METRVM
Emendabis Mensvram Silvarvm

ISSN 1519–5058

Um Algoritmo Matricial para Construção da "Curva de Acumulação de Espécies"

João Luís F. Batista
Ana Cristina Schilling

ABSTRACT

A Matricial Algorithm for Construction of the "Species Acumulation Curve"

Species acumulation curve is a graphical representation frequently used in ecology, phytosociology and forest inventory. This document presents a matrix algorithm for the construction of the species acumulation curve starting from a matrix of presence-absence of species per sampling unit. All steps in the algorithm are matrix operations and a fast processing is expected in vectorial processing based softwares. The construction of empirical "average" and "confidence interval" curves throught the permutation of the order of sampling units is presented.

APRESENTAÇÃO

A "curva de acumulação de espécies" é uma representação gráfica com grande utilização nas áreas de ecologia, fitossociologia e inventário florestal. Esse documento apresenta um algoritmo matricial para construção da curva de acumulação de espécies que parte da matriz de presença-ausência de espécies por unidade amostral. Como todos os passos do algoritmo são operações matriciais, espera-se que a computação da curva com base nesse algoritmo seja bastante rápida em softwares com processamento vetorial. O documento ilustra ainda que, com base nesse algoritmo, se torna bastante simples gerar curvas "médias" e intervalos de confiança empíricos através da permutação da ordem das unidades amostrais.

METRVM, n.3/abril/2006

METRVM (ISSN 1519-5058) é uma publicação on-line do Laboratório de Métodos Quantitativos do Depto. de Ciências Florestais, ESALQ, Universidade de São Paulo. Seu objetivo é disponibilizar documentos de carácter técnico produzidos pelo laboratório na área de mensuração, biometria e bioestatística florestal.

METRVM é produzido utilizando sistema TEX & LATEX sob o sistema operacional Linux.

http://lmq.esalq.usp.br/METRVM/

1 *Introdução*

A "curva de acumulação de espécies" é uma representação gráfica com grande utilização nas áreas de ecologia, fitossociologia e inventário florestal (Braun-Blanquet, 1932; Mueller-Dombois e Ellenberg, 1974; Colwell e Coddington, 1994). Seu gráfico consiste no número de espécies (grafado no eixo das ordenadas ou eixo y) contra o esforço amostral (grafado no eixo das abscissas ou eixo x). O esforço amostral pode ser medido de diversas formas: número de unidades amostrais, número de indivíduos amostrados, horas de observação, etc.

Nos estudos de fitossociologia de florestas e nos levantamentos florestais em geral, essa curva é freqüentemente designada por "curva do coletor", pois é utilizada como forma de definição da "*suficiência amostral*". A suficiência amostral é um conceito originário dos estudos fitossociológicos que exprime o tamanho de amostra necessário para se representar adequadamente uma comunidade vegetal (Mueller-Dombois e Ellenberg, 1974; Cain e Castro, 1959). Já nos estudos de biogeografica de ilhas, a curva de acumulação de espécies é chamada de "curva espécie-área", sendo normalmente construída com o número de espécies que formam uma comunidade em uma ilha, grafado contra com o tamanho (área) da ilha (Colwell e Coddington, 1994).

A curva de acumulação de espécies é uma expressão da diversidade de espécies numa comunidade. Comparando-se comunidades de maior e menor diversidade, a curva das primeiras terão uma inclinação mais acentuada e uma assíntota mais elevada. Uma variação da curva proposta como representação da diversidade é a curva do número de espécie por número de individuos amostrados (Colwell e Coddington, 1994; Gotelli e Colwell, 2001).

Nos trabalhos de fitossociologia, a curva do coletor é geralmente construída utilizando-se a ordem em que as parcelas foram medidas no campo como ordem "natural" para acumular o número de espécies e o número de parcelas ou a área amostrada. Nas curvas assim geradas, é comum que o pesquisador interprete rápidos crescimentos ou patamares temporários como atributos característicos da comunidade sendo estudada. Numa perspectiva de métodos de amostragem, no entanto, a comunidade sendo estudada é a população alvo e qualquer ordem das unidades amostrais é absolutamente arbitrária, inclusive a ordem em que as unidades foram medidas em campo. Exceção a esse conceito é quando se estuda não uma comunidade mas um gradiente ambiental onde ocorre uma mudança gradual de comunidade (Colwell e Coddington, 1994). Sendo qualquer ordem para construção da curva de acumulação de espécies arbitrária, o ideal é que várias ordens sejam utilizadas e uma "curva média" seja construída. A construção dessa curva pode ser também realizada através do método proposto por Coleman (1981), que apresenta uma fórmula baseada na abundância relativa de cada espécie na amostra.

Entretanto, em nossas pesquisas tornou-se necessário permutar a ordem das unidades amostrais para verificar o comportamento da curva de acumulação de espécies em diferentes situações. Para implementar a construção rápida de um grande número de curvas foi necessário um procedimento automático e de grande velocidade. Esse tra-

balho apresenta o desenvolvimento de um algoritmo matricial que permite a construção da curva de acumulação de espécies, sem a necessidade de incluir no código controles de fluxo condicionais ("*if* ... *then* ... ").

Inicialmente, o algoritmo matricial é apresentado num exemplo simples com 7 espécies e 7 unidades amostrais. A seguir, apresenta-se alguns exemplos de construção da curva de acumulação de espécies. Por fim, conclui-se monstrando a implementação do algoritmo na linguagem S, que é a linguagem utilizada nos softwares estatísticos R e Splus (Chambers, 1998).

2 *Apresentação do algoritmo*

O algoritmo se inicia com uma tabela de número de árvores por espécie por unidade amostral:

```
   especies            un. amostrais
                       1 2 3 4 5 6 7
-------------------------------------
Alchornea triplinervia 0 0 5 1 0 0 0
Bathysia meridionalis  0 0 1 1 0 0 1
Buchenavia kleinii     0 0 0 0 0 1 1
Hymenea courbaril      0 0 1 0 0 0 0
Cabralea canjerana     2 0 0 0 0 0 0
Campomanesia guaviroba 0 1 3 3 0 0 0
Cedrela fissilis       1 0 0 0 0 0 0
```

1° passo: Converter a tabela de número de árvores por espécie numa tabela de presença-ausência de espécies por unidade amostral.

```
                       1 2 3 4 5 6 7
Alchornea triplinervia 0 0 1 1 0 0 0
Bathysia meridionalis  0 0 1 1 0 0 1
Buchenavia kleinii     0 0 0 0 0 1 1
Hymenea courbaril      0 0 1 0 0 0 0
Cabralea canjerana     1 0 0 0 0 0 0
Campomanesia guaviroba 0 1 1 1 0 0 0
Cedrela fissilis       1 0 0 0 0 0 0
```

Essa tabela dá origem a uma matriz que chamaremos de A. Apesar de não apresentarmos mais as espécies e unidades amostrais, devemos lembrar que as *linhas* da matriz indicam as espécies, enquanto que as *colunas* indicam as unidades

amostrais.

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2° passo: Realizar uma soma acumulativa em cada linha da matriz A, gerando uma matriz B que indica, ao longo das linhas, o número de vezes que cada espécie aparece à medida que as unidades amostrais se acumulam.

$$B = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 & 2 & 2 \\ 0 & 0 & 1 & 2 & 2 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 3 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

3° passo: Gerar uma matriz C eliminando-se a última coluna da matriz B e convertendo todos os elementos não nulos (≥ 1) em zero e todos os elementos nulos em um (1).

$$B = \left[\begin{array}{cccccc|c} 0 & 0 & 1 & 2 & 2 & 2 & 2 \\ 0 & 0 & 1 & 2 & 2 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 3 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right] \longrightarrow C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Considerando que em cada linha da matriz C acumulam-se as unidades amostrais, os elementos nulos (0) indicam quando cada espécie já está presente na amostra.

4° passo: Gerar uma matriz D fazendo-se a diferença entre as colunas sucessivas da matriz B (elimina-se a primeira coluna).

$$B = \left[\begin{array}{c|cccccc} 0 & 0 & 1 & 2 & 2 & 2 & 2 \\ 0 & 0 & 1 & 2 & 2 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 3 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right] \longrightarrow D = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Os elementos unitários (1) da matriz D indicam em que unidade amostral cada espécie entra na amostra. Se numa dada linha há mais de um elemento unitário é porque a espécie foi encontrada em mais de uma unidade amostral.

5° passo: Gerar uma matriz E com o produto *elemento-a-elemento* das matrizes C e D.

$$C \times D = E = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Os elementos unitários (1) da matriz E indicam em que unidade amostral cada espécie entra na amostra *pela primeira vez, excetuando-se a primeira unidade amostral*.

6° passo: Gerar uma matriz F acrescentando à matriz E a primeira coluna da matriz A.

$$F = \left[\begin{array}{c|cccccc} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Essa matriz F indica a presença de cada espécie somente na unidade amostral (coluna) em que ela surge pela primeira vez. Isso pode ser notado comparando-a com a matriz original de presença-ausência:

MATRIZ ORIGINAL:

	unidades amostrais						
	1	2	3	4	5	6	7
Alchornea triplinervia	0	0	1	1	0	0	0
Bathysia meridionalis	0	0	1	1	0	0	1
Buchenavia kleinii	0	0	0	0	0	1	1
Hymenea courbaril	0	0	1	0	0	0	0
Cabralea canjerana	1	0	0	0	0	0	0
Campomanesia guaviroba	0	1	1	1	0	0	0
Cedrela fissilis	1	0	0	0	0	0	0

MATRIZ F:

	unidades amostrais						
	1	2	3	4	5	6	7
Alchornea triplinervia	0	0	1	0	0	0	0
Bathysia meridionalis	0	0	1	0	0	0	0
Buchenavia kleinii	0	0	0	0	0	1	0
Hymenea courbaril	0	0	1	0	0	0	0
Cabralea canjerana	1	0	0	0	0	0	0
Campomanesia guaviroba	0	1	0	0	0	0	0
Cedrela fissilis	1	0	0	0	0	0	0

7° passo: Somando-se as colunas da matriz F obtem-se o número de "novas" de espécies por unidade amostral, isto é, o número de "espécies novas" que cada unidade amostral acrescenta à lista de espécies.

unidades amostrais	1	2	3	4	5	6	7
número de espécies novas	2	1	3	0	0	1	0

8° passo: A soma acumulada do número de espécies novas por unidade amostral gera o número acumulado de espécies, necessário para se construir a curva do coletor.

unidades amostrais	1	2	3	4	5	6	7
número de espécies novas	2	1	3	0	0	1	0
número de acumulado de espécies	2	3	6	6	6	7	7

3 Exemplos com Aplicação do Algoritmo

Para enfatizar que o algoritmo apresentado independe do uso de parcelas como unidade amostral, o exemplo apresentado se refere a um levantamento onde se utilizou o método de pontos de Bitterlich, também chamado de método da enumeração angular, como método amostral. Foram locados 40 pontos em floresta ombrófila densa no Município de Sete Barras, SP, dispostos na forma de quatro linhas com 10 pontos em cada linha e distância de 100 m entre pontos. As linhas foram locadas perpendicularmente ao rio Taquarí em posições selecionadas aleatoriamente num trecho de 5 km ao longo do rio.

3.1 Curva Baseada na Relação Espécie-Unidade Amostral

A figura 1 mostra a curva de acumulação de espécies obtida utilizando-se pontos de Bitterlich como unidades amostrais. As unidades amostrais podem ser de qualquer tipo: parcelas de diferentes formas ou tamanhos, pontos (métodos de quadrantes e de Bitterlich), linhas (método de interceptação por linha) ou mesmo os indivíduos, embora esses últimos não sejam tecnicamente unidades amostrais.

No caso dos dados que geraram a figura 1, foram observadas 103 espécies arbóreas. Um aspecto ressaltado pela figura é a maneira de se grafar e suavizar a linha da curva. Tradicionalmente, a curva é grafada como uma linha contínua tanto no eixo Y (número de espécies), quanto no eixo X (número de unidades amostrais). Uma forma mais precisa de grafar a curva, seria assumir que os acréscimos se dão de forma discreta de uma unidade amostral para outra, o que resulta numa linha na forma de "escada". A figura 1 ilustra que os vértices superiores da linha na forma de "escada" são coincidentes com a linha contínua.

A construção da curva de acumulação de espécies com base em permutação da ordem das unidades amostrais gera uma curva média equivalente àquela proposta por Coleman (1981). Enquanto o método de Coleman gera intervalos de confiança teóricos, a construção da curva por permutação permite gerar intervalos de confiança *empíricos*, isto é, intervalos baseados em quantis da distribuição das curvas geradas nas permutações. A figura 2 apresenta esses intervalos com os mesmos dados apresentados na figura 1.

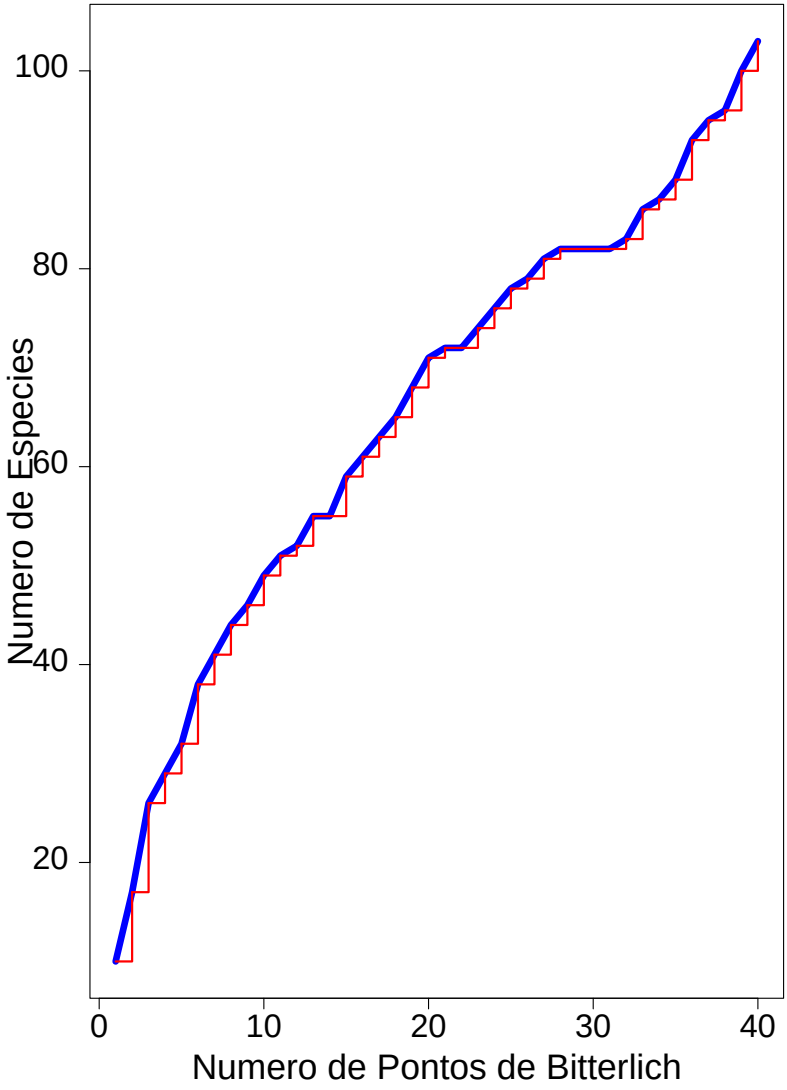

Figura 1: Curva de acumulação de espécies com base em **40 pontos de Bitterlich** como unidades amostrais em floresta ombrófila densa no Município de Sete Barras, SP.
A linha em azul é a forma tradicional de grafar a curva de acumulação de espécies, sendo obtida pela simples ligação dos pontos no gráfico. A curva em vermelho tem forma de "escada" e enfatiza o fato de que o aumento do número de espécies se dá de forma discreta de uma unidade amostral para outra.

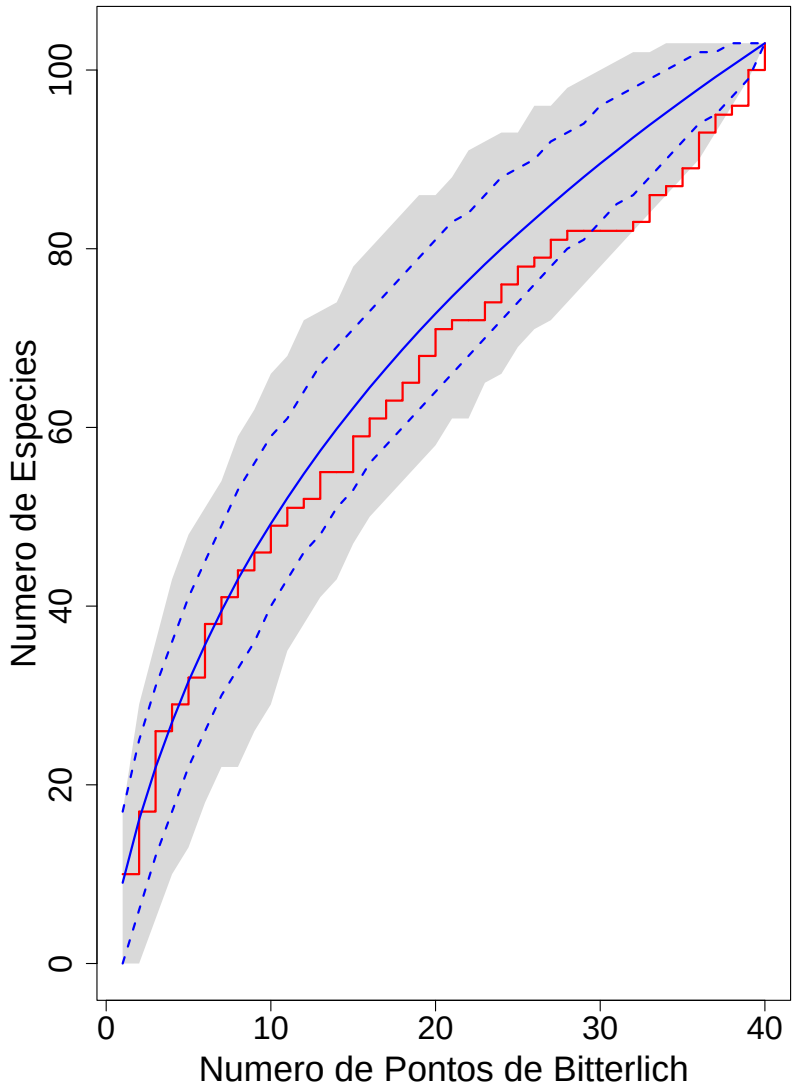

Figura 2: Curva de acumulação de espécies com base em **40 pontos de Bitterlich** como unidades amostrais em floresta ombrófila densa no Município de Sete Barras, SP. A área sombreada no gráfico indica a região ocupada pelas curvas geradas por 5000 permutações da ordem das unidades amostrais. As linhas no gráfico indicam a curva gerada pela ordem de medição no campo (linha vermelha), a curva média das 5000 permutações (linha sólida azul) e o intervalo de confiança empírico de 95% com base nas 5000 permutações (linhas tracejadas em azul).

É interessante notar que a curva média, tanto a curva construída pelo método de Coleman quanto a gerada por permutação da ordem das unidades amostrais, pressupõe que a localização dos índivíduos das diferentes espécies é completamente aleatória no espaço, o que é conhecido como *Hipótese da Localização Aleatória* (*Random Placement Hypothesis*). Nas florestas tropicais, essa hipótese não se verifica, sendo natural que a curva construída com base numa ordem arbitrária extrapole o intervalo gerado por permutação, como observado na figura 2.

3.2 Curva Baseada na Relação Espécie-Indivíduo Amostrado

Para ilustrar que a curva de acumulação de espécies pode ser construída também com base nas árvores individuais como unidades amostrais, a figura 3 apresenta a mesma curva das figuras anteriores mas utilizando as árvores como unidades amostrais. Note que quando as árvores são tomadas como unidades amostrais a linha contínua e a linha na forma de "escada" são coincidentes.

Também nesse caso é possível se construir intervalos de confiança empíricos com base em permutação da ordem das unidades amostrais. No caso das árvores individuais, a região gerada pelo intervalo de confiança empírico difere da região no caso de unidades amostrais em dois aspectos. Primeiramente, no caso das árvores, a região se inicia e finaliza num ponto, enquanto que no caso de unidades amostrais a região se inicia num intervalo e finaliza num ponto. O ponto de finalização dos intervalos é o número total de espécies pelo número total de árvores ou unidades amostrais. Já o início da região começa com um número variável de espécies no caso das unidades amostrais, mas começa sempre com uma única espécie no caso das árvores individuais. O segundo aspecto é que as linhas que definem o intervalo de confiança empírico, no caso das árvores individuais, é sempre em forma de "escada".

Note que no caso da curva de acumulação de espécies com base no número de árvores amostradas, a curva gerada pela ordem arbitrária de observação no campo sai do intervalo empírico gerado de forma bem menos marcante, ainda que o mesmo número de simulações (5000) tenha sido utilizado para gerar os intervalos. Esse efeito pode resultar do fato que, ao se utilizar indivíduos na construção da curva de acumulação de espécies, a estrutura espacial da floresta implícita nos dados é parcialmente perdida. Nesse caso, a informação sobre as espécies se encontra menos agregada que no caso das unidades amostrais, onde os indivíduos estão agrupados, resultando que a curva na ordem arbitrária se aproxime da curva gerada por permutação.

4 Implementação do Algoritmo na Linguagem S

O código na linguagem S necessário para gerar a curva de acumulação de espécies é bastante direto e compacto. A função `sac`, apresentada abaixo, assume como único argumento uma matriz de presença/ausência das espécies nas unidades amostrais, sendo que as unidades amostrais são representadas **nas linhas** e as espécies **nas colunas**.

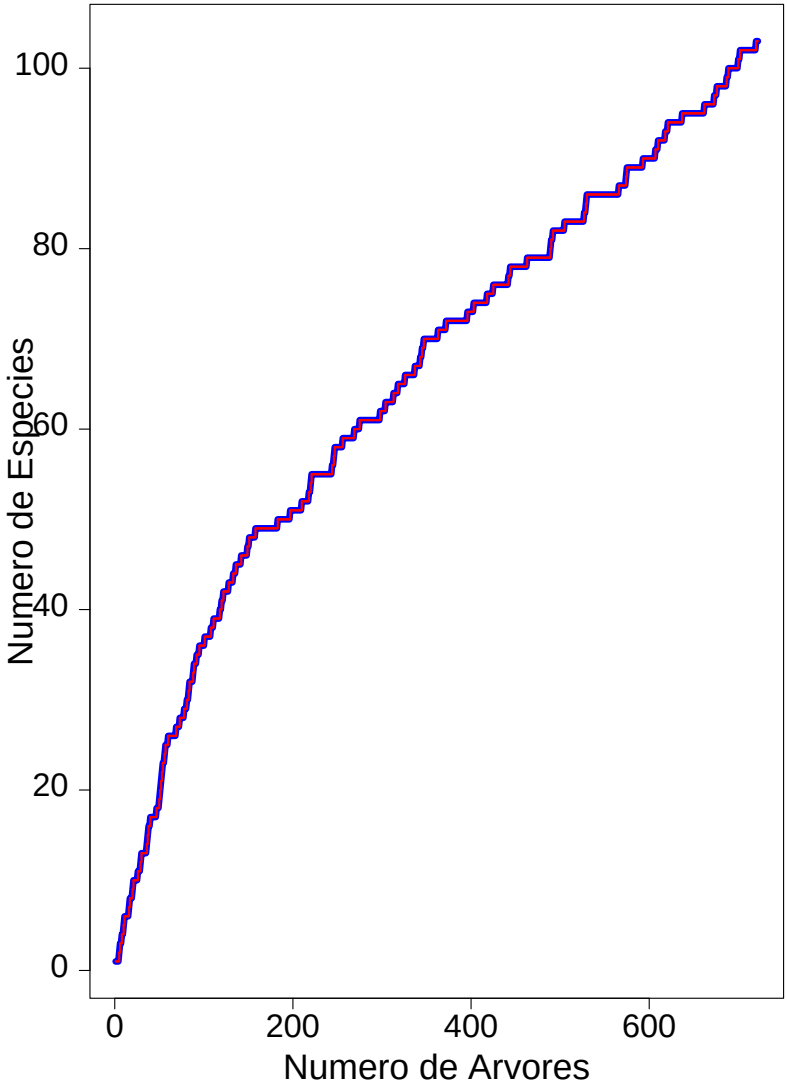

Figura 3: Curva de acumulação de espécies com base no **número de árvores** amostradas em floresta ombrófila densa no Município de Sete Barras, SP.
A linha em azul é a forma tradicional de grafar a curva de acumulação de espécies, sendo obtida pela simples ligação dos pontos no gráfico. A curva em vermelho tem forma de "escada" e enfatiza o fato de que o aumento do número de espécies se dá de forma discreta de uma unidade amostral para outra.

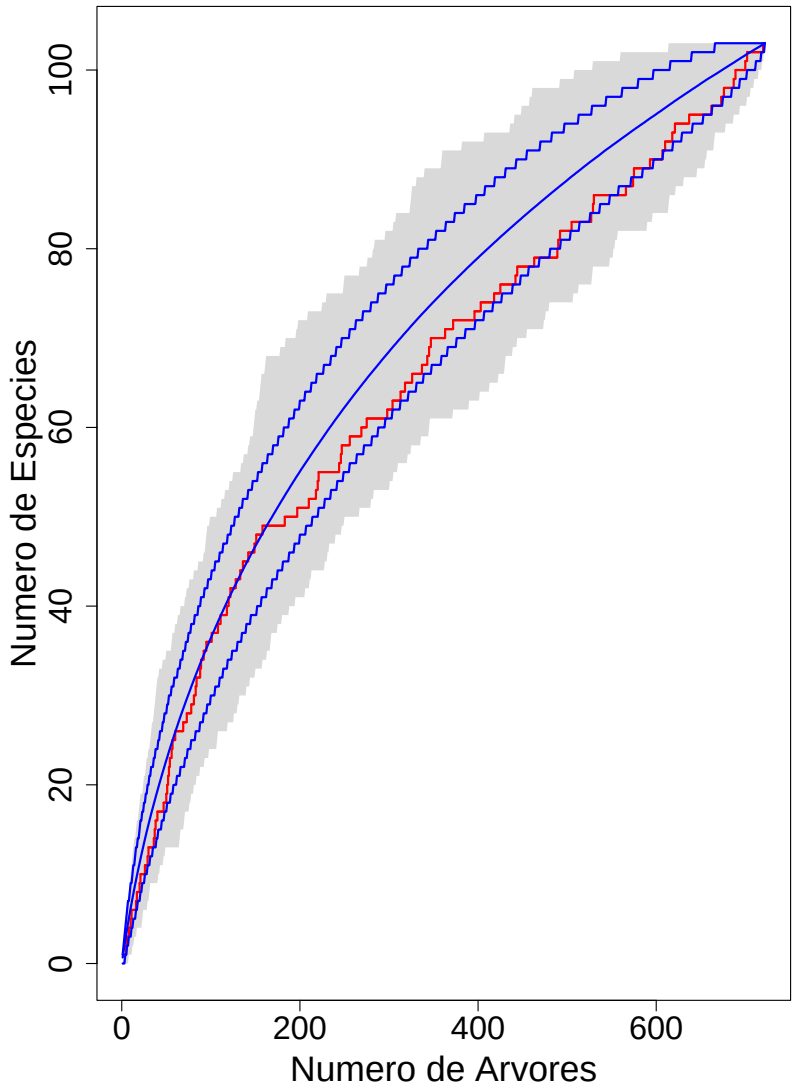

Figura 4: Curva de acumulação de espécies com base no **número de árvores** amostradas em floresta ombrófila densa no Município de Sete Barras, SP. A área sombreada no gráfico indica a região ocupada pelas curvas geradas por 5000 permutações da ordem das unidades amostrais. As linhas no gráfico indicam a curva gerada pela ordem de medição no campo (linha vermelha), a curva média das 5000 permutações (linha sólida azul) e o intervalo de confiança empírico de 95% com base nas 5000 permutações (linhas na forma de "escada" em azul).

```
sac <- function(x)
{
#
# Argumento "x"
#    -> matriz de presenca-ausencia das especies (passo no. 1):
#          linhas = unidades amostrais (parcelas, pontos, arvores)
#          colunas  = especies
#
   n.ua <- dim(x)[1]                # no. de unidades amostrais
   n.sp <- dim(x)[2]                # no. de especies
   acumula <- apply(x, 2, cumsum)              # Passo no. 2
   reverte <- acumula[ - n.ua, ] == 0          # Passo no. 3
   difere <- apply(acumula, 2, diff)           # Passo no. 4
   coletor <- rbind( x[1,], difere * reverte)  # Passos no. 5 e 6
   sac <- cumsum( apply(coletor, 1, sum) )     # Passos no. 7 e 8
   return(sac)
}
```

Referências Bibliográficas

Braun-Blanquet, J. *Plant Sociology: the study of plant communities*. New York/London: McGraw-Hill, 1932.

Cain, S. A. e Castro, G. M. O. *Manual of vegetation analysis*. New York: Harper & Brothers, 1959.

Chambers, J. *Programming with data: a guide to the S language*. New York: Springer, 1998.

Coleman, B. On random placement and species-area relations. *Mathematical Biosciences*, v.54, p.191-215, 1981.

Colwell, R. e Coddington, J. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of Royal Society of London, Series B*, v.345, p.101-118, 1994.

Gotelli, N. e Colwell, R. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, v.4, p.379-391, 2001.

Mueller-Dombois, D. e Ellenberg, H. *Aims and methods of vegetation ecology*. New York: John Wiley & Sons, 1974.