```julia
using Plots   ,PlutoUI   , StatsBase   ,Distributions   ,StatsPlots
,LaTeXStrings   ,Symbolics
```

负二项式分布 简介

参考: An Introduction to the Negative Binomial Distribution - Statology

Concept

A Negative binomial distribution describes the number of failures before the rth success in a sequence of independent Bernoulli trials. It is parameterized by r, the number of successes, and p, the probability of success in an individual trial.

负二项式分布描述了在一系列独立进行的伯努利实验中在r 次成功实验之前失败次数的分布

因为负二项分布在回归分析中有使用,所以这里详细介绍一下

在 Distribuitions.jl 中用

```julia
NegativeBinomial()    # 默认为成功一次,r=1, 成功概率为 p=0.5

NegativeBinomial(r, p)  # r 为成功次数, p 为成功概率
```

来实例化负二项式分布

1. 第一次抛出硬币正面之前可能性

日常生活中, 我们总是依靠运气, 有时运气十分不好, 极端差的运气到底有多差? 会不会有连续扔硬币 100次, 1 次正面也没出现. 根据经验, 这种情况不会出现,但是其实机会还是存在的. 这里是在假设硬币为公正的前提下进行的($p = 0.5$)

第一个最简单的负二项分布就是描述这个问题.

问题:扔出一次正面之前的一次硬币是正面的机会是多少?(或者说扔出一次正面之前一次硬币不是反面的机会是多少?)

这个实验很无聊, 但是为了说明后面的实验才列出来.

上述问题,根据负二项式分布定义为: 扔出一次正面硬币之前反面机会为 0 的概率是多少?

```
Distributions.NegativeBinomial{Float64}(r=1.0, p=0.5)
```

```
begin
    params1=1,0.5  # 1 表示第一次扔出硬币正面的事件, 0.5 表示硬币没有问题
    coin_head1=NegativeBinomial(params1...)  #根据需求定义分布
end
```

```
0.5
```

```
pdf(coin_head1,0)
```

上面的概率密度就是我们扔一次硬币, 结果第一次就是正面, 试验结束. 这种情况出现的概率为 0.5

```
0.25
```

```
pdf(coin_head1,1)|>d->round(d,digits=2)
```

上面的概率密度表示为:

当我们扔出一次正面, 前面失败一次, 实际至少进行了两次实验.

这个实验的概率表示为:第一次反面(失败)的概率乘以第二次正面(成功的)的概率

```
0.125
```

```
pdf(coin_head1,2)|>d->round(d,digits=3)
```

上面的概率密度表示为:

当我们扔出一次正面, 前面失败2次, 实际至少进行了3次实验.

这个实验的概率表示:第一次反面(失败)的概率乘以第二次反面(失败的)的概率 乘以第三次正面(成功的)概率

如果扔出一次正面(成功)之前 10 次都是反面(失败的) 机会是多少呢?

如果是100 次都是反面的机会呢?

如果已经扔了 9 次反面, 第 10 次会是什么结果呢?

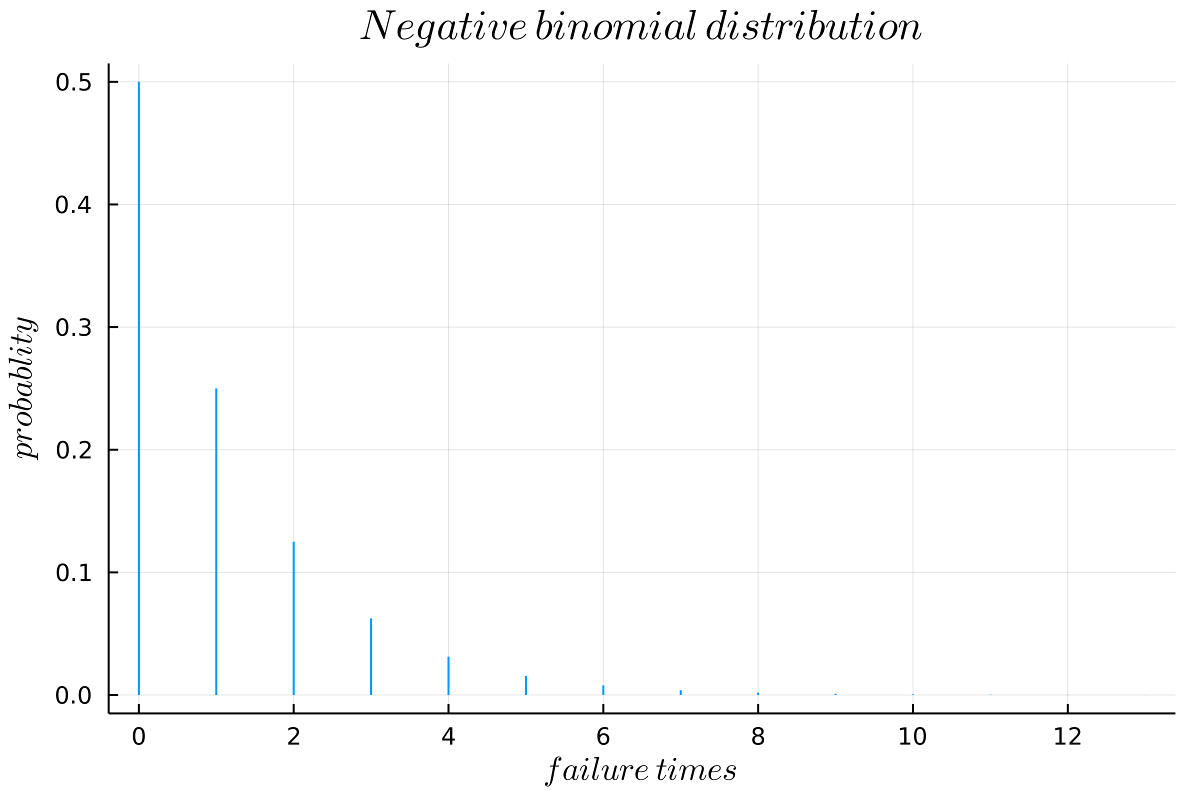
```
0.0004882812499999995
```

```
10|>failure_times->pdf(coin_head1,failure_times)  # 10 次反面的机会其实还算是挺大的,如果改为 100, 概率就相当低了, 但是机会并不是 0
```

```
0.0009765625
```

```
9|>failure_times->pdf(coin_head1,failure_times) # 9 次反面, 第 10 次会是什么结果呢?
```

下面是第一次出现硬币正面前失败次数的概率分布




```
plot(coin_head1,label=false,title=L"Negative \: binomial \: distribution",
    xlabel=L"failure \: times", ylabel=L"probablity")
```

Example

example 1

正反概率相当的硬币,3 次反面后连续 4 次正面的概率

抛出正面定义为成功, 概率相当$p = 0.5$,4次正面实验定义如下:

2. 多次实验的负二项式分布

如果成功次数大于1, 比上面的示例复杂一点, 因为成功的的实验掺杂在失败的实验中, 涉及到排列组合的问题, 具体可以看看参考文献里的定义.

下面我们就直接使用 `Distribuitions.jl` 定义的方法来计算例1 的结果

```
Distributions.NegativeBinomial{Float64}(r=4.0, p=0.5)
```

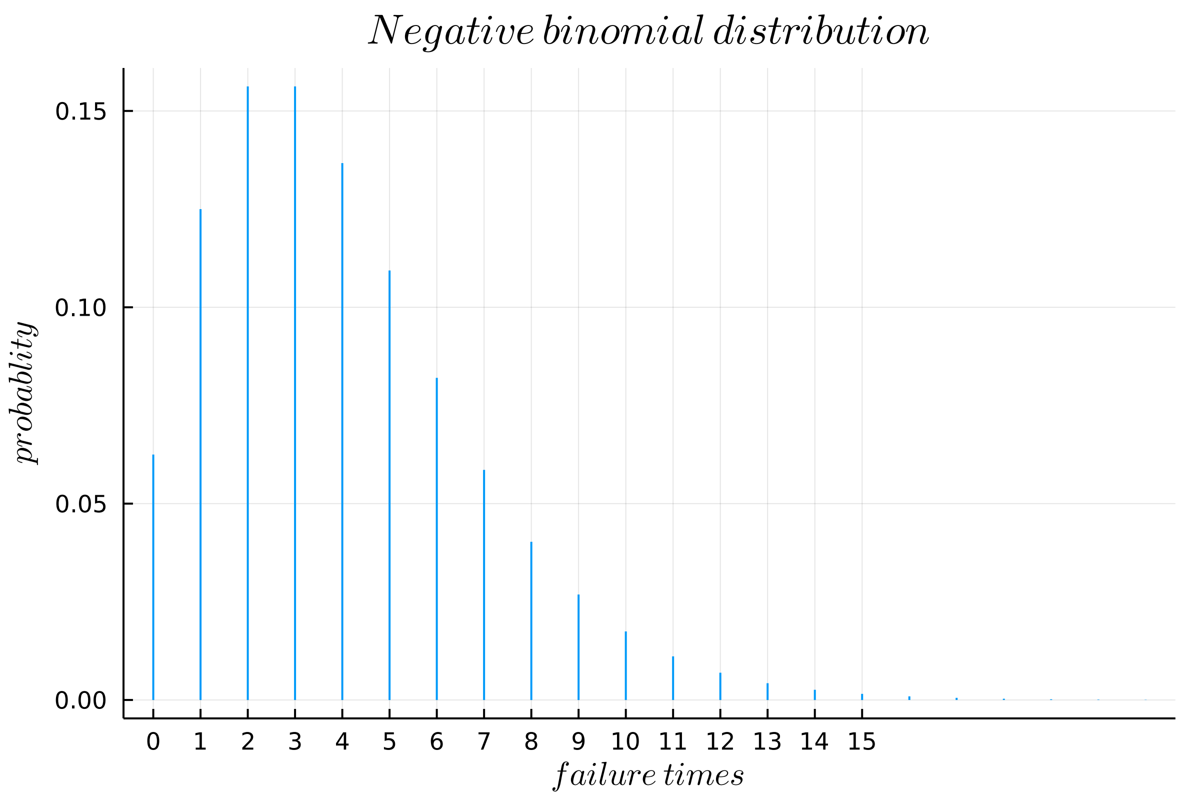
```
begin
    params=4,0.5
    coin_head4=NegativeBinomial(params...)  #定义 实验有 4 次成功的分布函数
end
```

```
0.15624999999999998
```

```
3|>failure_times->pdf(coin_head4,failure_times)
```

0.15624999999999998

```
pdf(coin_head4,3)
```




```
plot(coin_head4,label=false,title=L"Negative \: binomial \: distribution",
     xlabel=L"failure \: times", ylabel=L"probablity",xticks=0:15)
```

0.06250000000000003

```
pdf(coin_head4,0)   #起手连续扔 4 次正面的机会(反面不出现)
```

0.0625

```
0.5^4   # 因为每次出现正面的机会为 0.5 连续四次正面的概率为四个概率的乘积
```

coin_head10 = Distributions.NegativeBinomial{Float64}(r=10.0, p=0.5)

```
coin_head10=NegativeBinomial(10,0.5)
```

0.0009765625

```
pdf(coin_head10,0)

```

概率不为 0.5的情况.

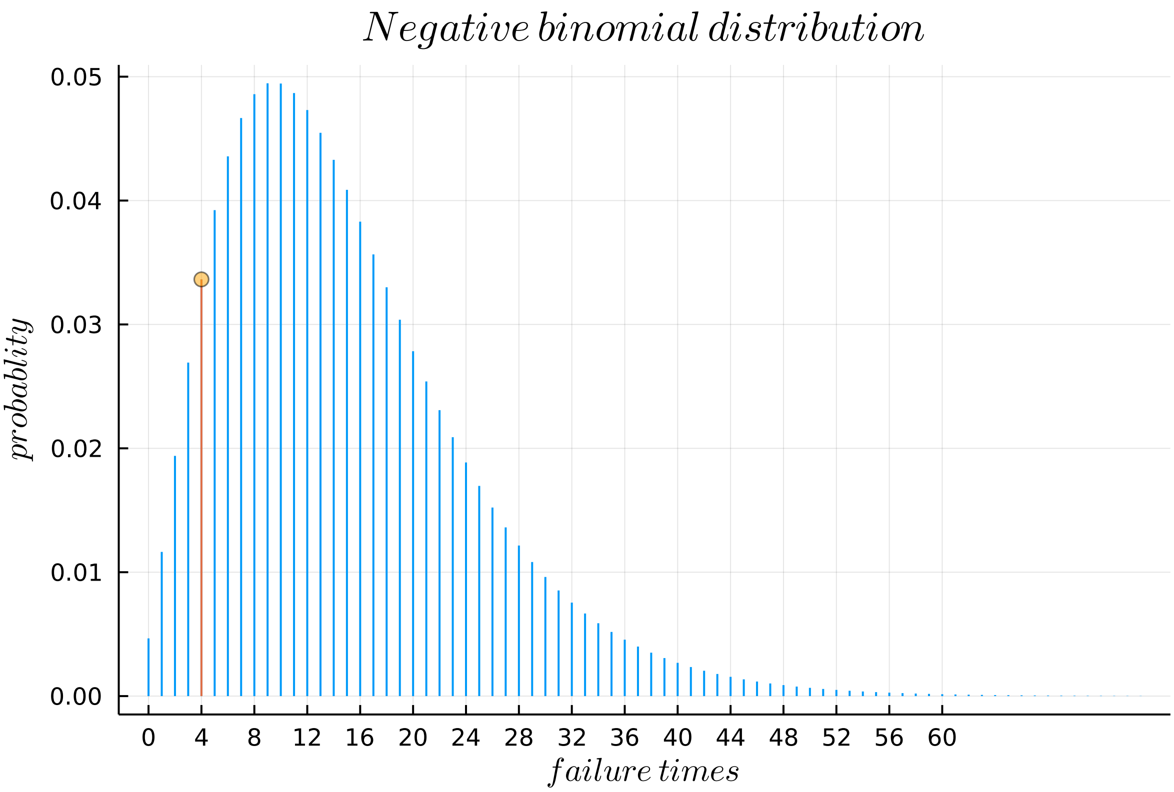
假设扔一个六面的色子(三面, 四面, 五面的都有), 扔三次5 点之前经历 4 次不是 5 点的概率为多少?

因为六面的色子每一面的出现概率都为0.167所以定义的负二项式分布函数为:

dice5 = Distributions.NegativeBinomial{Float64}(r=3.0, p=0.167)

```
dice5=NegativeBinomial(3,0.167)
```

概率分布图为:




```
begin
plot(dice5,label=false,title=L"Negative \: binomial \: distribution",
     xlabel=L"failure \: times", ylabel=L"probablity",xticks=0:4:60)
plot!([4,4],[0,pdf(dice5,4)] ,label=false)
scatter!([4],[pdf(dice5,4)],ms=4,ma=0.5, mc=:orange,label=false)
end
```

0.034

```
4|>failure_times->pdf(dice5,failure_times)|>d->round(d,digits=3) #扔出 3 次 5 点之前四次不是 5 点的概率
```

总结

这就是我日常最经常谈到的概率问题, 但是其实是很复杂的问题. 依照负二项式分布可以构建一下更为复杂的统计模型工具,所以需要理解.