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。

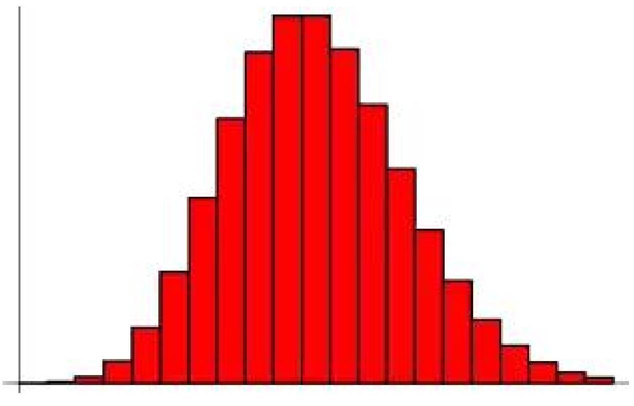
接下来两个小时，一个婴儿都不出生的概率是0.25%，基本不可能发生。

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

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。

泊松分布使用范围

Poisson分布主要用于描述在单位时间(空间)中稀有事件的发生数. 即需满足以下四个条件:

1、给定区域内的特定事件产生的次数，可以是根据时间，长度，面积来定义;

2、各段相等区域内的特定事件产生的概率是一样的;

3、各区域内，事件发生的概率是相互独立的;

4、当给定区域变得非常小时，两次以上事件发生的概率趋向于0。

例如:

1、放射性物质在单位时间内的放射次数；

2、在单位容积充分摇匀的水中的细菌数；

3、野外单位空间中的某种昆虫数等。

泊松分布的期望和方差

由泊松分布知E[N(t) − N(t0)] = D[N(t) − N(t0)] = λ(t − t0)

特别的，令t_0=0.由于假设N(0)=0,故可推知泊松过程的均值函数和方差函数分别为E[N(t)] = λt,D[N(t)] = λt,

泊松过程的强度lambda （常数）等于单位长时间间隔内出现的质点数目的期望值。即对泊松分布有：E(X) = D(X) = λ

泊松分布的特征

1、泊松分布是一种描述和分析稀有事件的概率分布。要观察到这类事件，样本含量n必须很大。

2、λ是泊松分布所依赖的唯一参数。λ值愈小，分布愈偏倚，随着λ的增大，分布趋于对称。

3、当λ = 20时，分布泊松接近于正态分布；当λ = 50时，可以认为泊松分布呈正态分布。在实际工作中，当$\lambda \geq 20$时就可以用正态分布来近似地处理泊松分布的问题。

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。如果下一个婴儿间隔时间 t ，就等同于 t 之内没有任何婴儿出生。

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

$$= e^{-\lambda t}$$

反过来，事件在时间 t 之内发生的概率（至少出生一个的概率），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%。

$$P(X \leq 0.25) = 1 - e^{-3 \times 0.25}$$

$$\approx 0.5276$$

接下来的15分钟到30分钟，会有婴儿出生的概率是24.92%。

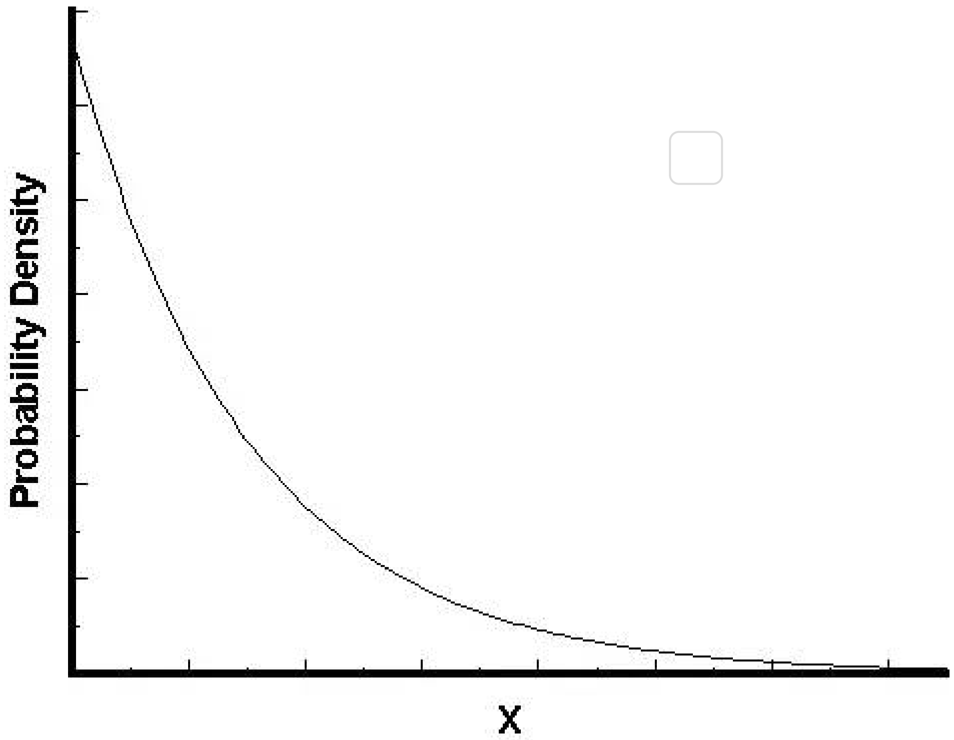
$$P(0.25 \leq X \leq 0.5) = P(X \leq 0.5) - P(X \leq 0.25)$$

$$= (1 - e^{-3\times 0.5}) - (1 - e^{-3\times 0.25})$$

$$= e^{-0.75} - e^{-1.5}$$

$$\approx 0.2492$$

指数分布的图形大概是下面的样子。



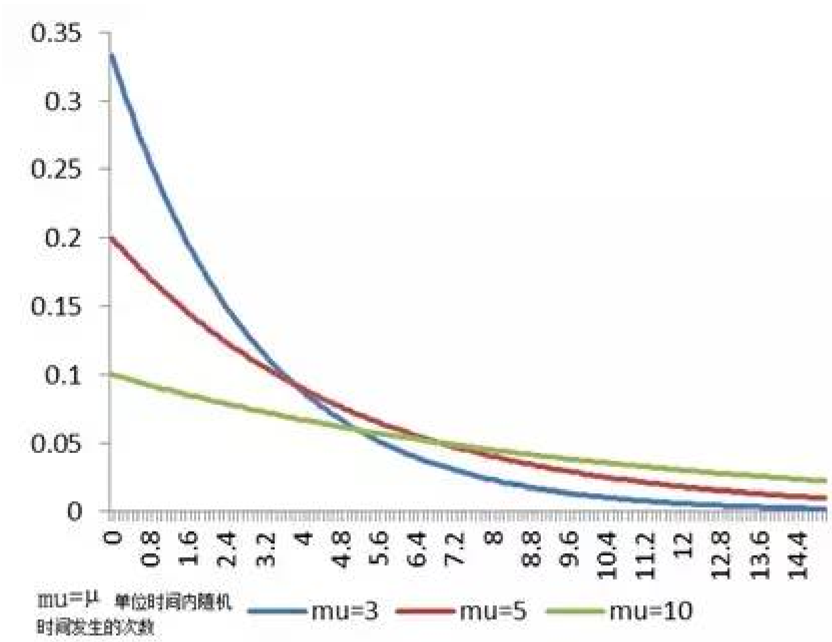
可以看到，随着间隔时间变长，事件的发生概率急剧下降，呈指数式衰减。想一想，如果每小时平均出生3个婴儿，上面已经算过了，下一个婴儿间隔2小时才出生的概率是0.25%，那么间隔3小时、间隔4小时的概率，是不是更接近于0?

指数分布的概率密度为：

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

式中：x是给定的时间；λ为单位时间事件发生的次数；e＝2.71828。

指数分布概率密度曲线如下图：



指数分布具有以下特征：

（1）随机变量X的取值范围是从0到无穷；

（2）极大值在x＝0处，即f(x)＝λ；

（3）函数为右偏，且随着x的增大，曲线稳步递减；

（4）随机变量的期望值和方差为$\mu = 1/\lambda$，$\sigma^2 = 1/\lambda^2$。

通过对概率密度函数的积分，就可以得到相应的概率，其表达式有两种

$P(X \geq x) = e^{-\lambda x}$

$P(X \le x) = 1 - e^{-\lambda x}$

例：某电视机生产厂生产的电视机平均10年出现大的故障，且故障发生的次数服从泊松分布。

问（1）该电视机使用15年后还没有出现大故障的比例；（2）如果厂家想提供大故障免费维修的质量担保，但不能超过全部产量的20%，试确定提供担保的年数。

解：

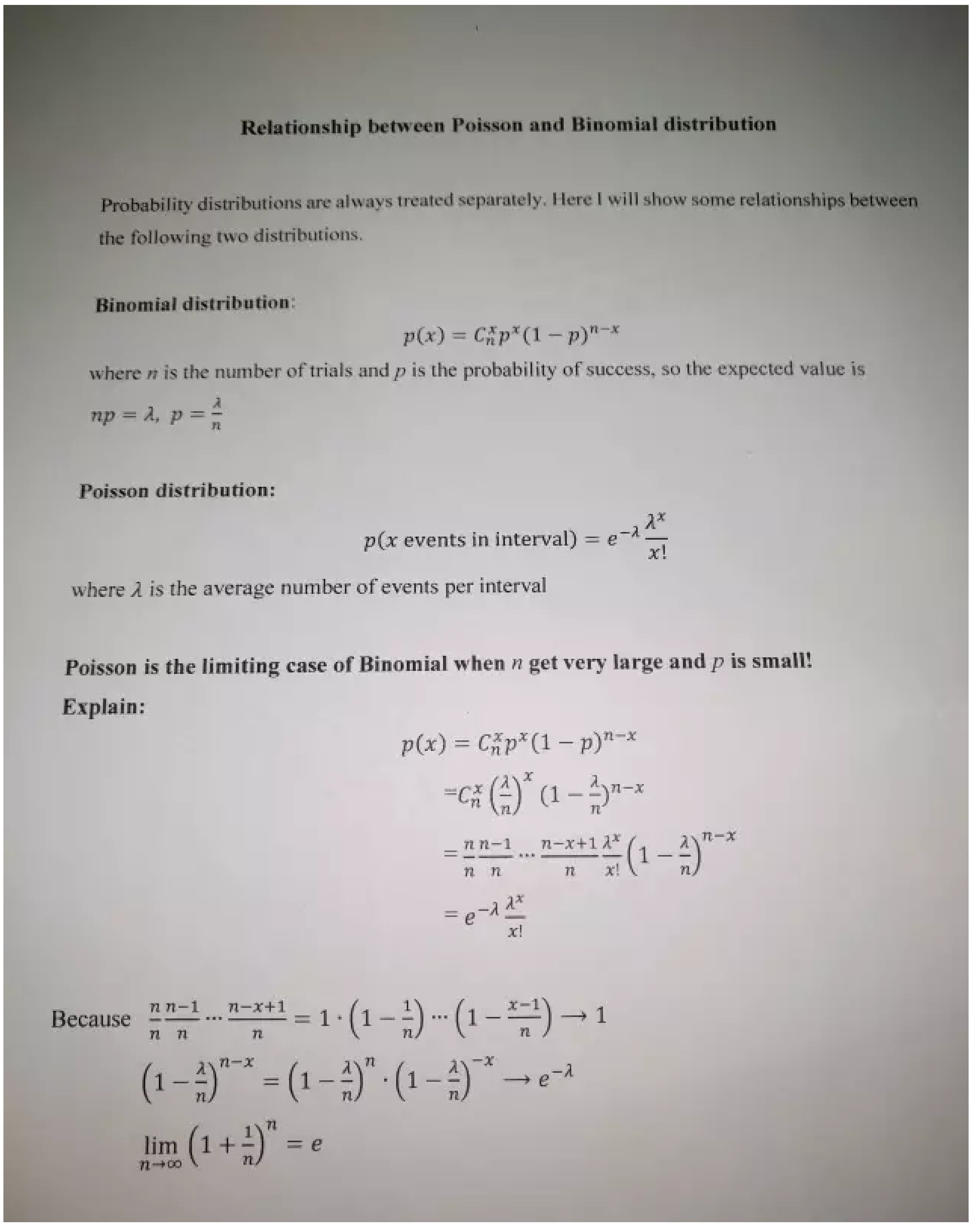
（1）设X为电视机出现大故障的时间。已知μ＝10年，则λ＝1/μ＝0.1，于是，$P(X \ge x) = e^{-\lambda x}$＝e-0.1*15≈0.223。则15年后，没有出现大故障的电视机约占22.3%。

（2）问题要求比例不超过20%，这是求X的右侧概率面积，现在根据公式确定适当的X值。

电视机各年累计出现的故障比例	
担保年数X	累计概率$P(X \le x) = 1 - e^{-\lambda x}$
1	0.095
2	0.181
3	0.259

从表中可以看到：担保2年时，出现大故障的比例是18.1%，不超过20%。担保3年时，出现大故障的比例为25.9%，已经超过20%。所以，厂家应以2年为担保期。

泊松分布是二项式分布的细分，当n→∞，p非常小的时候。



Relationship between Poisson and Binomial distribution

Probability distributions are always treated separately. Here I will show some relationships between the following two distributions.

Binomial distribution:

$$p(x) = C_n^x p^x (1-p)^{n-x}$$

where n is the number of trials and p is the probability of success, so the expected value is $np = \lambda,\ p = \frac{\lambda}{n}$

Poisson distribution:

$$p(x \text{ events in interval}) = e^{-\lambda} \frac{\lambda^x}{x!}$$

where λ is the average number of events per interval

Poisson is the limiting case of Binomial when n get very large and p is small!

Explain:

$$\begin{aligned} p(x) &= C_n^x p^x (1-p)^{n-x} \\ &= C_n^x \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= e^{-\lambda}\frac{\lambda^x}{x!} \end{aligned}$$

Because $\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n} = 1 \cdot \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right) \to 1$

$$\left(1-\frac{\lambda}{n}\right)^{n-x} = \left(1-\frac{\lambda}{n}\right)^{n} \cdot \left(1-\frac{\lambda}{n}\right)^{-x} \to e^{-\lambda}$$

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$$