一、泊松分布

日常生活中，大量事件是有固定频率的。

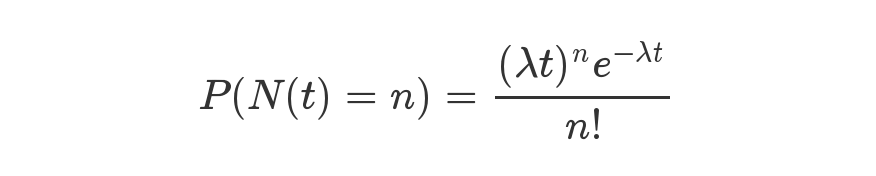
* 某医院平均每小时出生3个婴儿
* 某公司平均每10分钟接到1个电话
* 某超市平均每天销售4包xx牌奶粉
* 某网站平均每分钟有2次访问

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？



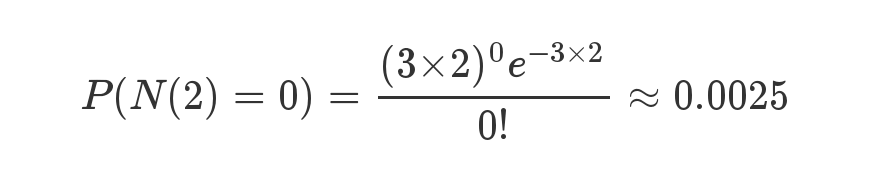
有可能一下子出生6个，也有可能一个都不出生。这是我们没法知道的。

**泊松分布就是描述某段时间内，事件具体的发生概率。**

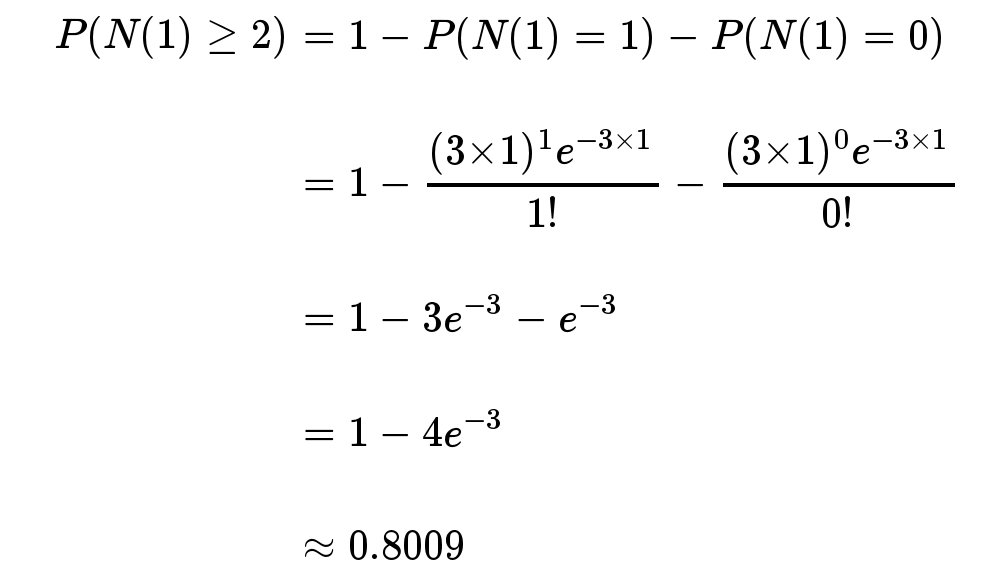


上面就是泊松分布的公式。等号的左边，P 表示概率，N表示某种函数关系，t 表示时间，n 表示数量，1小时内出生3个婴儿的概率，就表示为 P(N(1) = 3) 。等号的右边，λ 表示事件的频率。

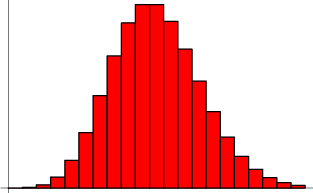
接下来两个小时，一个婴儿都不出生的概率是0.25%，基本不可能发生。



接下来一个小时，至少出生两个婴儿的概率是80%。



泊松分布的图形大概是下面的样子。



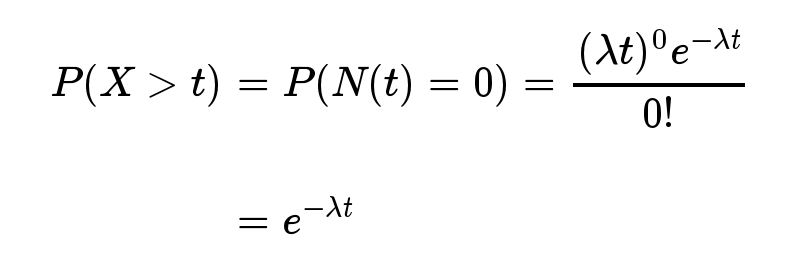
可以看到，在频率附近，事件的发生概率最高，然后向两边对称下降，即变得越大和越小都不太可能。每小时出生3个婴儿，这是最可能的结果，出生得越多或越少，就越不可能。

二、指数分布

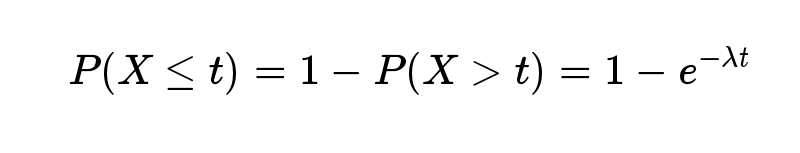
**指数分布是事件的时间间隔的概率。**下面这些都属于指数分布。

* 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

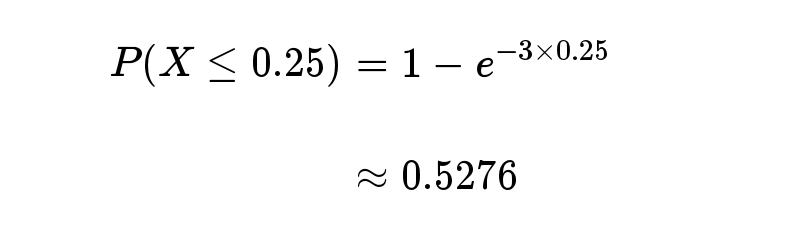
指数分布的公式可以从泊松分布推断出来。如果下一个婴儿要间隔时间 t ，就等同于 t 之内没有任何婴儿出生。



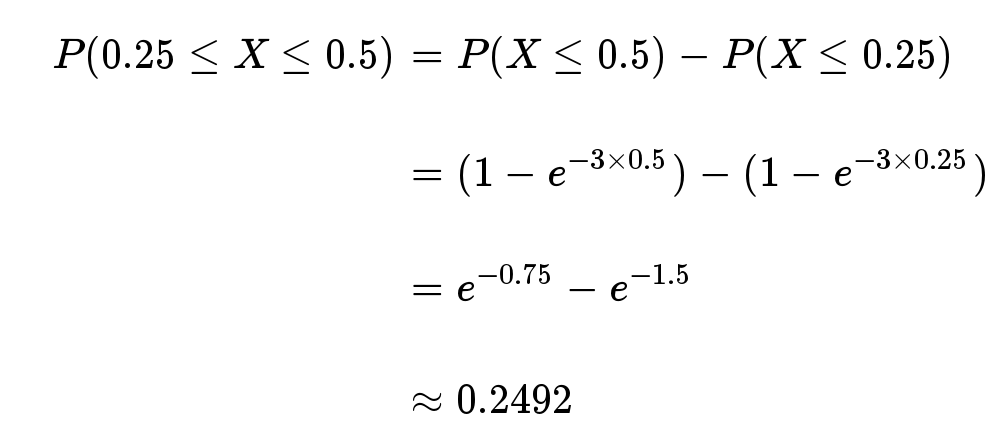
反过来，事件在时间 t 之内发生的概率，就是1减去上面的值。



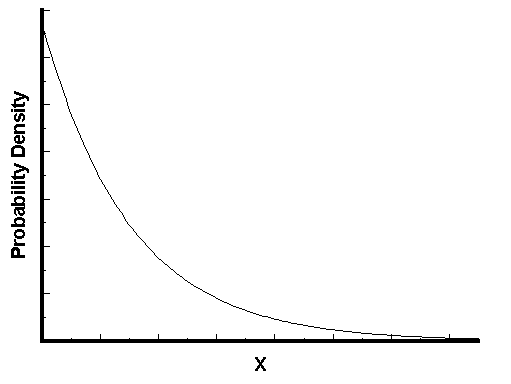
接下来15分钟，会有婴儿出生的概率是52.76%。



接下来的15分钟到30分钟，会有婴儿出生的概率是24.92%。



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。

1. A file server receives requests from a Poisson process at a rate of 30 requests/sec. Measurement data indicates that the coefficient of variation of the service time of a request at the file server is very close to 1. The average service time of a request is 15 msec. (25 marks total)

4.1 Briefly explain what type of queue this system represents. [5 marks]

Answer:

Since the measurement data indicate that the coefficient of variation of the service time of a request at the file server is very close to1, this system is an M/M/1 queue.

Correct answer will score 5 marks.

4.2 Calculate the utilization. [5 marks]

Answer:

Since it is an M/M/1 queue, the utilization of the server would be= rate \* average service time = 30\*15/1000 = 0.45

Correct answer will score 2 marks, correct calculation will score 3 marks.

4.3 Calculate the number of customers in the system. [5 marks]

Answer:

N = U/(1-U) = 0.45 / 0.55 = 0.818

Correct answer will score 2 marks, correct calculation will score 3 marks.

4.4 Calculate the average waiting time. [5 marks]

Answer:

W = N \*average service time = 0.818 \* 15 / 1000 = 0.0123 sec

Correct answer will score 2 marks, correct calculation will score 3 marks.

4.5 Calculate the average response time if the arrival rate of requests were to double. [5 marks]

Answer:

The new utilization of the server would be= rate \* average service time = 60\*15/1000 = 0.9

The new average response time would be T = average service time/(1-utilization) = (15/1000) /(1-0.9) = 0.15 sec

Correct answer will score 2 marks, correct calculation will score 3 marks.

A M/M/1 queue has an arrival rate of 4 jobs per second and it has a service rate of 8 jobs per second.

1.Explain what is a "M/M/1 queue" and its corresponsive meaning. (5 marks)

Answer:

A M/M/1 queue is a special case of G/G/1 queue(A G/G/1 queue is a single queue single server system). The first M indicates that the distribution of arrival times of customers will be exponentially distributed. The second M indicates that the distribution of service times are also exponentially distributed. The 1 indicates that there is only a single server.

Correct explanation of first "M" will score 2 marks, correct explanation of second "M" will score 2 marks, and correct explanation of the "1" will score 1 mark.

2. Calculate the utilization and the mean number of customers in the system. (8 marks)

Answer:

Utilization U = arrival rate / service rate = 4 / 8 = 0.5

mean number of customers N = U/(1-U) = 0.5/(1-0.5) =1

Each correct answer will score 2 marks. Each correct calculation will score 2 marks.

3. Calculate the mean customer delay and number of customers in the queue in the system. (8 marks)

Answer:

Mean customer delay T = N/ arrival rate = 1/4 = 0.25 sec

User waiting time W = T - 1/service rate = 0.25 - 1/8 = 0.125 sec

The number of customers in the queue = arriver rate \* W = 4 \* 0.125 = 0.5

Each correct answer will score 2 marks. Each correct calculation will score 2 marks.

4. If the arrival rate to this system increases by 50%, how much will the mean customer delay increase? (4 marks)

Answer:

The new arrival rate becomes = 4 (1 + 0.5) = 6 jobs/sec

So, the new Utilization U = arrival rate / service rate = 6 / 8 = 0.75

the new mean number of customers N = U/(1-U) = 0.75 / (1-0.75) = 3

Thus, the new mean customer delay T = N / arrival rate = 3 / 6 = 0.5 sec

The original mean customer delay is 0.25 sec, and the new one is 0.5 sec, so it will increase by 100%.

Correct percentage will score 2 marks. Correct calculation will score 2 marks.