# 学习重点

本章将要学习几种特殊的离散型随机变量, 它们分别服从以下的分布

- 二项分布
- 泊松分布
- 几何分布

# 二项分布

## 定义

若一个随机变量的取值为 $0, 1, 2, \cdots, n$, 且对 $k = 0, 1, \cdots, n$,

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

这里我们称 $X$ 服从 二项分布, 记为 $X \sim B(n, p)$. 特别的, 如果 $n = 1$, 那么 $X$ 只有 0 和 1 两个取值, 我们称 $X$ 服从 两点分布. 进一步, 若 $p = 1$ 或者 0, 那么两点分布退化为 单点分布.

# 二项分布的性质

若 $X \sim B(n, p)$, 则

- $P(X = k) \geq 0$,
- 由二项式定理,

$\sum_k C_n^k \cdot p^k \cdot (1-p)^{n-k} = (p+1-p)^n = 1$

//

$$\sum_{k=0}^{n} P(X = k) = 1,$$

- $$P(X \leq k) = \sum_{i=0}^{k} b(i; n, p)$$

# 二项分布的例子

记 $n$ 重伯努利试验中成功的次数为 $X$, 且一次成功的概率为 $p$, 那么

$$X \sim B(n, p)$$

那么 $n$ 次试验中成功 $k$ 次的概率为

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

# 例 2.2.1

设 $X \sim B(10, 0.9)$, 求 $P(X = 8)$, $P(X \le 8)$, $P(3 \le X \le 9)$.

$P(X=8) = C_{10}^{8} \cdot 0.9^{8} \cdot (1-0.9)^{10-8}$

$P(X \le 8) = \sum_{k=0}^{8} P(X=k)$

$P(3 \le X \le 9) = P(X \le 9) - P(X < 3)$

$= [P(X \le 8) + P(X=9)]$

$- [\quad 0 \quad , 1, \quad 2 \quad ]$

# 例 2.2.2

导弹击中目标的概率为 0.96。
问：要发射多少枚导弹保证目标被击中的概率至少 0.999?

设 $n$ 为发射的导弹数目（$n$ 重伯努利实验）.

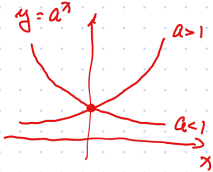
$P = 0.96$.

$X$：刚好命中的数目.

Q：求 $n$ 的取值范围

s.t. $P[X \geq 1] \geq 0.999$.

$$\begin{aligned} P[X \geq 1] &= 1 - \underbrace{P[X=0]} \\ &= 1-(1-0.96)^n \end{aligned}$$




解不等式方程 $1-(1-0.96)^n \geq 0.999$

$(0.04)^n \leq 0.001$

$n \cdot \log 0.04 \geq \log 0.001$

$n \geq \dfrac{\log 0.001}{\log 0.04}$ $\boxed{\approx \cancel{2.51}\ 2.14}$

# 课堂练习

- 设某种药物的治愈率是 0.6, 请问需要同时治疗多少位病人, 才能使得至少有一位病人的治愈率为 90%
- 设随机变量 $X \sim B(2, p)$, $Y \sim B(3, p)$, 若 $P(X \geq 1) = 5/9$, 求 $P(Y \geq 1)$.

设 $n$ 为治疗病人的数目.

$X$ 为刚好被治愈的数目.

则 $X \sim B(n, p)$ where $p = 0.6$

要求 $n$ 的取值范围

s.t. $P(X \geq 1) \geq 0.9$

得 $1-(1-0.6)^n \geq 0.9$

$0.1 \geq 0.4^n$

$\Rightarrow n \geq \frac{\log 0.1}{\log 0.4} \doteq 2.51$

$\because X \sim B(2, p)$

$\therefore P(X \geq 1) = 1 - P(X=0)$

$= 1-(1-p)^2$

解 $P(X \geq 1) = \frac{5}{9}$ 得 $p = \frac{1}{3}$.

$\therefore P(Y \geq 1) = 1 - P(Y=0)$

$= 1-(1-p)^3$

$= \frac{19}{27}$

# 二项分布的极值

## 定理 2.2.1

设 $X \sim B(n,p)$, 当 $(n+1)p$ 为整数 $m$ 时, $X$ 取 $m$ 和 $m-1$ 的概率最大, 且 $b(m;n,p) = b(m-1;n,p)$; 当 $(n+1)p$ 不为整数时, $X$ 取 $\lfloor (n+1)p \rfloor$ 的时候概率最大.

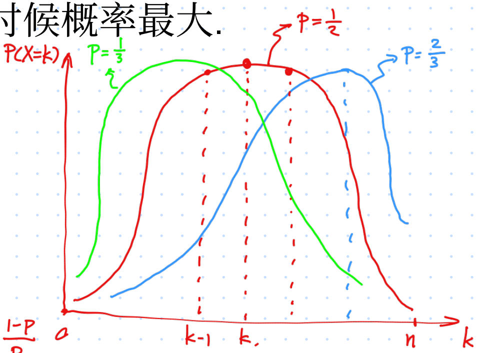
| $p=\frac{1}{2}$, $n$ | $(n+1)p$ | 最大 |
| --- | --- | --- |
| 2 | $\frac{2+1}{2}=\frac{3}{2}$ | 1 |
| 3 | $\frac{3+1}{2}=2$ | 1, 2 |

$$\frac{d}{dk}p(k) \underset{\approx}{\overset{\Delta k=-1}{}} \frac{p(k+\Delta k)-p(k)}{\Delta k} \quad \text{差分}$$

$$= p(k-1)-p(k) \to 0$$

$$\frac{p(k-1)}{p(k)} = \frac{C_n^{k-1}\cdot p^{k-1}(1-p)^{n-(k-1)}}{C_n^k\cdot p^k\cdot(1-p)^{n-k}} = \frac{\frac{n!}{(k-1)!(n-k+1)!}}{\frac{n!}{k!(n-k)!}} \times \frac{1-p}{p}$$

$$= \frac{k}{n-k+1} \times \frac{1-p}{p} \sim 1 \Rightarrow \frac{1}{p} \sim \frac{n+1}{k} \Rightarrow k \sim (n+1)p$$



# 例 2.2.3

为了估计一个鱼塘中鱼的总数, 我们做下面的试验

- 网起一兜鱼, 假设是 100 条, 给这些鱼做上标记.
- 把做过标记的鱼放回鱼塘中, 让这些放回去的鱼欢快的游一段时间.
- 再网起一兜, 假设是 80 条, 数出有标记的鱼的数目, 假设有 2 条.

怎样根据试验的结果, 估计鱼塘中鱼的总数呢?

# 例 2.2.3

定义随机变量 $X$ 表示 80 条鱼中有记号的鱼的个数, 那么 $X \sim B(80, \frac{100}{N})$, $X = 0, 1, 2, \cdots, 80$.
由于 小概率事件在一次试验中基本不发生, 即一次随机试验出现的结果是发生可能性大的, 或者是最大的, 那么根据定理 2.2.1, 有

$$2 = (80+1)\frac{100}{N}$$

即 $N = 4050$

Q: $A=\{$ 80条中刚好2条有标记 $\}$

找 $N$ s.t.

$P(N \mid A)$ 最大.

# 泊松分布

$X \sim B(n, p)$

$P(n=k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}$.

## 定义

如果一个随机变量只能取非负整数值, 并且

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots ,$$

则称 $X$ 服从 泊松 (Poisson) 分布, 记为
$X \sim Pois(\lambda)$.

泊松分布实际上是二项分布在 $p \to 0$, $n \to \infty$ 情形下的近似. 换句话说, 泊松分布描述的是 “稀有事件”

| $n$ | $P_n$ |
|---|---|
| 1 | 0.9 |
| 2 | 0.9/2 |
| 4 | 0.9/4 |
| $\vdots$ | $\vdots$ |

$X=k$.  $Y = \frac{X}{n}$.

# 泊松定理

## 泊松定理

设随机变量 $X \sim B(n, p_n)$, $0 < p_n < 1$ 是与 $n$ 有关的概率, 并且有 $\lim_{n\to\infty} np_n = \lambda$, 那么

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

# 泊松定理证明

记 $\lambda_n = np_n$, 那么

$$
\begin{aligned}
& b(k; n, p_n) \\
&= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\
&= \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \cdot
\end{aligned}
$$

$$
1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)
$$

# 泊松定理证明

固定 $k$, 有

$$\lim_{n\to\infty}\lambda_n^k=\lambda^k,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda}\left(hint:\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=e\right)$$

$$\left(\underbrace{\left(1+\frac{\lambda_n}{-n}\right)^{\frac{-n}{\lambda_n}}}_{\to e.}\right)^{-\lambda_n}$$

$$\frac{\lambda_n}{n}=\frac{1}{p_n}\to+\infty$$

以及

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{k-1}{n}\right)=1$$

故

$$\lim_{n\to\infty}b(k;n,p_n)=\frac{e^{-\lambda}\lambda^k}{k!}$$

# 泊松分布的应用

对于可以用二项分布 $X \sim B(n, p_n)$ 描述的随机变量, 如果同时满足

- $n \to \infty$, $p_n \to 0$
- $\lim_{n\to\infty} np_n = \lambda$

我们就可以用泊松分布代替二项分布.

Q: 如果 $n \to \infty$, $p_n \to 0$. s.t. $np_n \to \lambda$.

那么 $P_{pois} \sim P_b$

如果 $n \to \infty$, $p_n \to 1$

$\leadsto q_n = 1 - p_n \to 0$. s.t. $n \cdot q_n \to \lambda$

$P_{pois}(\overline{X = k}) = 1 - P( \quad )$

# 例 2.2.4

假如一个孕妇生三胞胎的概率是 $p = 10^{-4}$, 求在 100000 个孕妇中, 有 2 个人生三胞胎的概率.

解: 由于每个孕妇生几胞胎这个试验是相互独立的, 故定义随机变量 $X$ 为生三胞胎的人数, $X \sim B(100000, 10^{-4})$. 所以事件

$A_i = \{100000\text{个孕妇中有}i\text{个人生三胞胎}\}$

# 例 2.2.4

$$P(A_2) = P_{\text{binom}}(X = 2) = b(2, 10^5, 10^{-4})$$
$$= \binom{10^5}{2}(10^{-4})^2(1 - 10^{-4})^{10^5-2} = 0.002269293$$

注意 $p \ll 1$ 且 $n \gg 1$, $pn = 10$, 我们用泊松分布来计算一下

$$= \frac{e^{-10} \cdot 10^2}{2!}$$

$$P(A_2') = P_{\text{pois}}(X = 2) = 0.00226996$$

近似效果还是可以的, 绝对误差在第 5 位有效数字上.

# 例 2.2.5

同类设备有 80 台, 各台设备是否正常工作是独立事件, 每台设备的故障率是 $p = 0.01$, 一台设备发生故障时需要安排一人维修. 求

- 一人负责维修 20 台设备的时候, 设备发生故障无人维修的概率.

# 例 2.2.5

当一个人负责 20 台设备的时候, 只要同时有两台及以上发生故障的时候, 就有设备没有人去维修, 所以要求的事件是
$A = \{$至少有两台设备同时发生故障$\}$.
定义随机变量 $X = \{$同时发生故障的设备台数$\}$,
由试验条件, $X \sim B(20, p)$. 计算

$$\begin{aligned} P(A) &= P_{\text{binom}}(X \geq 2) \\ &= 1 - P_{\text{binom}}(X \leq 1) = 0.01685934 \end{aligned}$$

改用泊松分布的话, 可以求得

$$P(A) = P_{\text{pois}}(X \geq 2) = 0.0175231$$

# 例 2.2.5

同类设备有 80 台, 各台设备是否正常工作是独立事件, 每台设备的故障率是 $p = 0.01$, 一台设备发生故障时需要安排一人维修. 求

- 由三个人共同负责维修 80 台设备的时候, 设备发生故障无人维修的概率.

# 例 2.2.5

当三个人共同负责 80 台设备的时候, 只要有四台以及以上的设备同时发生故障, 就有设备没有人去维修, 定义事件
$A = \{$至少有四台设备同时发生故障$\}$.

$$P(A) = P_{\text{binom}}(X \geq 4) = 1 - P(X \leq 3) = 0.008659189$$

用泊松分布计算

$$P(A) = P_{\text{pois}}(X \geq 4) = 0.009079858$$

当 $n$ 不是很大的时候, 泊松分布对二项分布的近似并不理想.