概率

顾立平

01理解概率

- 概率是关于量化对尚未发生的事件的预测，而可能性是衡量已经发生的事件的频率。在统计学中和机器学习，我们通常使用数据形式的似然（过去）来预测概率（未来）。
- 所有可能性事件的可能互斥结果（意味着只有一个结果可以发生，而不是多次）的总和必须为1.0或100%。
- 将概率O(X)转换为比例概率P(X，可用这个公式：

$$P(X)\,\frac{O(X)}{1+O(X)}$$

02概率与统计学

- 有时，人们可以互换使用概率和统计学这两个术语，把这两个学科混为一谈是可以理解的，它们确实有区别。

➢ 概率纯粹是关于事件发生的可能性的理论，不需要数据。

➢ 另一方面，统计数据没有数据就不可能存在，使用它来发现概率，并提供描述数据的工具。

- 当我们处理事件P(X)的单个概率时，称为边际概率，这个想法相当简单。
- 但当我们开始组合不同事件的概率，它变得不那么直观了。

04联合概率

$$P(A\ \text{AND}\ B) = P(A) \times P(B)$$

$$P(\text{heads}) = \frac{1}{2}$$

$$P(6) = \frac{1}{6}$$

- 假设有一枚公平的硬币和一个公平的六面骰子。我们想在硬币和骰子上分别翻转头部和点数六。
- 这是两个单独事件的单独概率，我们希望找到这两个事件将同时发生的概率；即：联合概率。

$$P(\text{heads AND } 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} = .08\overline{333}$$

- 将联合概率想象为AND运算符。
- 硬币上有两个面，模具上有六个面，因此正面的概率是1/2，6的概率是1/6。两个事件发生的概率，就是简单地相乘。

04联合概率

- 许多概率规则通过生成来自以下区域的所有可能的事件组合：
- 当抛硬币和掷骰子时，有12种可能的结果。我们唯一感兴趣的是“H6”
- H1 H2 H3 H4 H5*H6*T1 T2 T3 T4 T5 T6
- 离散数学称为排列和组合。
- 我们可以再次使用乘法作为求联合概率的捷径；即：产品规则（product rule）：

$$P(A\ \text{AND}\ B) = P(A) \times P(B)$$

$$P(\text{heads AND } 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} = .08\overline{333}$$

05并集概率

- **互斥事件**：如果我掷一个骰子，我不能同时得到4和6。我只有一个结果。获得这些情况的并集概率是很容易的。我只是简单地将它们加在一起。
- 与之相反，这些事件如果可以同时发生，即：**非互斥事件**。

$$P(heads) = \frac{1}{2}$$

$$P(6) = \frac{1}{6}$$

$$P(A \text{ OR } B) = P(A) + P(B) - P(A) \times P(B)$$

$$P(\text{heads OR } 6) = \frac{1}{2} + \frac{1}{6} - \left(\frac{1}{2} \times \frac{1}{6}\right) = .58\overline{333}$$

05并集概率

- **联合概率**，即两个或多个事件的概率同时发生。但得到事件A或B的概率呢？
- 当我们处理具有概率的OR运算时，即：**并集概率**。

$$P(A\ \mathrm{OR}\ B) = P(A) + P(B) - P(A\ \mathrm{AND}\ B)$$

$$P(A\ \mathrm{OR}\ B) = P(A) + P(B) - P(A) \times P(B)$$

- 当两个或多个事件之间存在并集概率时，如果不是互斥的，请确保减去联合概率，以便没有概率重复计数。

06条件概率与贝叶斯定理

- 给定事件B发生的事件A发生的概率。

$$P(\mathrm{A\ GIVEN\ B})\ \text{or}\ P(\mathrm{A}|\mathrm{B})$$

- 假设一项研究声称85%的癌症患者喝咖啡。我们是否放弃喝咖啡呢？
- 首先，我们把它定义为条件概率：

$$P(\text{Coffee given Cancer})\ \text{or}\ P(\text{Coffee}|\text{Cancer})$$

06条件概率与贝叶斯定理

$$P(\text{Coffee}) = .65$$

$$P(\text{Cancer}) = .005$$

$$P(\text{Coffee}|\text{Cancer}) = .85$$

- 我们把癌症确诊人数的百分比与咖啡饮用者的百分比，做比较：
  - 癌症患者（根据cancer.gov，0.5%）和饮酒的比例。
  - 咖啡（根据statista.com，65%）。

- 在任何给定条件下，只有0.5%的人口患有癌症。
- 然而，65%的人口定期喝咖啡。如果咖啡引发癌症，那么我们的癌症数量应该比0.5%高得多，接近65%了？

06条件概率与贝叶斯定理

- 人们如此容易被条件概率混淆的原因是因为条件的方向很重要：”你是一个喝咖啡的人，有可能患癌“不同于”如果你患有癌症，那么你可能会去喝咖啡“。简单地说：喝咖啡的人很少有癌症，但许多癌症患者喝咖啡。
- 如果我们有兴趣研究咖啡是否会致癌，我们是对第一条件概率感兴趣：某人拥有癌症的概率，作为一位喝咖啡的人。
- 如何翻转条件？有一个强大的小公式，我们可以用它来翻转条件概率：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

06条件概率与贝叶斯定理

- 把已有信息代入上述公式中，我们可以求解：如果某人喝咖啡，那么他患癌症的可能性。

```
p_coffee_drinker = .65

p_cancer = .005

p_coffee_drinker_given_cancer = .85

p_cancer_given_coffee_drinker = p_coffee_drinker_given_cancer *
p_cancer / p_coffee_drinker

print(p_cancer_given_coffee_drinker)
```

07联合和并条件概率

- 联合概率，以及它们如何与条件概率相互作用。
- 我想知道有人喝咖啡得了癌症的可能性，要用哪一个？

Option 1:

$$P(\text{Coffee}) \times P(\text{Cancer}) = .65 \times .005 = .00325$$

Option 2:

$$P(\text{Coffee}|\text{Cancer}) \times P(\text{Cancer}) = .85 \times .005 = .00425$$

- 如果我们已经确定我们的概率仅适用于癌症患者，用P（Coffee | Cancer) 代替P（Coffee）咖啡有意义吗？

07联合和并条件概率

- 如果我们没有任何条件概率可用，那么能做的是：P(Coffee)乘以P(Cancer)答案是0.325%
- 如果P（Cancer）已经是我们联合概率的一部分。这意味着某人患癌症并喝咖啡的概率为0.425%。

$$P(\text{Coffee and Cancer}) = P(\text{Coffee}|\text{Cancer}) \times P(\text{Cancer}) = .85 \times .005 = .00425$$

$$P(\text{Cancer}|\text{Coffee}) \times P(\text{Coffee}) = .0065 \times .65 = .00425$$

- 如果事件A对事件B没有影响，这意味着P（B|A）=P（B），表示事件A的发生对事件B发生的可能性没有影响。因此，更新的联合概率公式：

$$P(A \text{ AND } B) = P(B) \times P(A|B)$$

07联合和并条件概率

- 最后，如果我们想计算A或B发生的概率，但A可能影响B的概率，我们得要更新规则：

  $$P(\ A\ OR\ B) = P(\ A\ ) + P(\ B) - P(\ A|B\ ) \times P(\ B\ )$$

- 这也适用于互斥事件。

  $$P(A|B) \times P(B)$$

08二项分布

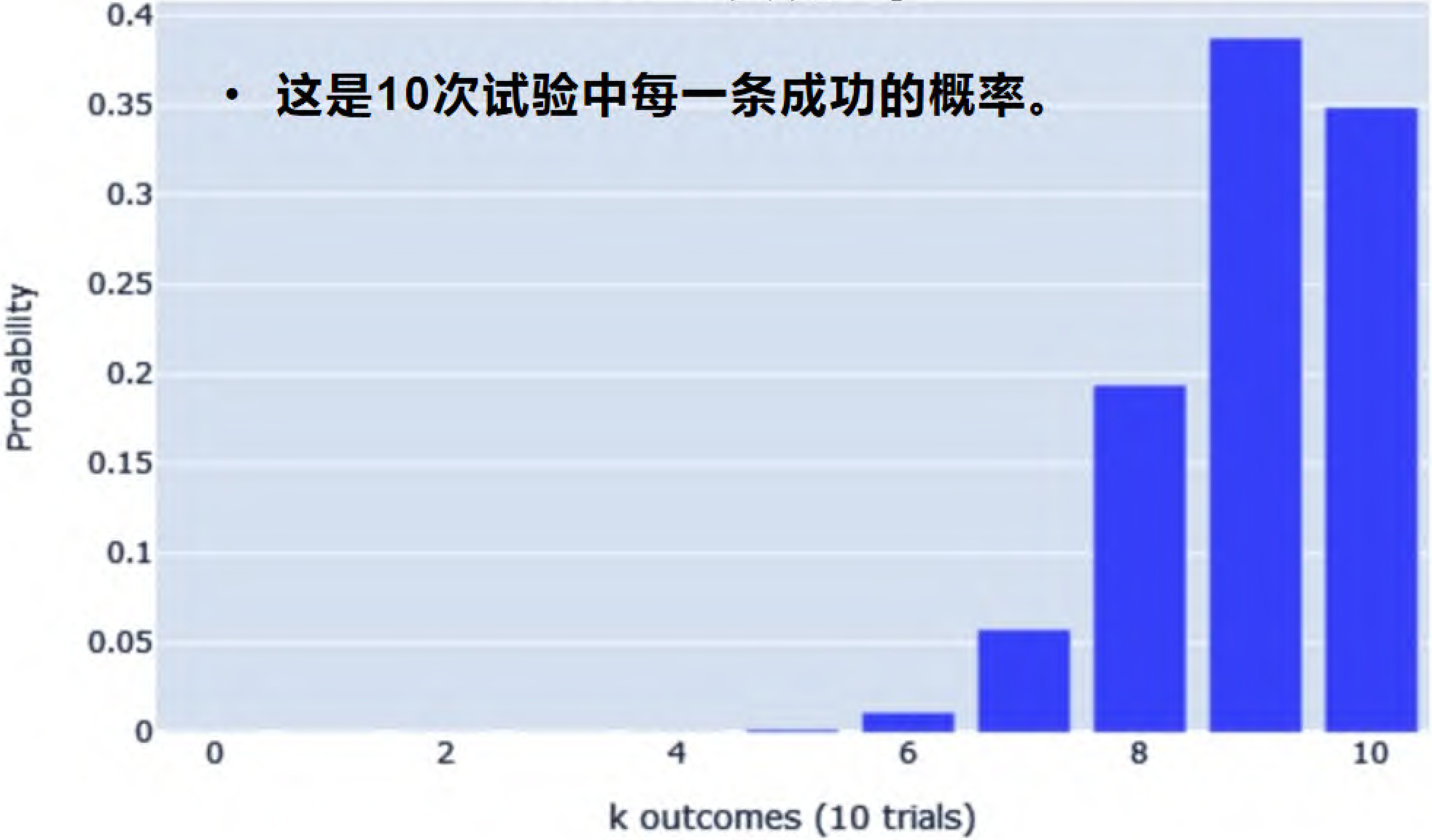
- 假设我们正在开发一个新的涡轮喷气发动机，进行了10次测试。结果：

  ✓ ✓ ✓ ✓ ✓ ✗ ✓ ✗ ✓ ✓

- 我们希望获得90%的成功率，但结论只有80%的成功。每个测试都很耗时，并且昂贵，因此决定图板的重新设计。然而，一位工程师坚持认为应该有更多的测试。“我们唯一的办法是要进行更多的测试”。毕竟，如果你把一枚硬币掷了10次，得到了8个头，这并不意味着硬币固定在80%。
- 二项式分布，用于：测量如何在给定p概率的情况下，n次试验中可能有k次成功。

08二项分布

- 这是10次试验中每一条成功的概率。



08二项分布

- 这个二项式分布假设概率p为90%，意即，存在.90（或90%）的机会以实现成功。

- 如果这是真的，那意味着有0.1937的概率，我们会在10次试验中获得8次成功。在10次试验中获得1次成功的概率是0.000000008999，极不可能。

```
from scipy.stats import binom
n = 10
p = 0.9
for k in range(n + 1):
    probability = binom.pmf(k, n, p)
    print("{0} - {1}".format(k, probability))
```

- 我们提供n作为试验的数量，p作为成功的概率对于每个试验，k是成功的次数

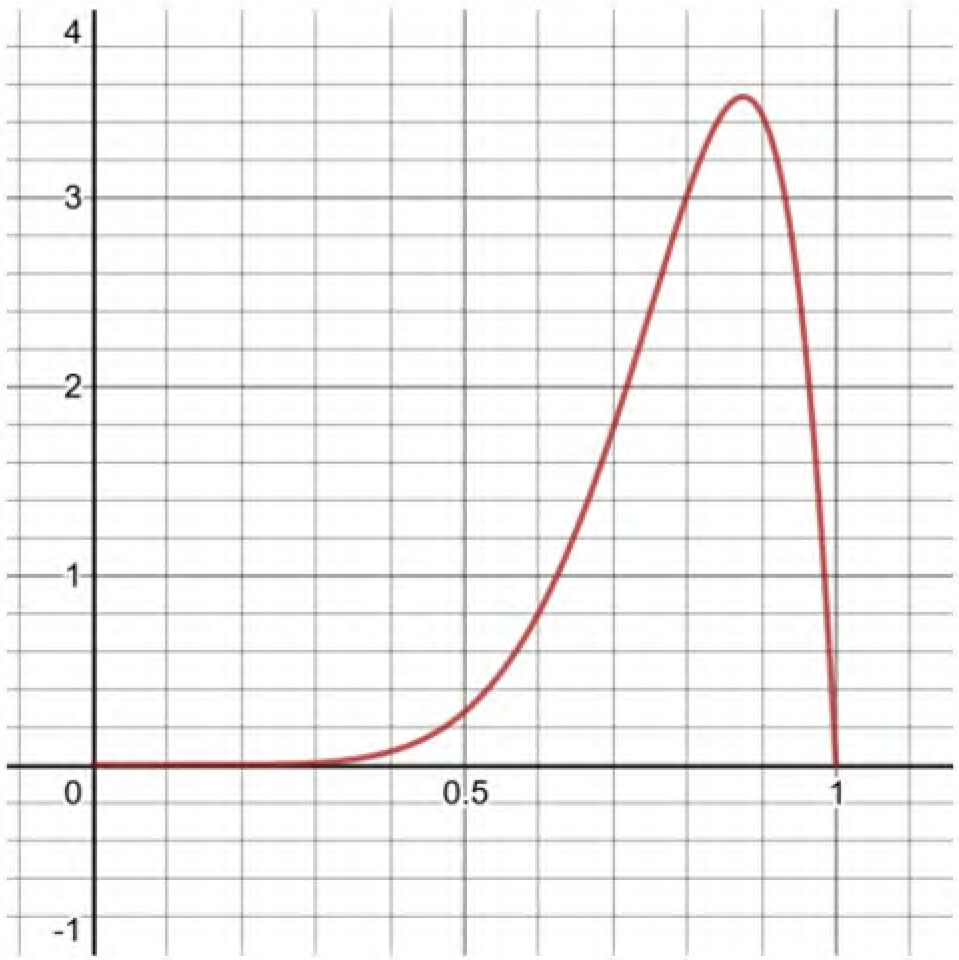
08二项分布

- 我们用相应的概率来迭代每个成功次数x将会看到许多成功。正如我们在输出中看到的，最可能的数字成功的总数是九。

- 但如果我们将八个或更少的成功概率相加，我们将得到0.2639。这意味着我们有26.39%的机会看到8个或更少的成功，即使潜在成功率为90%。所以也许工程师是对的：26.39%的可能性不是什么都没有，当然是可能的。

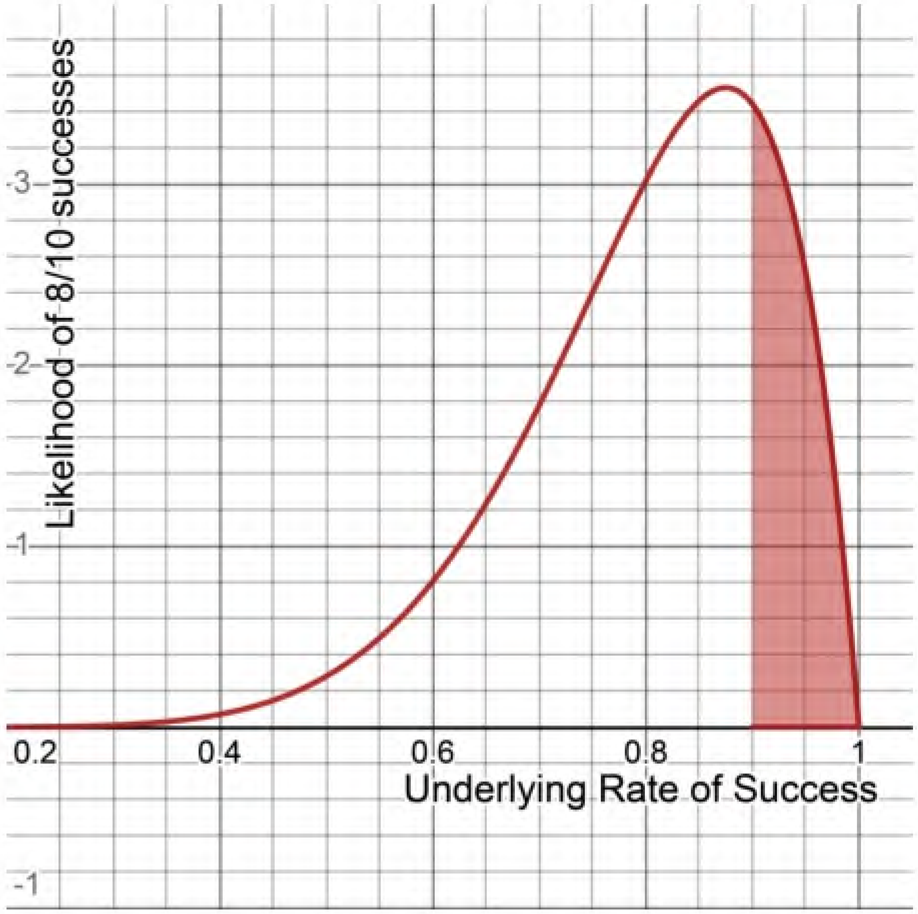
- 此处，可能有问题的是：我们假设成功率为90%（潜在的成功率）。根据模型，我们有26.39%的机会，在8个次少于10次就能试验成功；因为，基本成功率为90%嘛。但让我们翻转这个问题，并且考虑一下：如果存在其他基础利率，该怎么办？Beta分布。

09Beta分布



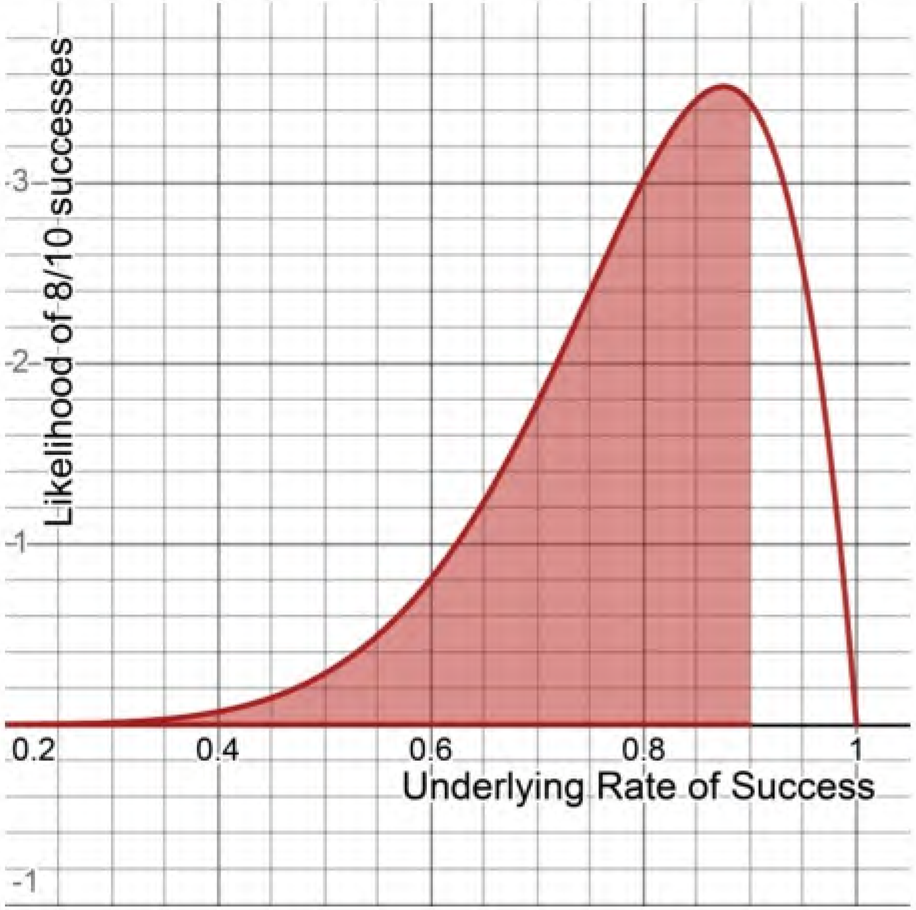
- Beta分布，没有创建无数的二项式分布，而是：给定alpha成功和beta失败，事件发生的不同潜在概率。
- 如左图所示，给定八次成功和两次失败的beta分布。
- 左图x轴表示从0.0到1.0（0%）的所有基本成功率到100%），y轴表示给定8的概率的可能性成功和两次失败。换句话说，beta分布允许我们看到给定8/10成功的概率的概率。其视为元概率。

09Beta分布



- beta分布是一个连续函数，它形成了一个十进制值的连续曲线。
- 因为y轴上的给定密度值不是概率。所以，我们使用曲线下面积计算概率。
- β分布是一种概率分布，它表示整个曲线为1.0或100%。要找到概率，我们需要找到一个范围。例如，如果我们想评估8/10成功的概率成功率为90%或更高，我们需要找到0.9到1.0之间的区域，如左图所示，这片区域为0.225。

09Beta分布

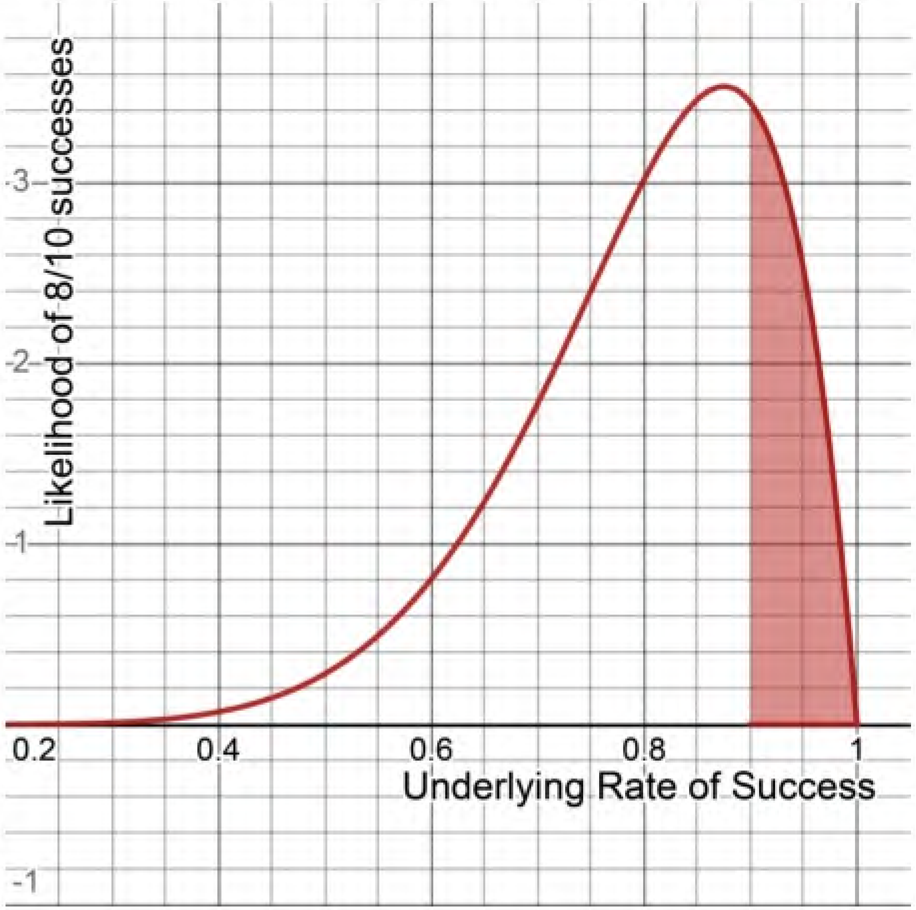


- 每个连续概率分布都有一个累积密度函（CDF），它计算达到给定x值的面积。假设我想计算面积高达90%（0.0至0.90），如左图所示。

```
from scipy.stats import beta

a = 8

b = 2

p = beta.cdf(.90, a, b)
```

- 根据计算，潜在概率为77.48%成功率为90%或更少。

09Beta分布

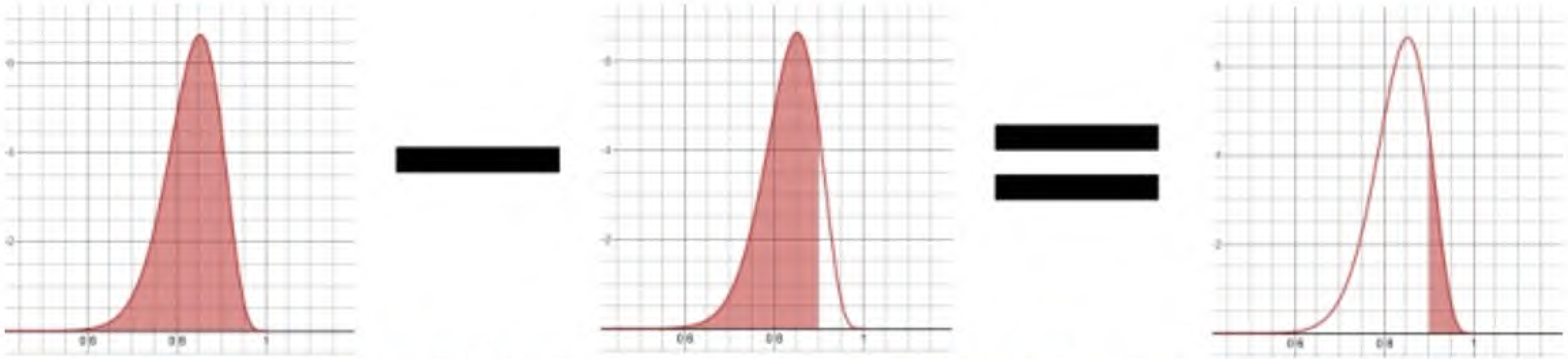


- 如何计算成功概率为90%或更高（如左图所示）？

```
from scipy.stats import beta
a = 8
b = 2
p = 1.0 - beta.cdf(.90, a, b)
print(p)
```

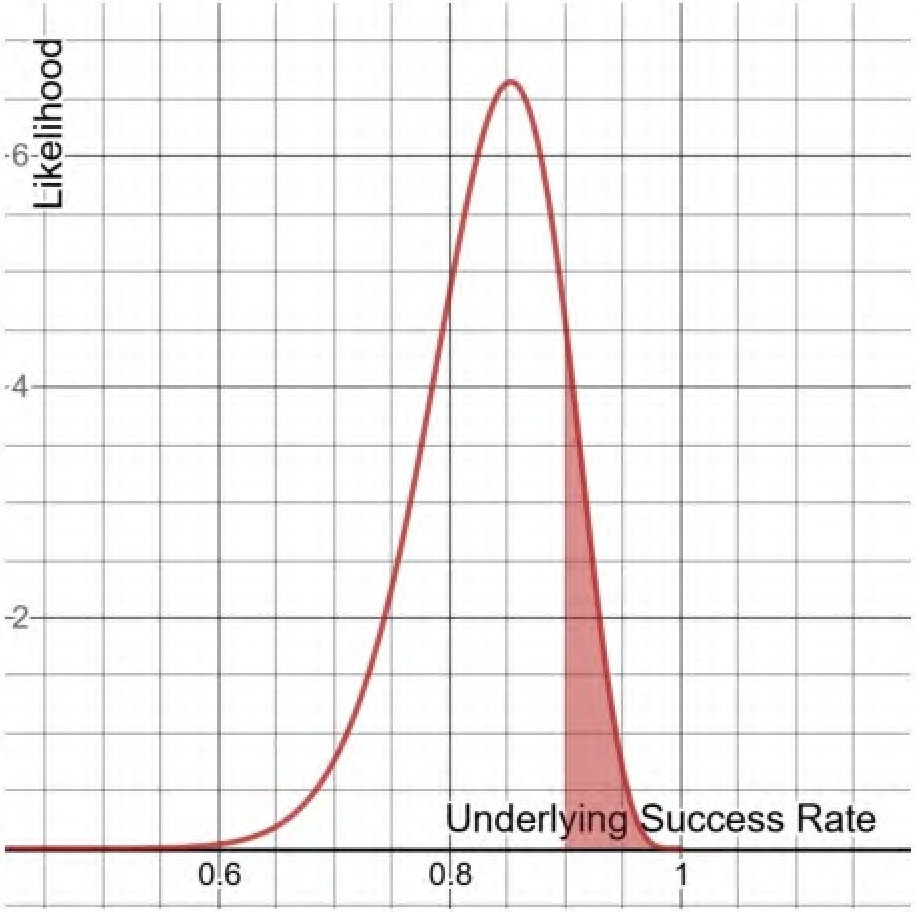
- 这似乎能够猜想得到。然而，真正理解这么做的含义（定义），往往更为重要。

09Beta分布



- 如果我们想找到事件的相反概率（大于0.90与小于0.90相反），只需从1.0中减去小于0.90的概率，剩余概率将大于0.90。
- 计算结果是0.22515902199999993这意味着在8/10个成功的发动机测试中，只有22.5%的机会潜在成功率为90%或更高。但有大约77.5%的几率是较少的超过90%。显然，测试结果不乐观，除非我们可以就22.5%的机会，赌一赌（做更多发动机的测试）。

09Beta分布

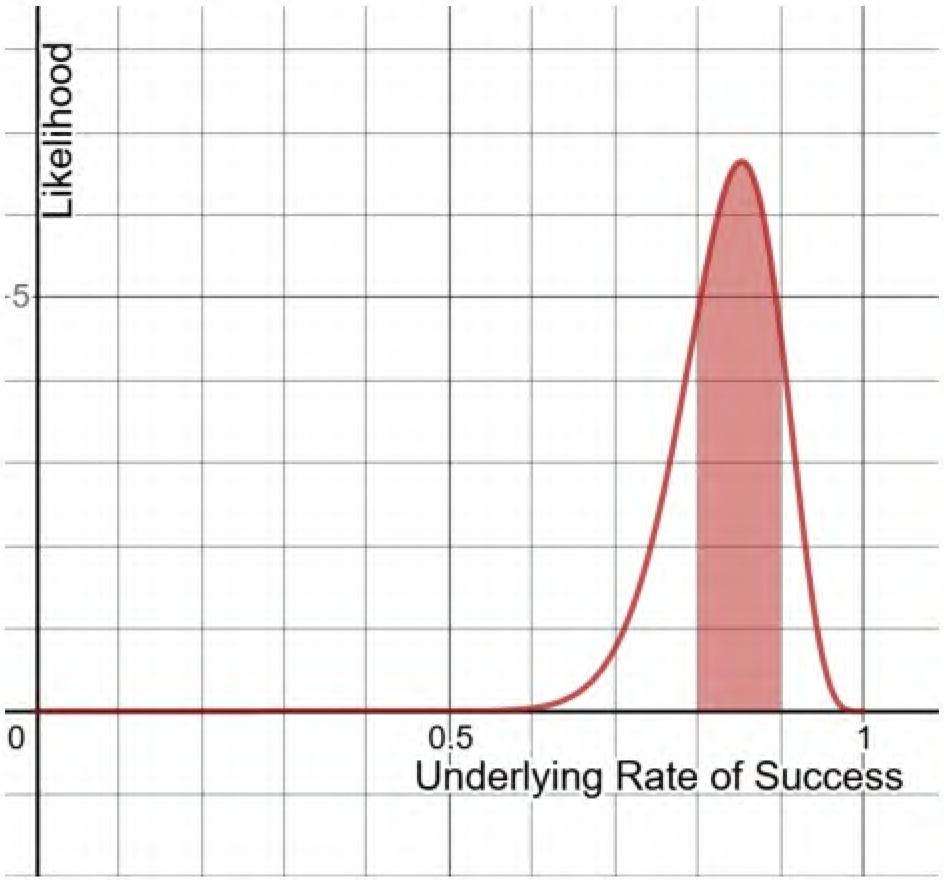


- 如果资金能够支持另外26个测试，从而导致30次成功，6次失败，则如左图所示。

```
from scipy.stats import beta
a = 30
b = 6
p = 1.0 - beta.cdf(.90, a, b)
print(p)
```

- 现在分布变得更窄，更确信潜在的成功率，坐落在较小的范围内。而且，要达到90%成功率的概率已从22.5%降至13.16%了。

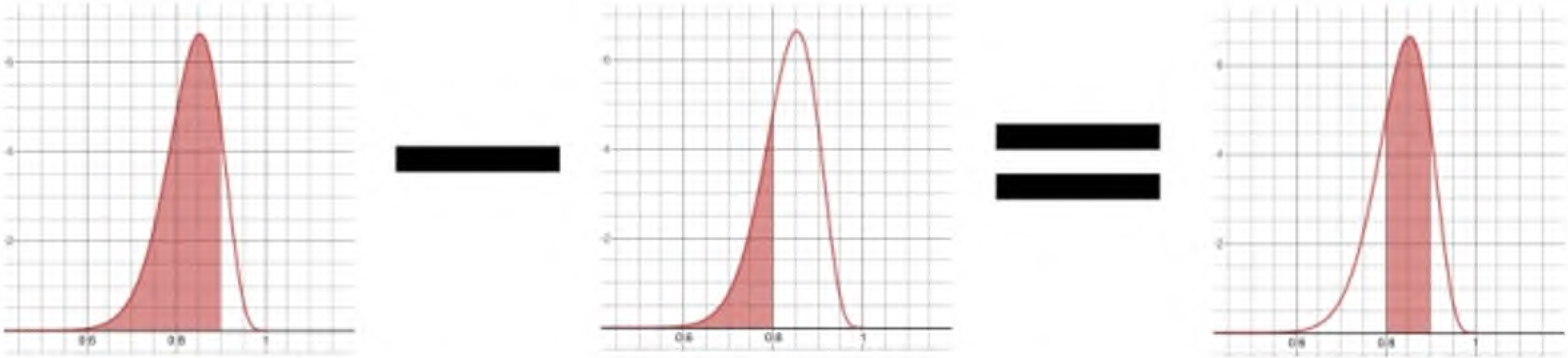
09Beta分布



- 同样重要的是，如何计算中间的面积？如果我们想要发现潜在成功率在80%到90%之间的概率，如左图所示。

```
from scipy.stats import beta
a = 8
b = 2
p = beta.cdf(.90, a, b) - beta.cdf(.80, a, b)
print(p)
```

09Beta分布



- 这能给我们0.80到0.90之间的面积。它的概率为0.3386到33.86%的区域。
- 贝塔分布是一种测量事件概率的迷人工具：发生与不发生，基于有限的一组观察。它允许我们解释概率的概率，我们可以在获得新数据时更新它。
- Beta分布也可用于假设检验，不过人们更加强调使用正态分布和T分布。

谢谢！

gulp@mail.las.ac.cn