

AGREGAÇÃO E VIÉS EM MEDIDAS DE DIVERSIDADE: UMA  
ABORDAGEM COMPUTACIONAL INTENSIVA

Davi Butturi-Gomes

Dissertação apresentada à Universidade  
Estadual Paulista “Julio de Mesquita Filho”  
para obtenção do título de Mestre em  
Biometria.

BOTUCATU

São Paulo – Brasil

Fevereiro – 2011

AGREGAÇÃO E VIÉS EM MEDIDAS DE DIVERSIDADE: UMA  
ABORDAGEM COMPUTACIONAL INTENSIVA

Davi Butturi-Gomes

Orientador: Prof. Dr. Miguel Petrere Junior

Dissertação apresentada à Universidade  
Estadual Paulista “Julio de Mesquita Filho”  
para obtenção do título de Mestre em  
Biometria.

BOTUCATU  
São Paulo – Brasil  
Fevereiro – 2011

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE

Butturi-Gomes, Davi.

Agregação e viés em medidas de diversidade: uma abordagem computacional intensiva / Davi Butturi-Gomes. - Botucatu, 2011

Dissertação (mestrado) – Instituto de Biociências, Universidade Estadual Paulista, 2011

Orientador: Miguel Petreire Junior

Capes: 10203001

1. Estatística matemática – Processamento de dados. 2. Probabilidades – Processamento de dados.

Palavras-chave: Amostragem por quadratins; Entropia de Tsallis; *Jackknife* e *bootstrap*; Modelos de espécie/abundância.

## Agradecimentos

Ao Prof. Dr. Miguel Petreire Junior, pela orientação, ensinamentos, apoio e amizade.

Aos professores Dr. Paulo de Marco Jr. e, especialmente, Dra. Luzia Ap. Trinca pelos ensinamentos e pela colaboração em todas as fases do desenvolvimento deste trabalho.

Aos colegas Eduardo Pacífico, Henrique Giacomini e Fábio Cop pela amizade, pela colaboração no desenvolvimento dos programas e pelas correções e sugestões.

À entidade financiadora deste estudo, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (processo nº 135117/2009-7), pelo apoio financeiro.

Aos alunos e amigos da Turma 2010 do Programa de Pós-Graduação em Biometria Andrea, Diego, Iuri, Juliana e Leandro pelo companheirismo, pelos ensinamentos e todo o trabalho em equipe.

Aos colegas do Curso de Verão da Biometria, especialmente, ao Henrique do Nascimento Silva, pelos ensinamentos e pela amizade.

Aos meus pais Milton e Rosa, meus irmãos Daniela e Danilo, minha sobrinha Natália e meus cunhados, pela paciência, apoio e carinho.

À minha amada Juliana Oler, pela dedicação, carinho e atenção.

*“Pouco se pode ver do futuro, porém o suficiente  
para saber que há muito que fazer”.*

Alan Turing

## Sumário

1. INTRODUÇÃO.....	11
2. OBJETIVOS.....	14
3. MATERIAL E MÉTODOS.....	15
4. RESULTADOS.....	22
4.1. Cenários simulados.....	22
4.2. Os vieses nas estimativas da família de índices de Tsallis.....	23
4.3. Precisão e acurácia nas estimativas <i>jackknife</i> e <i>bootstrap</i> .....	32
4.4. Amostragens em quadratins: qual a melhor unidade de reamostragem a considerada?.....	39
5. DISCUSSÃO.....	41
6. CONCLUSÕES.....	47
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	49

## Lista de Tabelas e Figuras

Tabela 1: Índices de diversidade reais (parâmetros populacionais) das comunidades simuladas, de acordo com as distribuições de espécie/abundância logsérie, lognormal e geométrica.....	22
Figura 1: Esquema representando os procedimentos computacionais realizados para simular as comunidades e para a amostragem.....	21
Figura 2: Família de índices de Tsallis reais das comunidades simuladas.....	23
Figura 3: Vieses percentuais para os índices $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem <i>jackknife</i> e <i>bootstrap</i> .....	25
Figura 4: Vieses percentuais para os índices $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem <i>jackknife</i> e <i>bootstrap</i> , considerando as distribuições espécie/abundancia lognormal, logsérie ou geométrica.....	27
Figura 5: Vieses percentuais para os índices $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem <i>jackknife</i> e <i>bootstrap</i> , considerando o nível de agregação populacional (ou raio de dispersão).....	29
Figura 6: Vieses percentuais para os índices $S_q$ estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem <i>jackknife</i> e	

*bootstrap*, considerando a forma de amostragem (aleatória ou em quadratins de tamanhos 1, 2, 3 ou 4)..... 31

Figura 7: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, pelos valores dos parâmetros  $q$  da fórmula de Tsallis..... 33

Figura 8: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap* sob as distribuições de espécie/abundância logsérie, lognormal e geométrica, pelos valores dos parâmetros  $q$  da fórmula de Tsallis..... 34

Figura 9: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, considerando diferentes níveis de agregação populacional, pelos valores dos parâmetros  $q$  da fórmula de Tsallis..... 36

Figura 10: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, nas diferentes formas de amostragem, pelos valores dos parâmetros  $q$  da fórmula de Tsallis..... 38

Figura 11: Avaliações da melhor unidade amostral a ser considerada para estimar a riqueza de espécies pelos métodos *jackknife* e *bootstrap*, em comunidades amostradas por quadratins..... 40



# AGREGAÇÃO E VIÉS EM MEDIDAS DE DIVERSIDADE: UMA ABORDAGEM COMPUTACIONAL INTENSIVA

Autor: DAVI BUTTURI GOMES

Orientador: Prof. Dr. MIGUEL PETRERE JUNIOR

## RESUMO

É bastante comum no estudo de ecologia de comunidades o uso de índices de diversidade. Historicamente, os principais índices são os derivados por Shannon e por Simpson. Atualmente, existe uma família de índices, derivada da entropia de Tsallis, que inclui os índices supracitados. Igualmente importantes para descrever os padrões observados na natureza, são os modelos de distribuição de espécie/abundância. Sabe-se também que a amostragem em comunidades naturais geralmente é realizada por quadratins e que as populações das espécies na natureza exibem um padrão agregado de distribuição espacial, ao invés de uma distribuição aleatória. Esta dissertação tem como objetivo central avaliar os possíveis efeitos que a estrutura da comunidade imposta por meio de modelos clássicos, como as séries logarítmica e geométrica e a distribuição lognormal, que a agregação populacional e que o método de amostragem (por quadratins ou aleatória de indivíduos) podem ter sobre a estimação dos índices da família de Tsallis. Foram realizadas simulações baseadas no método modificado de Thomas (seguindo outros estudos na área), nas quais foram gerados diferentes cenários, representando comunidades naturais, com os três tratamentos: (i) distribuição de espécie/abundância, (ii) nível de agregação populacional e (iii) método de amostragem. Então foi

verificado o desempenho dos métodos de estimação de máxima verossimilhança (EMV), *jackknife* e *bootstrap*. O desempenho desses métodos foi medido pelo viés percentual calculado em relação ao parâmetro populacional, pela amplitude do intervalo de confiança e também pela porcentagem de casos em que o intervalo de confiança cobriu o verdadeiro valor do parâmetro populacional. Foi observado que o EMV, em todos os cenários, produziu estimativas mais enviesadas do que os métodos intensivos de estimação. Também se notou que a distribuição geométrica é o modelo de espécie/abundância, entre os avaliados, que menos causa efeito sobre o viés nas estimativas, sobretudo na estimação da riqueza de espécies. Com relação aos níveis de agregação populacional, observou-se que raios de dispersão intermediários foram aqueles enviesaram menos os estimadores. Por fim, a análise da forma de amostragem demonstrou fraco efeito do tamanho do quadratim sobre o viés nas estimativas de diversidade, porém tal forma de amostragem fere consideravelmente o pressuposto de amostragem aleatória de indivíduos. De maneira geral, o método *jackknife* se mostrou mais preciso do que o método *bootstrap* e, para índices que abordam a riqueza de espécies, também se mostrou mais acurado. Entretanto, para índices de diversidade que enfocam a dominância dentro das comunidades, é bastante recomendado o uso do método *bootstrap*.

Palavras-chave: modelos de espécie/abundância, amostragem por quadratins, *jackknife* e *bootstrap*, entropia de Tsallis.

# AGGREGATION AND BIAS IN DIVERSITY MEASURES: A COMPUTER-INTENSIVE APPROACH

Author: DAVI BUTTURI GOMES

Adviser: Prof. Dr. MIGUEL PETRERE JUNIOR

## ABSTRACT

Community ecology is often studied by the use of diversity indices. Historically, the most important indices are derived from Shannon and from Simpson's formulas. Presently, there is a family of diversity indices, derived from Tsallis entropy, which include the above ones. The species/abundance models are also important to describe the observed patterns in nature. It is well known that, when sampling in natural communities, quadrats are used and that species populations usually exhibit an aggregated pattern of spatial distribution, instead of a random one. The main objective of this dissertation is to evaluate the possible effects that the community structure imposed by the classic species/abundance models, such as the logarithmic and geometric series and the lognormal distribution, the level of aggregation and the sampling method (by quadrats or individual sampling) would interfere with the estimation of the Tsallis diversity indices family. Simulations based on the modified Thomas method were performed (following other studies on the field), in which different scenarios were computationally created, representing the biological communities, with three treatments: (i) the species/abundance model, (ii) the aggregation level and, (iii) the sampling method. Then the

performances of the maximum likelihood (MLE), *jackknife* and bootstrap estimators were evaluated. The performances are measured by the relative bias around the population parameter, the length of the confidence intervals and also by the relative number of cases in which the confidence intervals covered the population parameter. It is observed that the MLE, in every scenario, generated more biased estimations than the two evaluated resampling methods. Also it is observed that the geometric series is the species/abundance distribution which caused less bias among the tested models, especially for species richness estimation. Concerning the aggregation patterns, it is observed that the medium levels of aggregation were those which produced lesser biased estimators. Thus, the sampling method analyses have shown a low intensity effect of the quadrat size over the bias in the diversity estimations, although such sampling method goes strongly against the principle of random sampling of individuals. Generally, the *jackknife* method is most precise than the *bootstrap* and, for indices that approach the species richness, the *jackknife* is also more accurate. Nevertheless, for those diversity indices which are more influenced by dominant species, it is strongly recommended the use of the *bootstrap* method of estimation.

Keywords: species/abundance models, quadrat sampling, *jackknife* and *bootstrap*, Tsallis entropy.

## 1. INTRODUÇÃO

A pesquisa em ecologia de comunidades vem tentando descrever assembléias biológicas por meio de índices de diversidade, que levam em conta a riqueza e a abundância de espécies (MAGURRAN, 1988). Historicamente, os dois índices mais importantes são o de Shannon ( $H'$ ) e o de Simpson ( $D$ ). O primeiro é baseado na Teoria da Informação e mede a “diversidade média” entre amostras da comunidade (KREBS, 1999). Já o índice de Simpson mede a probabilidade de dois indivíduos amostrados aleatoriamente pertencerem à mesma espécie (SIMPSON, 1949).

Para os estimadores dos índices de diversidade citados acima já foram calculados os respectivos erros-padrão teóricos, por Basharin (1959) e Bowman et al. (1971) com relação ao  $H'$ , e por Simpson (1949) para o  $D$ . Em ambos os casos, foi assumida uma amostragem aleatória de indivíduos. Ainda que possível, isto é dificilmente atingido pelos ecólogos. Normalmente são utilizadas amostras em quadratins ou algum dos diversos tipos de amostragens por grupos, conhecidas por conglomerados (ROUTLEDGE, 1980), para estimar a proporção de indivíduos pertencentes a cada espécie, utilizando o estimador de máxima verossimilhança (EMV) para os índices de diversidade. Entretanto, tal estimador tem suas propriedades garantidas apenas no caso de amostragens aleatórias de indivíduos.

Existem, na verdade, diversos estimadores desses dois índices (GOOD, 1953; MILLER, 1974; PIELOU, 1975; ZAHL, 1977) e há autores que observam vantagens no uso do índice de Simpson (ROUTLEDGE, 1980) e outros que criticam seu uso (HELTSHE e FORRESTER, 1985). Segundo Peet (1974), entretanto, a importância dos índices de diversidade reside mais nas respostas que essas medidas têm de acordo com a mudança na estrutura das comunidades do que nas propriedades instantâneas que os índices possam ter,

de forma que é difícil escolher qual índice seja “melhor” ou mais representativo de uma assembléia biológica.

Uma abordagem mais nova e completa dos índices de diversidade em Ecologia é apresentada por Mendes et al. (2008), utilizando a família de índices de entropia de Tsallis (TSALLIS, 1988), nomeado  $S_q$ .

Esse índice é de especial relevância, pois incorpora os principais índices comumente utilizados. Substituindo-se o parâmetro  $q$  por valores reais não-negativos, é possível explorar uma série de índices de diversidade: quando  $q \rightarrow 1$ , observa-se que  $S_q \rightarrow H'$ ; quando  $q = 2$ , temos que  $S_q = 1 - D$ . Assim, se  $q < 1$ , enfatiza-se a riqueza de espécies e, se  $q > 1$ , aborda-se a dominância dentro da comunidade (MENDES et. al, 2008).

Uma outra família de índices foi estudada por Kempton (1979), com relação ao viés na estimação. Esta família de índices, atribuída por Kempton (1979) a Hill (1973), foi primeiramente derivada por Rényi (1961). Entretanto, nem Rényi (1961) e nem Hill (1973) haviam dado uma abordagem tão ampla à estimação desses índices. Como a família de índices de Rényi (1961) trata de funções dos índices de diversidade mais utilizados ( $H'$  e  $D$ ) e a família de índices de Tsallis (1988) aborda diretamente tais índices, é mais conveniente se usar a proposta de Tsallis (1988).

Além da tentativa de descrever comunidades biológicas por meio de medidas unidimensionais (índices de diversidade), os pesquisadores em Ecologia tentam ajustar os dados da frequência de espécies a modelos de distribuição de espécie/abundância (PIELOU, 1975; KREBS, 1999). Entre os “modelos clássicos” estão o modelo lognormal (PRESTON, 1948; BULMER, 1974), a série logarítmica (FISHER et al., 1943), a série geométrica e o *broken-stick* (MACARTHUR, 1957). Uma revisão completa, incluindo modelos mais recentes, pode ser encontrada em Ferreira e Petrere-Jr (2008).

São poucos, entretanto, os trabalhos que investigam a relação existente entre os modelos de espécie/abundância e índices de diversidade. He e Hu (2005) trazem à tona esse tema e sugerem estimar o parâmetro do modelo neutro de espécie/abundância por meio da estimação do índice de Simpson. Ainda assim, não foram investigados os efeitos que a própria estrutura da

distribuição de espécie/abundância pode ter sobre a estimação dos índices de diversidade.

Além dos possíveis efeitos que a distribuição de espécie/abundância pode ter sobre os estimadores dos índices de diversidade, a organização espacial das populações dentro de uma comunidade pode ser outra fonte de viés, de imprecisão ou falta de acurácia ao se estimar esses índices. Uma vez que o padrão agregado é a forma mais comum de distribuição espacial dos organismos (TAYLOR et al., 1978), uma amostragem aleatória de quadratins pode frequentemente produzir amostras não aleatórias de indivíduos. Dessa forma, a agregação pode influenciar o estimador dos índices de diversidade diminuindo a sua precisão e sua acurácia. Se um indivíduo se localiza próximo a outros da mesma espécie, a proporção de indivíduos da espécie dominante pode ser superestimada ou subestimada, dependendo da frequência com que as grandes agregações são incluídas na amostra.

Nesta dissertação foram comparados os desempenhos dos estimadores dos índices de diversidade representados por variações do índice unificado de Tsallis, quando estimados em comunidades com diferentes padrões de agregação populacional e com diferentes distribuições de espécie/abundância. Dentro da variação imposta ao índice unificado, estarão incluídos os índices  $H'$  de Shannon e  $D$  de Simpson, por meio dos valores  $q \rightarrow 1$  e  $q = 2$ . Além disso, foram utilizados métodos computacionais intensivos (*jackknife* e *bootstrap*) para produzir intervalos de confiança das estimativas e tentar remover o seu viés. Dessa forma, foi possível comparar esses métodos quanto à acurácia e precisão.

Este estudo tem importância estratégica, visto que os índices de diversidade têm sido estimados de forma muitas vezes inapropriada, apesar de uma longa história de alertas sobre seus vieses e seus pressupostos de amostragem (Z AHL, 1977; KEMPTON, 1979; HELTSHE & FORRESTER, 1985), além de serem desconhecidos os efeitos que as distribuições de espécie/abundância e que a agregação populacional dentro das comunidades podem exercer sobre a estimação desses índices.

## 2. OBJETIVOS

O objetivo desta dissertação foi verificar o viés dos índices de diversidade em níveis crescentes de agregação populacional, em comunidades com distribuições espécie/abundância lognormal, logsérie e geométrica, seguindo métodos de amostragens comumente utilizados em campo, tendo como hipóteses:

(i) Para um esforço amostral fixo, a presença de um maior número de espécies raras, como é o caso de comunidades com distribuição logsérie, deve implicar em maior viés nas estimativas dos índices do que comunidades que apresentam distribuição geométrica ou lognormal de espécie/abundância

(ii) A metodologia de estimação dos índices supõe que os indivíduos das diferentes espécies estão aleatoriamente distribuídos no espaço, agregados populacionais devem viciar as estimativas desses índices, de forma que maiores níveis de agregação populacional implicam em maior viés.

(iii) Outra fonte de viés dos índices pode ser resultante do tipo de amostragem adotado. Apesar dos estimadores dos índices terem como suposição a amostragem aleatória de indivíduos, isto é de difícil aplicação em campo e normalmente são utilizadas amostras em quadratins. Espera-se, portanto, que as amostras através de quadratins (que podem sortear nenhum, ou um ou mais indivíduos dentro de cada quadratim simultaneamente) resultem em maior viés nas estimativas dos índices do que naquelas amostras que consistem em sorteio de indivíduos.



### 3. MÉTODOS

Para o desenvolvimento desta dissertação foi utilizado o método de simulação de Monte Carlo. As simulações foram feitas visando gerar indivíduos de diferentes espécies, alocados em *grids* de  $100 \times 100$ , segundo diferentes distribuições de relação espécie/abundância e diferentes níveis de agregação populacional. Após gerar a comunidade, foram aplicadas amostragens aleatórias de indivíduos e por quadratins e, utilizando os dados provenientes dessas amostras, foram estimados os índices de diversidade pelo método da máxima verossimilhança (EMV) e também pelos métodos *jackknife* e *bootstrap*. As simulações foram repetidas 100 vezes em cada cenário. Todas as simulações foram programadas na plataforma Matlab® 6.1.0.450 (R 12) (THE MATHWORKS INC., 1992), software devidamente registrado nos computadores do Laboratório de Computação, Departamento de Bioestatística, Instituto de Biociências da UNESP-Botucatu (Campus Rubião Junior). As análises e gráficos foram produzidos no *software* R (R Development Core Team, 2009).

Para os padrões de espécie/abundância, foram utilizadas as distribuições logsérie (PIELOU, 1975; KREBS, 1999), lognormal (PIELOU, 1975; KREBS, 1999) e geométrica (PIELOU, 1975; KREBS, 1999). O modelo de Broken-Stick não foi simulado, pois é bastante próximo ao modelo geométrico. Todas as simulações foram realizadas por processos estocásticos. Com relação aos parâmetros das distribuições, foram utilizados aqueles referentes ao ajuste dos dados do censo de plantas realizado na Ilha de Barro Colorado por Condit et al. (2002), cujos dados estão disponíveis no pacote “vegan” da plataforma R (R Development Core Team, 2009). Tendo como base os dados do censo realizado por esse censo, os números de espécies e indivíduos que se objetivou produzir para as três distribuições foram 225 e 10.000, respectivamente. Vale salientar que o número total de indivíduos encontrados no referido censo foi

aproximadamente 20.000 indivíduos e aqui esse número foi reduzido pela metade, devido aos custos computacionais das simulações.

Para produzir os padrões agregados foi utilizada uma modificação do método de Thomas apud Diggle et al. (1976), seguindo outros trabalhos na área (DIGGLE et al., 1976; HELTSHE e BITZ, 1979; HELTSHE e FORRESTER, 1985). De acordo com o procedimento adotado, os “indivíduos pais” foram aleatoriamente distribuídos no *grid* de 100x100. Então, os “indivíduos filhos” foram dispostos em torno dos pais uniformemente entre ângulos de 0 a 360° e a uma distância normalmente distribuída com média zero e variância  $\sigma^2$ . Se, eventualmente, a simulação produziu um indivíduo fora do *grid* de 100 × 100, ele foi descartado e re-simulado. O número total de “indivíduos pais” depende da frequência da população à qual estão associados, de forma que o mínimo de “indivíduos pais” em uma população foi 1, se 10% do total da população foi menor ou igual a 1. O número de filhos associados a cada pai seguiu uma distribuição de Poisson, com média igual a 10% do total da população. O nível de agregação é inversamente proporcional a  $\sigma^2$ . Assim, valores pequenos de  $\sigma^2$  estão associados a indivíduos filhos próximos aos pais. Foram utilizadas variâncias, denominadas raios de dispersão ou níveis de agregação, iguais a 100, 50, 10 e 5 e para investigar as propriedades dos estimadores em padrões crescentes de agregação.

A amostragem foi efetuada de duas formas: (i) os indivíduos foram enumerados e sorteados, com mesma chance, até atingir 15% do total; (ii) a amostragem foi composta por quadratins de 1, 4, 9 ou 16 unidades espaciais (dentro as 10.000 do *grid*). Todos os indivíduos contidos nos quadratins fizeram parte da amostra. Os quadratins foram obtidos independentemente ao se sortear coordenadas cartesianas. Esta coordenada sorteada sempre se referiu ao vértice superior esquerdo do quadratim, levando em conta a não superposição e que toda a área estivesse dentro do *grid*. A soma das áreas das unidades amostrais foi restrita a cerca de 14,4% da área total (a escolha dessa porcentagem se deve ao fato de que escolher o mínimo múltiplo comum dos valores de área utilizados, isto é 144, facilita a programação). Vale salientar que não há necessidade de levar em conta a influência do número de quadratins em uma amostra de acordo com Heltshe e Forrester (1985). Um resumo

esquemático da geração da comunidade e da amostragem está disponível na Figura 1.

Para cada simulação, foram estimados os índices de diversidade de Tsallis (Equações 1 a 3) (TSALLIS, 1988; MENDES et. al., 2008), com o parâmetro  $q$  assumindo os valores de 0 a 2,5, em intervalos de 0,01, diretamente pelo estimador de máxima verossimilhança (EMV) e também pelos métodos *bootstrap* e *jackknife* de primeira ordem.

O *jackknife* pode ser definido como um método de reamostragem sistemática, no qual são eliminadas as unidades amostrais uma a uma e, para cada ciclo, são estimados os índices de interesse (Equações 4 e 5). Tal como o *jackknife*, o método *bootstrap* não paramétrico (Equações 6 e 7) é caracterizado pela reamostragem, mas possui uma abordagem aleatória, na qual se sorteia com reposição e com igual probabilidade, uma amostra do mesmo tamanho da original a ser usada para a estimação do índice de interesse (MILLER, 1974; EFRON, 1979; SMITH e BELLE, 1984; MANLY, 2007). Este procedimento deve ser repetido um número pré-fixado de vezes e, neste trabalho, o sorteio foi repetido 500 vezes.

O índice de Tsallis é assim definido (TSALLIS, 1988):

$$S_q = w \frac{1 - \sum_{i=1}^S p_i^q}{q - 1} \quad (1)$$

onde,  $p_i = N_i/N$  é a proporção da espécie  $i$ ,  $N_i$  é a abundância da espécie  $i$ ,  $N = \sum N_i$  é o número total de indivíduos,  $q$  é um parâmetro real, usualmente não-negativo,  $S$  é o número total de espécies presentes na comunidade e  $w$  é uma constante relacionada à grandeza da unidade.

Ao trazer o índice de Tsallis para a Ecologia, Mendes et al. (2008) fizeram  $w = -1$  e obtiveram:

$$S_q = \frac{1 - \sum_{i=1}^S p_i^q}{1 - q} \quad (2)$$

Esta versão modificada do índice de Tsallis é a que foi utilizada nesta dissertação.

O EMV deste índice, derivado do modelo multinomial, foi apresentado por Mendes et al. (2008)

$$\widehat{S}_q = \frac{1 - \sum_{i=1}^s \hat{p}_i^q}{1 - q} \quad (3)$$

onde,  $\hat{p}_i = n_i/n$  é a proporção da espécie  $i$  na amostra,  $n_i$  é a abundância da espécie  $i$  na amostra,  $n = \sum n_i$  é o número total de indivíduos presentes na amostra,  $q$  é um parâmetro real, usualmente não-negativo e  $s$  é o número total de espécies presentes na amostra.

As estimativas *jackknife* foram realizadas segundo uma adaptação da fórmula de Miller (1974) para os índices de Tsallis:

$$S_q^{\widehat{Jackknife}} = k \cdot \widehat{S}_q - \frac{(k-1)}{k} \cdot \sum_{j=1}^k \widehat{S}_{q-j} \quad (4)$$

onde  $S_q^{\widehat{Jackknife}}$  é a estimativa *jackknife* do índice  $S_q$ ,  $k$  é o número total de unidades amostrais,  $\widehat{S}_q$  é o EMV do índice  $S_q$  e  $\widehat{S}_{q-j}$  é o EMV de  $S_q$ , quando retiramos a unidade amostral  $j$ .

A variância amostral das estimativas *jackknife* pode ser obtida, numa abordagem tratando os pseudo-valores de cada reamostragem aproximadamente independentes e identicamente distribuídos (TUKEY, 1958 apud MILLER, 1974; SMITH e VAN BELLE, 1984; EFRON e TIBSHIRANI, 1993), via a fórmula:

$$V(S_q^{\widehat{Jackknife}}) = \frac{\sum_{j=1}^k (\widehat{S}_{q-j} - S_q^{\widehat{Jackknife}})^2}{k(k-1)} \quad (5)$$

As estimativas *bootstrap* foram realizadas segundo uma adaptação da fórmula de Efron (1979) para os índices de Tsallis:

$$S_q^{\widehat{bootstrap}} = 2 \cdot \widehat{S}_q - \frac{1}{B} \cdot \sum_{b=1}^B \widehat{S}_{q_b} \quad (6)$$

onde  $S_q^{\widehat{bootstrap}}$  é a estimativa *bootstrap* não paramétrica e corrigida para o viés do índice  $S_q$ ,  $B$  é o número total de repetições *bootstrap*,  $\widehat{S}_q$  é o EMV do índice  $S_q$  e  $\widehat{S}_{q_b}$  é o EMV de  $S_q$ , quando é realizada uma repetição *bootstrap*.

A variância das estimativas *bootstrap* pode ser obtida via a fórmula (EFRON, 1979; SMITH e VAN BELLE, 1984; EFRON e TIBSHIRANI, 1993):

$$V(S_q^{\widehat{bootstrap}}) = \frac{\sum_{b=1}^B (\widehat{S}_{q_b} - S_q^{\widehat{bootstrap}})^2}{(B - 1)} \quad (7)$$

O índice de Tsallis é de especial relevância, pois incorpora os principais índices de diversidade comumente utilizados. Substituindo-se o parâmetro  $q$  por valores reais não-negativos, é possível explorar uma série de índices de diversidade: quando  $q = 0 \Rightarrow S_q = S - 1$ , quando  $q \rightarrow 1$ , observa-se que  $S_q \rightarrow H'$ , o índice de Shannon (Equações 8 a 10); quando  $q = 2$ , temos que  $S_q = 1 - D$ , o simétrico do índice de Simpson (Equações 11 a 14). Assim, se  $q < 1$ , enfatiza-se a riqueza de espécies e, se  $q > 1$ , aborda-se a dominância dentro da comunidade (MENDES et al., 2008).

O índice de Shannon é assim definido (KREBS, 1999):

$$H' = - \sum_{i=1}^S p_i \log p_i \quad (8)$$

O EMV do índice de Shannon pode ser obtido por (PIELOU, 1975):

$$\widehat{H}' = - \sum_{i=1}^S \hat{p}_i \log \hat{p}_i \quad (9)$$

A variância do EMV do índice de Shannon pode ser aproximada por (ZAR, 1999):

$$V(\hat{H}') = \frac{\sum \hat{p}_i \log^2 \hat{p}_i - (\sum \hat{p}_i \log \hat{p}_i)^2 / n}{n^2} \quad (10)$$

O índice de Simpson é assim definido (KREBS, 1999):

$$D = \sum_{i=1}^s p_i^2 \quad (11)$$

Vale notar que o parâmetro  $D$  é um índice de dominância, ou “concentração”, como observou Pielou (1975); a diversidade é geralmente expressa por seu simétrico, ou complementar  $1 - D$  (KREBS, 1999).

O EMV do índice de Simpson pode ser obtido por (PIELOU, 1975):

$$\hat{D} = \sum_{i=1}^s \hat{p}_i^2 \quad (12)$$

Porém Pielou (1969 apud KREBS, 1999) demonstrou que para uma população finita, o estimador apropriado é:

$$1 - \hat{D} = \sum_{i=1}^s \frac{n_i(n_i - 1)}{n(n - 1)} \quad (13)$$

A variância do EMV do índice de Simpson é obtida ao se calcular (SIMPSON, 1949):

$$V(\hat{D}) = \frac{4n(n - 1)(n - 2) \sum_i^s \hat{p}_i^3 + 2n(n - 1) \sum_i^s \hat{p}_i^2 - 2n(n - 1)(2n - 3)(\sum_i^s \hat{p}_i^2)^2}{[n(n - 1)]^2} \quad (14)$$

Para as estimativas *jackknife* e *bootstrap* dos índices de diversidade, foram relatados a média e os intervalos a 95% de confiança (IC a 95%), seguindo aproximações pela distribuição  $t$ , para o *jackknife*, e pela distribuição  $Z$ , para o *bootstrap* (MILLER, 1974; ZAHL, 1977; EFRON, 1979; SMITH E BELLE, 1984).

Para as estimativas realizadas pelo EMV não foram calculados os intervalos de confiança, pois não é conhecida a variância generalizada da família de índices de Tsallis.

Foi também calculado a estimativa do viés percentual, que é definido como o valor real menos o valor estimado, divididos pelo valor real e multiplicado por 100.

A comparação entre os índices e a avaliação do desempenho do *jackknife* e *bootstrap* foi feita diretamente por meio do viés percentual e das medidas dos intervalos de confiança, permeando as variações impostas pelo diferentes tratamentos: distribuição espécie/abundância, nível de agregação e procedimento de amostragem.

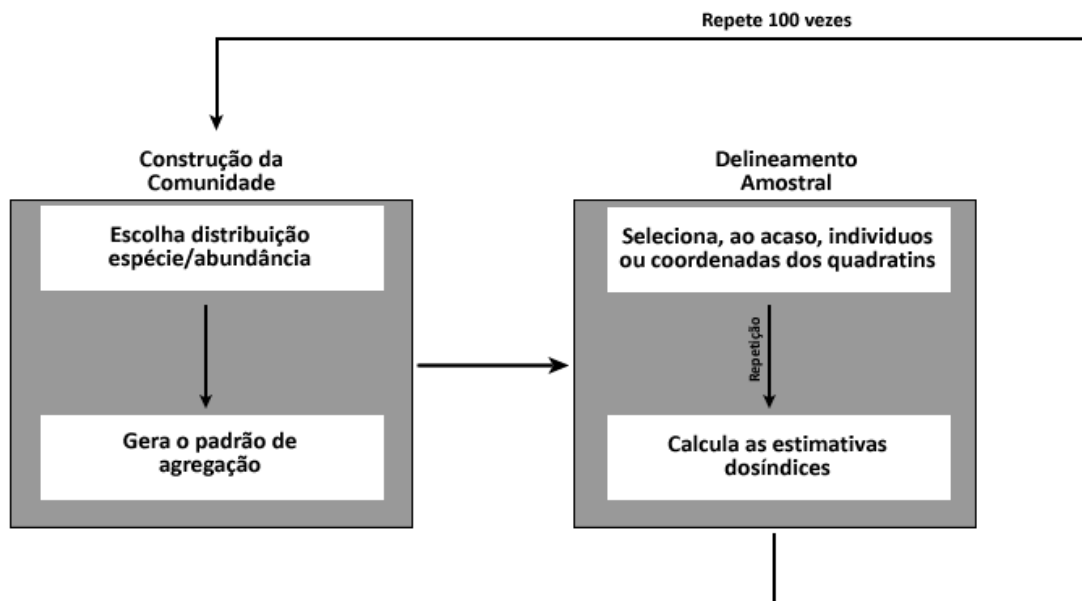


Figura 1: Esquema representando os procedimentos computacionais realizados para simular as comunidades e para a amostragem.

## 4. RESULTADOS

### 4.1 Cenários simulados

Os dados mais relevantes dos cenários produzidos para amostragem podem ser visualizados na Tabela 1. Pode-se observar como varia a diversidade das comunidades simuladas na Figura 2.

Tabela 1: Índices de diversidade reais (parâmetros populacionais) das comunidades simuladas, de acordo com as distribuições de espécie/abundância logsérie, lognormal e geométrica. Nas colunas estão destacados os índices principais: S0 referente à riqueza de espécies (S-1), S1 referente ao índice de Shannon (H') e S2 referente ao complementar do índice de Simpson (1-D).

	Índice*		
	S <sub>0</sub>	S <sub>1</sub>	S <sub>2</sub>
<b>Logsérie</b>	224 ± 0	4,28 ± 0,11	0,98 ± 0,0055
<b>Lognormal</b>	224 ± 0	3,94 ± 0,36	0,94 ± 0,0404
<b>Geométrica</b>	224 ± 0	4,99 ± 0,03	0,99 ± 0,0005

\* média ± desvio padrão de 400 réplicas.



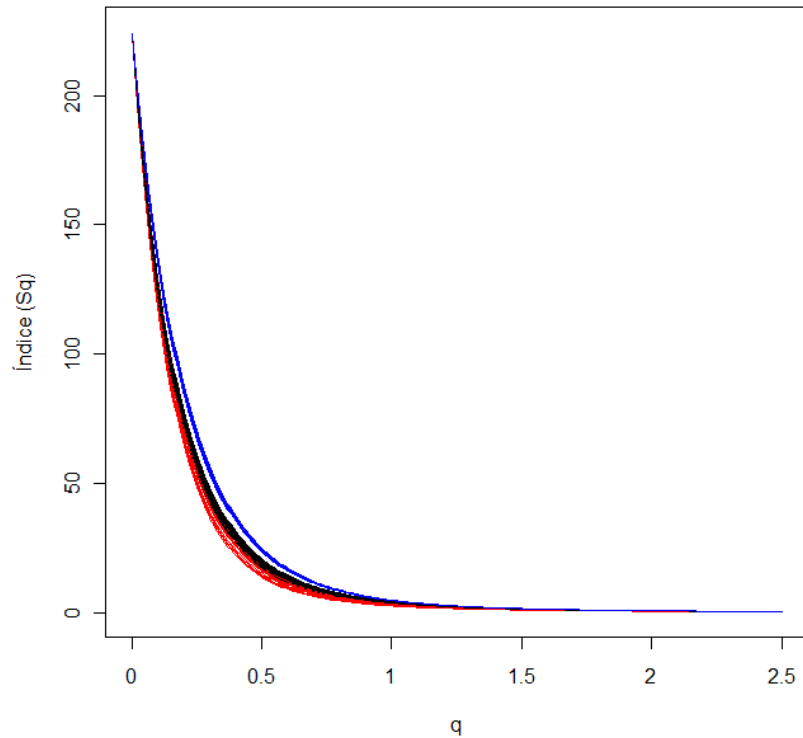


Figura 2: Família de índices de Tsallis. No eixo das coordenadas, os valores reais dos índices  $S_q$  e no eixo das abscissas, os valores dos parâmetros  $q$ . As distribuições lognormal, logsérie e geométrica estão representadas pelas linhas em vermelho, preto e azul, respectivamente.

#### 4.2 Os vieses nas estimativas da família de índices de Tsallis

A análise que se segue é referente ao índice  $S_q$ , para o qual o valor de  $q$  é arbitrariamente escolhido. Os resultados apresentados se referem ao viés encontrado quando as estimativas dos índices são realizadas pelos métodos da máxima verossimilhança (EMV), *jackknife* e *bootstrap*. Para cada um desses métodos, também são apresentados os resultados obtidos dos possíveis efeitos que a distribuição de espécie/abundância da comunidade, a agregação populacional e o método de amostragem podem exercer sobre os vieses nas estimativas.

#### 4.2.1 Métodos de estimação

Para a análise do EMV e dos métodos *jackknife* e *bootstrap*, foram tomadas as médias e as variâncias do viés percentual (Figura 3), sem considerar nenhuma das outras variáveis (método de amostragem, distribuição espécie-abundância ou nível de agregação populacional).

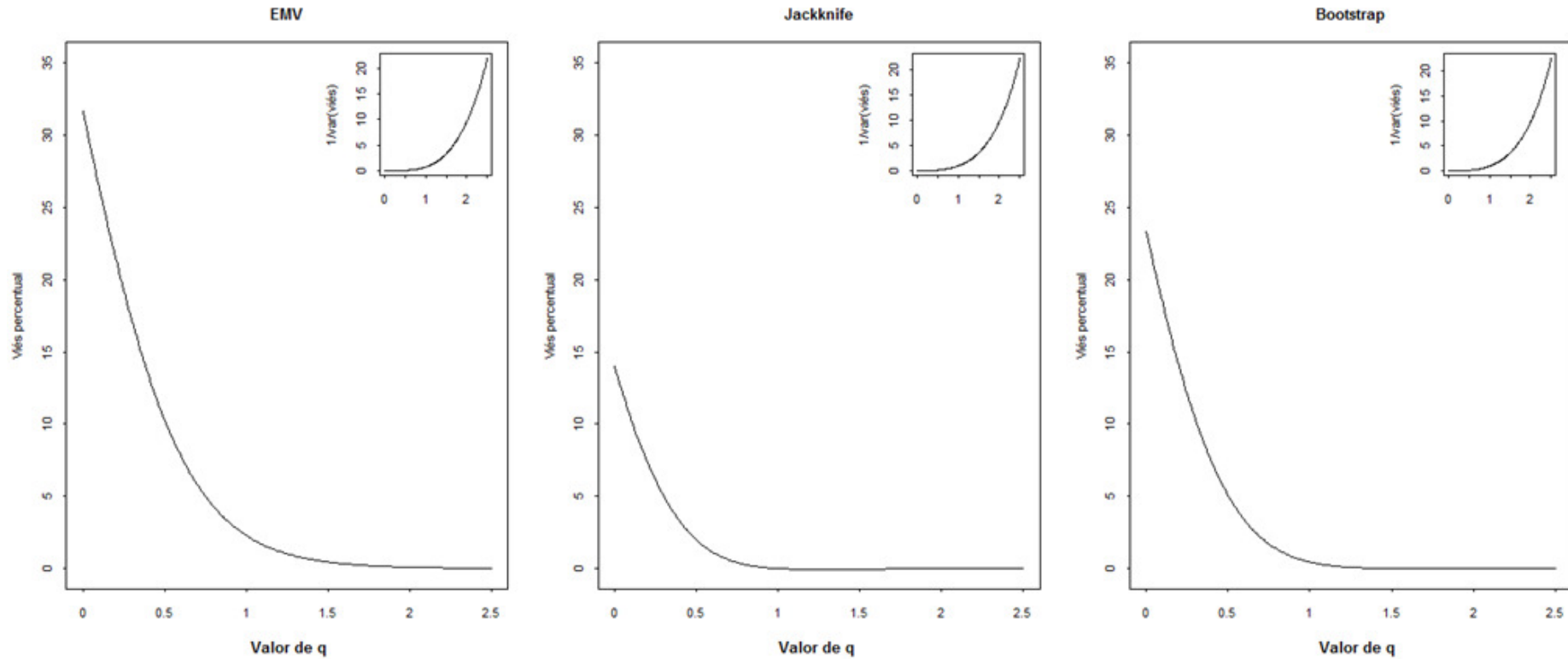


Figura 3: Vieses percentuais para os índices  $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem *jackknife* e *bootstrap*. Podem ser visualizadas as médias, calculados a partir de todas as amostras, sem distinção de distribuição espécie/abundância, método de amostragem ou raio de dispersão (agregação populacional). Valores negativos de viés percentual se referem à superestimação do índice real e valores positivos de viés à subestimação. Os gráficos menores, em destaque, se referem à relação entre  $1/var(viés)$ , que é uma medida de precisão associada ao viés.

#### 4.2.2 Distribuições de espécie/abundância

São apresentadas na Figura 4 as estimativas do viés percentual dos índices  $S_q$ , obtidas para as distribuições de espécie/abundância logsérie, lognormal e geométrica. Esses cálculos foram realizados sem levar em conta a forma de amostragem ou o nível de agregação populacional.

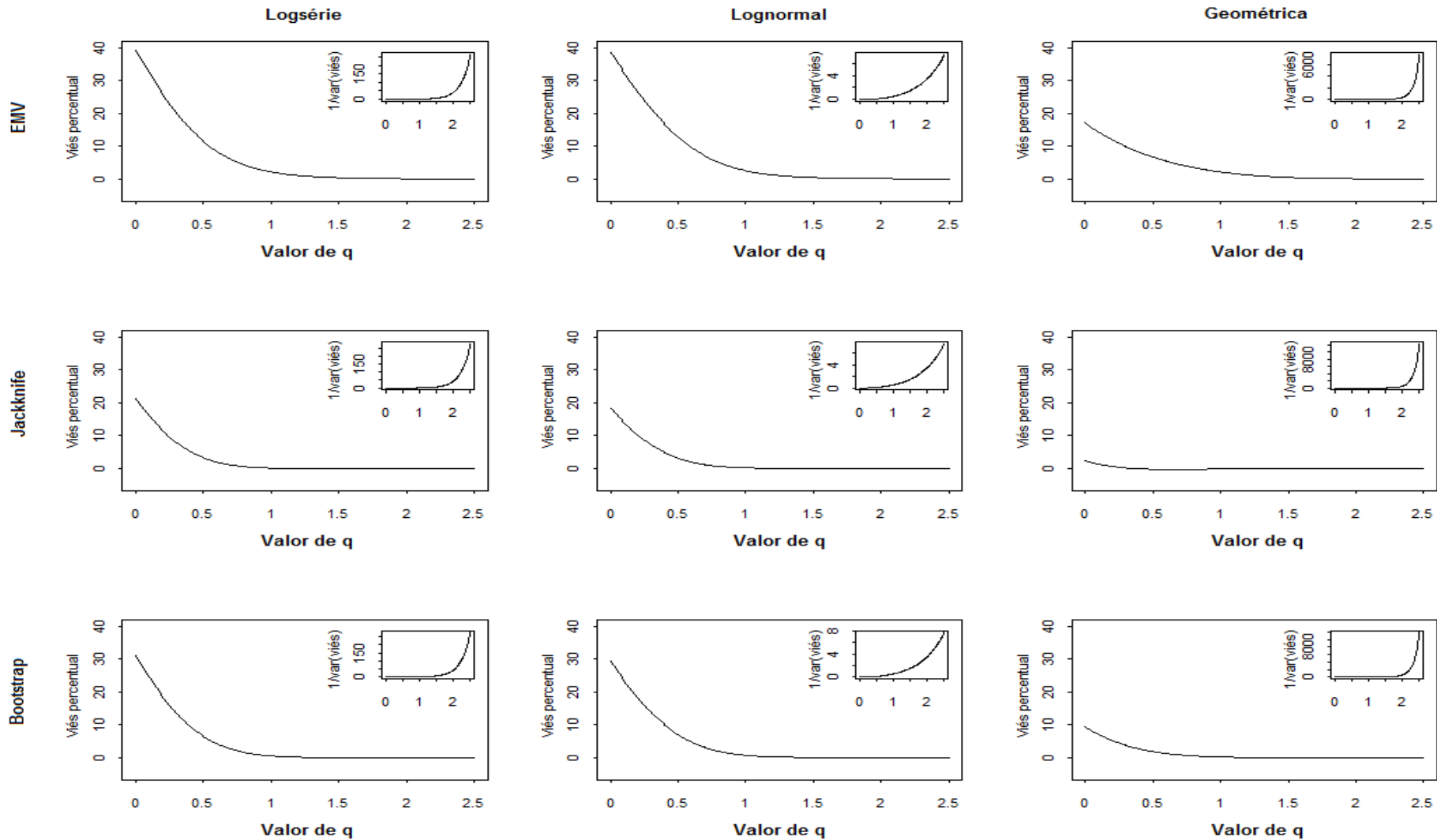


Figura 4: Vieses percentuais para os índices  $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem *jackknife* e *bootstrap*, considerando as distribuições espécie/abundância lognormal, logsérie ou geométrica. Podem ser visualizadas as médias, calculadas a partir de todas as amostras, sem distinção de método de amostragem ou raio de dispersão (nível de agregação populacional). Valores negativos de viés percentual se referem à superestimação do índice real e valores positivos de viés à subestimação. Os gráficos menores, em destaque, se referem à relação entre  $1/var(viés)$ , que é uma medida de precisão associada ao viés.

### **4.2.3 Agregação populacional**

O enfoque desta seção é a agregação populacional dentro das comunidades, assim aqui os dados não foram divididos segundo a distribuição espécie/abundância ou forma de amostragem.

São apresentadas as medidas do viés percentual para os quatro níveis de agregação são apresentados na Figura 5.

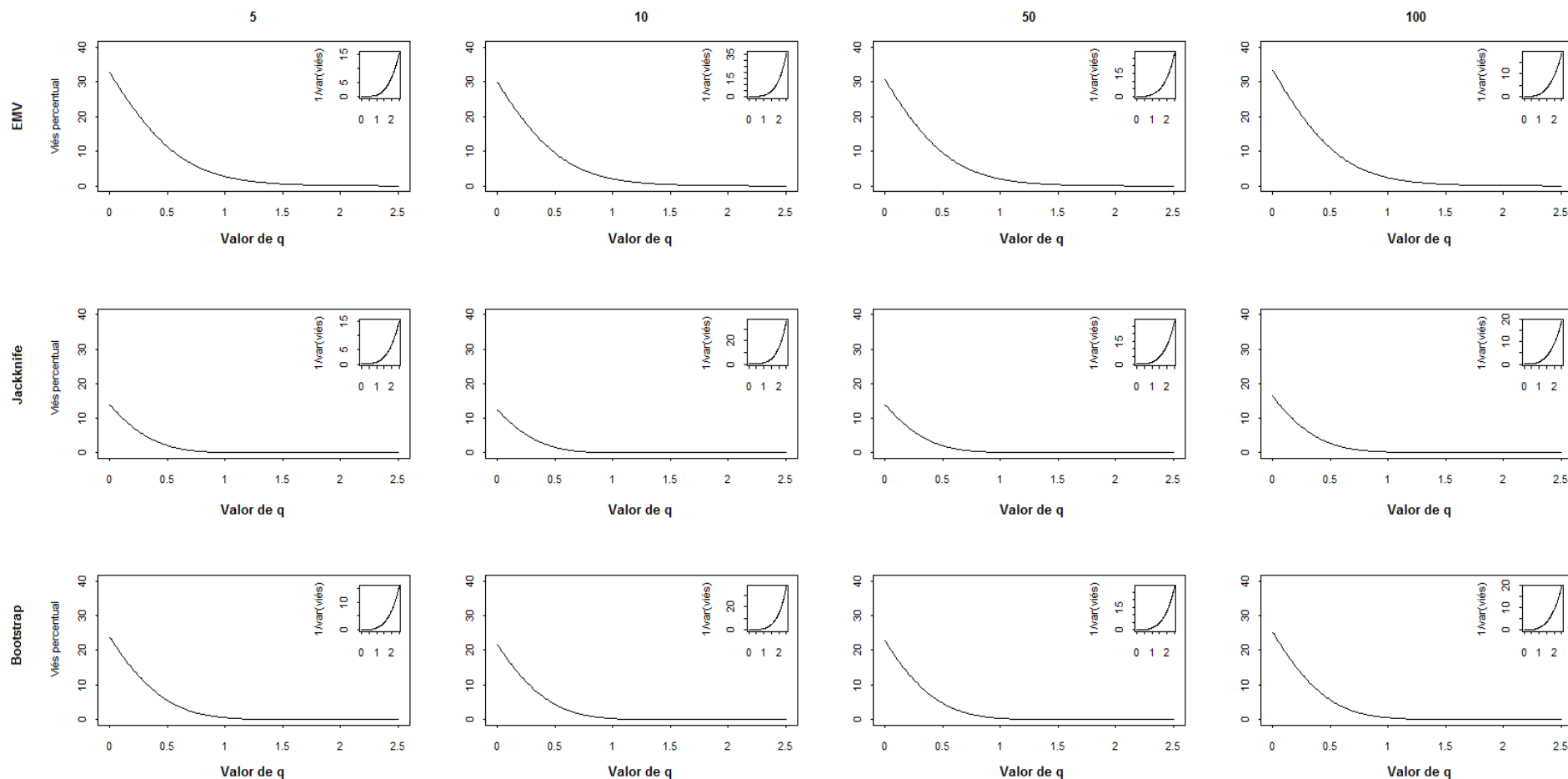


Figura 5: Vieses percentuais para os índices  $S_q$ , estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem *jackknife* e *bootstrap*, considerando o nível de agregação populacional (ou raio de dispersão). Podem ser visualizadas as médias, calculadas a partir de todas as amostras, sem distinção de método de amostragem ou distribuição de espécie/abundância. Populações mais agregadas são aquelas com menor raio de dispersão, assim, da esquerda para a direita, populações mais até menos agregadas. Valores negativos de viés percentual se referem à superestimação do índice real e valores positivos de viés à subestimação. Os gráficos menores, em destaque, se referem à relação entre  $1/\text{var}(\text{viés})$ , que é uma medida de precisão associada ao viés.

#### **4.2.4 Métodos de amostragem**

Nesta seção, são apresentados os resultados obtidos enfocando os métodos de amostragem. Aqui, as variáveis distribuição espécie/abundância e raio de agregação populacional não são consideradas.

Os dados com relação às estimativas dos vieses percentuais dos estimadores, para cada método de amostragem, estão expostos na Figura 6.



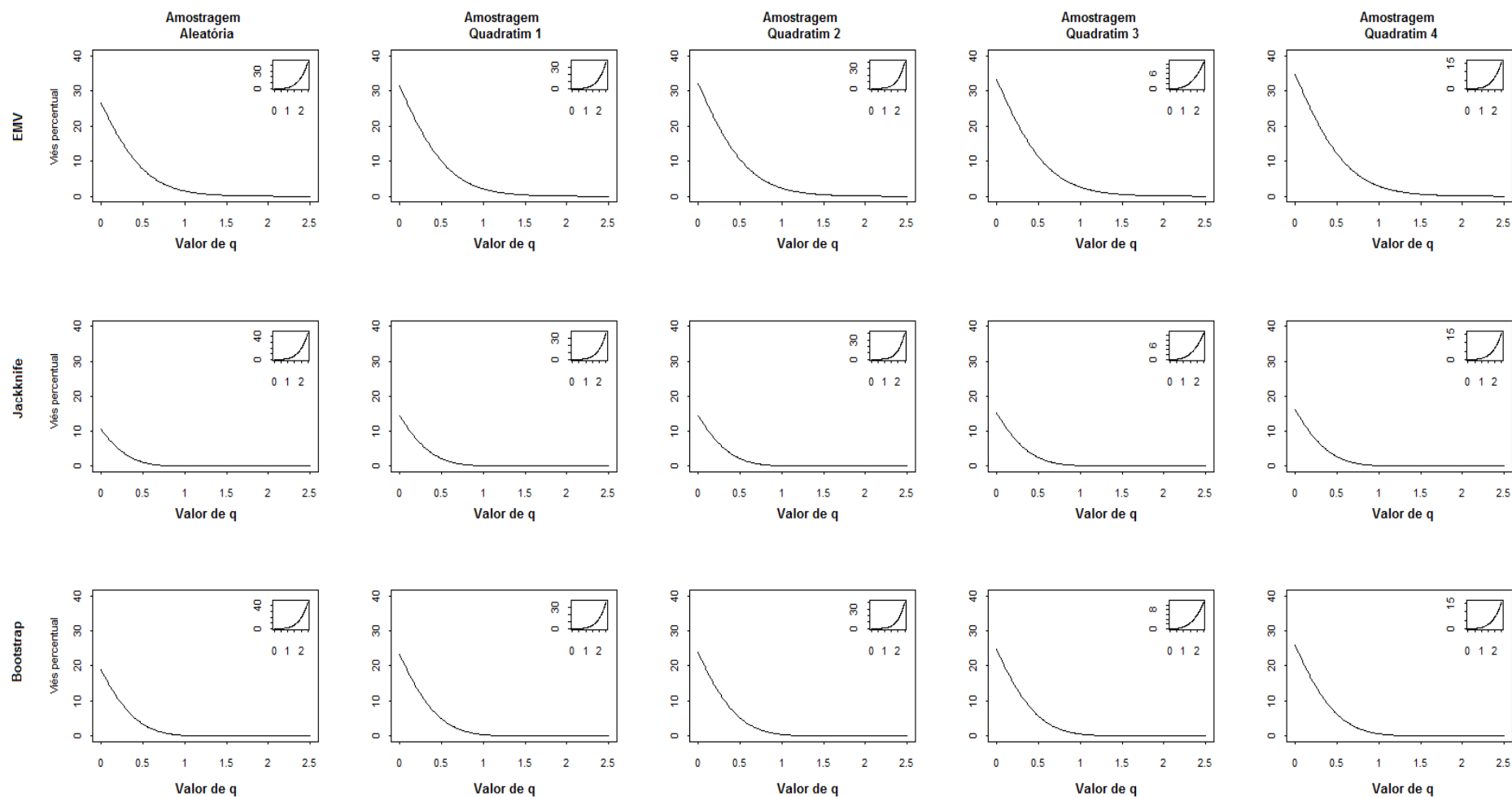


Figura 6: Vieses percentuais para os índices  $S_q$  estimados pelos métodos da máxima verossimilhança (EMV) e pelos métodos de reamostragem *jackknife* e *bootstrap*, considerando a forma de amostragem (aleatória ou em quadratins de tamanhos 1, 2, 3 ou 4). Podem ser visualizadas as médias, calculadas a partir de todas as amostras, sem distinção de distribuição espécie/abundância, ou raio de dispersão (agregação populacional). Valores negativos de viés percentual se referem à superestimação do índice real e valores positivos de viés à subestimação. Os gráficos menores, em destaque, se referem à relação entre  $1/\text{var}(\text{viés})$ , que é uma medida de precisão associada ao viés.

### 4.3 Precisão e acurácia nas estimativas *jackknife* e *bootstrap*

A seguir, são apresentados os resultados dos métodos de estimação e dos possíveis efeitos que a distribuição de espécie/abundância, nível de agregação e o método de amostragem podem exercer sobre a precisão e acurácia das estimativas realizadas pelos métodos *jackknife* e *bootstrap*.

A estrutura seguida é a mesma daquela realizada na seção 4.2.

#### 4.3.1 Métodos de estimação

A amplitude do intervalo de confiança (IC a 95%) e a frequência relativa em que o IC cobre os valores reais podem ser visualizadas na Figura 7. Note que a apresentação desses dados é realizada sem levar em conta a distribuição de espécie/abundância, o nível de agregação das populações ou o método de amostragem.

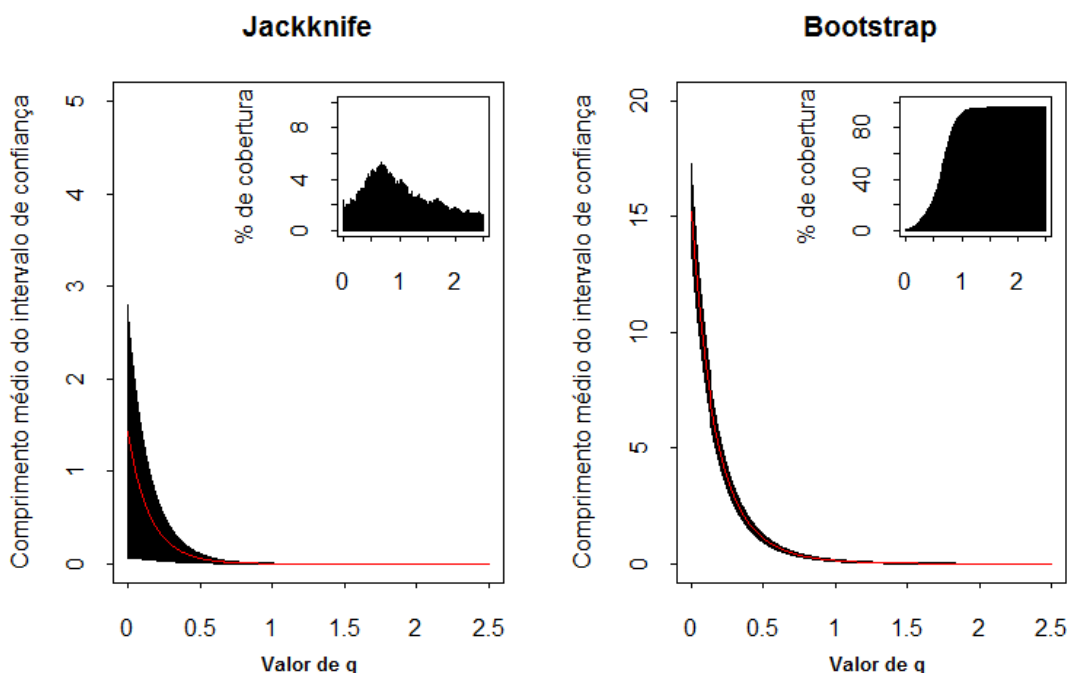


Figura 7: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, pelos valores dos parâmetros  $q$  da fórmula de Tsallis. A linha em vermelho representa o comprimento médio e, em preto,  $\pm$  desvio-padrão. Os gráficos menores, na parte superior direita dos gráficos maiores, trazem, no eixo das coordenadas, a porcentagem de casos em que o IC a 95% cobriu o verdadeiro valor do parâmetro e, no eixo das abscissas, os valores de  $q$ . Note que não foram feitas distinções entre distribuição espécie/abundância, nível de agregação populacional ou método de amostragem.

#### 4.3.2 Distribuições espécie/abundância

Os dados que foram utilizados para explorar os efeitos das distribuições de espécie abundância sobre amplitude do IC a 95% e a frequência em que o IC a 95% cobriu o verdadeiro valor do parâmetro estão apresentados na Figura 8 e não fazem distinção entre os níveis de agregação populacional e nem entre os métodos de amostragem.

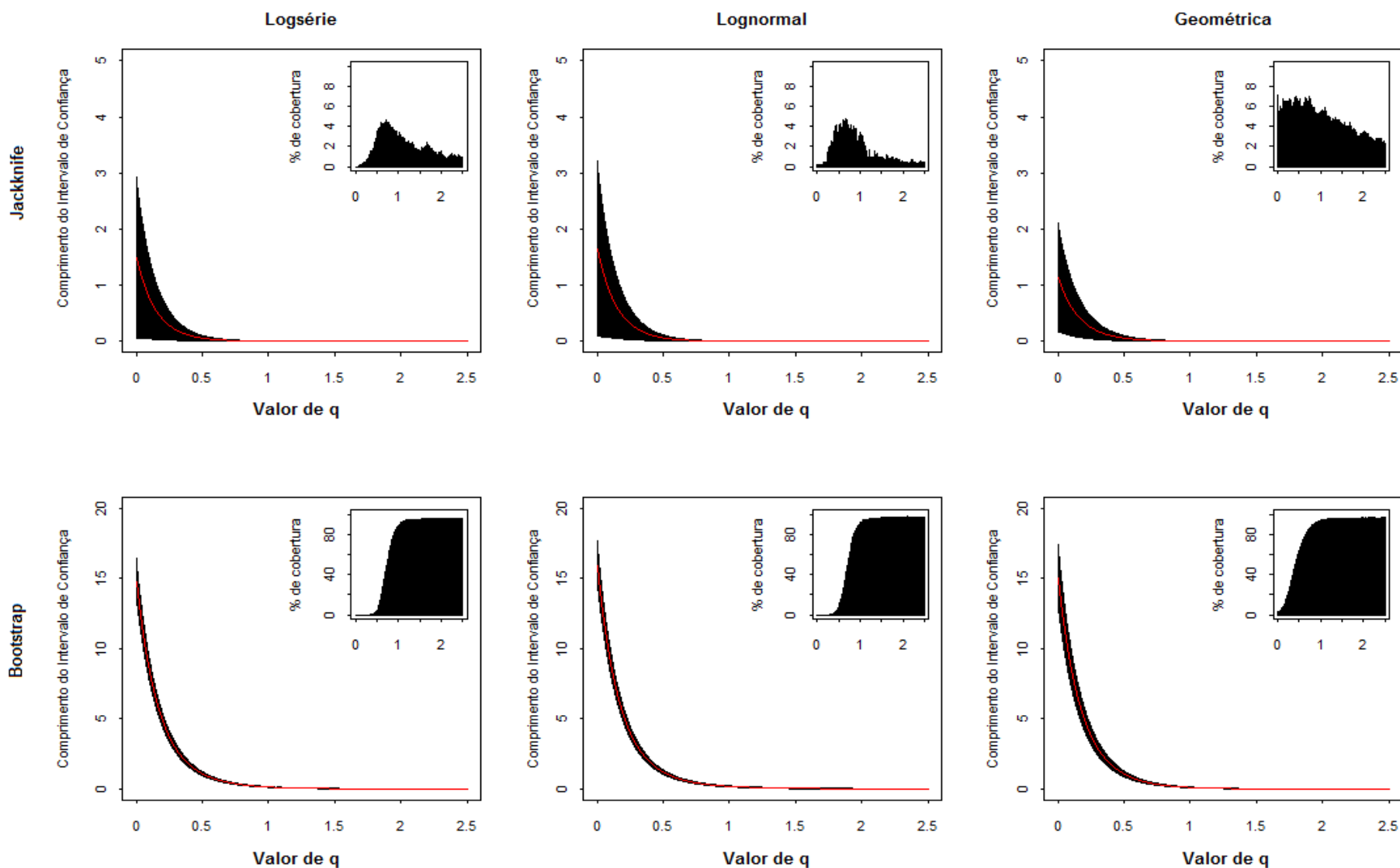


Figura 8: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, pelos valores dos parâmetros  $q$  da fórmula de Tsallis. A linha em vermelho representa o comprimento médio e, em preto,  $\pm$  desvio-padrão. Os gráficos menores, na parte superior direita dos gráficos maiores, trazem, no eixo das coordenadas, a porcentagem de casos em que o IC a 95% cobriu o verdadeiro valor do parâmetro e, no eixo das abscissas, os valores de  $q$ . Note que não foram feitas distinções entre nível de agregação populacional ou método de amostragem.

### 4.3.3 Agregação populacional

Nesta seção será considerado o efeito do raio de dispersão da população, uma variável inversamente proporcional à agregação populacional, nas estimações dos índices da família de Tsallis de acordo com cada método de estimação. Para tal, as variáveis distribuição de espécie/abundância e método de amostragem não foram consideradas.

Os dados referentes ao IC a 95% para o *jackknife* e para o *bootstrap* podem ser visualizados na Figura 9.

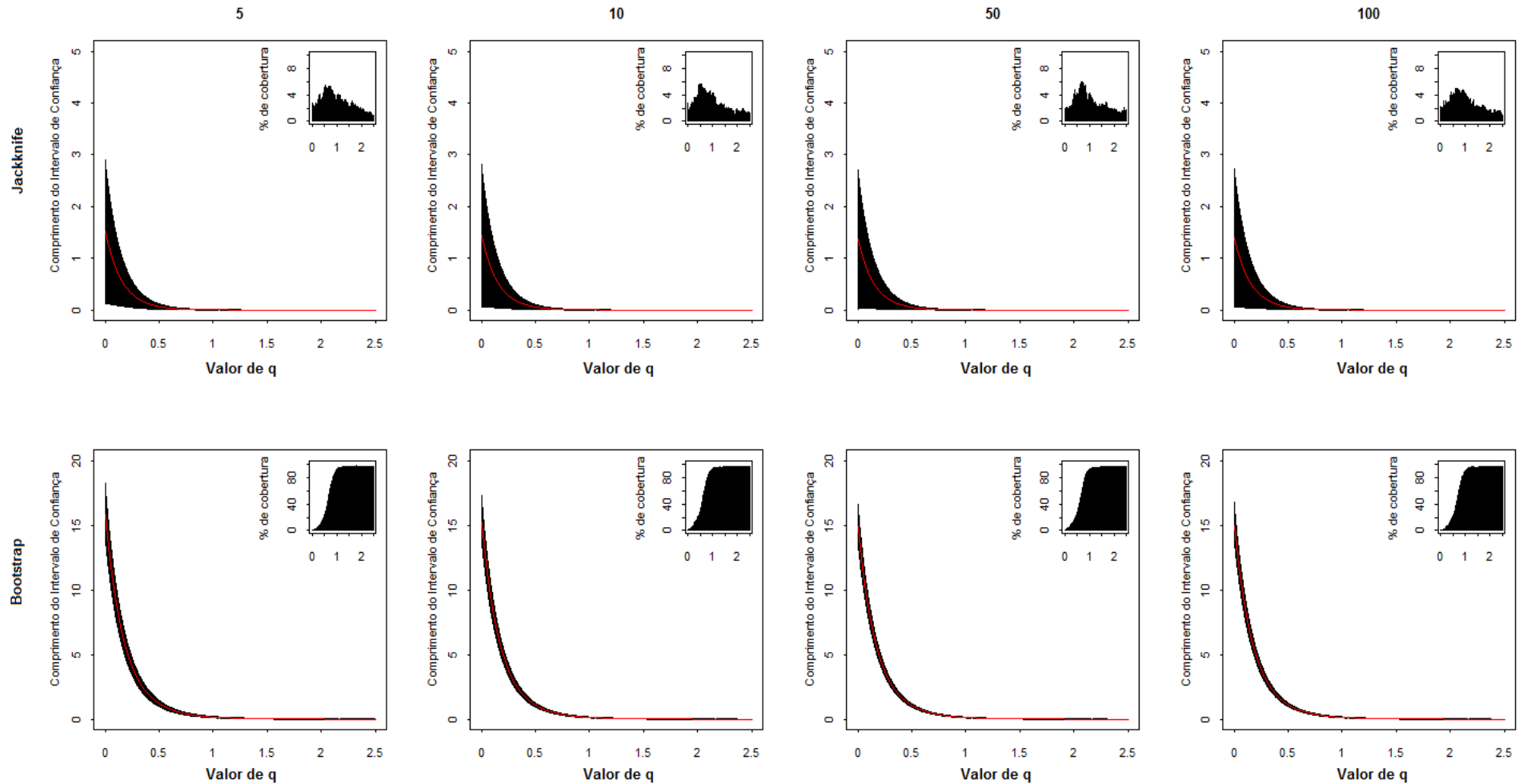


Figura 9: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, pelos valores dos parâmetros  $q$  da fórmula de Tsallis. A linha em vermelho representa o comprimento médio e, em preto,  $\pm$  desvio-padrão. Populações mais agregadas são aquelas com menor raio de dispersão, assim, da esquerda para a direita, populações mais até menos agregadas. Os gráficos menores, na parte superior direita dos gráficos maiores, trazem, no eixo das coordenadas, a porcentagem de casos em que o IC a 95% cobriu o verdadeiro valor do parâmetro  $e$ , no eixo das abscissas, os valores de  $q$ . Note que não foram feitas distinções entre distribuição de espécie/abundância ou método de amostragem.

#### **4.3.4 Métodos de amostragem**

Sem se ater ao nível de agregação ou à distribuição de espécie/abundância, são apresentados os dados para os métodos de amostragem, de acordo com cada método de estimação. A Figura 10 resume as informações sobre os IC a 95%.

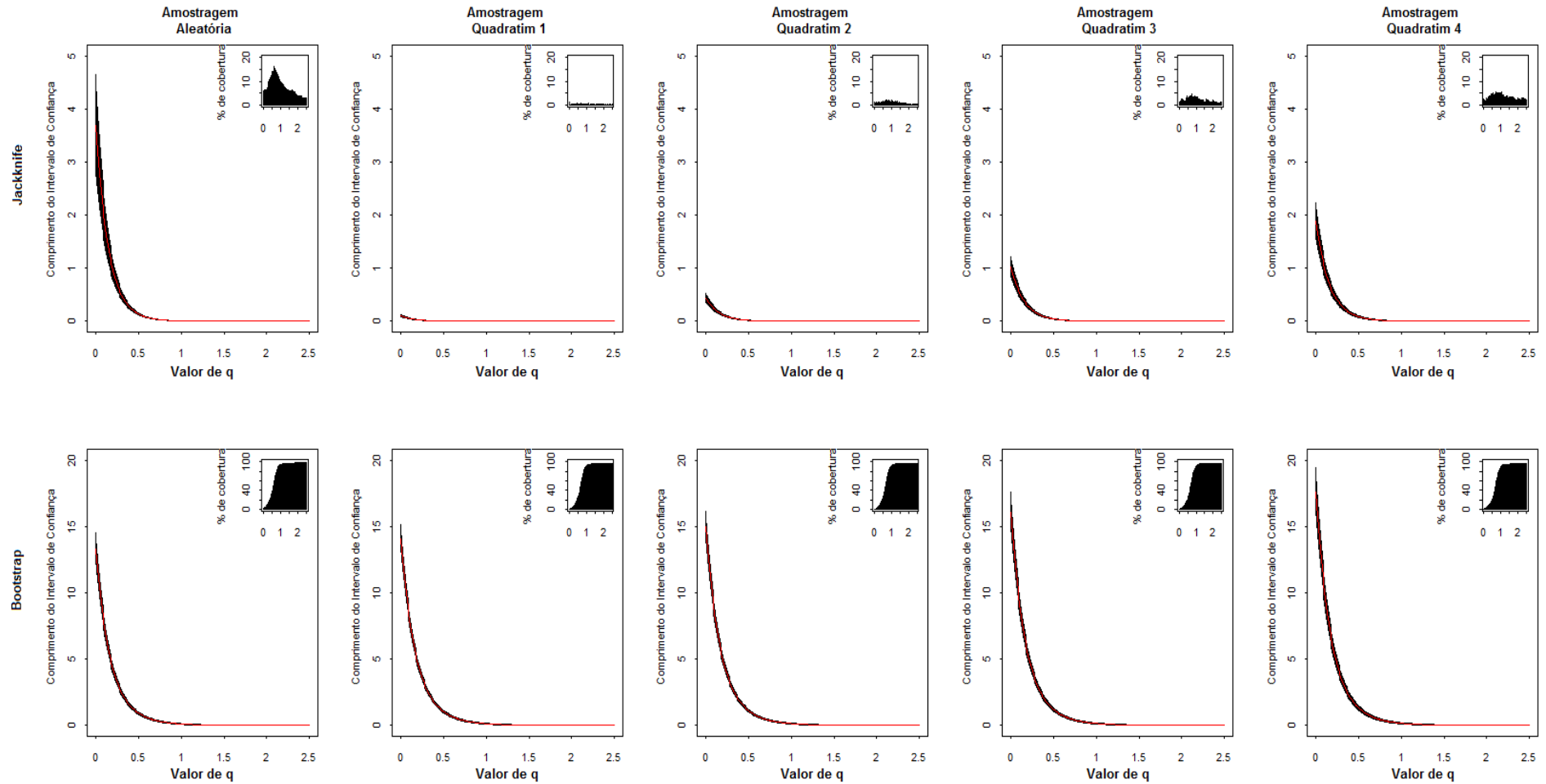


Figura 10: Amplitudes dos intervalos de confiança (IC a 95%) produzidos pelos métodos *jackknife* e *bootstrap*, pelos valores dos parâmetros  $q$  da fórmula de Tsallis. A linha em vermelho representa o comprimento médio e, em preto,  $\pm$  desvio-padrão. Os gráficos menores, na parte superior direita dos gráficos maiores, trazem, no eixo das coordenadas, a porcentagem de casos em que o IC a 95% cobriu o verdadeiro valor do parâmetro  $e$ , no eixo das abscissas, os valores de  $q$ . Note que não foram feitas distinções entre distribuição de espécie/abundância ou nível de agregação populacional.



#### 4.4 Amostragens em quadratins: qual a melhor unidade de reamostragem a ser considerada?

Usualmente, amostragens em comunidades reais são realizadas em quadratins, que constituem, portanto, a unidade amostral natural em um procedimento de análise estatística. Entretanto, como os verdadeiros objetos de interesse da amostragem em comunidades biológicas são os indivíduos das diferentes espécies, foram comparados os desempenhos dos métodos *jackknife* e *bootstrap*, ora considerando quadratins ora considerando os indivíduos como unidades de reamostragem para todo o conjunto de dados produzido nesta dissertação.

Estão representados na Figura 11 os resultados do comprimento do IC a 95%, a porcentagem de cobertura do IC a 95% e o viés percentual obtido para os métodos *jackknife* e *bootstrap* ao estimar a riqueza de espécies na comunidade. Em busca de generalidade, não foram feitas distinções entre nível de agregação populacional ou distribuição de espécie/abundância.

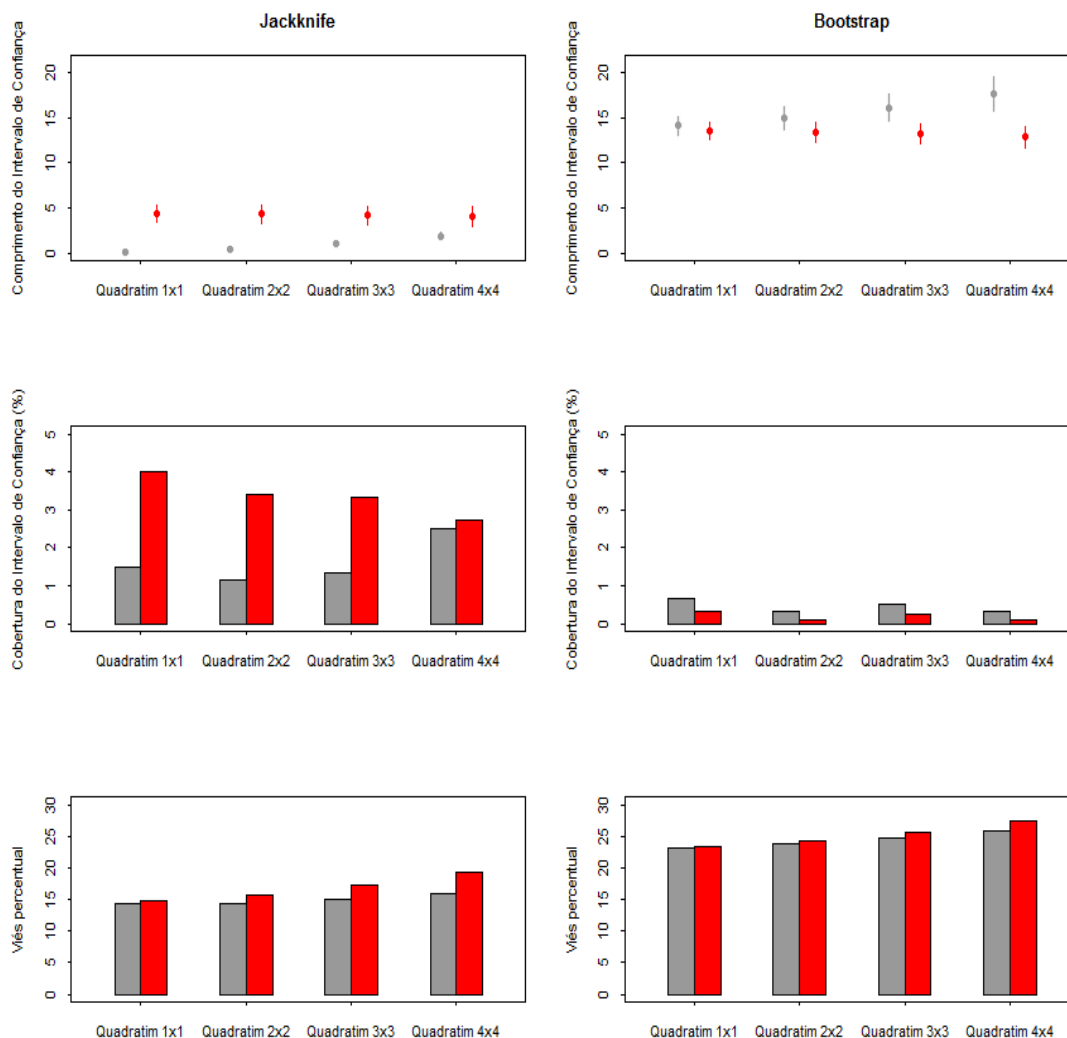


Figura 11: Avaliação da melhor unidade de amostragem a ser considerada para estimar a riqueza de espécies pelos métodos *jackknife* e *bootstrap*, em comunidades amostradas por quadratins. Todos os elementos gráficos em cinza representam que a unidade de amostragem foi o quadratim. Já os elementos em vermelho, representam que a unidade de amostragem foi o indivíduo.

## 5. DISCUSSÃO

Com relação aos cenários simulados, os mesmo padrões encontrados na Figura 2 foram descritos por Mendes et al. (2008). Uma vez que o número de espécies (Tabela 1) e também o número de indivíduos foram fixados, a diferença na diversidade entre as réplicas e entre as distribuições de espécie abundância se deu provavelmente à diferença na equidade, o que é esperado em simulações estocásticas.

Já iniciando com a abordagem estatística das estimativas, ao longo dos crescentes valores de  $q$ , o viés percentual de todos os métodos de estimação é decrescente, seguindo uma curva semelhante à Figura 2, que remete à própria diversidade. Este padrão decrescente de viés corrobora as afirmações de Routledge (1980), que observa vantagens no uso do índice  $S_2$ , alegando que nenhum estimador de  $S_1$  entre os diversos já propostos (GOOD, 1953; MILLER, 1974; PIELOU, 1975; ZAHL, 1977) é menos viesado do que o simples EMV de  $S_2$ .

Além disso, os dados desta dissertação corroboram os resultados observados por Rényi (1961), que ao avaliar a família de índices de diversidade proposta por Kempton (1979), também verificou a presença de vieses decrescentes à medida que esses índices passavam a dar mais peso às espécies dominantes.

Entre os estimadores, como esperado, o EMV foi o que apresentou maior viés percentual em comparação aos métodos *jackknife* e *bootstrap*, chegando a subestimar a riqueza de espécies em mais de 30% (Figura 3). As estimativas corrigidas para viés dos métodos *jackknife* e *bootstrap* reduziram o viés na riqueza de espécies para aproximadamente 15% e 25%, respectivamente. Vale notar que estes resultados representam o viés esperado em estimativas de diversidade em comunidades das quais não se conhece informação alguma sobre a distribuição de espécie abundância ou nível de

agregação das populações, e nas quais não tenha sido realizado qualquer planejamento amostral.

Assim, quando a informação da distribuição de espécie/abundância (Figura 4) se torna disponível, os vieses esperados para os diferentes estimadores são alterados, demonstrando o forte efeito que a estrutura da comunidade tem sobre as estimativas de diversidade, apesar de não haver alteração na comparação de desempenho dos estimadores, isto é, o método de EMV é o que apresenta maior viés, seguido por *bootstrap* e depois o *jackknife*.

Para o EMV, em comunidades com distribuições, tanto logsérie quanto lognormal, o viés percentual esperado na estimativa da riqueza de espécies fica em torno de 40%; já em comunidades com distribuição geométrica, o viés percentual esperado na estimativa de riqueza pelo EMV cai para menos de 20%. Para o *jackknife*, em comunidades com distribuição logsérie, o viés percentual esperado é de pouco mais de 20% e, em comunidades com distribuição lognormal, pouco menos de 20%; já em comunidades com distribuição geométrica, o viés percentual esperado na estimativa de riqueza pelo *jackknife* cai para menos de 5%. Para o *bootstrap*, em comunidades com distribuição logsérie, o viés percentual esperado é de pouco mais de 30% e, em comunidades com distribuição lognormal, pouco menos de 30%; já em comunidades com distribuição geométrica, o viés percentual esperado na estimativa de riqueza pelo *bootstrap* cai para menos de 10%.

Como pode ser observado, é esperado menor viés ao se estimar a diversidade em comunidades com distribuição geométrica do que em comunidades com distribuições logsérie e lognormal, que apresentaram estimativas com viés semelhante. Isto provavelmente se deve à presença de um menor número de espécies raras em comunidades com distribuição geométrica, de forma que existe maior probabilidade de incluir todas as espécies em uma amostragem.

Com relação aos níveis de agregação populacional (Figura 5), foi observado um efeito fraco sobre as estimativas de viés, porém com padrão bastante marcante: apesar das diferenças entre os vieses esperados nos crescentes níveis de agregação populacional não serem grandes, foi observado que comunidades que apresentam populações com raios intermediários de

dispersão (10 e 50) é esperado menor viés nas estimativas de diversidade do que em comunidades que apresentam raios extremos de dispersão (5 e 100), isto é, independentemente do método de estimação utilizado, comunidades nas quais as populações se apresentem extremamente agregadas ou extremamente dispersas é esperado que as estimativas de diversidade apresentem maior viés. Comumente, sem qualquer suporte teórico, assumimos que, primeiro, fortes agregados populacionais produzem maior viés nas estimativas de diversidade, principalmente quando são realizadas amostragens em quadratins; e, segundo, que quando as populações se apresentam muito dispersas, espera-se redução do viés. Entretanto, populações espacialmente dispersas não corresponde a uma distribuição aleatória de indivíduos no espaço, o que, segundo Taylor et al. (1978) é uma disposição bastante rara e, segundo Pielou (1975), isto é um importante pressuposto para a estimação dos índices de diversidade. Assim, proceder como sugerido por He e Hu (2005), que argumentam que é possível realizar a estimativa de parâmetros das distribuições de espécie/abundância utilizando a estimativa de máxima verossimilhança de  $S_2$ , pode comprometer essa estimação, se negligenciados os pressupostos do EMV.

Ao abordar o efeito do método de amostragem (Figura 6), notou-se, para todos os métodos de estimação, que a amostragem aleatória provoca menor viés nas estimativas do que amostragens por quadratins. Para o EMV, notou-se também que há aumento no viés percentual esperado na medida em que o tamanho do quadratim aumenta. Entretanto, os métodos de reamostragem, *jackknife* e *bootstrap*, se mostraram robustos com relação ao tamanho do quadratim.

A comparação da precisão e da acurácia dos métodos *jackknife* e *bootstrap* (Figura 7) via o comprimento do intervalo de confiança (IC a 95%) e a cobertura percentual do intervalo, demonstrou que, independentemente da distribuição de espécie/abundância, método de amostragem ou nível de agregação populacional, o método *jackknife* é consistentemente mais preciso (IC a 95% mais curtos) do que o método *bootstrap*, sobretudo para valores pequenos de  $q$ . Além disso, os IC a 95% produzidos pelo método *jackknife* cobriram percentualmente o valor real da riqueza de espécies ( $q = 0$ ) em mais casos do que os IC a 95% produzidos pelo *bootstrap*. Entretanto, a riqueza de

espécies constitui uma exceção quanto à acurácia do método *jackknife*: para  $q > 0$ , o método *bootstrap* mostrou-se muito mais acurado, chegando a cobrir os valores dos parâmetros populacionais em mais de 80% dos casos.

Também é interessante notar o formato que tomaram os gráficos da cobertura percentual do IC a 95%. Para o método *jackknife*, existem picos de cobertura para  $q = 0$  e para  $q \in ]0,5 ; 1[$ . Para o método de bootstrap, o formato se assemelha ao de uma curva logística. Estes formatos característicos podem fornecer ao pesquisador critérios de decisão ao optar por um ou outro método de estimação, dependendo do índice de diversidade no qual se esteja interessado em estimar. Os autores Heltshe e Forrester (1985) argumentam, entretanto, que coberturas do IC a 95% muito grandes podem ser indicativos de desvios-padrões muito grandes, o que pode diminuir o poder de detecção da diferença entre áreas que estejam sob monitoramento.

Ao se observar o efeito das distribuições de espécie/abundância sobre o comprimento e a cobertura do IC a 95% (Figura 8), pode-se notar que, pelo o método *jackknife*, nas distribuições lognormal e geométrica foram gerados os maiores e menores IC a 95%, respectivamente. Similarmente, na distribuição geométrica é que os IC a 95% gerados pelo *jackknife* cobriram em mais casos o valor real do parâmetro, sobretudo para os menores valores de  $q$ . O método de bootstrap se mostrou robusto, não tendo sido observado forte efeito das distribuições de espécie/abundância avaliadas, tanto sobre o comprimento quanto sobre a cobertura do IC a 95%, além de um ligeiro aumento na cobertura do IC a 95% gerado pelo *bootstrap* na riqueza de espécies, quando as comunidades estavam geometricamente distribuídas.

Com relação ao efeito do nível de agregação populacional (Figura 9), ambos os métodos de estimação se mostraram relativamente robustos, tanto em precisão quanto em acurácia. Apesar disso, nota-se um efeito fraco de raios de dispersão curtos (5 e 10) sobre a cobertura do IC a 95% gerado pelo método *jackknife*, com picos de cobertura do parâmetro real ligeiramente maiores (para  $q = 0$  e para  $q \in ]0,5 ; 1[$ ) do que raios de dispersão mais longos (50 e 100).

As formas de amostragem (Figura 10) produziram um forte efeito sobre a precisão de ambos os métodos, *jackknife* e *bootstrap*. Para o *jackknife*, os maiores IC a 95% observados ocorreram em amostragens aleatórias.

Diferentemente, para o bootstrap, os menores IC a 95% observados ocorreram nesse método de amostragem. Nas amostragens em quadratins, os IC a 95% gerados, tanto pelo *jackknife* quanto pelo *bootstrap*, aumentam conforme aumenta o tamanho do quadratim.

Com relação à acurácia, o *bootstrap* se mostrou robusto, não tendo sido notada diferença na porcentagem de cobertura dos IC a 95% gerados por este método, independente da forma de amostragem aplicada. Já a cobertura dos IC a 95% gerados pelo *jackknife*, existe uma marcante diferença entre as porcentagens de cobertura do parâmetro real quando são aplicadas amostragens aleatórias ou em quadratins: para qualquer valor de  $q$ , os IC a 95% gerados pelo *jackknife* cobrem em mais casos o parâmetro populacional quando são realizadas amostragens aleatórias; e, à medida que se aumenta o tamanho do quadratim, também aumenta o número de casos em que o IC a 95% gerado pelo método *jackknife* cobre o valor real do parâmetro.

A última análise realizada nesta dissertação tratou da unidade amostral que deveria (ou poderia) ser considerada ao realizar os procedimentos *jackknife* e *bootstrap* para melhorar as estimativas da riqueza de espécies em comunidades biológicas, dado que a forma de amostragem adotada foi por meio de quadratins (Figura 11).

A relação do viés percentual se mantém para ambos os métodos de estimação independentemente da unidade amostral considerada, isto é, quanto maior o quadratim, mais enviesada é a estimativa da riqueza de espécies. Ainda assim, tanto para o *jackknife* quanto para o *bootstrap*, a estimativa de  $S_q$  quando  $q = 0$  se torna mais enviesada se a unidade amostral considerada para os procedimentos de reamostragem são indivíduos.

Pode-se notar que, com relação aos comprimentos dos IC a 95%, o método *jackknife* é robusto ao tamanho do quadratim quando a unidade amostral considerada são indivíduos, mas que este método gera IC a 95% maiores à medida que o tamanho do quadratim aumenta, quando as unidades amostrais consideradas para o procedimento são os quadratins, como já foi observado acima. Apesar da robustez, é importante salientar que a alternativa do uso de indivíduos como as unidades amostrais para o procedimento

*jackknife* resultou sempre em IC a 95% maiores do que quando foram considerados os quadratins, tornando o método *jackknife* menos preciso.

Também já foi explicitado que os IC a 95% produzidos pelo método *bootstrap* aumentam conforme aumenta o tamanho do quadratim, quando as unidades amostrais são os próprios quadratins, entretanto, é interessante observar que, quando as unidades amostrais consideradas para o procedimento de bootstrap são os indivíduos, o comprimento do IC a 95% produzido diminui ao se aumentar a área do quadratim e, além disso, os IC a 95% gerados considerando indivíduos as unidades amostrais para o procedimento de reamostragem gera IC a 95% menores, o que pode indicar uma melhoria na precisão no método *bootstrap* quando esta alternativa é aplicada.

Por fim, com relação à porcentagem de cobertura do IC a 95%, pode-se observar que, apesar do método *jackknife* ter sua precisão diminuída ao se utilizar indivíduos como unidades de reamostragem, ao se utilizar esta alternativa, a acurácia do método de estimação é aumentada, visto as maiores porcentagens de cobertura, sobretudo para tamanhos pequenos de quadratins. Como esperado, foi observado o oposto para o método de bootstrap: uma vez que a utilização de indivíduos como unidades de reamostragem aumentou a precisão do método, agravou sua acurácia, que, de qualquer forma, já não apresentou alta porcentagem de cobertura do parâmetro real. Ainda para o *bootstrap*, vale notar que os dados nesta dissertação não forneceram uma relação clara entre tamanho do quadratim e porcentagem de cobertura do verdadeiro número de espécies das comunidades, independentemente da unidade de reamostragem, o que sugere que deve existir independência entre tamanho do quadratim e a acurácia do método *bootstrap*.



## 6. CONCLUSÕES

Nesta dissertação foi primeiramente proposta a hipótese de que as distribuições de espécie/abundância logsérie, lognormal e geométrica deveriam, nesta ordem, resultar em vieses decrescentes nas estimativas dos índices de diversidade. Os dados corroboram em parte esta hipótese, pois, apesar da distribuição geométrica de fato implicar em menores vieses ainda não ficou esclarecido como as distribuições logsérie e lognormal podem exercer efeitos semelhantes sobre o viés nas estimativas.

Com relação à segunda hipótese proposta, pode-se concluir que enquanto que os dados obtidos nesta dissertação corroboram com o pressuposto comumente assumido de que maiores vieses são gerados em comunidades cujas populações estejam mais agregadas, o mesmo não pode ser afirmado para o pressuposto de que comunidades nas quais as populações estejam amplamente dispersas no espaço gerem menor viés nas estimativas de diversidade, estabelecendo uma nova hipótese de que raios intermediários de dispersão populacional é que são responsáveis por menores vieses nas estimativas.

Para a terceira hipótese, conclui-se que amostragens aleatórias de indivíduos, como esperado pelos pressupostos dos índices de diversidade, reduzem o viés nas estimativas. Com relação aos tamanhos dos quadratins, a hipótese de que quadratins maiores produzem maior viés é corroborada, quando as estimativas são realizadas pelo EMV; entretanto para os métodos intensivos de estimação testados, o tamanho do quadratim não exerce forte influência sobre o viés nas estimativas.

Ao abordar a precisão e acurácia dos métodos *jackknife* e *bootstrap* para estimar a família de índices de Tsallis, foi concluído que, para índices que refletem mais riqueza de espécies ( $0 \leq q < 1$ ), o método *jackknife* é mais preciso e mais acurado que o método *bootstrap*, entretanto para índices que

refletem a dominância na comunidade ( $1 \leq q \leq 2.5$ ), o *bootstrap* é o método mais indicado. Além disso, apesar dessas considerações a respeito do *jackknife*, faz-se necessária ainda uma comparação do desempenho do procedimento desenvolvido nesta dissertação com aquele proposto por Heltshe e Forrester (1983), ao se estimar a riqueza de espécies diretamente pelo índice  $S_0$ .

Ainda com relação à estimativa de  $S_0$ , a alternativa do uso de indivíduos como unidades de reamostragem nos procedimentos computacionalmente intensivos, mesmo quando as comunidades foram amostradas por quadratins, pode ser utilizada, pois, na maioria dos casos, houve melhora no desempenho dos estimadores. Tendo em mente, entretanto, que a amostragem em quadratins pode não produzir amostragens aleatórias de indivíduos, recomenda-se cautela no uso desta alternativa, ficando a critério do pesquisador optar pelo melhor desempenho ou pelo rigor estatístico.

Por fim, vale ressaltar que os resultados e conclusões desta dissertação podem ser verificados em um estudo ainda mais generalizado, variando-se o tamanho das comunidades e o esforço amostral, abordando também a estimação de possíveis efeitos de interação, além de outros modelos de distribuição de espécie/abundância.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

BASHARIN, G.P. (1959) On a statistical estimate for the entropy of a sequence of independent random variables. *Theory Probability and its Applications*, 4, 333-36.

BOWMAN, K.O.; HUTCHESON, K.; ODUM, E.P.; SHENTON, L.R. (1971) Comments on the distribution of indices of diversity. In *Statistical Ecology*, G.P. Patil, E.C. Pielou and W.E. Waters (eds), The Pennsylvania State University Press, pp. 315-59.

BULMER, M.G. (1974). On fitting the Poisson Lognormal distribution to species-abundance data. *Biometrics*, 9, 460-466.

CONDIT, R.; PITMAN, N.; LEIGH, E.G.; CHAVE, J.; TERBORGH, J.; FOSTER, R.B.; NUÑEZ, P.; AGUILAR, S.; VALENCIA, R.; VILLA, G.; MULLER-LANDAU, H.C.; LOSOS, E.; HUBBELL, S.P. (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666-669.

DIGGLE, P.J.; BESAG, J.E.; GLEAVES, J.T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, 32, 659-67.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1), 1-26.

EFRON, B.; TIBSHIRANI, R. (1993). *An introduction to bootstrap*. Chapman & Hall, 436.

FERREIRA, F.C.; PETRERE-JR, M. (2008). Comments about some species abundance patterns: classic, neutral and niche partitioning models. *Braz. J. Biol.*, 68, 1003-1012.

FISHER, R.A.; CORBET, A.S.; WILLIAMS, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, 12, 42-58.

GOOD, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40, 237-264.

HE, F.; HU, X.S. (2005). Hubbell's fundamental biodiversity parameter and the Simpson diversity index. *Ecology Letters*, 8, 386-390.

HELTSHE, J.F.; BITZ, D.W. (1979) Comparing Diversity Measures in Sampled Communities. In *Ecological Diversity in Theory and Practice*, J.F. Grassle, G.P. Patil, W. Smith and C. Taillie (eds), International Co-operative Publishing House, Fairland, Maryland, 133-44.

HELTSHE, J. & FORRESTER, N.E. (1983). Estimating species richness using the *jackknife* procedure. *Biometrics* 39, 1-11.

HELTSHE, J.F.; FORRESTER, N.E. (1985) Statistical evaluation of the *Jackknife* estimate of diversity when using quadrat samples. *Ecology*, 66, 107-111.

HILL, M.O. (1973). Diversity and Evenness: a unifying notation and its consequences. *Ecology*, 54, 427-432.

KEMPTON, R.A. (1979). The Structure of Species Abundance and Measurement of Diversity. *Biometrics*, 35, 307-321.

KREBS, C.J. (1999). *Ecological methodology*. 2<sup>a</sup> ed. Addison Wesley Longman.

MACARTHUR, R.H. (1957). On the relative abundance of bird species. Proc. Nat. Acad. Sci. Wash., 43, 293-295.

MAGURRAN, A.E. (1988). Ecological Diversity and its Measurement, Cambridge University Press, London.

MANLY, B. F. J. (2007). Randomization, *bootstrap* and Monte Carlo methods in Biology. Chapman & Hall/ CRC. 455p.

MENDES, R. S.; EVANGELISTA, L.R.; THOMAZ, S.M.; AGOSTINHO, A.A.; GOMES, L.C. (2008) A unified index to measure ecological diversity and species rarity. Ecography, 32, 450-456.

MILLER, R.G. (1974). The *jackknife* – a review. Biometrika, 61, 1-15.

PEET, R.K. (1974). The measurement of species diversity. Ann. Rev. Ecol. Syst., 5, 285-307.

PIELOU, E.C. (1975). Ecological diversity. New York: John Wiley. 165p.

PRESTON, F.W. (1948) The commonness and rarity of species. Ecology, 29, 254-283.

R Development Core Team. (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RÉNYI. A. (1961). On measures of entropy and information. In Berkeley symposium on mathematical statistics and probability, J. Neyman (ed), Berkeley, 4<sup>a</sup> ed., 547-561.

ROUTLEDGE, R.D. (1980) Bias in estimating the diversity in large, uncensused communities. Ecology, 61, 276-81.

SIMPSON, E.G. (1949). Measurement of Diversity. *Nature*, 163, 688.

SMITH, E.P.; VAN BELLE, G. (1984). Nonparametric estimation of species richness. *Biometrics*, 40(1), 119-129.

TAYLOR, L.R.; WOIWOD, I.P.; PERRY, J.N. (1978). The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology*, 47, 383-406.

THE MATHWORKS. (1992). Matlab®, software para computação numérica de alta performance. The MathWorks Inc., Natick, MA, USA

TSALLIS, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, 52, 479-487.

ZAHL, S. (1977). Jackknifing an index of diversity. *Ecology*, 58, 907-913.