \frac{1-p}{p^2}$. 因此, 几何分布可由期望唯一确定.

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: 例 0.79

例 0.79 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 $p = 0.5$, 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 0.79

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 $p = 0.5$, 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
- $P(X = n) = p(1 - p)^{n-1} = \frac{1}{2^n}$
- $P(X \geq n) = \sum_{k \geq n} p(1 - p)^{k-1} = \frac{1}{2^{n-1}}$
- 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 $E_{\text{男}} = p + p^2 + \cdots \rightarrow 1$; 生女孩的期望 $E_{\text{女}} = (1 - p) + (1 - p)^2 + \cdots \rightarrow 1$. 因此不会造成男女失衡.

负二项分布 (Pascal 分布)

定义 0.26 在多重伯努利试验中, 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 $r, r+1, r+2, \dots$, 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 $X \sim \text{Pascal}(r, p)$.

- 直观: 第 n 次成功 $\times p$; 前 $n-1$ 次成功了 $r-1$ 次: $\times \binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$
- 若随机变量 $X \sim \text{Pascal}(r, p)$, 则有 $\mathbb{E}(X) = \frac{r}{p}$ 和 $\text{VAR}(X) = \frac{r(1-p)}{p^2}$

小结: 离散变量分布

分布	分布列	期望	方差
0-1 分布	$P(X = 1) = p, P(X = 0) = 1 - p$	p	$p(1 - p)$
二项分布	$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$	np	$np(1 - p)$
泊松分布	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k \geq 0)$	λ	λ
几何分布	$P(X = k) = (1 - p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布	$P(X = n) = \binom{n-1}{r-1} p^{r-1} (1 - p)^{n-r} \cdot p \quad (n \geq r)$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

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 $1, 2, \dots, n$. 盟军在战斗中任意击毁了 k 辆坦克, 被击毁的坦克编号为 $x_1, x_2, \dots, x_k$, 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.

- 概率建模: 该问题变为“从 $\{1, 2, \dots, n\}$ 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

$$P(X = i) = \frac{\binom{i-1}{k-1}}{\binom{n}{k}} \quad \text{离散型随机变量的分布列}$$

- 计算期望

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

- 统计估计: 如今采样得到 k 辆坦克的实际编号: $x_1, x_2, \dots, x_k$. 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

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

- 所以

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

- 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: $p_1 = 1$
 - 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 $X_1, X_2, \dots, X_n$ 有

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

- 或者

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