

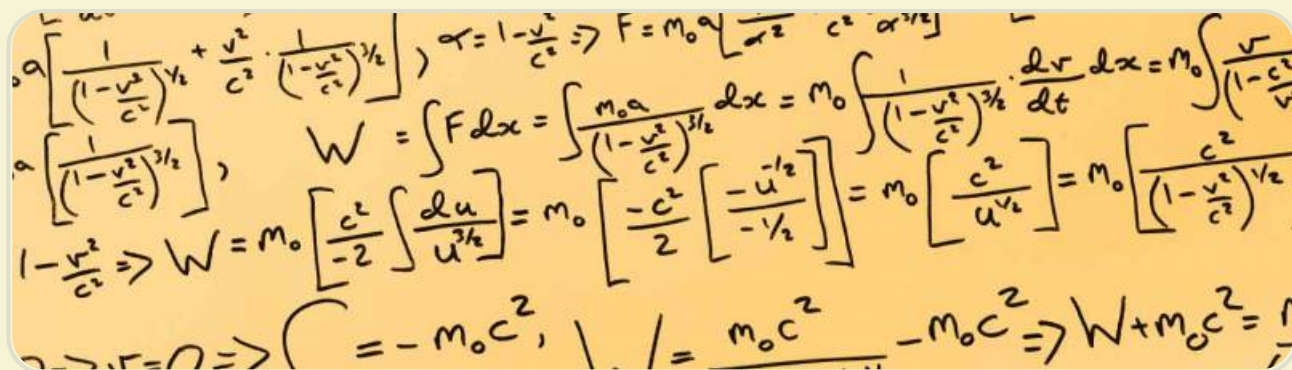
泊松分布和指数分布：10分钟教程

作者：阮一峰

日期：2015年6月10日

大学时，我一直觉得统计学很难，还差点挂科。

工作以后才发现，难的不是统计学，而是我们的教材写得不好。比起高等数学，统计概念其实容易理解多了。



The image shows handwritten mathematical derivations on a yellow background. The top part shows the derivation of the relativistic energy formula $E = \gamma m_0 c^2$ from the definition of force $F = \frac{dp}{dt}$ and work $W = \int F dx$. It starts with $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$ and $p = \gamma m_0 v$. The work W is calculated as $\int F dx = \int \frac{dp}{dt} dx = \int \frac{dp}{dv} \frac{dx}{dt} dv = \int \frac{dp}{dv} v dv$. Then it shows $\frac{dp}{dv} = \frac{d}{dv} (\gamma m_0 v) = m_0 \frac{d\gamma}{dv}$. The final result is $W = \gamma m_0 c^2 - m_0 c^2$, which is the relativistic energy minus the rest energy.

我举一个例子，什么是泊松分布和指数分布？恐怕大多数人都说不清楚。

我可以在10分钟内，让你毫不费力地理解这两个概念。

一、泊松分布

日常生活中，大量事件是有固定频率的。

- 某医院平均每小时出生3个婴儿
- 某公司平均每10分钟接到1个电话
- 某超市平均每天销售4包xx牌奶粉
- 某网站平均每分钟有2次访问

它们的特点就是，我们可以预估这些事件的总数，但是没法知道具体的发生时间。已知平均每小时出生3个婴儿，请问下一个小时，会出生几个？



有可能一下子出生6个，也有可能一个都不出生。这是我们没法知道的。

泊松分布就是描述某段时间内，事件具体的发生概率。

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

上面就是泊松分布的公式。等号的左边，P 表示概率，N表示某种函数关系，t 表示时间，n 表示数量，1小时内出生3个婴儿的概率，就表示为 $P(N(1) = 3)$ 。等号的右边， λ 表示事件的频率。

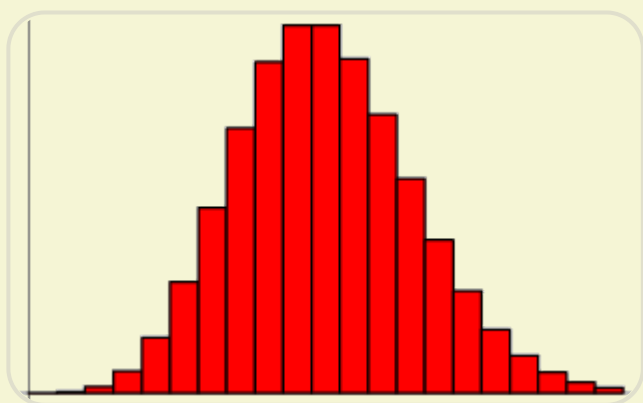
接下来两个小时，一个婴儿都不出生的概率是0.25%，基本不可能发生。

$$P(N(2) = 0) = \frac{(3 \times 2)^0 e^{-3 \times 2}}{0!} \approx 0.0025$$

接下来一个小时，至少出生两个婴儿的概率是80%。

$$\begin{aligned} P(N(1) \geq 2) &= 1 - P(N(1) = 1) - P(N(1) = 0) \\ &= 1 - \frac{(3 \times 1)^1 e^{-3 \times 1}}{1!} - \frac{(3 \times 1)^0 e^{-3 \times 1}}{0!} \\ &= 1 - 3e^{-3} - e^{-3} \\ &= 1 - 4e^{-3} \\ &\approx 0.8009 \end{aligned}$$

泊松分布的图形大概是下面的样子。



可以看到，在频率附近，事件的发生概率最高，然后向两边对称下降，即变得越大和越小都不太可能。每小时出生3个婴儿，这是最可能的结果，出生得越多或越少，就越不可能。

二、指数分布

指数分布是事件的时间间隔的概率。下面这些都属于指数分布。

- 婴儿出生的时间间隔
- 来电的时间间隔
- 奶粉销售的时间间隔
- 网站访问的时间间隔

指数分布的公式可以从泊松分布推断出来。如果下一个婴儿要间隔时间 t ，就等同于 t 之内没有任何婴儿出生。

$$\begin{aligned} P(X > t) &= P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \\ &= e^{-\lambda t} \end{aligned}$$

反过来，事件在时间 t 之内发生的概率，就是1减去上面的值。

$$P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}$$

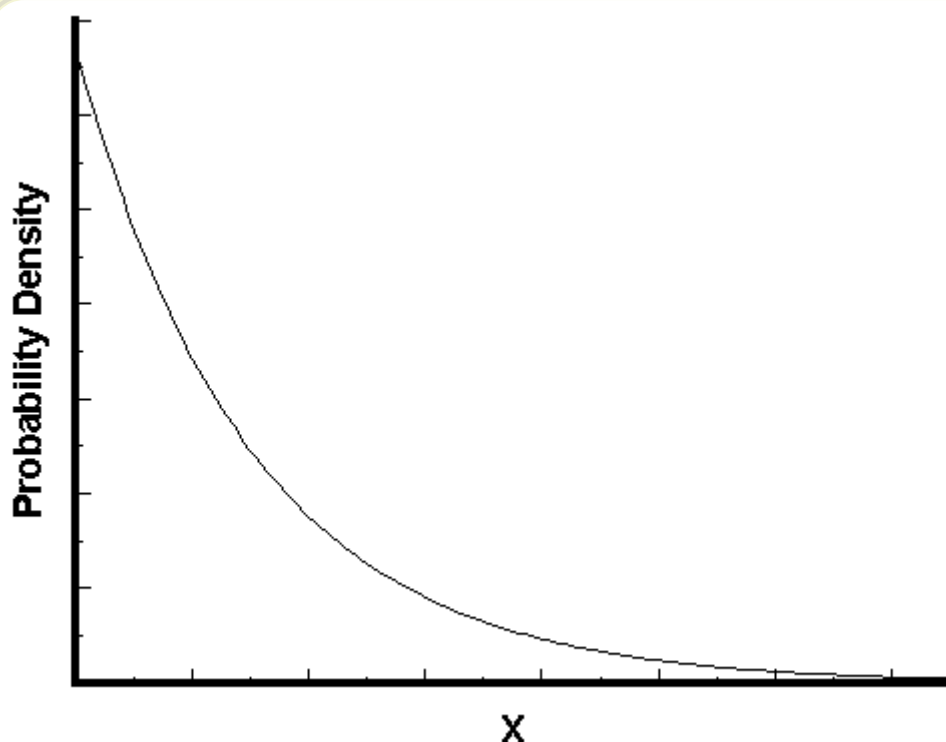
接下来15分钟，会有婴儿出生的概率是52.76%。

$$\begin{aligned} P(X \leq 0.25) &= 1 - e^{-3 \times 0.25} \\ &\approx 0.5276 \end{aligned}$$

接下来的15分钟到30分钟，会有婴儿出生的概率是24.92%。

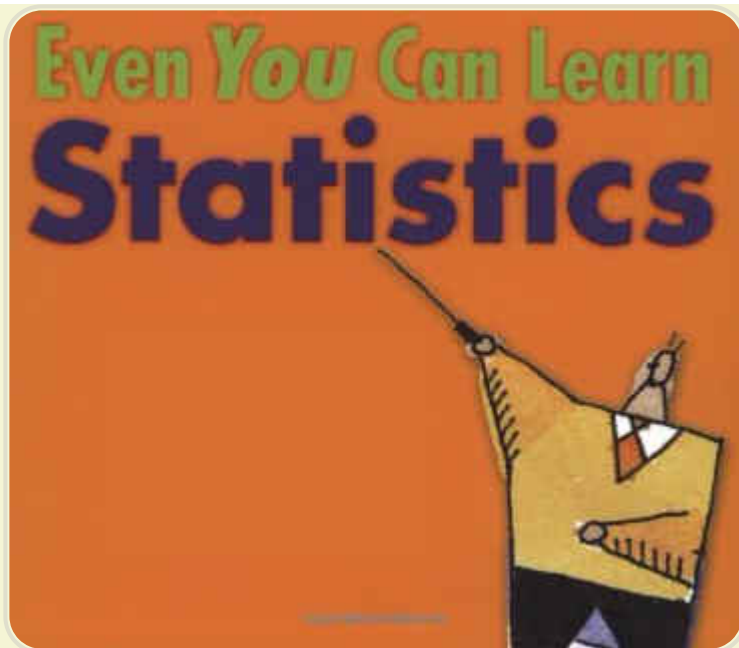
$$\begin{aligned}P(0.25 \leq X \leq 0.5) &= P(X \leq 0.5) - P(X \leq 0.25) \\&= (1 - e^{-3 \times 0.5}) - (1 - e^{-3 \times 0.25}) \\&= e^{-0.75} - e^{-1.5} \\&\approx 0.2492\end{aligned}$$

指数分布的图形大概是下面的样子。



可以看到，随着间隔时间变长，事件的发生概率急剧下降，呈指数式衰减。想一想，如果每小时平均出生3个婴儿，上面已经算过了，下一个婴儿间隔2小时才出生的概率是0.25%，那么间隔3小时、间隔4小时的概率，是不是更接近于0？

三、总结



一句话总结：泊松分布是单位时间内独立事件发生次数的概率分布，指数分布是独立事件的时间间隔的概率分布。

请注意是"独立事件"，泊松分布和指数分布的前提是，事件之间不能有关联，否则就不能运用上面的公式。

[说明] 本文受到 [nbviewer 文档](#) 的启发。

（正文完）

=====

以下为广告部分。欢迎大家在我的网络日志[投放广告](#)，推广自己的产品。

【赞助商广告】

帮最好的程序员 发现更好的 OFFER

100offer
程序员拍卖

在我之前的博客中，给大家推荐过"帮最优秀的工程师找工作"的 [100offer](#) 。时隔半年，我收到了一封朋友来信，讲自己的职业选择，文中也提到了 [100offer](#) 。读完后，不禁有些感触，推荐给大家。

08年大学毕业，迄今七年，共有两份工作经历，都在大公司。2011年社招进入华为南研所（南京研究所），从Java+Flex开发到MDE，再后来转SE，在华为来说，这个节奏不算多快但也不算慢，走得挺稳当，与领导、同事相处也很融洽。

只是，渐渐会有一些不安，得空反思，发觉自己在做的这些事，于个人成长而言价值不大。譬如我会用jQuery，它简洁、高效，但我并不知道为何可以这样；再如HashMap、HashTable主要区别在于线程安全与否，但原理是什么我还不清楚??

今年年初，决心离开南研所，并离开南京这座城市.....（[点击此处阅读原文](#)）

同时决定文尾再帮我的赞助商 [100offer](#) 做次推广，1次申请，10个优质offer，1份更好的工作。有需求的程序员们去注册申请拍卖吧。

（完）

文档信息

- 版权声明：自由转载-非商用-非衍生-保持署名（创意共享3.0许可证）
- 发表日期：2015年6月10日

相关文章

■ 2021.09.22: [俄罗斯总理的几何题](#)

9月1日是俄罗斯的知识节，因为这一天是各级学校的开学日，象征进入知识宝库的日子。

■ 2019.01.28: [Prolog 语言入门教程](#)

Prolog 是一种与众不同的语言，不用来开发软件，专门解决逻辑问题。比如，"苏格拉底是人，人都会死，所以苏格拉底会死"这一类的问题。

■ 2018.09.05: [哈希碰撞与生日攻击](#)

一、哈希碰撞是什么？所谓哈希（hash），就是将不同的输入映射成独一无二的、固定长度的值（又称"哈希值"）。它是最常见的软件运算之一。

■ 2017.12.13: [图像与滤波](#)

我对图像处理一直很感兴趣，曾经写过好几篇博客（1，2，3，4）。



weibo.com/yifengruan | twitter.com/yifengruan | github.com/yifengruan

Email: yifeng.ruan@gmail.com