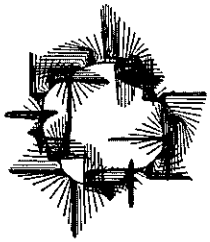
生物多様性と統計数理



島谷　健一郎

1. 生物多様性研究から生まれた統計数理

生物多様性（biodiversity）に関するデータ解析においても，機械学習のテクニックは使われ始めている．画像からの種同定，種組成データに基づく生物群集の分類，センサーや動画で得られるデータに基づく動物の行動分類などはその典型である．実際，2018年3月に行われた日本生態学会では，人工知能（AI）に関する企画が3件もあった．ただ，その多くは確立された手法を生物多様性研究に応用したもので，固有の修正や工夫などの新規性は伴うものの，生物多様性研究から生まれたデータ解析法というと言い過ぎの感がある．

一方，生物多様性研究を一つの発祥の地とする統計数理がある．ここでは，その例として，多様性指数，観測されなかった種数の推定，多様性の分解の3つを紹介する．いずれも，分野を跨いで発展してきている統計数理で，本稿を契機に他分野への適用や普及並びに異分野研究者の生物多様性の統計数理への参与を促したい．同時に，機械学習という先端技術を導入する際に，ともすれば疎かにしかねない，対象に関する背景概念やその定量化に関する理解と再検討の重要性を示す例の提供も図る．

最初に基礎概念や記法に関して簡単に説明し，それから3つの事例について3–5節で解説する．

2. 群集の種多様性

2.1 生物多様性と群集生態学

単純にいうと，生物多様性とは，「様々な生物が相互に関係し合いながら共存するサマ」を広く含有する概念である．それはしばしば，（種内の）遺伝的多様性，群集の種多様性，生態機能多様性，ランドスケープ多様性，といった階層に区分される．本稿では，この中の，群集の種多様性に焦点を当てる．なお，他の階層の多様性についても類似の統計数理が適用できる．

群集（community）およびこれを研究対象とする群集生態学には，植物–草食動物–肉食動物のような食物網を対象とするものと，森林の樹木，草原の草本，干潟を訪れる鳥類のように，特定の生物種に限ったものがある．本稿で取り上げる群集は後者を扱う．

2.2 群集データ

群集データの一つの基本形は，観察された種のリストと各種の個体数からなるものである．こうしたデータから様々な生物多様性指数を計算し，どういう環境で多様性が高いか，環境傾度と多様性の相関や回帰分析，多様性の低い群集についてその要因は何か，等々を考察する．

なお，多様性指数の多くは，数学の言葉で定義された，直接測ることのできない数量で定義されており，データから計算できるのはその統計的推定量である．そこでまず，群集生態学の言葉で生物多様性指数を定義することにする．

2.3 代表的な生物多様性指数

一般に最も広く用いられている生物多様性の指標は，異なる生物種の数（種数，species richness）である．

しかし，種数が同じでも，一つの種が突出して多いと，実質，その1種からなる群集とあまり多様性は変わらない．両者を区別するには，構成種の均一度を考慮した指数が必要となる．

ある群集からランダムに1個体サンプリングしたとき，それが種 s である確率（相対頻度，relative frequency）を p_s とする（$0 \le p_s \le 1$，種 s がいないなら $p_s = 0$）．以下の2つの指数は種数と並んで広く用いられている（log は自然対数）．

$$\text{Gini–Simpson 指数}：D = 1 - \sum_s p_s^2,$$

$$\text{Shannon 指数}：H = -\sum_s p_s \log p_s.$$

ちなみに，種数を数式で表すと，$S = \sum_s I(p_s > 0)$ ($I(Z)$ は Z が正しいとき 1，正しくないとき 0) となる．

p_s^2 は（繰り返しを許して）ランダムに選んだ 2 個体が共に種 s である確率だから，$\sum_s p_s^2$ は選んだ 2 個体が同種である確率，Gini–Simpson 指数 D は異種である確率を表す．D は少数の種が優占していると低く，各種が均一に近いと高い（後の図 1 参照）．すべての種が均等のとき（種数を S とすると $p_s = 1/S$），D は最大値 $1 - 1/S$ をとる．

Shannon 指数 H は情報量や統計物理のエントロピー（乱雑さ）でよく知られているが，それはある個体の種を予測しにくい状況に対応するので多様性が高いと評価する．D と同じく，種構成が特定の種に片寄っていると H は低く，均一だと高い．全種が均等のとき，H は最大値 $\ln S$ をとる．

D も H も相対頻度の均一具合の定量化だが，概して D は優占する種の均一さに大きく依存し，相対頻度の低い種の影響をほとんど受けない．一方，H はマイナー種の存在の影響を（D よりは）受ける．ちなみに，種数 S は 1 個体しかいない種も優占種も同じように 1 と数えるため，マイナー種の存在に 3 者の中で最も敏感である．

種数，Gini–Simpson 指数，Shannon 指数のどれが「最も優れた」生物多様性指数であるかを議論する場面（どの指数の高い地域を優先的に保全すべきか，等々）を見かけるが，それは無意味である．優占種がバランスよく生息するという側面を重視するなら Gini–Simpson 指数，稀少種の存在が重要なら種数で評価することが望ましい．群集生態学研究が目的なら，データからの推定値と理論値との比較で理論を検証できるような指数が望ましい．多様性を数値化する目的が何で，それにはどの側面を重視すべきか．この点をおさえた上で，適切な指数を選ぶことになる．

3. 生物多様性指数が満たすべき条件

とはいうものの，やはり多様性の尺度として満たしていてほしい性質というものはある．

その第一は，多様性は加えられるという点である．Shannon 指数がその典型であるが，多様性指数の多くは，乱雑さや予測しにくさなどを表す数式を援用している．確かに，多様性が高いと，何が出るか予測しにくいのでこの援用は妥当である．しかし，多様性は加えられるという点で概念として異なっている．

表 1　予測しにくさと多様性の違い．共通種を持たない 2 つの群集があるとき，予測しにくさの指数を加えても 2 群集を合わせた群集の予測しにくさにはならない（左）．一方，適切に定義された多様性指数（ヒル数）は単純和になる（右）．

群集	種構成	Gini–Simpson 指数	ヒル数
A	△△△ ○○	0.48	1.92
B	××× ＊＊	0.48	1.92
A + B	△△△ ○○ ××× ＊＊	0.74	3.85

例えば，共通種を持たない 10 種と 20 種の群集があるとき，2 つを合わせた群集には $10 + 20 = 30$ の種がいる．

一方，10 種の群集と 20 種の群集の予測しにくさを，単純に足すことはできない．もちろん合わせた群集の予測しにくさは数値化できるし，それはそれぞれの予測しにくさより上である．しかし，それぞれの単純な「和」とはならない（表 1 参照）．

次に，すべての種が同じ相対頻度なら種数だけで多様性を評価できるが，一般に相対頻度は不均一で，均一なときより多様性は低いと評価するのが自然である．つまり，均一なら種数に一致し，一般にはそれより低い値になっているという性質である．

一方，古くからの試みの一つに，様々な指数を一つの数式に統一するというものがある．すなわち，パラメータの値に応じて Shannon 指数になったり Gini–Simpson 指数になったりする数式である．ただし，単に数式としてそうなればいいというものではない．パラメータは何を意味しているのか．数式全体の意味は何か．生物学的な解釈を伴うことが望まれる．

以上を踏まえると，多様性指数の数式は，次の 3 つの条件を満たしていてほしい．

1. 複数の指数を統一：代表的な指数を特別な場合として含む．
2. 有効な種数 (effective number)：種数が S ですべての種の相対頻度が等しく $1/S$ なら値は S となる．
3. 加法性 (replication principle)：互いに共通する

種を持たない L 個の群集が等しい多様性指数 Q を持つとき，L 個を合わせた群集の多様性指数は LQ になる．

当然の条件のように感じられるかもしれないが，Jost (2006) とこれに続く論文で指摘されるまで，生態学ではエントロピーなどを援用することで多様性を定量化することをとりわけ問題視してこなかった．そして，Jost の問題提起は，以下の 5 節で示すように，単なる多様性指数の数式の形だけでなく，その概念自体にも深く切り込む研究の発端となったのである．

3.1 ヒル数

上の 3 条件を満たす数式として，Jost (2006) は以下の式（ヒル数，Hill number）を提案した．

$$ {}^qD = \left(\sum_{s=1}^{S} p_s^q \right)^{\frac{1}{1-q}} \quad (q \geq 0). $$

$q = 0$ のとき，${}^0D$ は種数 S になる．

$q = 2$ のとき，${}^2D = 1/(1 - D)$ となり，Gini–Simpson 指数に対応する．

$q = 1$ のときは不定形の極限となるが，対数をとることで e^H という形で Shannon 指数と対応することを示せる．

このように，ヒル数は，代表的な 3 つの指数を統合しており最初の条件を満たす．また，0 と 1 の間の数 p_s を q 乗するため，q が 0 に近いと p_s の小さいマイナー種の存在が大きい貢献度を示し，q が 1 を超すと，逆に p_s の大きい優占種の貢献度が高くなる（$q = 1$ はある意味で中庸）．従って，q はマイナー種と優占種をどのくらいのバランスで重要視するかという尺度と解釈できる．

2 つ目の条件は，$p_s = 1/S$ を代入することで ${}^qD = S$ を確かめられる．

群集 j に種 s_j から s_{j+S_j-1} までの S_j 種がいて，かつ，${}^qD$ はすべての群集で等しい ${}^q\overline{D}$ とする．L 個の群集を合わせたときの種 s の相対頻度は p_s/L なので，全体の多様性指数は，

$$ \begin{aligned} {}^qD &= \left(\sum_{j=1}^{L} \sum_{s=s_j}^{s_j+S_j-1} \left(\frac{p_s}{L} \right)^q \right)^{\frac{1}{1-q}} \\ &= \left(\frac{L}{L^q} \cdot \sum_{s=s_j}^{s_j+S_j-1} (p_s)^q \right)^{\frac{1}{1-q}} \\ &= (L^{1-q})^{\frac{1}{1-q}} \cdot {}^q\overline{D} = L\,{}^q\overline{D} \end{aligned} $$

となり，3 つ目の条件も満たす．

	群集 A	B	C
種 1	0.6	0.7	0.46
種 2	0.2	0.1	0.37
種 3	0.1	0.1	0.17
種 4	0.1	0.05	0
種 5	0	0.05	0

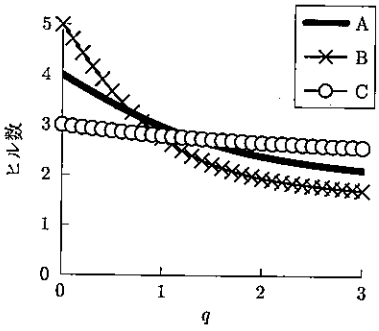


図 1　様々な q の値におけるヒル数のグラフの例．上に 3 つの群集 A, B, C における 5 種の相対頻度を示した．C は種数は最小だが優占種の均一性は高い．B は種数は多いが種 1 が突出している．B と C は q が 0 と 1（種数と Shannon 指数に対応する）の間で交叉している．A と C は 1 と 2（Shannon 指数と Gini–Simpson 指数に対応する）の間で交叉している．

生物多様性指数については，どの指数が優れているかという論争から，優劣論争は無意味といった主張を経て，（今まで提唱された中では）ヒル数の数式が一番妥当である，ということで生態学者の間でほぼ見解は一致するに至っている（Ellison (2010)）．

実際の群集に対しては，様々な q について（実際には $q = 3$ くらいまでで十分な場合が多い）${}^qD$ を計算し，横軸に q をとってグラフ（diversity profile）を描く（図 1）．ある曲線がすべての q で上にいたら，その群集は多様性が高いと評価できる．曲線が交叉していたら，マイナー種の存在と優占種の均一さのどちらを重視するかにより，多様性の評価は異なることになる．どのくらいのバランスで評価が反転するかは，交点の q の値で定量化される．

言い換えると，群集の相対頻度の均一さを考慮した多様性は，1 個の数値でなく 1 本の曲線で評価する．

3.2 簡単そうで厄介な問題をはらむデータからの推定

前節の数式は群集の構成種の真の相対頻度を用いて表したものである．もちろん，実際の野外群集で全構成種の相対頻度がわかっていることなど，滅多にない．

データから推定する必要がある．

群集からランダムに選んで調べた n 個体のサンプル中に種 s が x_s 個体だったら（$n=\sum_s x_s$），相対頻度を $\hat{p}_s = x_s/n$ として ${}^q\hat{D} = (\sum_{s=1} \hat{p}_s^q)^{\frac{1}{1-q}}$ という式で何ら問題ないように感じられる．

しかし，この推定には深刻な問題が潜んでいる．例えば動物の場合，目視で確認できた個体にせよ，罠にかかった個体にせよ，観察できなかった種がいて当然である．すなわち，観察された種数は常に真の種数を過小推定する．また，観察されなかった種の相対頻度を 0 とすると，観察された種の相対頻度は常に過大推定となる．

さらに，こうした問題を意識するようになると，観察できた種で群集のどれだけを把握したのか，その見当すらついていないという現実を認識する．それで，観察できた種は実は群集全体の 1 割にも満たない割合でしかないのではないか，といった不安感に苛まれるようになる．

統計学の役割の一つに，サンプルからの全体（母集団）の推定がある．しかし，情報がない状況では原理的に不可能である．観察されなかった種についての情報はないので，それらのリストを作る（推定する）ことは不可能である．

ところが，目標を「観察されなかった種」でなく，「観察されなかった種数」および「観察された種の全体における割合」に転換すると，データからある程度の推定が可能となるのである．次の 4 節では，この問題を解説する．

3.3 観察されなかった種数と想定外の科学

ところで，同じ問題は，生物多様性に限らず広く人間社会に見られるのではなかろうか．

例えば，事故である．どれだけ過去の事故のデータを分析しても，「まだ起こっていないがいつ起こってもおかしくない事故」というものがあるはずである．これは観察されなかった種と類似する．

起こっていない事故を予測することは不可能なので，「まだ起こったことのない事故を想定して対策を立てることは不可能です」と言われると納得してしまいがちである．しかし，もし既に体験した事故は起こり得る事故の何割くらいなのかがわかるなら，想定外の事故がまだどのくらいあるのか察しがつき，初めて体験する事故に対する対策も整えられるのではないか．これは，観察された種で全体の何割程度かという問題と対応する．

4. 観察されなかった種数の推定

4.1 群集からのランダムなサンプルと 2 項分布

n 回のサンプリング（反復を許す）が毎回独立とすると，種 s の個体数 x_s は調査のたびにランダムに変動する確率変数の実現とみなすことができる（観察されなかったら $x_s=0$）．相対頻度が p_s なら，k 個体観察される確率は 2 項分布を用いて

$$P(X_s=k) = {}_n\mathrm{C}_k\, p_s^k (1-p_s)^{n-k}$$

と書ける．

観察された個体数が k である種数を f_k で表す（$0 \le k \le n$）．f_0 は 0 個体観察された，すなわち観察されなかった種数を表す．なお，f_k は調査によって毎回変化するから，確率変数として $F_k = \sum_{s=1}^{S} I\,(X_s = k)$ と表すと，その期待値は $\mathbf{E}[F_k] = \sum_{s=1}^{S} {}_n\mathrm{C}_k\, p_s^k (1-p_s)^{n-k}$ となる．$k=0,1,2$ として，$\mathbf{E}[F_0] = \sum_{s=1}^{S}(1-p_s)^n$，$\mathbf{E}[F_1] = \sum_{s=1}^{S} np_s(1-p_s)^{n-1}$，$\mathbf{E}[F_2] = \sum_{s=1}^{S}(n(n-1)/2)p_s^2(1-p_s)^{n-2}$ を得る．

4.2 観察されなかった種数

よく知られたシュヴァルツの不等式

$$\sum_s \alpha_s^2 \sum_s \beta_s^2 \ge \left(\sum_s \alpha_s\beta_s\right)^2$$

に $\alpha_s = (1-p_s)^{\frac{n}{2}}$，$\beta_s = p_s(1-p_s)^{\frac{n}{2}-1}$ を代入する．

$$\sum_{s=1}^{S}(1-p_s)^n \sum_{s=1}^{S} p_s^2(1-p_s)^{n-2} \ge \left(\sum_{s=1}^{S} p_s(1-p_s)^{n-1}\right)^2.$$

両辺を $\sum_{s=1}^{S} p_s^2(1-p_s)^{n-2}$ で割ると左辺は $\mathbf{E}[F_0]$ に等しくなる．右辺に $\mathbf{E}[F_1]$, $\mathbf{E}[F_2]$ の式を代入して，

$$\mathbf{E}[F_0] \ge \frac{\left(\sum_{s=1}^{S} p_s(1-p_s)^{n-1}\right)^2}{\sum_{s=1}^{S} p_s^2(1-p_s)^{n-2}} = \frac{n-1}{n}\cdot\frac{(\mathbf{E}[F_1])^2}{2\mathbf{E}[F_2]}$$

という不等式が得られる．

これは，観察されなかった種数の期待値 $\mathbf{E}[F_0]$ が，1 個体および 2 個体観察される種数の期待値 $\mathbf{E}[F_1]$ と $\mathbf{E}[F_2]$ を用いて下から押さえられるということを意味する．

そこで，観察値と期待値はある程度近いと考えて置き換えると，

$$\hat{f}_0 = \frac{n-1}{n}\cdot\frac{f_1^2}{2f_2}$$

という形で，観察されなかった種数の推定量が得られ

る．近似不等式ではあるが，観察されなかった種数を，1個体だけ観察された種および2個体観察された種に関する情報から推定している．経験知から想定外を予測する数式といえるだろう．

4.3 観察されなかった種の割合

k 個体観察された種の相対頻度の平均を $a_k = \sum_{s=1}^{S} p_s I(x_s = k)/f_k$ とする．a_0 は観察されなかった種の相対頻度の平均だから $a_0 f_0$ はその総和，すなわち観察されなかった種の相対頻度の合計である．これはまさしく観察されなかった種の群集における割合を表すので，$a_0 f_0$ の推定量を導くことにする．

まず，F_k の期待値の式の2項分布を書き下して整理し，期待値を実測値 f_k で近似することで，f_k と a_k に関する近似の漸化式 $\frac{r+1}{n-r} \cdot f_{k+1} = \frac{a_k}{1-a_k} \cdot f_k$ を導く．それを a_k について解いてから $k=0$ とし，f_0 に推定量 $\hat{f}_0 = \frac{n-1}{n} \cdot \frac{f_1^2}{2f_2}$ を代入することで，$a_0 f_0$ の推定量

$$\hat{a}_0 \hat{f}_0 = \frac{2f_2}{(n-1)f_1 + 2f_2} \frac{n-1}{n} \frac{f_1^2}{2f_2} = \frac{f_1}{n} \cdot \frac{(n-1)f_1}{(n-1)f_1 + 2f_2}$$

が得られる（導出の詳細は Chiu *et al.* (2014) などを参照）．

右辺は，n が大きいとほぼ f_1/n なので，労して得たデータで群集の $1 - f_1/n$ くらいを把握できているという目安を立てられるわけである．これも，経験知から想定外を予測する一例といえる．

ちなみに，1個体だけ観察された種の相対頻度を $1/n$ で推定すると，それらが f_1 種あるので全体で占める割合は f_1/n となる．ところが，これは観察されなかった個体の占める割合の推定だと主張しているのである．

相対頻度の過大推定をどう修正してヒル数を推定するかについては，Chao and Jost (2012) や Chiu *et al.* (2014) で研究が進められている．

5. 多様性の分解：α-, β-, γ-多様性

群集が広がりを有するとき，対照的な次の2つのパターンが考えられる（表2）．一つは，ある部分（局所群集）はある特定種が多く別な部分では別な特定種が多く，一部分だけを見ると多様性は低い（局所群集間の差異は大きい）．もう一つは，どの局所群集も全体と同じような相対頻度で全体と同じくらい多様性は高い（局所群集間の差異は小さい）．

各局所群集の多様性を **α-多様性**，局所群集間の差異を **β-多様性**，全体の多様性を **γ-多様性**という．この概念は古く1960年代に遡る．単純には α-多様性は局所群集の多様性指数の平均，γ-多様性は全局所群集のデータをプールして算出し，β-多様性は群集間の類似度指数（各群集の種数の平均と全体種数の比など）で評価してきた．そこへ3節同様，α-, β-, γ-多様性指数が満たすべき条件の定式化を Jost (2007) は問題として提起した．何らかの数式で定められる α-多様性，β-多様性，γ-多様性を D_α, D_β, D_γ とし，局所群集の数を N とする．

表2　群集全体の構造は同じだが，部分部分は対照的な例．左では局所群集における優占種はそれぞれ異なっている．右ではすべて全体と同じ種組成になっている．

	群集1		群集2
局所群集	種構成	局所群集	種構成
A	△△△△△ ○	A	△△ ○○ □□
B	○○○○○ □	B	△△ ○○ □□
C	△ □□□□□	C	△△ ○○ □□
群集全体	△△△△△△ ○○○○○○ □□□□□□	群集全体	△△△△△△ ○○○○○○ □□□□□□

1. 3者は**乗法的** $D_\gamma/D_\alpha = D_\beta$，または**加法的に** $D_\gamma - D_\alpha = D_\beta$，という関係式で結ばれる．
2. 不等式 $D_\alpha \le D_\gamma \le N D_\alpha$．全体の多様性は，部分の多様性以上で部分の多様性の総和以下である．
3. **独立性**：D_α と D_β は独立に与えられる．この意味を巡って議論が紛糾したが，Chao *et al.* (2010) は，D_α が与えられたとき D_β はある数値を超えないなどの制限が出てこない，という意味で解決を図った．
4. **β-多様性の加法性**（replication principle）：互いに共通種を持たない局所群集が N 個あったとき，$D_\beta = N$．言い換えると，D_β は「有効な局所群集数」と解釈できる．
5. D_β は（何らかの変換を経たら）よく使われている群集間の類似度（差異）指数になる．
6. 3者とも相対頻度以外の因子を含む多様性指数へ拡張しやすい．

Chao *et al.* (2010) や Chao and Chiu (2016) が主

張している多様性の分解は以下のようなものである.

1. ヒル数を基本とする.
2. D_γ は各局所群集の（真の）相対頻度の重みづけ平均を用いる.

$$ {}^qD_\gamma = \left\{ \sum_{s=1}^{S} \left(\sum_{j=1}^{N} w_j p_{sj} \right)^q \right\}^{\frac{1}{1-q}}. $$

3. D_α は

$$ {}^qD_\alpha = \frac{1}{N} \left\{ \sum_{s=1}^{S} \sum_{j=1}^{N} (w_j p_{sj})^q \right\}^{\frac{1}{1-q}} $$

という形の重みづけ平均を用いる.

4. $D_\beta = D_\gamma / D_\alpha$ で D_β を定義する（乗法性を満たすように定める）.

多様性指数の拡張として相対頻度以外でよく取り入れられているのが，種間差異（類似度）である．例えば，針葉樹ばかり3種の森と広葉樹も混ざった3種の森では後者のほうが多様性が高い．種間差異は，進化系統樹における距離や生態的機能などがよく用いられている．なお，種間差異を取り入れた指数への拡張も既に提唱されているが，Chao and Chiu (2016)，Botta-Dukát (2018) などからうかがえるように，多様性の分解と合わせて今なお優劣論争が続いている．

多様性を分解して観ることで，社会でよく指摘される人材の多様性などでも新たな視点を広げられるに違いない．

6. むすび

観察できなかった種数の推定は，古く戦争時における Turing の暗号に関する研究に遡る（Good (2000) などを読むと，その時点で観察されない種との関連を認識していたようである）ことが示すように，社会の様々な分野で活用できるものである．対象やデータの性質に応じて適切な仮定を設けることで，本稿で紹介した数式以外の推定量もいろいろ開発できるに違いない（観察されなかった種数についても Tuvo *et al.* (2017) などで本稿にはない推定法が提唱されている）.

国内では「想定外」という醜い言い訳が2011年3月から流行しているが，観察されなかった種数の推定法を修正して発展させていくことで，経験知から想定外の何がしかを推定できるに違いない．今後はそうしたデータ解析を行わずに唱えられた「想定外」は，言い訳として通用しない時代になってしかるべきである．

多様性は加えられるという点でエントロピーや予測しにくさとは異なる．この Jost (2006) の問題提起は，多様性の分解や種間差異を取り入れた指数など，生物多様性の統計数理だけでなく概念そのものにも多大な進歩をもたらした．画像からの種同定など，機械学習により，これから生物多様性データはより効率よく，より正確に，より大量に蓄積されていく．そのとき，背景にある統計数理を知っているか否かで，データの活用や収集過程で格段の差が生じるだろう．先端技術を取り入れるばかりで諸概念の検討や理解を疎かにしていたのでは，せっかくの技術革新がもたらすはずの恩恵も小さなものになってしまいかねない.

参考文献

1) Botta-Dukát, Z. (2018) The generalized replication principle and the partitioning of functional diversity into independent alpha and beta components. *Ecography* **41**: 40–50.
2) Chao A., Chiu C.H. and Hsieh T.C. (2010) Proposing a resolution to debates on diversity partitioning. *Ecology* **93**: 2037–2051.
3) Chao A and Jost L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93**: 2533–2547.
4) Chao A. and Chiu C-H. (2016) Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution* **7**: 919–928.
5) Chiu C.H., Wang Y.T., Walther B.A. and Chao A. (2014) An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics* **70**: 671–682.
6) Ellinson, A.M. (2010) Partitioning diversity. *Ecology* **91**: 1962–1963.
7) Good, I.J. (2000) Turing's anticipation of empirical Bayes in connection with the cryptanalysis of the naval Enigma. *Journal of Statistical Computation and Simulation* **66**: 101–111.
8) Jost L. (2006) Entropy and diversity. *Oikos* **113**: 363–375.
9) Jost L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427–2439.
10) Tuvo, A. Suweis, S., Formentin, M., Favretti, M. Volkov, I., Banavar, J.R., Azaele, S. and Maritan, A. (2017) Upscaling species richness and abundances in tropical forests. *Science Advances* **3**. Doi: https://doi.org/10.1126/sciadv.1701438.

（しまたに・けんいちろう，統計数理研究所）