

Universidade Federal de Goiás
Campus Avançado de Catalão
Departamento de Matemática

Métodos Numéricos para a Probabilidade Binomial e
Hipergeométrica

por

Rosiane Evangelista Borges

Catalão - GO

2003

Rosiane Evangelista Borges

Métodos Numéricos para a Probabilidade Binomial e
Hipergeométrica

Monografia apresentada ao Curso de Especialização em Matemática do campus de Catalão, da Universidade Federal de Goiás, como parte dos requisitos necessários para a obtenção do título de Especialista em Matemática.

Orientador: Dr. Donald Mark Santee.

Catalão - GO

2003

Rosiane Evangelista Borges

**Métodos Numéricos para a Probabilidade Binomial e
Hipergeométrica**

Monografia apresentada e aprovada em 31 de julho de 2003, pela Banca Examinadora constituída pelos professores.

Prof. Dr. Donald Mark Santee

Prof. Dr. Carlos Alberto Pereira dos Santos

Prof. Ms. Cleves Mesquita Vaz

Agradecimentos

Primeiramente a Deus pela força e oportunidade que me deu de conhecer e conviver com pessoas tão especiais e importantes, aos quais externo aqui os mais profundos agradecimentos:

Ao Prof. Dr. Donald Mark Santee, pela orientação e amizade;

Ao meu esposo Nelson, pelo companheirismo e carinho;

Ao meu pai Jair e às minhas irmãs Regiane e Rosirene, pela compreensão que sempre pude contar;

Aos meus amigos pelo apoio demonstrado, carinhosamente à Renata, pela grande ajuda na digitação desta monografia.

A minha mãe Valdivina, in memoriam.

Resumo

Essa monografia apresenta diversos artifícios para contornar o problema do estouro aritmético no cálculo dos fatoriais e das combinações que são aparecem na definição das probabilidades Binomial e Hipergeométrica. Entre esses artifícios estão a fórmula de Lanczos para aproximação da função Gama, a Fórmula de Stirling e sua dedução, fórmulas de recorrência e, finalmente a aproximação de uma probabilidade por outra.

Abstract

This monograph describes several tricks to avoid the overflow problems that can occur during the factorial and combinatorial computations which appear at the Binomial and Hypergeometric probability definitions. These tricks include Lanczos formula for the approximation of the Gamma function, Stirling's formula and its development, recurrence formulas and, finally the approximation of one probability by another one.

Sumário

1	INTRODUÇÃO.....	1
2	OS MODELOS DE PROBABILIDADE.....	2
2.1	A PROBABILIDADE BINOMIAL.....	2
2.2	A PROBABILIDADE HIPERGEOMÉTRICA.....	4
2.3	A PROBABILIDADE NORMAL.....	5
3	O CÁLCULO DO FATORIAL.....	7
3.1	O LOGARITMO DO FATORIAL.....	7
3.2	A FUNÇÃO GAMA E A APROXIMAÇÃO DE LANZOS.....	8
3.3	A APROXIMAÇÃO DE STIRLING.....	13
4	O CÁLCULO DA COMBINAÇÃO.....	15
4.1	COMBINAÇÃO E LOGARITMO.....	15
4.2	FÓRMULAS DE RECORRÊNCIA.....	16
5	O CÁLCULO DAS PROBABILIDADES.....	18
5.1	FÓRMULAS DE RECORRÊNCIA.....	19
5.2	APROXIMAÇÃO DA HIPERGEOMÉTRICA PELA BINOMIAL.....	21
5.3	APROXIMAÇÃO DA BINOMIAL PELA NORMAL (DEMOIVRE-LAPLACE).....	21
6	CONCLUSÃO.....	23
7	BIBLIOGRAFIA.....	23

1 Introdução.

O experimento aleatório chamado “amostragem” é um dos procedimentos mais comuns que se tem na estatística. Escolher ao acaso uma amostra e dela se tirar conclusões sobre a população de onde a amostra foi retirada e quase que a justificativa da existência do estudo da probabilidade e estatística.

Quando uma amostragem é feita, quer seja para inspecionar a qualidade das peças de uma linha de produção (população de tamanho indefinido), quer seja para se fazer uma pesquisa eleitoral (população de tamanho finito), a distribuição de probabilidades subjacente são a distribuição binomial e a hipergeométrica.

As fórmulas para o cálculo dessas probabilidades são relativamente simples e envolvem a determinação de fatoriais. Isso por sua vez esbarra em um problema de ordem prática. Um computador, ou calculadora, representa um número real no sistema denominado *aritmética de ponto flutuante*. Neste sistema, um número é representado na forma

$$\pm(0.d_1d_2d_3\dots d_t) \times \beta^e \quad (1)$$

Onde β é a base em que a máquina opera (normalmente 2);

t é o número de dígitos na mantissa; e

e é o expoente, que deve estar no intervalo $[l,u]$.

Dessa forma em qualquer máquina, apenas um subconjunto dos números reais é representado exatamente. Números que não cabem no molde determinado pela expressão (1) são truncados ou arredondados para caberem.

Existe, contudo, uma situação em que um número não pode ser ajustado para cabem na forma (1). Isso ocorre quando o número é tão grande que o seu expoente ultrapassa o limite

superior u . Esse problema é denominado de *overflow* ou *estouro aritmético*. Quando ocorre um estouro aritmético o número em questão não pode ser armazenado na máquina.

Esse é um problema comum no cálculo do fatorial de um número. O que pode impossibilitar o cálculo das probabilidades binomial e hipergeométrica para tamanhos de população que, em termos da Estatística, são relativamente pequenas.

A presente monografia endereça esse problema, apresentando diversos artifícios que tentam contornar o problema do estouro aritmético no cálculo dessas probabilidades.

2 Os Modelos de Probabilidade.

Os métodos da estatística são, normalmente, baseados em modelos matemáticos de probabilidade. Apresentamos nessa seção uma descrição dos modelos binomial e hipergeométrico, que são a motivação dessa monografia, e a distribuição normal, que levará a um dos artifícios importantes para o cálculo das outras probabilidades.

2.1 A Probabilidade Binomial.

Considere um experimento aleatório com um espaço amostral S . Considere, ainda, que seja definido um evento A , e que esse evento tenha uma probabilidade p de ocorrer. Ao executar o experimento, podemos dizer que houve um sucesso quando o evento A ocorrer, e um fracasso quando o evento A não ocorrer. Considere ainda que esse experimento será executado uma quantidade fixa, n , de vezes, e que estamos interessados apenas no número total, X , de sucessos e não na ordem em que elas ocorrem. O número de sucessos poderá ser qualquer um dos números $0, 1, \dots, n$. Se as n repetições forem executadas de forma

independente, chama-se de Probabilidade Binomial às probabilidades correspondentes a cada valor de X . Diz-se, também que X tem distribuição binomial, ou que X é uma variável aleatória binomialmente distribuída.

Teorema 1. Seja X uma variável aleatória binomialmente distribuída, baseada em n repetições com probabilidade p de sucesso. Então, a probabilidade de X assumir um determinado valor k é dado por

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n \quad (2)$$

Demonstração. Considere-se um particular elemento do espaço amostral satisfazendo a condição $X=k$. Um resultado como esse poderia surgir, por exemplo, se nas primeiras k repetições ocorresse A , enquanto nas últimas $n-k$ repetições ocorresse o seu complementar \bar{A} , isto é,

$$\underbrace{A \ A \ A \ A \ \dots \ A}_k \quad \underbrace{\bar{A} \ \bar{A} \ \bar{A} \ \bar{A} \ \dots \ \bar{A}}_{n-k}$$

Figura 1 – uma maneira de A ocorrer k vezes.

Como todas repetições são independentes, a probabilidade desta seqüência particular ocorrer seria $p^k(1-p)^{n-k}$, mas exatamente essa mesma probabilidade seria associada a qualquer outro resultado para o qual $X=k$. O número total de tais resultados é igual à permutação de todos os A com os \bar{A} , contando que se tem k elementos A repetidos e $n-k$ elementos \bar{A} a quantidade de permutações é

$$\frac{n!}{k!(n-k)!} = \binom{n}{k} \quad (3)$$

Onde se pode observar que é a combinação de n elementos tomados k a k .

Assim a probabilidade da união de todos os eventos em que $X=k$ é dado por

$$\binom{n}{k} p^k (1-p)^{n-k} \quad (4)$$

e isso completa a demonstração.

2.2 A Probabilidade Hipergeométrica.

Suponha-se que tenhamos uma coleção de N objetos, r dos quais tenham uma certa característica A , e $(N - r)$ das quais não tenham essa característica. Suponha que o experimento aleatório a ser realizado seja de escolher, sem reposição e ao acaso, n objetos dessa coleção ($n \leq N$). Em seguida conta-se o número, X , de objetos com a característica A . X poderá ser 0, 1, ..., n . Se a probabilidade de cada objeto ser escolhido for a mesma, a probabilidade associada a cada um desses resultados chama-se Probabilidade Hipergeométrica. Diz-se também que X tem *Distribuição Hipergeométrica*.

Teorema 2. Seja X uma variável aleatória com distribuição hipergeométrica, baseada na retirada de n objetos a partir de uma coleção de N objetos, sendo que r das quais possui uma característica A . Então, a probabilidade de X assumir um determinado valor k é dado por

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots, n \quad (5)$$

Demonstração.

O experimento aleatório possui um espaço amostral finito cujos eventos simples são formados por n objetos diferentes. Assim o número total de eventos simples do experimento é dado por

$$\binom{N}{n} \quad (6)$$

Pois retiram-se n objetos de um total de N .

$$\underbrace{A_{j1} A_{j2} A_{j3} A_{j4} \dots A_{jk}}_k \quad \underbrace{\bar{A}_{jk+1} \bar{A}_{jk+2} \bar{A}_{jk+3} \bar{A}_{jk+4} \dots \bar{A}_{jn}}_{n-k}$$

Figura 2 – um evento simples do espaço amostral onde k elementos possuem a característica A ocorrer k vezes.

Os eventos simples que compõe o evento $X=k$ apresentam a forma geral da figura 2. Para obter a quantidade de eventos simples retiram-se k objetos dos r que possuem a característica A , e, em seguida, retiram-se $n-k$ objetos dos $N-r$ restantes que não possuem a característica A , assim

$$\binom{r}{k} \binom{N-r}{n-k} \quad (7)$$

Como os eventos simples são igualmente prováveis, para calcular a probabilidade basta dividir a equação (7) pela (6). Isso completa a demonstração.

2.3 A Probabilidade Normal.

A probabilidade normal tem um papel importante dentro da teoria de probabilidade pois, de acordo com o Teorema Central do Limite, é o limite de uma soma muito grande de outras variáveis aleatórias. Sua descrição é apresentada a seguir.

Definição: A variável X , tem uma distribuição normal (ou gaussiana) quando sua função de densidade de probabilidade é da forma

$$f(x) = \frac{1}{\sqrt{2ps}} \exp\left(\frac{-1}{2} \left[\frac{x-m}{s}\right]^2\right), -\infty < x < \infty \quad (8)$$

Os parâmetros m e S devem satisfazer às condições $-\infty < m < \infty$, $S > 0$.

Usualmente emprega-se a seguinte notação: X terá distribuição $N(m, S^2)$ se e somente se, sua distribuição de probabilidade for dada pela equação acima.

Teorema 3. Se X tiver a distribuição $N(m, S^2)$, e se $Y = aX + b$, então Y terá a distribuição $N(am + b, a^2 S^2)$.

Demonstração: O fato de que $E(Y) = am + b$ e $V(Y) = a^2 S^2$ decorre imediatamente das propriedades do valor esperado e da variância.

Em conseqüência, se g for a função de densidade de probabilidade de Y , teremos

$$g(y) = \frac{1}{\sqrt{2ps}} \exp\left(\frac{-1}{2s^2} \left[\frac{y-b}{a} - m\right]^2\right) \left|\frac{1}{a}\right| \quad (9)$$

$$g(y) = \frac{1}{\sqrt{2ps|a|}} \exp\left(\frac{-1}{2s^2 a^2} [y - am + b]^2\right), \quad (10)$$

que, representa a função de densidade de probabilidade de uma variável aleatória com distribuição $N(am + b, a^2 S^2)$

Corolário: Se X tiver distribuição $N(m, S^2)$ e se $Y = (X - m) / S$, então Y terá distribuição $N(0,1)$.

Demonstração: É evidente que Y é uma função linear de X , e por isso, o teorema 3, se aplica.

3 O Cálculo do Fatorial.

Um dos problemas básicos que causam estouro aritmético é o cálculo do fatorial de um número. Mesmo que um número seja pequeno, em comparação com a capacidade de uma máquina de representar números, o fatorial desse número pode ser tão grande a ponto de causar um estouro aritmético.

O cálculo da função fatorial requer, portanto, uma análise cuidadosa para que seja reduzido tanto o esforço computacional quanto os problemas de estouro aritmético.

3.1 O Logaritmo do Fatorial.

Quando o fatorial é usado apenas nos cálculos intermediários, pode ser conveniente calcular o logaritmo do fatorial. Dessa forma um número grande pode ser armazenado em uma máquina como um número real e relativamente pequeno. O cálculo do logaritmo do fatorial é relativamente simples pois

$$\ln(n!) = \ln(1 \cdot 2 \cdot \mathbf{K} \cdot n) = \ln(1) + \ln(2) + \mathbf{L} + \ln(n) \quad (11)$$

Assim pode-se calcular o logaritmo do fatorial pela fórmula

$$\ln(n!) = \sum_{i=1}^n \ln(i) \quad (12)$$

Que pode ser facilmente programada. Entretanto é importante lembrar que esse método não é computacionalmente eficiente, uma vez que requer o cálculo do logaritmo n vezes.

3.2 A Função Gama e a Aproximação de Lanczos.

Outra maneira de calcular a função fatorial, é através da função gama. Esse método traz grandes vantagens do ponto de vista do esforço computacional.

Vamos inicialmente, introduzir a função gama. Representada pelo símbolo Γ ela é definida assim:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \quad p > 0 \quad (13)$$

Pode-se demonstrar que essa integral imprópria existe (converge) sempre que $p > 0$. Se integrarmos por partes, fazendo:

$$e^{-x} dx = dv \quad \text{e} \quad x^{p-1} = u, \quad (14)$$

obteremos:

$$\Gamma(p) = -e^{-x} x^{p-1} \Big|_0^{\infty} - \int_0^{\infty} [-e^{-x}(p-1) x^{p-2} dx] \quad (15)$$

$$= 0 + (p-1) \int_0^{\infty} e^{-x} x^{p-2} dx \quad (16)$$

$$= (p-1) \Gamma(p-1) \quad (17)$$

Desse modo, mostramos que a função gama obedece a uma interessante relação de recorrência. Suponha-se que p seja um inteiro positivo, digamos $p = n$. Então, aplicando-se a equação (17) repetidamente, obteremos:

$$\Gamma(n) = (n-1) \Gamma(n-1) \quad (18)$$

$$= (n-1) (n-2) \Gamma(n-2) = \dots = (n-1) (n-2) \dots \Gamma(1) \quad (19)$$

Porém, $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$ e por isso, teremos

$$\Gamma(n) = (n-1)! \quad (20)$$

Ou ainda

$$n! = \Gamma(n+1) \quad (21)$$

Se n for um inteiro positivo. Portanto poderemos considerar a função gama como uma generalização da função fatorial.

Há uma variedade dos métodos que podem ser usados para calcular a função $\Gamma(n)$ numericamente, mas, segundo Press (Press et al, 1997), nenhum é tão bom quanto a aproximação desenvolvida por Lanczos (Lanczos, 1964). Esta aproximação é específica para a função gama e depende de uma série de coeficientes c_1, c_2, \dots, c_n . A aproximação da função gama proposta por Lanczos é:

$$\Gamma(n) = (n + 5,5)^{n+0,5} e^{-(n+5,5)} \frac{\sqrt{2p}}{n} \left[c_0 + \frac{c_1}{n+1} + \frac{c_2}{n+2} + \dots + \frac{c_n}{n+6} \right] \quad (22)$$

Para o conjunto dos c 's apresentado na tabela 1, o erro é menor do que, 2×10^{-10} .

$c_0 =$	1,000000000190015
$c_1 =$	76,18009172947146

$c_2 =$	-86,50532032941677
$c_3 =$	24,01409824083091
$c_4 =$	-1,231739572450155
$c_5 =$	$0,1208650973866179 \times 10^{-2}$
$c_6 =$	$-0,5395239384953 \times 10^{-5}$

Tabela 1 – coeficientes para a aproximação de Lanczos da função gama.

Para evitar problemas de estouro aritmético, é conveniente que não se calcule a função gama diretamente, mas sim o logaritmo da função gama. O logaritmo da função gama pode ser usada como uma aproximação para o logaritmo do fatorial seguindo a fórmula (21), assim a expressão fica:

$$\ln(n!) = (n + 1,5) \ln(n + 6,5) - (n + 6,5) + \ln \left(\frac{\sqrt{2p}}{n+1} \left[c_0 + \frac{c_1}{n+2} + \frac{c_2}{n+3} + \mathbf{L} + \frac{c_n}{n+7} \right] \right) \quad (23)$$

A fórmula (23) pode ser facilmente programada. A listagem 1 mostra o código fonte de uma página da internet que calcula o fatorial de um número dado. Aparece em negrito a função que programa a fórmula (23). A figura 3 mostra a disposição dos controles da página.

Listagem 1 – Página que calcula o fatorial usando a aproximação de Lanczos.

```
<html>
<head>
<title>Cálculo de ln(n!) e n!</title>
<script language=javascript><!--

function TArray() {return this}
```

```

var c = new TArray()

c[0]=1.000000000190015
c[1]=76.18009172947146
c[2]=-86.50532032941677
c[3]=24.01409824083091
c[4]=-1.231739572450155
c[5]=0.00120865097386617
c[6]=-0.000005395239384953

function lnfactorial(n) {
  n = parseInt(n)

  var tmp = 0.0
  var soma = 0.0
  var y = 0.0

  tmp = n+6.5
  tmp = tmp-((n+1.5)*Math.log(tmp))

  y = n+2
  soma = c[0]
  for(var j=1; j<=6; j++, y++) {
    soma += c[j]/y
  } //end for j

  return -tmp+Math.log(2.5066282746310005*soma/(n+1))
} //end function lnfactorial

function factorial(n) {
  return Math.round(Math.exp(lnfactorial(n)))
} //end function factorial

!--></script>
</head>
<body bgcolor=lightyellow><center>
<h1>Cálculo de ln(<i>n</i>!) e <i>n</i>!</h1><hr>
<br>
<form name=form1><table>
<tr><td colspan=2 align=center>Número:
<input
  type=text
  name=num

```

```
        value=1><br><br>
</td></tr>
<tr valign=top><td>
    <input
        type=button
        value="Ln do Fatorial"
        onClick="form1.lng.value=lnfatorial(form1.num.value)">
</td><td>
    <input
        type=text
        name=lng><br><br>
</td></tr>
<tr valign=top><td>
    <input
        type=button
        value="Fatorial"
        onClick="form1.fat.value=fatorial(form1.num.value)">
</td><td>
    <input
        type=text
        name=fat><br><br>
</td></tr>
</table></form></center></body>
</html>
```

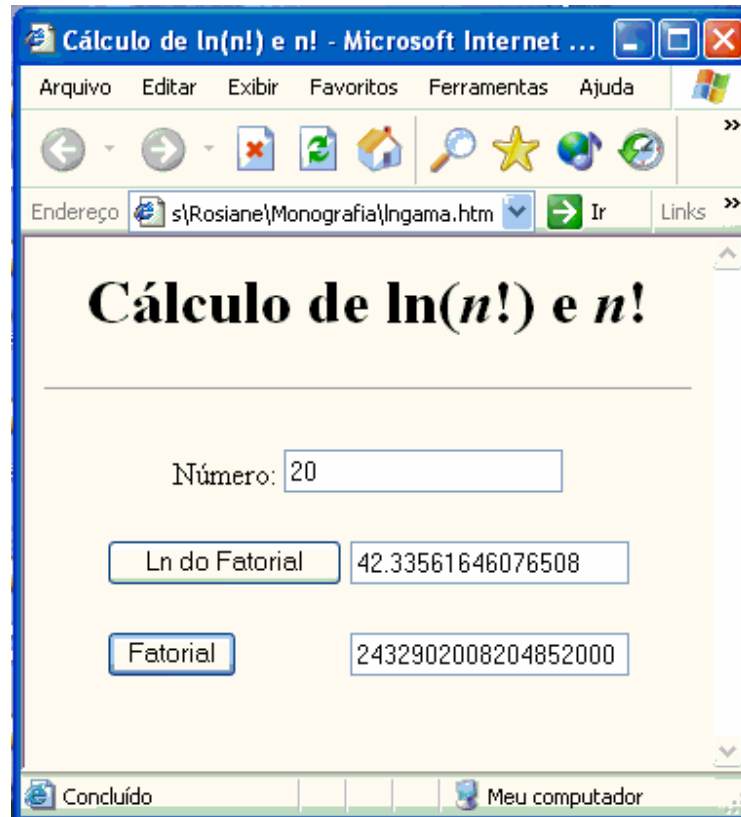


Figura 3 – Página da internet que calcula o Fatorial pela fórmula de Lanczos

3.3 A Aproximação de Stirling.

Uma ferramenta importante da teoria analítica das probabilidades está contida no teorema clássico conhecido pelo nome de Fórmula de Stirling

$$n! \cong n^n e^{-n} \sqrt{2\pi n} \quad (24)$$

Esse resultado é útil, não apenas para a teoria, mas também para a obtenção de excelentes aproximações numéricas. Embora o *erro absoluto*, que é a diferença entre os dois lados da equação não cresça com o de n , o *erro relativo*, que se comete ao utilizar a aproximação diminui. Esse erro decresce regularmente e a precisão da aproximação de Stirling

é notável, mesmo para valores pequenos de n . De fato, o lado direito da equação nos dá para $1!$ O valor 0,9221, para $2!$ 1,919 e para $5!$ (= 120) 118, 019. Os erros percentuais são 8,4 e 2 por cento respectivamente. Para $10!$ = 3.628800 a aproximação nos dá 3598600 com um erro de 0,8%. Para $100!$ O erro é somente 0,08%.

A dedução da fórmula de Stirling parte da sua aproximação pela função gama, assim:

$$\Gamma(n+1) = n! = \int_0^{\infty} x^n e^{-x} dx = \int_0^{\infty} e^{n \ln(x) - x}, \quad n > 0 \quad (25)$$

Fazendo-se a seguinte mudança de variáveis

$$x = n + y\sqrt{n} \quad (26)$$

Tem-se

$$n! = \int_{-\sqrt{n}}^{\infty} e^{n \ln(n+y\sqrt{n}) - n - y\sqrt{n}} \sqrt{n} dy \quad (27)$$

Para valores de n grandes pode-se fazer a seguinte aproximação pela fórmula de Taylor:

$$\ln(n + y\sqrt{n}) = \ln(n) + \ln\left(1 + \frac{y}{\sqrt{n}}\right) \cong \ln(n) + \frac{y}{\sqrt{n}} - \frac{y^2}{2n} + \mathbf{L} \quad (28)$$

Substituindo-se a aproximação (28) em (27) tem-se

$$n! = \int_{-\sqrt{n}}^{\infty} e^{n\left(\ln(n) + \frac{y}{\sqrt{n}} - \frac{y^2}{2n}\right) - n - y\sqrt{n}} \sqrt{n} dy \quad (29)$$

Que se simplifica para

$$n! = e^{n \ln(n) - n} \sqrt{n} \int_{-\sqrt{n}}^{\infty} e^{-\frac{y^2}{2}} dy = n^n e^{-n} \sqrt{n} \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy - \int_{-\infty}^{-\sqrt{n}} e^{-\frac{y^2}{2}} dy \right) \quad (30)$$

A primeira integral do lado direito da equação (30) vale $\sqrt{2p}$ (Ávila, vol. III) e a Segunda integral tende a zero quando n tende cresce. Assim chega-se à fórmula dada por (24).

Seguindo o mesmo procedimento, é possível chegar-se a uma expressão mais precisa tomando-se mais termos da série de Taylor, chegando-se a uma expressão do tipo:

$$n! \cong n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n} + \frac{1}{288n^2} + \mathbf{L} \right) \quad (31)$$

4 O Cálculo da Combinação.

No cálculo das probabilidades Binomial e Hipergeométrica, a função fatorial aparece quando se avaliam as combinações, uma vez que a fórmula para o cálculo da combinação é dada por

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (32)$$

Podem haver problemas de estouro aritmético no cálculo dos fatoriais, mesmo que o resultado final da combinação não tenha esse problema. Para contornar esse problema pode-se usar as técnicas de aproximação de função fatorial conforme descrito a seguir.

4.1 Combinação e Logaritmo.

Tomando-se o logaritmo da combinação tem-se

$$\ln \binom{n}{k} = \ln(n!) - \ln(k!) - \ln((n-k)!) \quad (33)$$

Assim a combinação pode ser calculada fazendo-se

$$\binom{n}{k} = e^{\ln(n!) - \ln(k!) - \ln((n-k)!) } \quad (34)$$

Onde o logaritmo do fatorial pode ser calculado pela fórmula (12) ou preferencialmente pela fórmula de Lanczos (23), reduzindo-se dessa forma o problema do estouro aritmético nos cálculos intermediários.

4.2 Fórmulas de Recorrência

Uma técnica interessante para calcular as combinações é com o uso de fórmulas de recorrência. O ponto mais importante nessa técnica é que as operações aritméticas são executadas apenas com número inteiros, não havendo, portanto, problemas de arredondamento, levando sempre a valores exatos, o que não ocorre quando se tiram o logaritmo dos fatoriais.

A seguir são apresentados dois teoremas que permitem contornar o problema do estouro aritmético nos cálculos intermediários usando fórmulas de recorrência.

Teorema 4.

$$\binom{n+1}{k} = \frac{n+1}{n-k+1} \binom{n}{k} \quad (35)$$

Demonstração. Pelas propriedades do fatorial tem-se que

$$(n+1)! = (n+1) \cdot n! \quad (36)$$

e

$$(n-k+1)! = (n-k+1) \cdot (n-k)! \quad (37)$$

A combinação expressa por (32) pode ser calculada com

$$\binom{n+1}{k} = \frac{(n+1)!}{k!(n-k+1)!} \quad (38)$$

Que usando-se as relações (36) e (37) pode ser escrita como

$$\binom{n+1}{k} = \frac{n+1}{n-k+1} \frac{n!}{k!(n-k)!} \quad (39)$$

Onde a segunda fração do lado direito da expressão é a combinação de n , k a k , o que completa a demonstração.

Teorema 5.

$$\binom{n}{k+1} = \frac{n-k}{k+1} \binom{n}{k} \quad (40)$$

Demonstração. Das propriedades da função fatorial tem-se

$$(k+1)! = (k+1) \cdot k! \quad (41)$$

e

$$(n-k)! = (n-k) \cdot (n-k-1)! \Rightarrow (n-k-1)! = \frac{(n-k)!}{(n-k)} \quad (42)$$

A combinação desejada pode ser calculada como

$$\binom{n}{k+1} = \frac{n!}{(k+1)!(n-k-1)!} \quad (43)$$

Substituindo-se as relações (41) e (42) chega-se a

$$\binom{n}{k+1} = \frac{n-k}{k+1} \frac{n!}{k!(n-k)!} \quad (44)$$

Onde a Segunda fração do lado direito é a combinação de n , k a k . Isso completa a demonstração.

As fórmulas apresentadas nos teoremas 4 e 5 são particularmente úteis se o problema a ser resolvido é o de calcular uma seqüência de combinações, como, por exemplo, todas as probabilidades associadas a um experimento binomial.

As fórmulas (35) e (40) permitem que se calcule uma combinação a partir de outra já conhecida fazendo-se apenas algumas operações de soma, subtração, multiplicação e divisão de inteiros. Uma combinação inicial simples é

$$\binom{n}{0} = 1 \quad (45)$$

Partindo-se dessa combinação pode-se calcular uma seqüência de combinações até que se chegue à combinação desejada.

5 O Cálculo das Probabilidades.

Por definição, o valor de uma probabilidade é sempre um número real entre 0 e 1. Nesse sentido mesmo que os cálculos intermediários para se chegar ao valor da probabilidade levem a problemas de estouro aritmético ou qualquer outro tipo de instabilidade, o resultado final será um número comum (entre zero e um). Pode ocorrer que, no cálculo da probabilidade binomial ou hipergeométrica resultado do cálculo da combinação leve a um estouro aritmético. Para contornar esse problema apresentamos algumas aproximações são feitas diretamente sobre a probabilidade, evitando-se ao máximo o cálculo das combinações.

5.1 Fórmulas de Recorrência.

Teorema 6 : Seja

$$P(X=k)=B(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (46)$$

Então

$$B(k+1) = \frac{n-k}{(k+1)(1-p)} \cdot B(k), \text{ para } k=1..n \quad (47)$$

$$\text{com } B(0) = (1-p)^n = e^{n \ln(1-p)}$$

Demonstração: A partir de (46) tem-se que

$$B(K+1) = \binom{n}{K+1} p^{K+1} \cdot (1-p)^{n-K-1} \quad (48)$$

Usando-se o teorema 5 e dividindo-se e multiplicando a expressão por $(1-p)$ tem-se que

$$B(k+1) = \frac{n-k}{K+1} \binom{n}{k} p^k \cdot \frac{(1-p)^{n-k}}{1-p} \quad (49)$$

Que pode ser escrita como

$$B(k+1) = \frac{n-k}{(k+1)(1-p)} \binom{n}{k} p^k (1-p)^{n-k} \quad (50)$$

Como a parte direita do lado direito da expressão é a própria expressão $B(k)$, isso completa a demonstração.

Teorema 7. Seja

$$P(X=k) = H(k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}} \quad (51)$$

Então

$$H(k+1) = \frac{(r-k)}{(k+1)} \cdot \frac{(n-k)}{(N-r-n+k+1)} \cdot H(k), \text{ para } k=1..n \quad (52)$$

$$\text{com } H(0) = \frac{(N-r)! (N-n)!}{(N-r-n)! N!} = e^{\ln(N-r)! + \ln(N-n)! - \ln(N-r-n)! - \ln N!}$$

Demonstração: A partir de (51) tem-se

$$H(k+1) = \frac{\binom{r}{k+1} \binom{N-r}{n-k-1}}{\binom{N}{n}} \quad (53)$$

Ao aplicar a relação (44) em cada uma das combinações chega-se a

$$= \frac{\frac{r-k}{k+1} \binom{r}{k} \frac{n-k}{N-r-n+k+1} \binom{n-r}{n-k}}{\binom{N}{n}} \quad (54)$$

Pode-se observar que as combinações formam $H(k)$. Isso completa a demonstração.

5.2 Aproximação da Hipergeométrica pela Binomial

Teorema 8: Admita-se que X tenha distribuição hipergeométrica. Se o tamanho da população N for muito maior que o tamanho da amostra n então teremos:

$$P(X = K) = \binom{n}{k} \binom{r}{N}^k \left(1 - \frac{r}{N}\right)^{n-k} \quad (55)$$

Demonstração.

A distribuição binomial é aplicável quando fazemos amostragem com reposição (visto que, neste caso, a probabilidade de obter um elemento com a característica estudada permanece constante), enquanto a distribuição hipergeométrica é aplicável quando fazemos amostragem sem reposição. Se o tamanho da população N for grande, não fará muita diferença se fizermos ou não retornar ao lote uma peça determinada, antes que a próxima seja escolhida. Assim se o tamanho da população N for suficientemente grande, a distribuição de X poderá ser aproximada pela distribuição binomial. Onde a probabilidade, p , de que seja escolhido um elemento com a característica estudada é de r/N . Isso completa a demonstração.

5.3 Aproximação da Binomial pela Normal (DeMoivre-Laplace)

Teorema 9. Se X tiver uma distribuição binomial com parâmetros n e p , e se n for grande, então:

$$P(a \leq X \leq b) = F\left(\frac{b+0,5-np}{\sqrt{np(1-p)}}\right) - F\left(\frac{a-0,5-np}{\sqrt{np(1-p)}}\right) \quad (56)$$

Onde $f(z)$ é a função de distribuição acumulada da função normal padrão.

Demonstração: Se X tem distribuição binomial com parâmetros n e p , então o Valor Esperado e a Variância são respectivamente $E(X)=np$ e $V(X)=np(1-p)$. Como X é a soma de n provas de Bernouilli, pelo Teorema Central do Limite X terá distribuição aproximadamente normal com média np e variância $np(1-p)$. Conforme o corolário do teorema 3

$$Y = \frac{X - np}{[np(1-p)]^{1/2}} \quad (57)$$

terá uma distribuição aproximadamente $N(0,1)$ quando n for suficientemente grande.

Na prática esta aproximação será válida para valores de $n > 10$, desde que p seja próximo de $\frac{1}{2}$. Se p for necessário de 0 ou 1, n deverá ser um tanto maior, para garantir uma boa aproximação.

Ao empregar da aproximação normal à distribuição binomial, estaremos aproximando a distribuição de uma variável aleatória discreta com a distribuição de uma variável contínua. Por isso, algum cuidado deve ser tomado como os pontos extremos dos intervalos considerados. Por exemplo, para uma variável aleatória contínua, $P(X=3)=0$, enquanto para uma variável aleatória discreta esta probabilidade pode ser não nula.

Segundo Meyer, as seguintes correções de continuidade melhoram a aproximação acima:

$$a) P(X = K) = P\left(k - \frac{1}{2} \leq x \leq k + \frac{1}{2}\right), \quad (58)$$

$$b) P(a \leq x \leq b) = P\left(a - \frac{1}{2} \leq x \leq \frac{1}{2} + b\right)$$

Essas correções estão consideradas na expressão (56).

6 Conclusão.

Apesar das limitações das máquinas de calcular, incluindo aqui os computadores, é possível calcular as probabilidades binomial e hipergeométrica com eficiência e precisão. Aproximações como a fórmula de Taylor ou frações contínuas, como a que deu origem à fórmula de Lanczos, reduzem o esforço computacional dramaticamente, fazendo com que máquinas de pequeno porte realizem operações que, de outra forma, poderiam ser imprecisas e demoradas.

O cálculo numérico, por sua vez, tem contribuído grandemente para a aplicação dos modelos matemáticos a problemas reais. Desta forma, acreditamos que o cálculo numérico juntamente com noções de programação devem ser conteúdos importantes na formação de matemáticos e cientistas que utilizam modelos matemáticos como ferramenta para suas decisões.

7 Bibliografia.

1. ABRAMOWITZ, M. e STEGUN, I. *A Handbook of Mathematical Functions*. Nova Iorque:Ed. Dover,1972.
2. Ávila, Geraldo S. S. *Cálculo Diferencial e Integral Volume III*. Livros Técnicos e Científicos Editora S.A. Rio de Janeiro, 1979
3. Boas, Mary, L. *Mathematical Methods in the Physical Sciences* 2ª edição. Ed. John

Wiley & Sons – New York, 1983 pp.472-473

4. COOKE, D., CRAVEN, A. H. e CLARKE, G. M. *Basic Statistical Computing*. Londres: Ed. Edward Arnold Publ. Co., 1981
5. DACHS, J. Norberto W. *Estatística Computacional*, Rio de Janeiro:Ed. LTC,1988.
6. FELLER, William. *Probability Theory and its Applications*. Nova Iorque: Ed. John Wiley and Sons, Inc, 1950.
7. Lanczos, C. *SIAM Journal on Numerical Analysis*, ser. B, vol. 1, pp. 86-96 - 1964
8. MEYER, Paul L. *Probabilidade e aplicações à Estatística*. Rio de Janeiro: LTC – Livros Técnicos e Científicos, S.A, 1983.
9. PEREIRA, Wilson. *Elementos de Estatística*. São Paulo:McGraw-Hill do Brasil, 1984.
10. PRESS, William H., TEUKOLSKY, Saul A., VETTERLING, William T., e FLANNERY, Brian P. *Numerical Recipes in C*. Cambridge, 1997.
11. SPIEGEL, Murray Ralp. *Probabilidade e Estatística*. São Paulo: McGraw-Hill do Brasil, 1978.