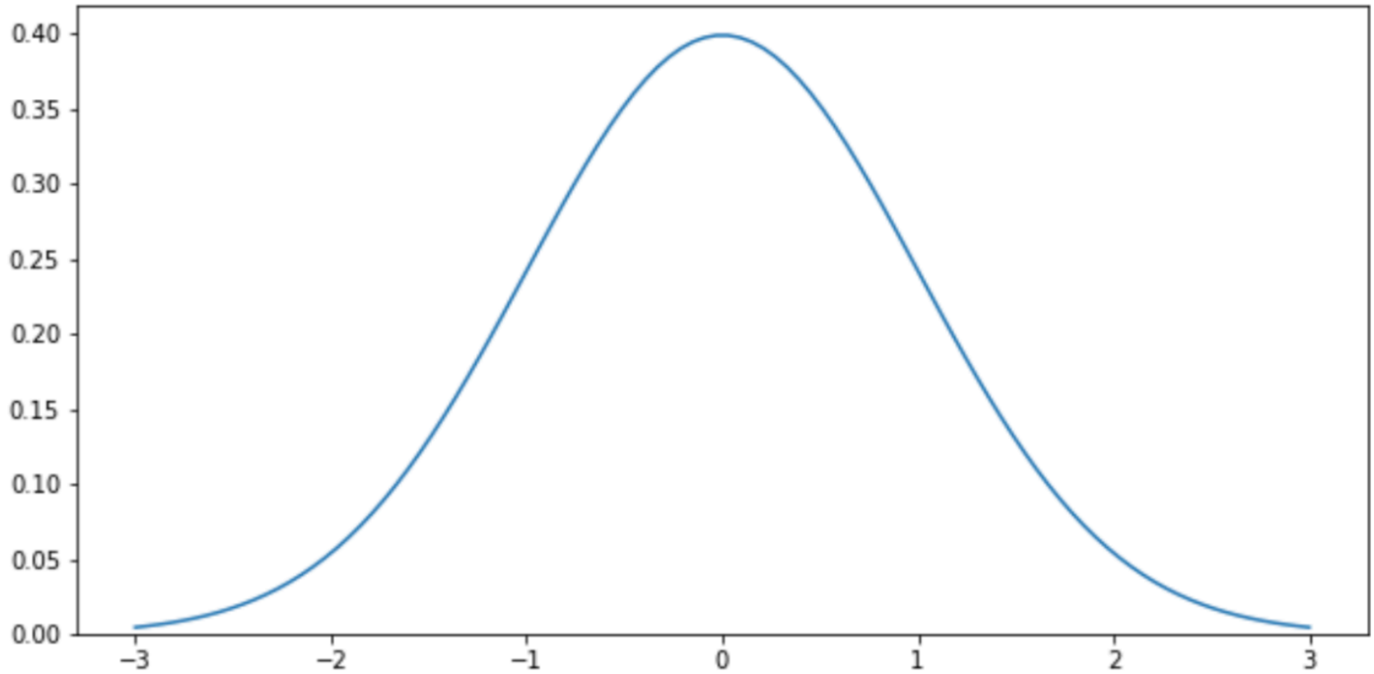
泊松分布

一、概率分布

- 假设Z是一些随机变量。Z 的概率分布函数将概率分配给 Z。从图形上看，概率值与曲线的高度成正比。



我们可以将随机变量分为三类：

- 离散discrete：离散随机变量只能假设指定列表中的值。人口、电影收视率等都是离散的随机变量。
- 连续continuous：连续随机变量可以取任意精确值。例如，温度、速度、时间等都是连续的随机变量。
- 混合mixed：混合随机变量将概率分配给离散和连续随机变量，即它是上述两个类别的组合。

二、期望

- 期望值（Expected value ，EV）是概率中最重要的概念之一。给定概率分布的 EV 可以描述为"来自该分布的许多重复样本（随机变量）的平均值。

三、泊松分布

- **泊松分布**是一种离散概率分布，它表示给定数量的事件在固定的时间或空间间隔内发生的概率，如果这些事件以已知的恒定平均速率发生，并且独立于自上次事件以来的时间。
- 例如，呼叫中心每天 24 小时平均每小时接听 180 个电话。调用是独立的；收到一个不会改变下一个何时到达的概率。
- 如果Z符合泊松分布，则：

$$P(Z=k)=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\ldots,\lambda \in R$$

其中，λ 被称为分布的参数，它控制分布的形状。对于泊松分布，λ 可以是任何正数。通过增加 λ，我们为较大的值添加更多概率，相反，通过减少 λ，我们为较小的值添加更多概率。可以将λ描述为泊松分布的强度。

k 必须是非负整数，即 k 必须取值 0、1、2 等。

- 泊松分布的一个有用特性是它的期望值等于它的参数，即：

  $E[Z\,|\,\lambda]=\lambda$

- 现在我相信更清楚为什么增加λ会为更大的值分配更多的概率，反之亦然。

四、实例

- FIFA世界杯比赛的平均进球数约为2.5，泊松分布模型是合适的。因为平均事件发生率为每场比赛 2.5 个进球，即 $\lambda = 2.5$。

$$P(k \text{ goals in a match})=\frac{2.5^k e^{-2.5}}{k!}$$

$$P(k=0 \text{ goals in a match})=\frac{2.5^0 e^{-2.5}}{0!}=\frac{e^{-2.5}}{1}\approx 0.082$$

$$P(k=1 \text{ goal in a match})=\frac{2.5^1 e^{-2.5}}{1!}=\frac{2.5e^{-2.5}}{1}\approx 0.205$$

$$P(k=2 \text{ goals in a match})=\frac{2.5^2 e^{-2.5}}{2!}=\frac{6.25e^{-2.5}}{2}\approx 0.257$$

- 在上面的例子中，我使用了假设 Poisson 模型是正确的短语，现在问题出现了，这个模型什么时候失败或者不能做出假设。让我们看另一个例子：
- 每分钟到达学生会的学生人数可能不会服从泊松分布，因为该比率不是恒定的（上课时间的比率低，上课时间之间的比率高），并且个别学生的到达不是独立的（学生往往成群结队）。
- 如果一次大地震增加了类似震级余震的概率，则一个国家每年发生 5 级地震的次数可能不遵循泊松分布。

- 因此，在假设泊松模型为真之前，请先问自己两个问题：
  - **事件的发生率是否在一定程度上是恒定的**
  - **事件是否相互独立地发生**