

# Practice of AI

## C2: 数据分布

Jim Xie

2020/3/6

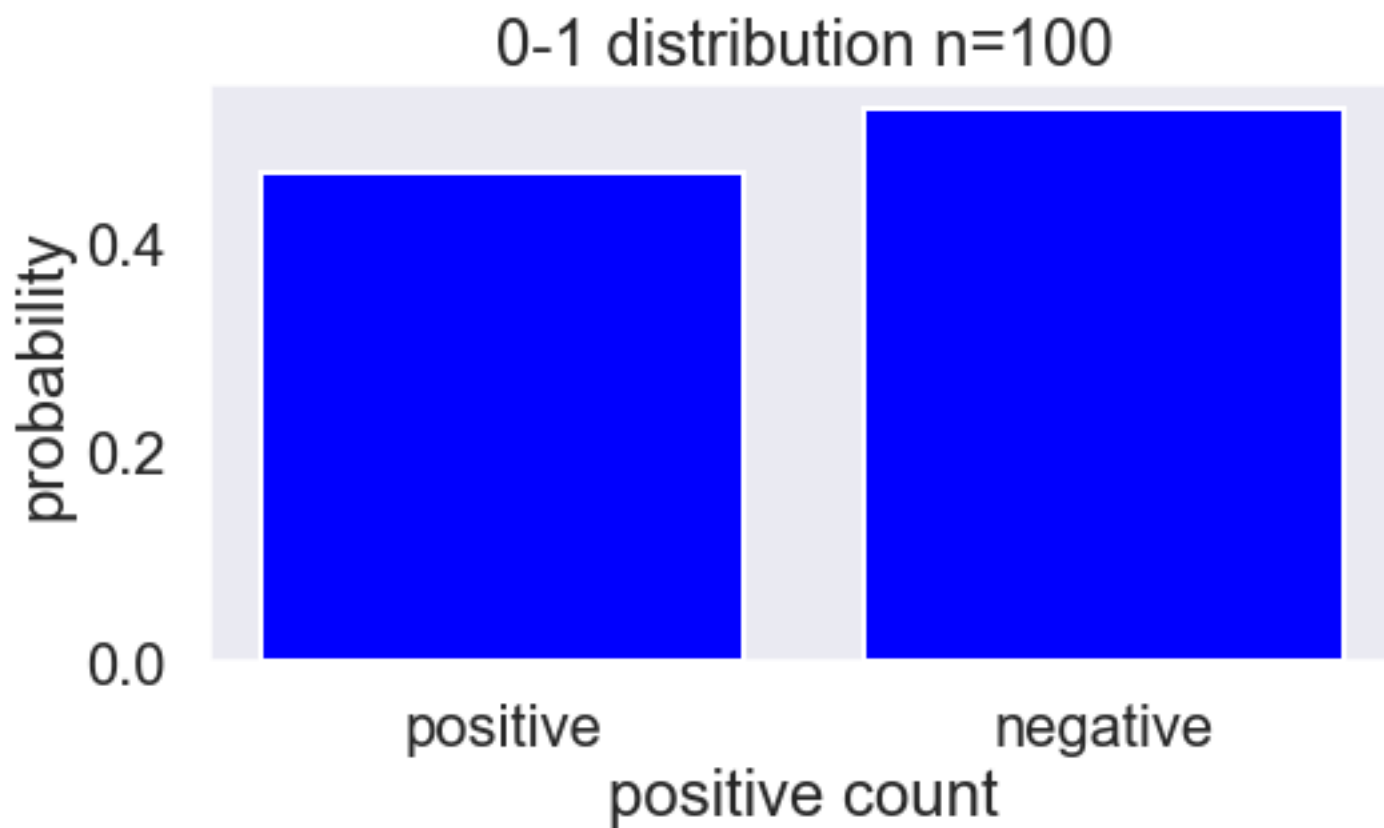
# 数据分布

---

- 二项分布
- 泊松分布
- 正太分布
- 几何分布
- 指数分布

# 二值分布

---



# 二项分布

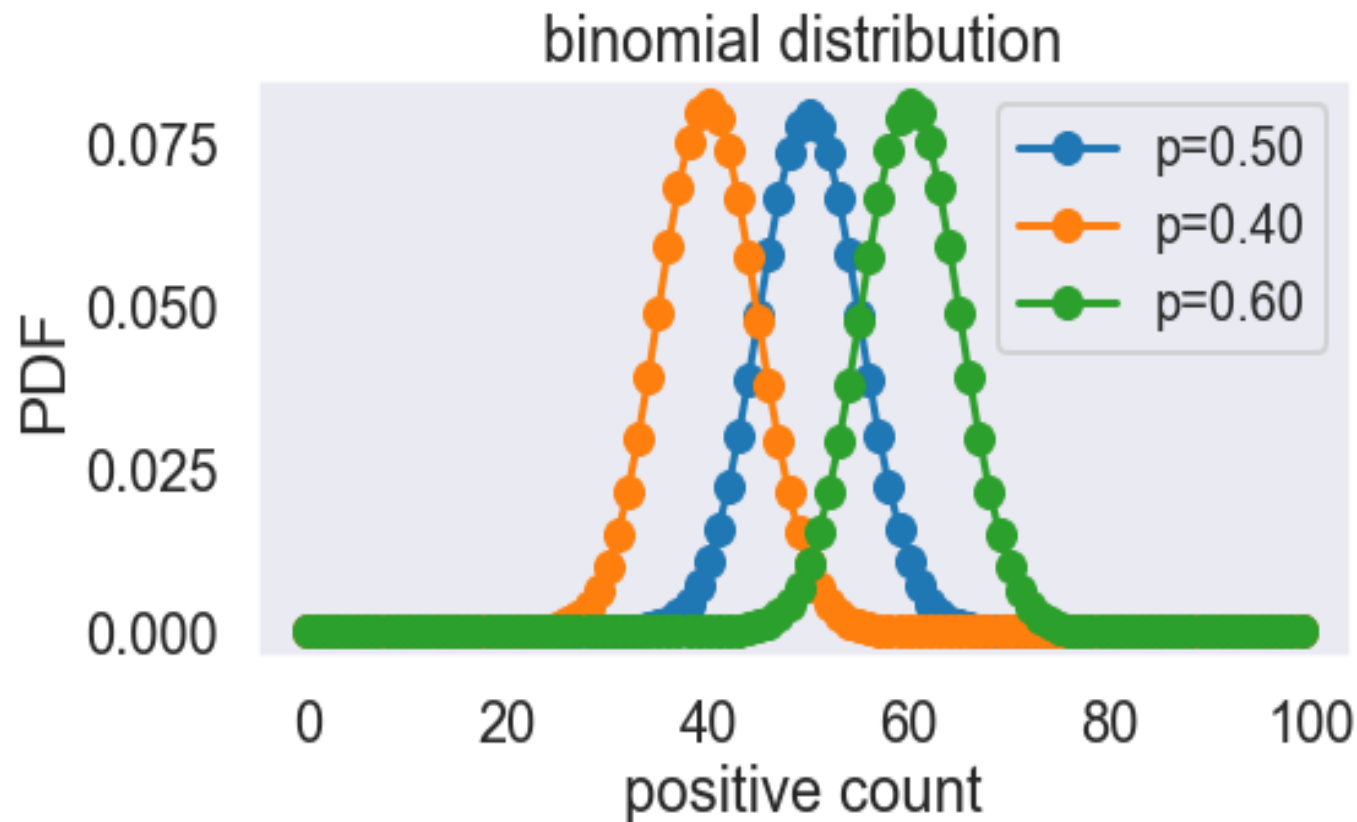
抛100次硬币，正面朝上的次数

举例:

最近每天有0.5个case  
这个月会接到多少个case?

$$E(X) = np$$

$$D(X) = np(1 - p)$$



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

# 几何分布

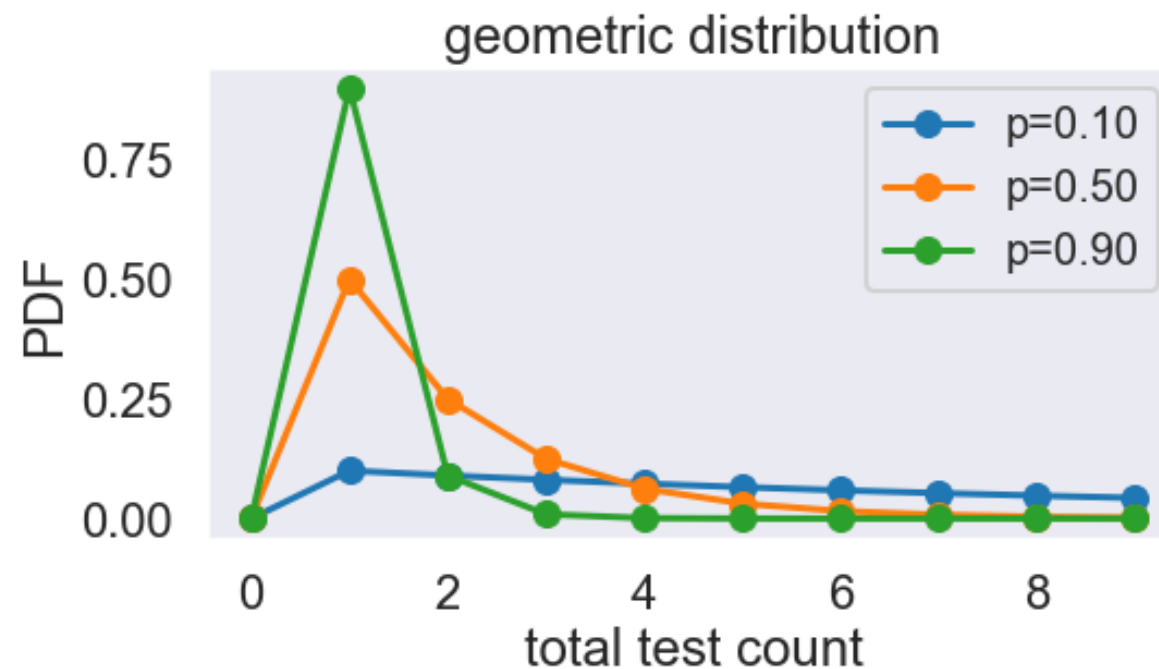
最少抛几次硬币，会出现正面朝上

举例:

客户产品上线，多久可以detect到virus?

$$E(x) = \frac{1}{p}$$

$$D(x) = \frac{1}{p^2} - \frac{1}{p}$$



$$P(X = k) = (1 - p)^{k-1} p, k = 1, 2, \dots$$

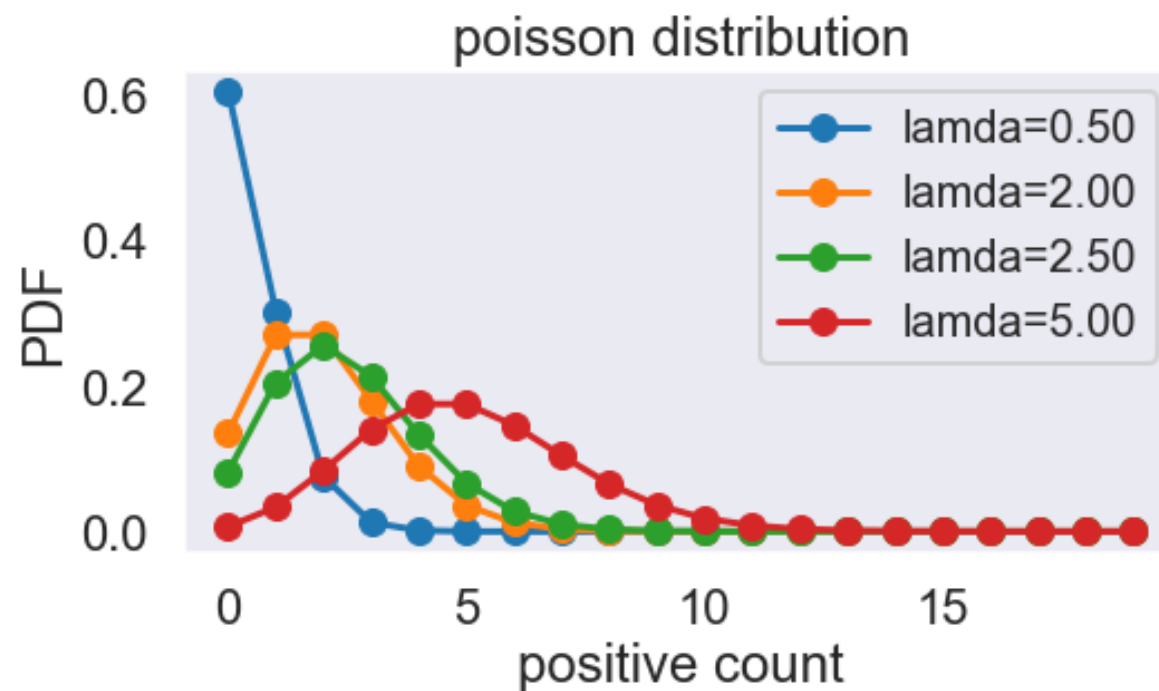
# 泊松分布

单位时间内，内发生k次事件的概率是

举例:

最近每天有0.5个case  
明天会有多少个case ?

$$E(x) = D(x) = \lambda$$



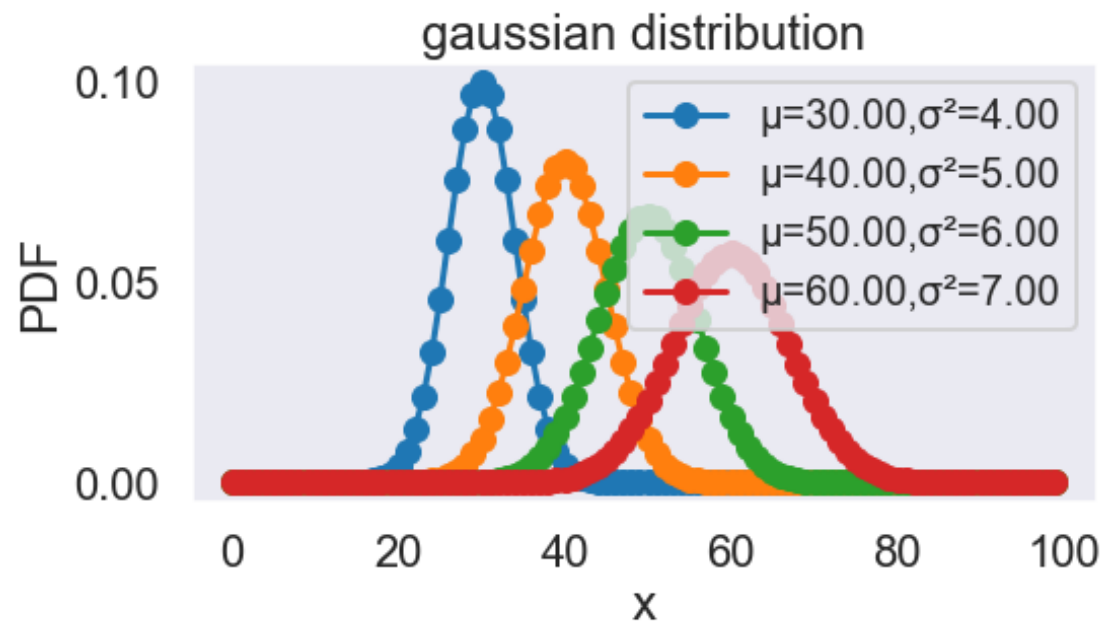
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots$$

# 正太分布

描述事件发生的概率

应用:

已知期望与标准差，进行预测  
多种AI算法的基础，  
线性回归的前提



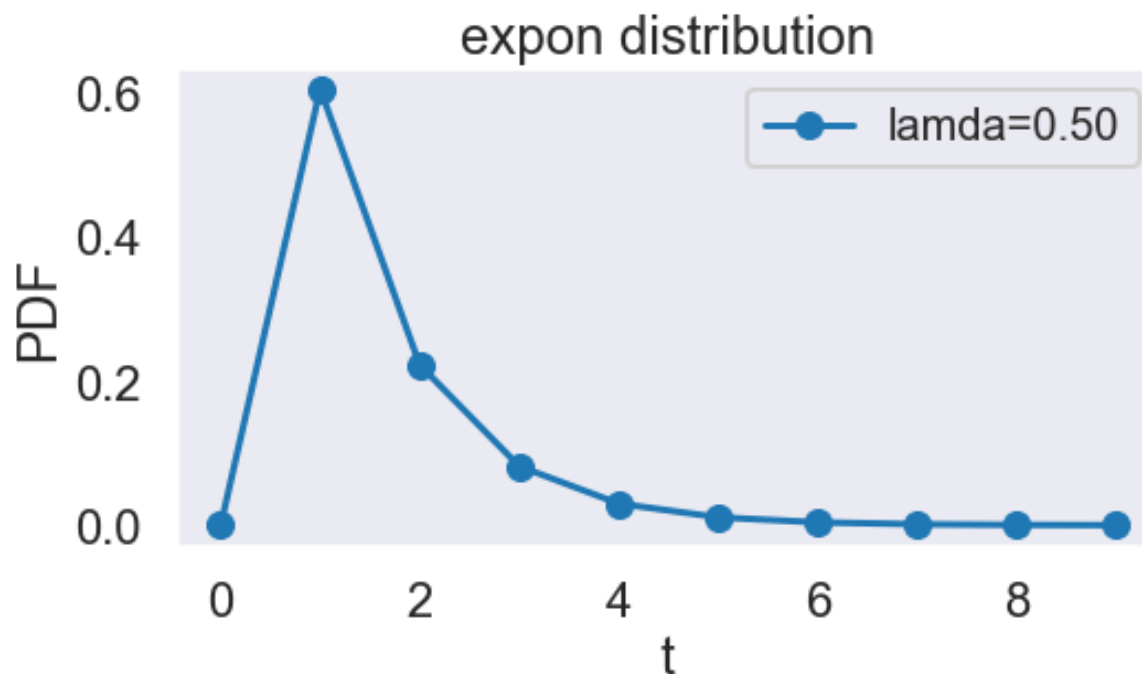
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

# 指数分布

事件的时间间隔的概率

应用:

婴儿出生的时间间隔  
来电的时间间隔



$$E(x) = \frac{1}{\lambda}$$
$$D(x) = \frac{1}{\lambda^2}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0, & x < 0, \end{cases}$$