Variáveis Discretas e Distribuições de Probabilidade

Distribuição Binomial

Experimentos aleatórios e variáveis aleatórias:

1. Jogue uma moeda ao ar 10 vezes. X é o número de caras obtidas.
2. Uma máquina produz 1% de partes defeituosas. X é o número de partes defeituosas.
3. Cada amostra de ar tem uma chance de 10% de conter uma molécula rara. X é o número de amostras de ar que contém a molécula rara nas próximas 18 amostras analisadas.
4. De todos os bits transmitidos através de um canal de transmissão digital, 10% são recebidos com erro. X é o número de bits recebidos com erro nos próximos 5 bits transmitidos.
5. Um teste de múltipla escolha contém 10 questões, cada uma com 4 opções e o candidato deve escolher uma. X é o número de questões respondidas corretamente.
6. Nos próximos 20 nascimentos em um hospital, X é o número de bebês com sexo feminino.
7. De todos os pacientes sofrendo de uma particular doença, 35% experimentam melhora após uso de uma dada medicação. Nos próximos 100 pacientes tratados com a medicação, X é o número de pacientes que apresentam melhora.

Experimentos aleatórios anteoriores:

Somente dois possíveis resultados:

Sucesso

Falha

Amostragem de Bernoulli

Suposições:

- Amostragens são independentes
- Probabilidades de sucesso são constantes

Exemplo. A chance de que um bit transmitido através de um canal de transmissão digital seja recebido com erro é 0.1. Suponha que as amostragens de transmissão são independentes. Seja X o número de bits em erro nos próximos 4 bits transmitidos. Determine *P(A)*.

E – bit em erro
O – bit OK

Típico resultado: OEOE

Correspondentes valores de *x* :

Outcome	x	Outcome	x
$OOOO$	0	$EOOO$	1
$OOOE$	1	$EOOE$	2
$OOEO$	1	$EOEO$	2
$OOEE$	2	$EOEE$	3
$OEOO$	1	$EEOO$	2
$OEOE$	2	$EEOE$	3
$OEEO$	2	$EEEO$	3
$OEEE$	3	$EEEE$	4

Por exemplo, o evento *X* = *2* está associado aos seis resultados:

{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE}

Como as amostragens são independentes temos,

$$P(EEOO) = (0.1)^2(0.9)^2 = 0.0081$$

$$P(X = 2) = 6(0.1)^2(0.9)^2 = 6 \times 0.0081 = 0.0486$$

Por exemplo, o evento *X = 1* está associado aos 4 resultados:

{EOOO, OEOO, OOEO, OOOE}

Temos então

$$P(EOOO) = (0.1)^1(0.9)^3 = 0.0729$$

$$P(X = 1) = 4(0.1)^1(0.9)^3 = 4 \times 0.0729 = 0.2916$$

Em geral,

$$P(X = x) = N(0.1)^x(0.9)^{N-x}$$

onde N é o número de resultados com x erros

N – número de modos de partição de 4 objetos em dois grupos, um de tamanho x .

$$\binom{4}{x} = \frac{4!}{x!(4 - x)!}$$

$$P(X = x) = \binom{4}{x}(0.1)^x(0.9)^{4-x}$$

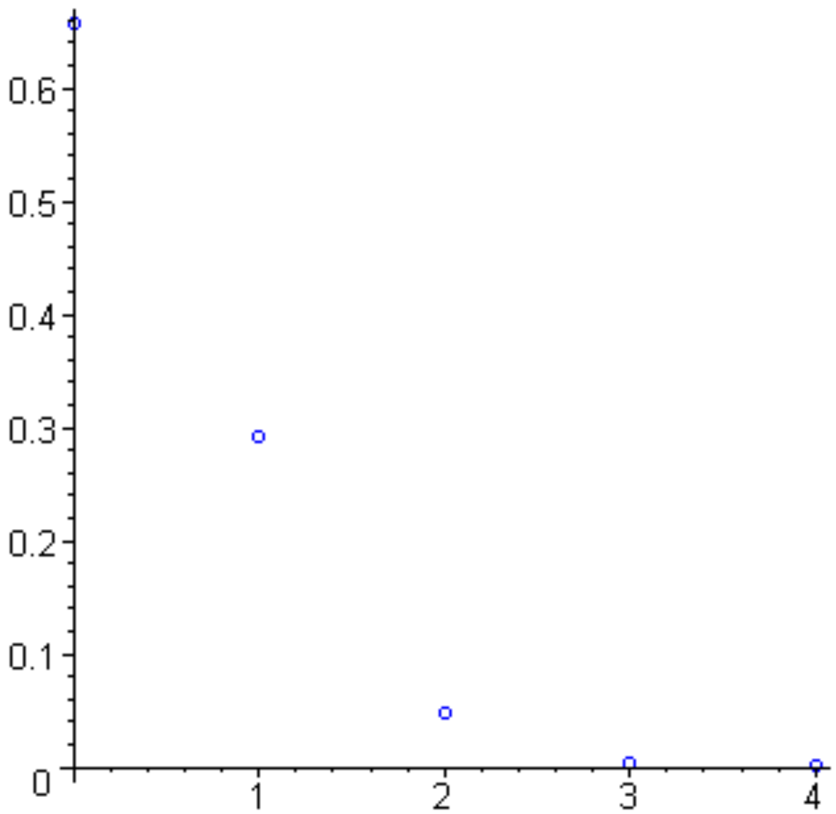
```
> restart:
> with(Statistics):
> for x from 0 to 4 do
   p[x]:=binomial(4,x)*0.1^x*0.9^(4-x);
od:
> xdata:=[seq(x,x=0..4)];
```

$$xdata := [0, 1, 2, 3, 4]$$

```
> ydata:=[seq(p[x],x=0..4)];
```

$$ydata := [0.65610, 0.2916, 0.0486, 0.0036, 0.00010]$$

```
> PointPlot(ydata,xcoords=xdata, color=blue, symbol=circle);
```



Definição

Um experimento aleatório que consiste de n amostragens é definido da seguinte forma:

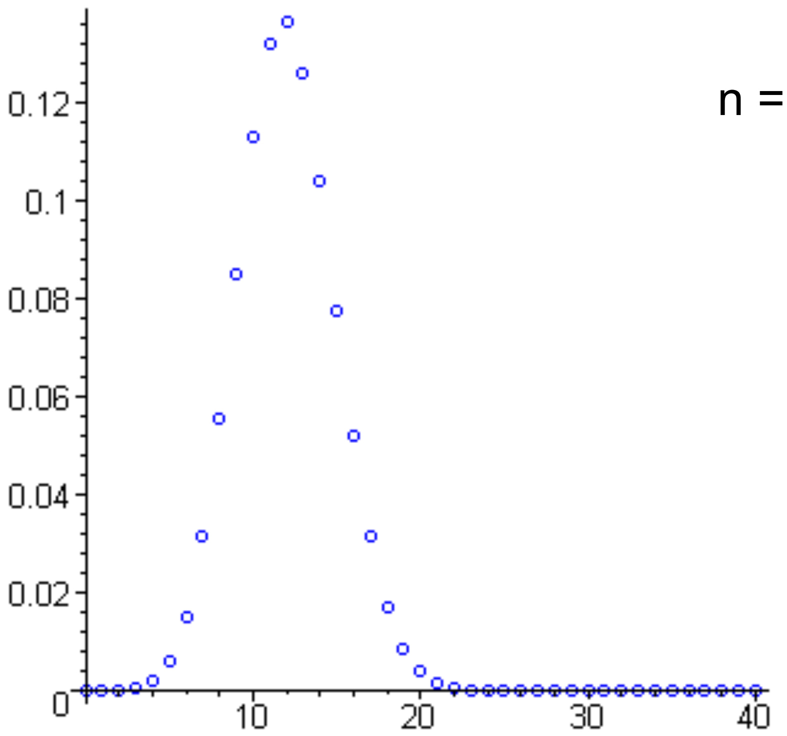
- As amostragens são independentes.
- Cada amostragem pode ter somente dois resultados, denominados *sucesso* e *falha*.
- A probabilidade de sucesso em cada amostragem, denotada por p, permanece constante.

A variável aleatória X que é igual ao número de amostragens que resultam em sucesso é denominada **binomial**, com parâmetros $0 < p < 1$ e $n = 1, 2, 3, \ldots$

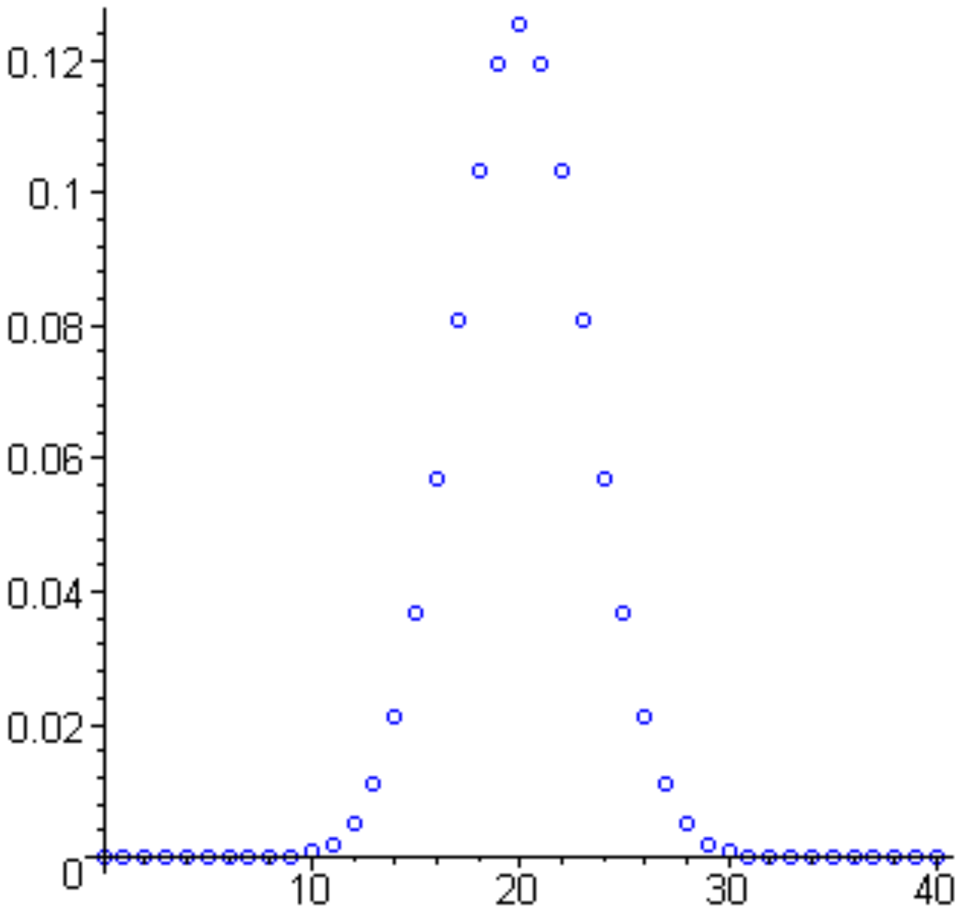
A função de massa de probabilidade de X é

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$

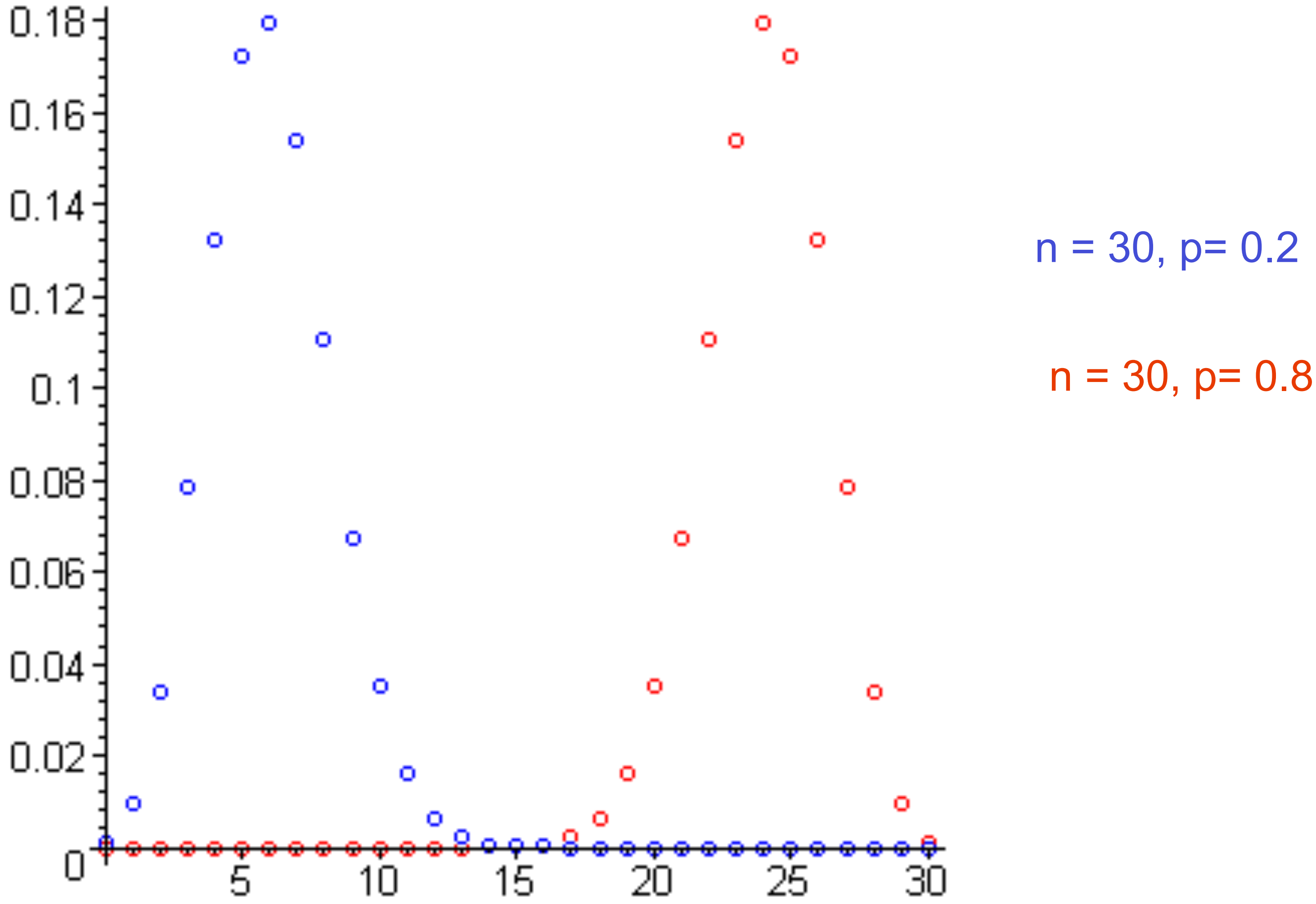
```
> restart:
> with(Statistics):
> n:=40:p:=0.3:
> for x from 0 to n do
      P[x]:=binomial(n,x)*p^x*(1-p)^(n-x);
  od:
> xdata:=[seq(x,x=0..n)]:
> ydata:=[seq(P[x],x=0..n)]:
> PointPlot(ydata,xcoords=xdata, color=blue, symbol=circle):
```



n = 40, p= 0.3



n = 40, p= 0.4


0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
10
15
20
25
30
n = 30, p= 0.2
n = 30, p= 0.8

Exemplo. Cada amostra de água tem 10% de chance de conter um particular poluente orgânico. Suponha que as amostras são independentes com relação à presença do poluente. Determine a probabilidade que nas próximas 18 amostras, exatamente 2 contenham o poluente.

X - número de amostras que contêm o poluentes nas próximas 18 amostras examinadas

X é uma variável aleatória binomial com n = 18 e p = 0.1

$$P(X=2)=\binom{18}{2}(0.1)^2(0.9)^{18-2}=0.284$$

Determine a probabilidade de que pelo menos 4 amostras contenham o poluente.

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x} (0.1)^x (0.9)^{18-x} = 0.098$$

ou

$$P(X \geq 4) = 1 - P(X < 4)$$

$$= 1 - \sum_{x=0}^{3} \binom{18}{x} (0.1)^x (0.9)^{18-x} = 0.098$$

Maple:

```
> 1-Sum(binomial(18,x)*0.1^x*0.9^(18-x),x=0..3);
```

$$1-\left(\sum_{x=0}^{3} \text{binomial}(18, x)\, 0.1^{x}\, 0.9^{(18-x)}\right)$$

```
> value(%);
```

$$0.0981968414$$

Determine a probabilidade de que

$$3 \leq X < 7$$

$$P(3 \leq X < 7) = \sum_{x=3}^{6} \binom{18}{x} (0.1)^x (0.9)^{18-x} = 0.265$$

Se X é uma variável aleatória binomial com parâmetros *p* e *n*,

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = np(1-p)$$

Problemas propostos: **http://fsasse.t35.com/stat2008.html**

Distribuições Binomiais Negativas e Geométricas

Distribuição geométrica

Suponhamos que temos uma série de amostragens de Bernoulli (amostragens independentes com probabilidade constante p de sucesso cada uma).

Agora, em vez de usarmos um número fixo de amostragens, elas serão realizadas até que o sucesso seja obtido.

X – número de amostragens até que o sucesso seja obtido.

Exemplo. A probabilidade de que um bit transmitido através de um canal de transmissão digital seja recebido com erro é 0.1. Suponha que as transmissões são eventos independentes.

X – número de bits transmitidos até o primeiro erro

$$P(X=5)=P(OOOOE)=(0.9)^4(0.1)=0.06$$

Em geral,

$$P(X=x)=f(x)=p^{x-1}(1-p)$$

Variável aleatória geométrica

Se X é uma variável aleatória geométrica com parâmetro p,

$$\mu = E(x) = 1/p$$

$$\sigma^2 = V(x) = \frac{1-p}{p^2}$$

Distribuição Binomial Negativa

Generalização da distribuição geométrica onde a variável aleatória é o número de amostragens de Bernoulli necessárias para obter r sucessos.

Exemplo. Suponha que a probabilidade de que um bit transmitido através de um canal digital de comunicação seja recebido com erro é 0.1. Suponha que as transmissões são eventos independentes.

X – quantidade de bits transmitidos até o quarto erro.

P(X=10) - probabilidade de que exatamente três erros ocorram nas primeiras nove amostragens e então a amostragem 10 resulta no quarto erro.

Probabilidade de que exatamente três erros ocorram nas primeiras nove amostragens : distribuição binomial

$$\binom{9}{3}(0.1)^3(0.9)^6 = 0.04464$$

Probabilidade de que a décima amostragem resulte num bit com erro: 0.1

Resultado:

$$\binom{9}{3}(0.1)^3(0.9)^6(0.1) = \binom{9}{3}(0.1)^3(0.9)^6(0.1) = 0.004464$$

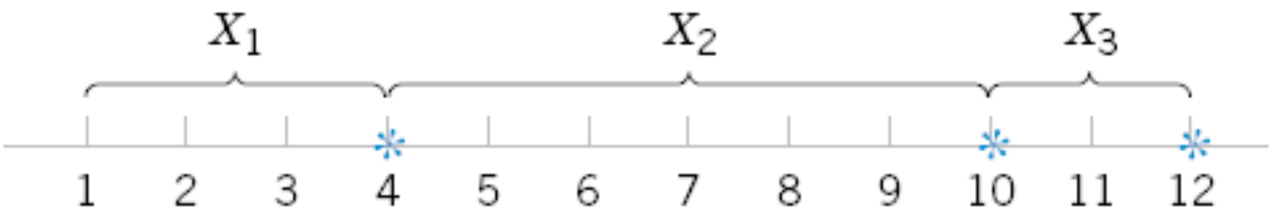
Definição

Em uma série de amostragens de Bernoulli, seja X o número de amostragens até que r sucessos ocorram. Então X é uma *variável aleatória binomial negativa,* com parâmetros
$0 < p < 1$, $r = 1,2,3,\ldots$, e

$$f(x) = \binom{x-1}{r-1}(p)^r (1-p)^{x-r} \qquad x = r, r+1, r+2, \ldots$$

Série binomial negativa é uma soma de séries geométricas

$$X = X_1 + X_2 + X_3$$



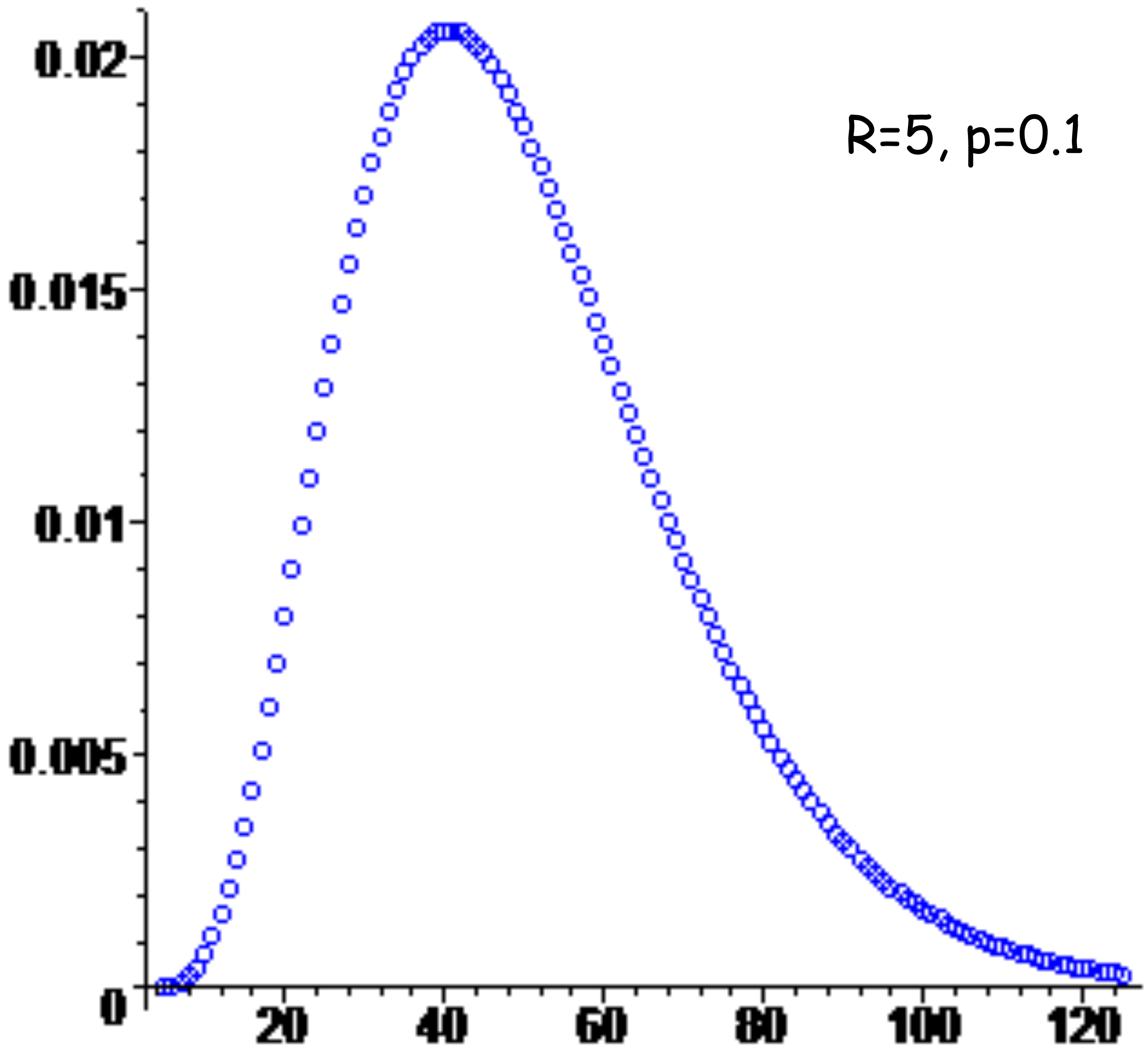
Série binomial: número de amostragens é pré-determinado, número de sucessos é aleatório.

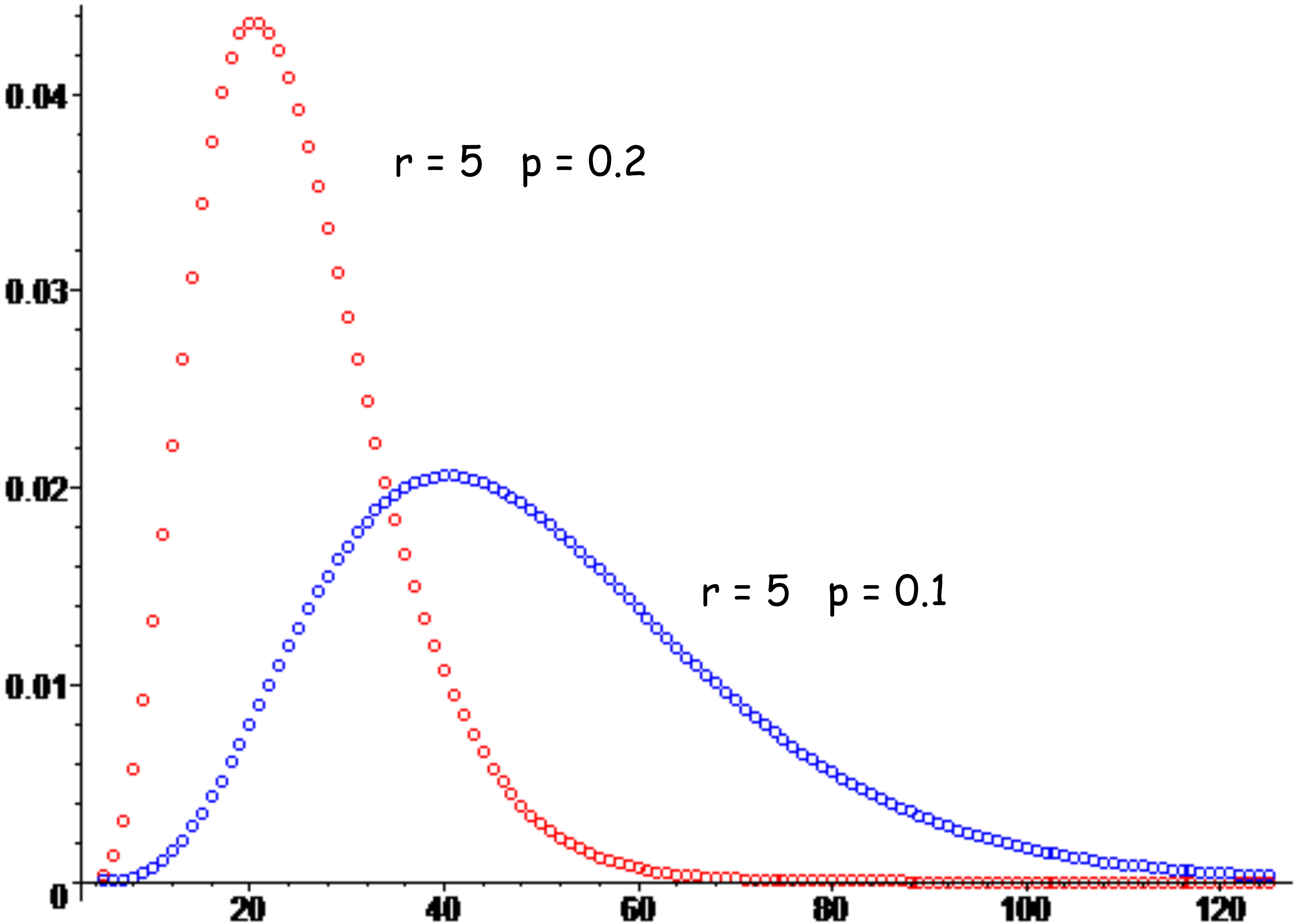
Série binomial negativa: número de sucessos é pré-determinado, número de amostragens é aleatório.

```
> restart:
> with(Statistics):with(plots):
> r:=5:p:=0.1:N:=120:
> for  x from r  to r+N do
      P[x]:=binomial(x-1,r-1)*p^r*(1-p)^(x-r);
  od:
> xdata:=[seq(i,i=r..r+N)]:
> ydata:=[seq(P[k],k=r..r+N)]:
> PL:=PointPlot(ydata,xcoords=xdata, color=blue, symbol=circle):
> display([PL]);
```

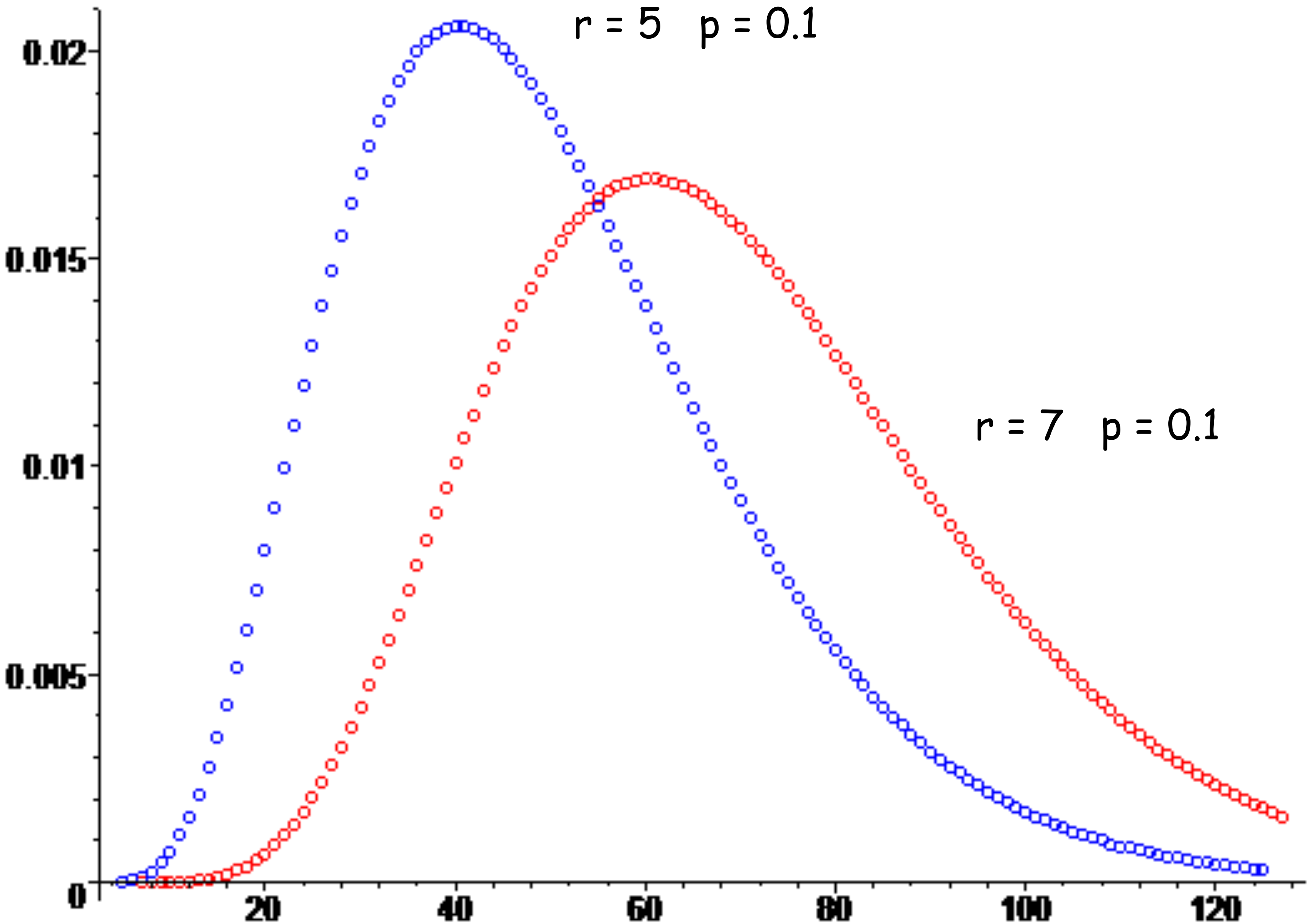
```
> sum(P[k],k=r..r+N):
```

0.9961414048


0.02
0.015
0.01
0.005
0
20
40
60
80
100
120
R=5, p=0.1



r = 5 p = 0.2
r = 5 p = 0.1
0.04
0.03
0.02
0.01
0
20
40
60
80
100
120


r = 5 p = 0.1
r = 7 p = 0.1
0.02
0.015
0.01
0.005
0
20
40
60
80
100
120

Se X é uma variável aleatória binomial negativa, então

$$\mu = E(x) = \frac{r}{p}$$

$$\sigma^2 = V(X) = \frac{r(1-p)}{p^2}$$

Exemplo. Um site da web possui três indênticos servidores. Um é utilizado para operar o site e os outros dois são reservas que devem ser ativados no caso do sistema falhar. A probabilidade de falha do computador principal (ou reserva) na requisição de um serviço é 0.0005. Supondo que cada requisição representa uma amostragem independente, qual é o número médio de requisições até que aconteça uma falha dos três servidores ?

X : número de requisições até que os três servidores falhem.

$$X = X_1 + X_2 + X_3 \qquad p = 0.5$$

$$r = 3$$

falha no servidor 1

$$\mu = E(x) = \frac{r}{p} = \frac{3}{0.0005} = 6000$$

Qual é a probabilidade de que os três servidores falhem em cinco requisições ?

$$P(X = x) = f(x) = \binom{x-1}{r-1}(p)^r (1-p)^{x-r}$$

$$P(X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X \leq 5) = \sum_{x=3}^{5} \binom{x-1}{3-1}(0.0005)^3 (1-0.0005)^{x-3}$$

$$= 0.1249062688 \times 10^{-8}$$

```
> p:=0.0005: r:=3:
> P(X<=5):=sum(binomial(x-1,2)*p^r*(1-p)^(x-r),x=3..5);
```

Distribuição Hipergeométrica

Exemplo: Uma linha de produção produz 50 partes defeituosas para cada lote de 850 partes. Duas partes são selecionadas aleatoriamente sem substituição e analisadas

X – número de partes defeituosas

$$P(X=0)=P(OO)=\frac{800}{850}\frac{799}{849}=0.886$$

$$P(X=1)=P(DO)+P(OD)$$

$$=\frac{50}{850}\frac{800}{849}+\frac{800}{850}\frac{50}{849}=0.111$$

$$P(X=2)=P(DD)=\frac{50}{850}\frac{49}{849}=0.003$$

Definição. Um conjunto de N objetos contém
K objetos classificados como sucessos
N-K objetos classificados como falhas
Uma amostra de n objetos é selecionada aleatoriamente, sem substituição dos N objetos

$$K \leq N \qquad\qquad n \leq N$$

X – variável aleatória que denota o número de sucessos
- variável aleatória hipergeométrica

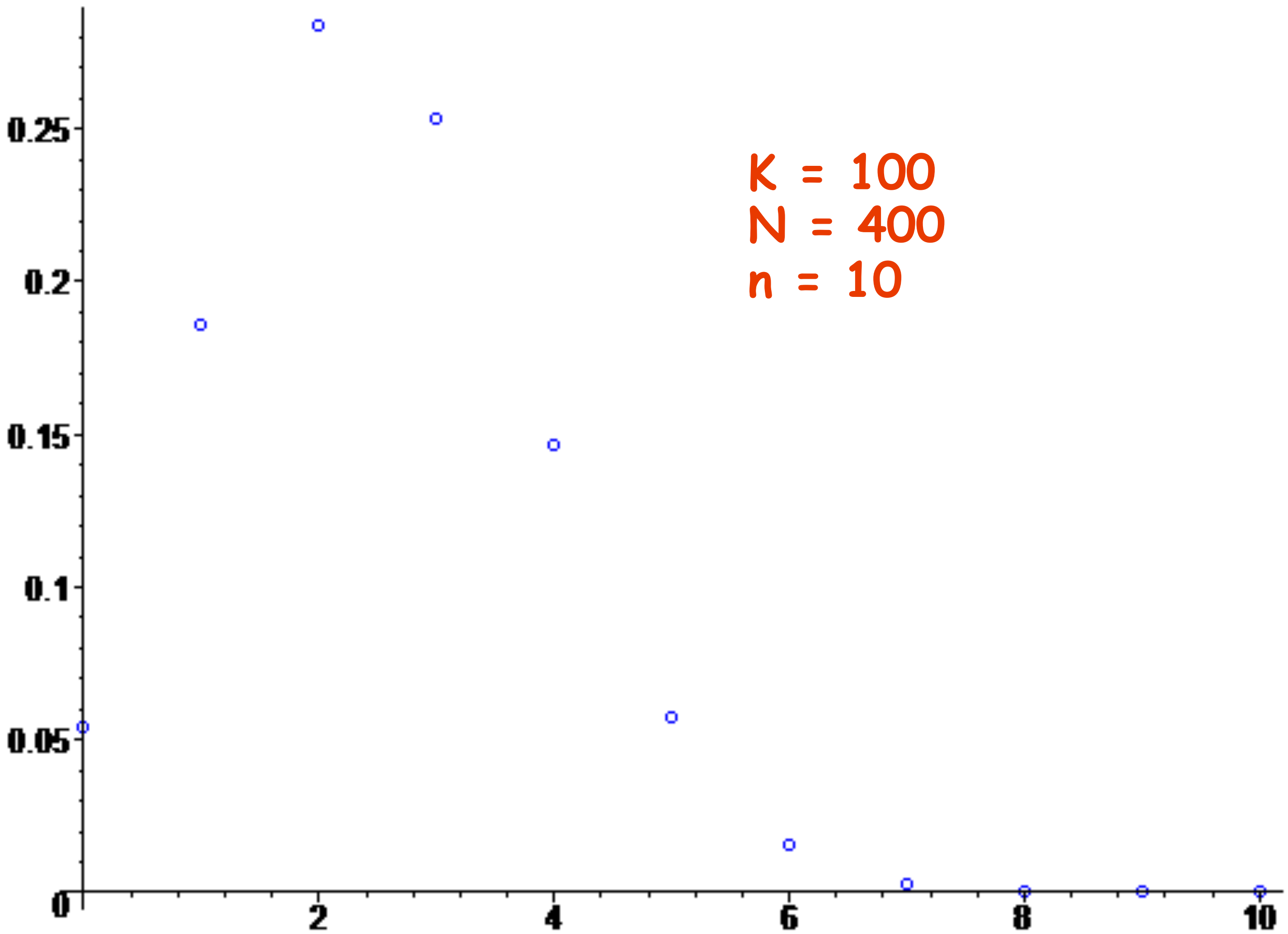
$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad\qquad x = \max\{0, n+K = N\}..\min\{K, n\}$$

$$f(0) = P(X = 0) = \frac{\binom{50}{0}\binom{850-50}{2-0}}{\binom{850}{2}} = 0.886$$

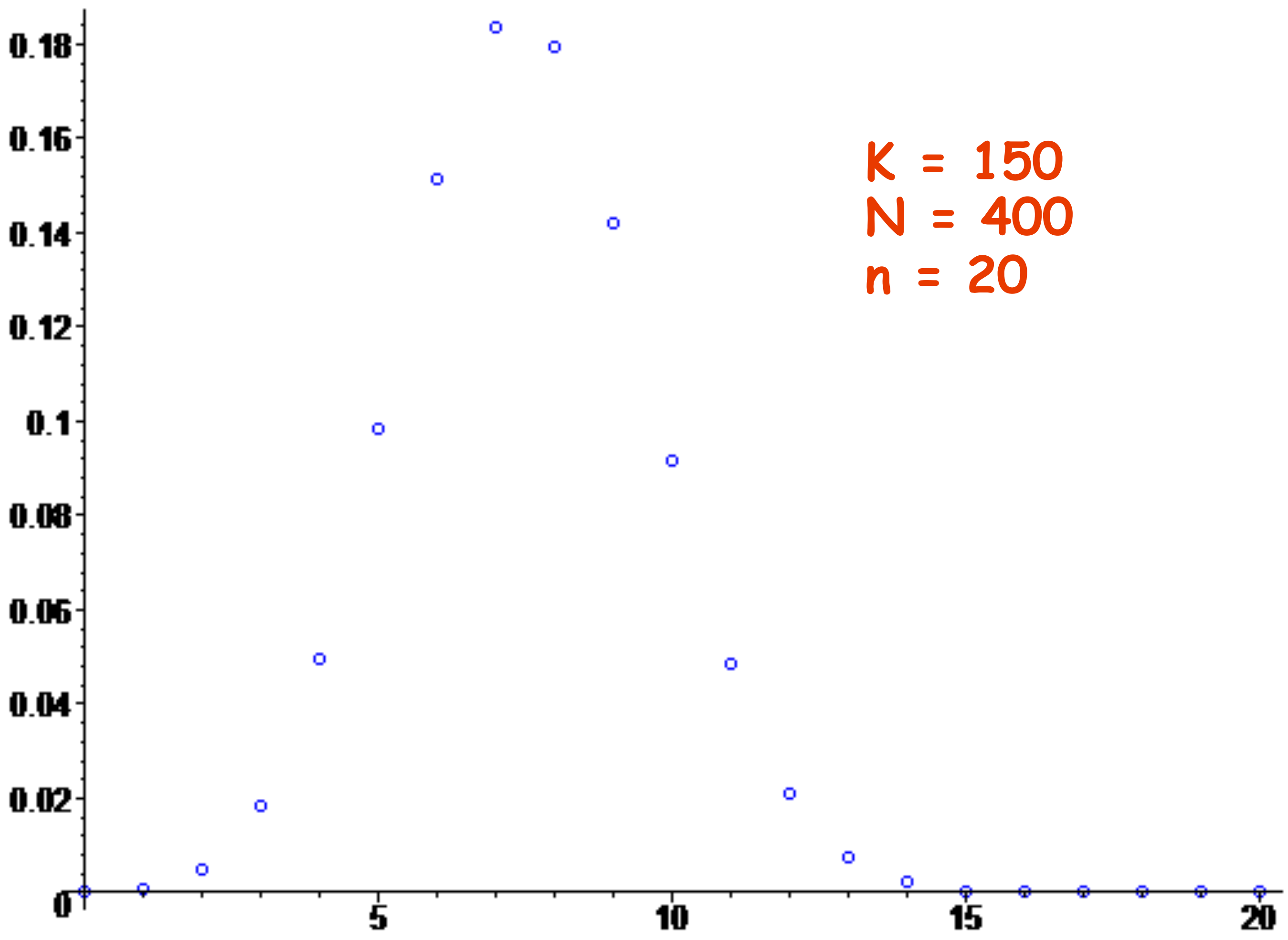
$$f(1) = P(X = 1) = \frac{\binom{50}{1}\binom{850-50}{2-1}}{\binom{850}{2}} = 0.111$$

$$f(2) = P(X = 2) = \frac{\binom{50}{2}\binom{850-50}{2-2}}{\binom{850}{2}} = 0.003$$

```
> restart:
> K:=100: N:=400:n:=10:
> with(Statistics):with(plots):
> xmin:=max(0,n+K-N):xmax:=min(K,n):
> for  x from xmin  to xmax do
     P[x]:=binomial(K,x)*binomial(N-K,n-x)/binomial(N,n);
  od:
> xdata:=[seq(i,i=xmin..xmax)]:
> ydata:=[seq(P[k],k=xmin..xmax)]:
> PL1:=PointPlot(ydata,xcoords=xdata, color=blue,
symbol=circle):
```



0.25
0.2
0.15
0.1
0.05
0
2
4
6
8
10
K = 100
N = 400
n = 10


0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
10
15
20
K = 150
N = 400
n = 20

Exemplo. Um lote contém 100 partes de um fornecedor brasileiro e 200 partes de um fornecedor chinês. Se quatro partes são selecionadas aleatoriamente, sem substituição, qual é a probabilidade de que sejam todas elas de um fornecedor brasileiro ?

X: número de partes de um fornecedor brasileiro.

$$P(X=4)=\frac{\binom{100}{4}\binom{200}{4-4}}{\binom{300}{4}}=0.0119$$

Qual é a probabilidade de que duas ou mais partes na amostra sejam de um fornecedor brasileiro ?

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{\binom{100}{2}\binom{200}{4-2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{4-3}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{4-4}}{\binom{300}{4}}$$

$$= 0.408$$

Qual a probabilidade de que pelo menos uma parte seja de um fornecedor brasileiro ?

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - \frac{\binom{100}{0}\binom{200}{4-0}}{\binom{300}{4}} = 0.804$$

Se X é uma variável aleatória geométrica com parâmetros N, K e n, então

$$\mu = E(X) = np$$

Fator de correção de população finita

$$\sigma^2 = V(X) = np(1-p)\boxed{\left(\frac{N-n}{N-1}\right)}$$

$$p = \frac{K}{N}$$

Proporção de sucessos no conjunto de N objetos

Se n é pequeno relativamente a N, a correção é pequena e a distribuição hipergeométrica é similar à binomial. Neste caso, a distribuição binomial pode ser usada como aproximação.

Exemplo. Uma lista de clientes em uma corporação contém 1000 nomes. Destes, 700 compraram pelo menos um produto da corporação nos últimos 3 meses. Para avaliar o design de um produto, 50 clientes são escolhidos aleatoriamente desta lista. Qual é a probabilidade de que mais do que 45 clientes selecionados tenham feito compras na corporação nos últimos 3 meses ?

A amostragem é feita sem substituição. No entanto, como o tamanho da amostra (50) é pequeno relativamente ao número total de clientes (1000), a probabilidade de cada seleção é aproximadamente constante.

X – variável aleatória hipergeométrica com parâmetros:

N = 1000,   n = 50,   K = 700

Com p = K/N = 0.7, usando a distribuição binomial

$$P(X = x) = f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

temos

$$P(X > 45) = \sum_{x=45}^{50} \binom{50}{x} 0.7^x (1 - 0.7)^{50-x}$$

```
> p:=0.7:n:=50:
> P[X>45]:=sum(binomial(n,x)*p^x*(1-p)^(n-x),x=46..50);
```

$$P_{45<X} := 0.0001719273855$$

Usando a distribuição hipergeométrica,

$$P(X=x) = f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

temos

$$P(X \geq 45) = \sum_{x=45}^{50} \frac{\binom{700}{x}\binom{1000-700}{50-x}}{\binom{1000}{50}}$$

Note que o resultado obtido aqui é algo diferente daquele da distribuição binomial:

```
> N := 1000.:   n := 50.:   K := 700.:
> xmin:=max(0,n+K-N):xmax:=min(K,n):
> P:=sum(binomial(K,x)*binomial(N-K,n-x)/binomial(N,n),x=46..xmax);
```

$$P := 0.0001275856944$$

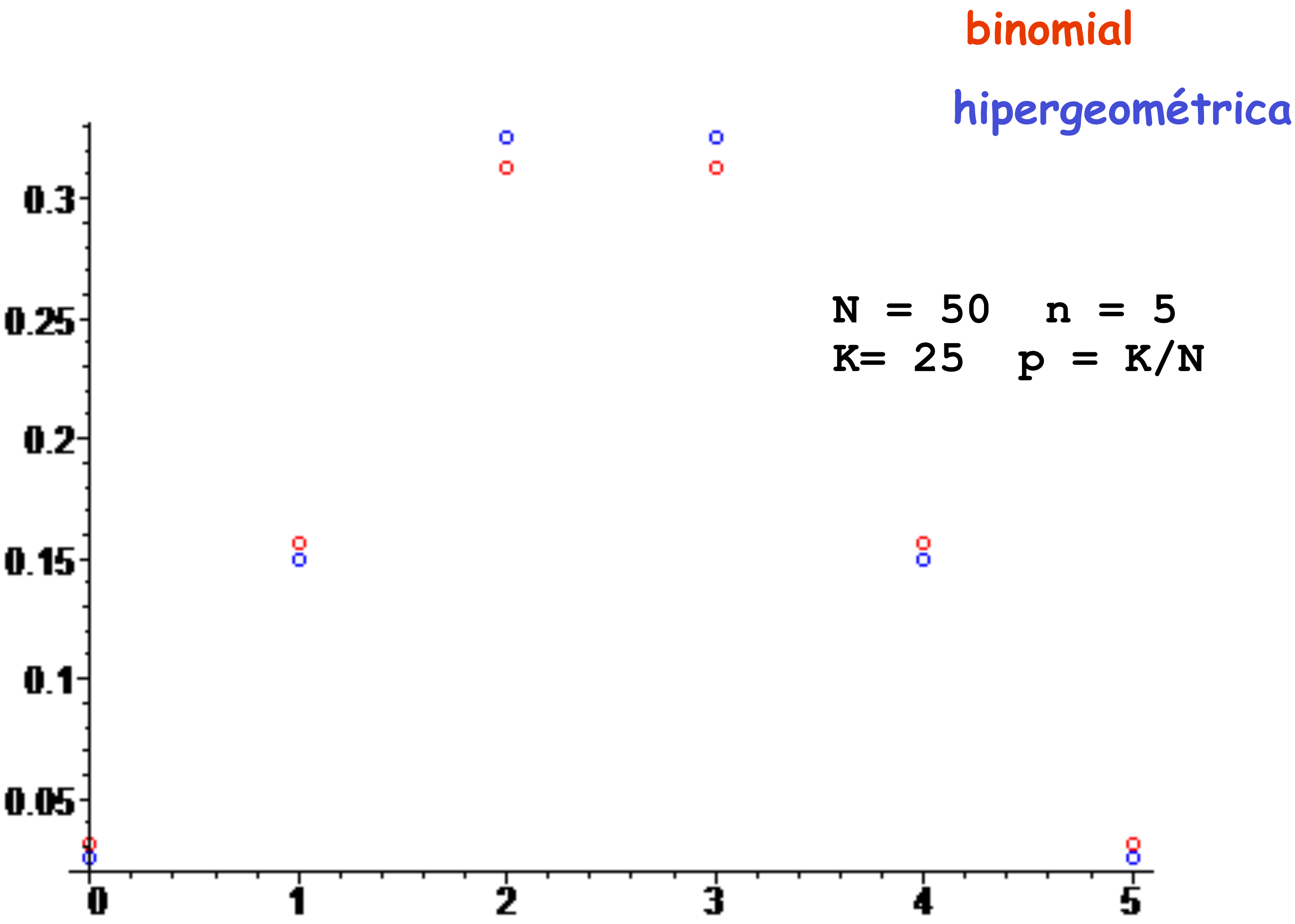
Observe que os termos numéricos que aparecem na fórmula são muito grandes para uma calculadora, por exemplo,

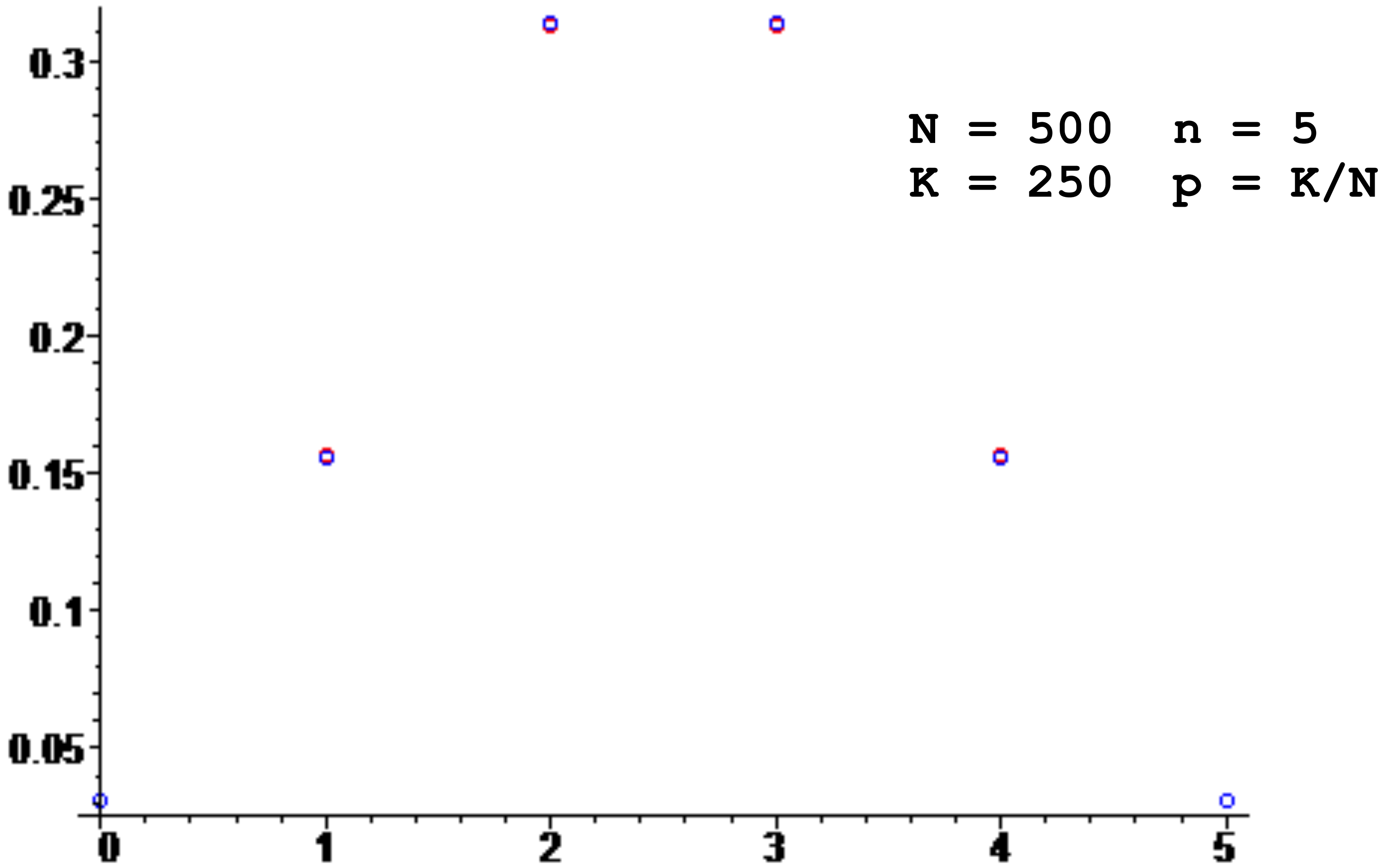
```
> binomial(1000,50);
```

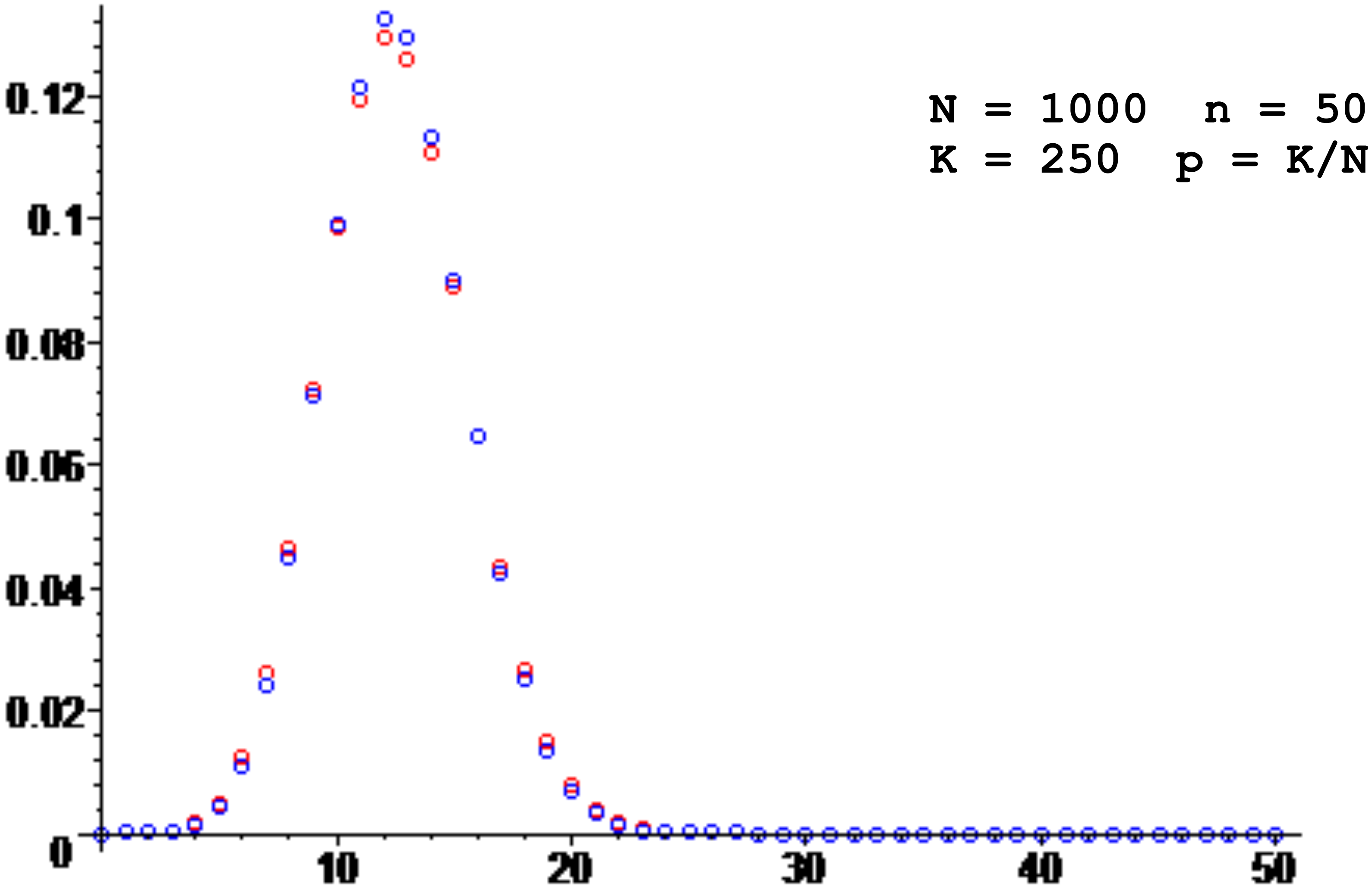
94604610175852178460637222777280449187296940016686540647935693213432526\
97198115263280

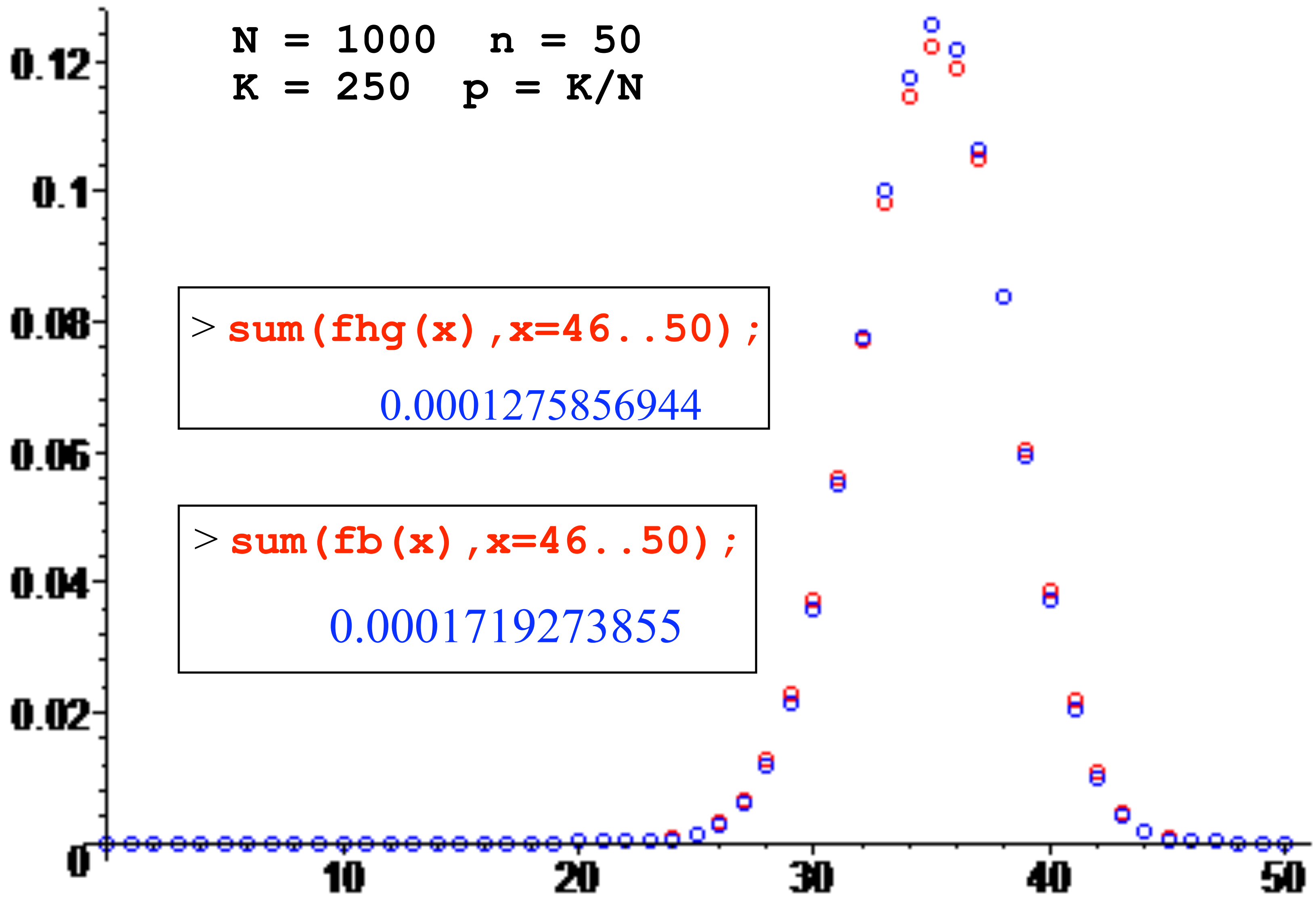
mas são facilmente manipulados pela aritmética de ponto flutuante (software – precisão arbitrária) do Maple.

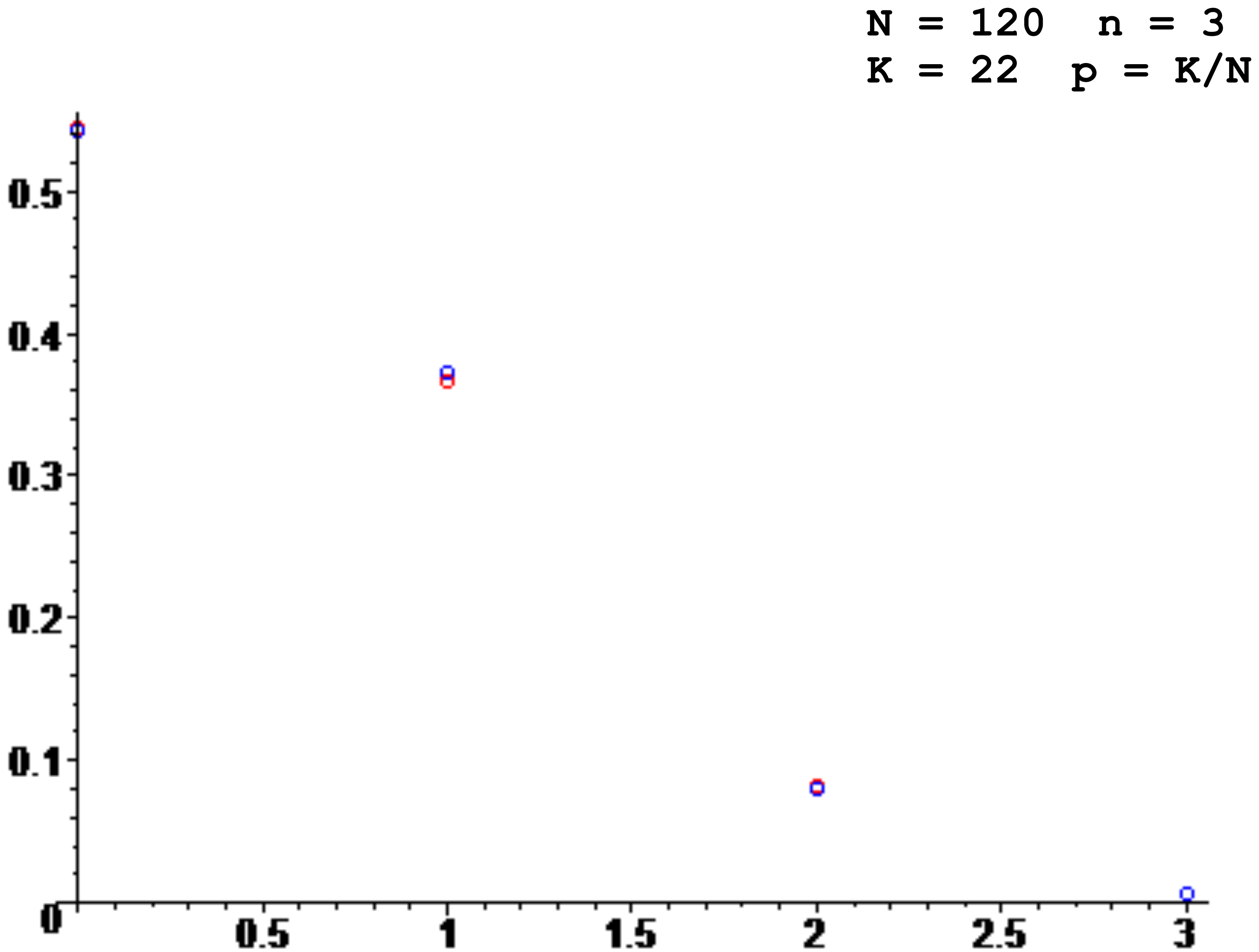
```
> restart:
> N := 50:   n := 5:   K := 25:p:=K/N;
> fhg:=x->evalf(binomial(K,x)*binomial(N-K,n-x)/binomial(N,n));
> fb:=x->evalf(binomial(n,x)*p^x*(1-p)^(n-x));
> with(Statistics):with(plots):
> xdata:=[seq(x,x=0..n)];
> ydata_hg:=[seq(fhg(x),x=0..n)];
> ydata_b:=[seq(fb(x),x=0..n)];
> PL_hg:=PointPlot(ydata_hg,xcoords=xdata, color=blue,
symbol=circle):
> PL_b:=PointPlot(ydata_b,xcoords=xdata, color=red,
symbol=circle):
> display([PL_hg,PL_b]);
```


binomial
hipergeométrica
N = 50 n = 5
K= 25 p = K/N
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5


N = 500 n = 5
K = 250 p = K/N
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5


N = 1000 n = 50
K = 250 p = K/N
0.12
0.1
0.08
0.06
0.04
0.02
0
10
20
30
40
50


N = 1000 n = 50
K = 250 p = K/N
> sum(fhg(x),x=46..50);
0.0001275856944
> sum(fb(x),x=46..50);
0.0001719273855
0.12
0.1
0.08
0.06
0.04
0.02
0
10
20
30
40
50


N = 120 n = 3
K = 22 p = K/N
0.5
0.4
0.3
0.2
0.1
0
0.5
1
1.5
2
2.5
3

```
> sum(fhg(x),x=1..3);
```

0.4584247258

```
> sum(fb(x),x=1..3);
```

0.4553287036

```
> 1-fhg(0);
```

0.4584247258

```
> 1-fb(0);
```

0.4553287037