

Clark JS, Mohan J, Dietze M and Ibanez I. 2003. Coexistence: How to identify trophic trade-offs. *Ecology* 84:17-31.

講釈：久保拓弥 kubo@ees.hokudai.ac.jp

1. Introduction

観測データからどうやって trophic (栄養資源の使いかたの) trade-off を推定するか?

- trophic trade-off が植物の多種共存に重要だろうと考えられてる (種によって異っていると考えられている例: 複数の栄養要求の下限, 光-光合成曲線, などなど)
- しかしこれまで「発見」された trophic trade-off はアヤしい— 観測データからそれを推定する統計的手法が不備であるからだ (後述するさまざまなばらつきを考慮していないから)
- (必要なばらつきを無視した) 従来の統計学的方法は偏った・不正確な・「言いすぎ」の結果をだすので, ここではばらつきの階層性を考慮したモデルの推定結果も示し, これまでの方法と比較する.

/* 久保注〜 じつは題材は trophic trade-off でなくてもよい……つまりどんなデータ解析にも共通する問題をとりあつかっている */

2. The data

その場の「明るさ」と実生の高さ成長の関係は樹種によって異なるのか? をみるデータセット

- 単純な対象を例として扱う: *Acer rubrum* (アメリカハナノキ) と *Liriodendron tulipifera* (ユリノキ). 同所に共存. 調査地は Duke Forest, Orange County, North Carolina, USA. 1-m² plot 多数を 3 回 census.
- 調査地の階層性: plot (location) $j \in \{1, \dots, m\}$; $m = 37$ の明るさの観測値 x_j (全天写真より推定した global site fraction, GSF — きちんと書いてないけど同一場所で複数回観測) と j 内の樹木実生個体 j の高さ成長量 y_{ij} (二回成長ぶん) を観測した ($i \in \{1, \dots, n_j\}$).

3. Why traditional methods can be misleading

実生の高さ成長は明るさ依存で光飽和, という簡単なモデルを例に考える

- plot j における高さ成長の平均は $\mu_j = G \left(\frac{x_j - x_0}{\theta + x_j} \right)$ とする (しかし推定計算には同値の Eq.1 を使う). パラメーター G は最大成長速度, x_0 は「最低」光強度, θ は「半飽和」光強度をあらわす. 個体の高さ成長は正規分布 $N(\mu_j, \sigma^2)$ の “i.i.d” (独立同分布) にしたがう (traditional model in FIG. 1). likelihood (尤度) は $p(\mathbf{y}|\mathbf{x}, \mathbf{b}, \theta, \sigma^2) = \prod_{j=1}^m \prod_{i=1}^{n_j} N(y_{ij}|\mu_j, \sigma^2)$, これを最大化してパラメーターの最尤推定値を得る.

3-1. Confidence (credible) intervals

- (高さ成長量の平均) μ_j などパラメーターを「定数」だと考えてよいのか? データのばらつきを考えるとこれは現実的ではない仮定で, μ_j の推定値の信頼区間は従来どおりの方式で計算できない. そもそも $y_{ij} = \mu_j + \varepsilon_{ij}$ として ε_{ij} を “error” term という呼ぶのがヘン.
- 1990 年代以降の計算機統計学の進歩によって Markov Chain Monte Carlo (MCMC) 法が普及してきた. これによってパラメーターのばらつきを考慮した Bayes 推定が可能になった.
- Bayes 推定での信頼区間を credible interval とよぶ.

3-2. When parameters are not identifiable

- これまでの方法では「飽和」パラメーターや制約つき推定が難しい (FIG. 2 の *Acer* へばい推定).
- (土壌含水量など) 環境データの種類を増やしてもダメ (FIG. 2 右上のハコ).

3-3. Uncertain data

- これまでのモデルでは plot j における明るさ (GSF) を x_j が「真の値」だと仮定してしまっている．実際にはこれは正確な値ではなく (真の値など誰も知らない) 推定結果に影響を与える．

3-4. Individuals differ

- 明るさ x_j への (高さ成長の) 応答に個体差があった場合，この方式では推定できない．

3-5. The consequences of inappropriate models

- 以上のような弱点があるので，次の節では (1) 資源そのものがばらついており (2) 個体の応答もまたばらつきがある，という状況の階層構造のある (Bayesian な) 統計モデリングについて検討する．

4. Accommodating the stochasticity

/* 久保注〜 この節で階層的 Bayesian モデルが定式化される— 本論文でもっとも重要なところ */

4-1. Uncertain resources

- (FIG. 1 method 1 に対応) 「ホント」の明るさ x_j のもとで，明るさの観測値 $x_j^{(o)}$ と高さ成長の観測値 y_{ij} を得る確率の同時分布を尤度方程式として定式化する．
- plot j ごとに明るさのばらつきをいれる．明るさの prior (事前分布) はベータ分布 $\text{Beta}(x_j|a_j, b_j)$ ，パラメーター a_j と b_j はモーメント法で求める (plot j ごとに!).
- さらに平均成長量 μ_{ij} に関連するパラメーターと個体ごとのばらつき σ^2 の事前分布も導入する: Eq.1 の (β_0, β_1) は二次元正規分布 $\mathcal{N}_2(\mathbf{b}|\mathbf{b}_0, \mathbf{V}_b)$ ，半飽和光パラメーター θ はベータ分布， σ^2 のばらつきについては (非負値をとる) 逆ガンマ分布とする．

4-2. Growth varies by location

- (FIG. 1 method 2 に対応) method 1 に加えて，plot j ごとに実生の高さ成長のばらつき σ^2 に差がある σ_j^2 であるとする．事前分布を j ごとに用意する．

4-3. The response varies among individuals

- (FIG. 1 method 3 に対応) method 2 に加えて，個体 (plot j の i) ごとに光に対する応答が異なる (さらに高さ測定誤差も含む) 場合のモデル．
- method 1 で導入した事前分布の hyperprior (超事前分布) を導入する．平均成長量 μ_{ij} に影響するパラメーターが個体ごとに異っている $\mathcal{N}_2(\mathbf{b}_{ij}|\mathbf{b}_0, \mathbf{V}_b)$ としてこれらのパラメーターがそれぞれ二次元正規分布 $\mathcal{N}_2(\mathbf{b}_0|\mathbf{h}, \mathbf{V}_b)$ と二次元 Wishart 分布 $W_2(\mathbf{V}_b^{-1}|(w\mathbf{R})^{-1}, w)$ とする．これら超事前分布は noninformative prior (無情報事前分布) になっている (\mathbf{h} だの w だのはてきとー，ということ)．
- このあたりの計算は Gibbs sampler を使ってできる (MCMC 計算には Gibbs sampler を使う方法と Metropolis-Hastings sampler による方法があり，一般に前者のほうが計算が簡単で速い)．
- あらっぱいモデル選択規準として BIC を改良した DIC をもちいる． $\text{DIC} = 2\overline{D(\phi)} - D(\overline{\phi})$ ， $\overline{D(\phi)}$ は deviance (= $-2 \times$ 対数尤度) の平均値， $D(\overline{\phi})$ はパラメーターの平均値で計算した deviance．

5. Results

(FIG. 3 - 6 はユリノキに関する推定結果，FIG. 7 は両樹種)

5-1. The traditional model

- FIG. 3a: 推定された平均的な明るさ $\rightarrow \mu$ の関係 (実線) と平均 μ の信頼区間 (すごくせまい)．最大成長速度 G ，最低明るさ x_0 ，半飽和明るさ θ の分布もしめされている．

- FIG. 3b (FIG. 3a とほぼ同じ図): 成長量 y の信頼区間 (むちゃくちゃひろい $\Leftrightarrow \varepsilon$ すなわち residual (残差) の大きさ)

5-2. Uncertain resources (FIG. 1 method 1 に対応)

- FIG. 4: 37 plot の明るさ x_j をベータ二項分布で組み込んだモデルの推定結果 (MCMC 計算). 残差 ε の大きさはあいかわらず.

5-3. Growth uncertainty varies by location (FIG. 1 method 2 に対応)

- FIG. 5: さらに場所ごとに成長のばらつきを組み込んだモデルの推定結果 (MCMC 計算).

5-4. The response varies among individuals (FIG. 1 method 3 に対応)

- FIG. 6: さらに個体ごとの明るさ応答性の差位を組み込んだモデルの推定結果 (MCMC 計算). とうぜんながら高さ成長量の平均値 μ_{ij} の信頼区間はひろがり, 残差 ε は小さくなる.
- (図のまちがい) メインのグラフの下で “Minimum light x_0 ” には一番上の “Light x_j ” は含まれず (これは plot j での明るさ x_j の分布), “Individual” (個体ごとの x_0 の事後分布) と “Marginalized” (その周辺分布) と “Hyperparameter” (事後分布間にばらつきもたらずパラメーターの事後分布) からなる.

/* 久保注〜 めまいのする作図 */

5-5. Comparing tropic responses

- 明るさ・場所差・個体差を組みこんだモデルの推定結果 FIG. 7a (FIG. 6 にくわえてアメリカハナノキの推定結果も加えたもの) と従来からのモデルの推定結果 FIG. 7b (FIG. 2 と FIG. 3b) の比較.
- FIG. 7a では両樹種の μ_{ij} の信頼区間は大きく重なっており, いっぽう FIG. 7b では重複がない.

6. Discussion

- 現実が FIG. 7b みたいだったらただちにアメリカハナノキは競争に敗北することになる. しかし FIG. 7a の結果は場所によって勝ち負けが変わることを示している.
- 今回の推定方法は生態学者が考えているばらつきの理解と矛盾するものではない.
- 近ごろの研究では trophic trade-off が重要だと強調し, たしかにそういう結果も得られているが, 本研究では trophic interaction のばらつきがきわめて大きいものである, と観測データから示した.
- “trade-off” はあるんだろうけど, その寄与は理論研究ならびに従来型データ解析の予測するほど共存には重要ではない.

6-1. Alternative structures

- すべてを組み込んでないけど, 本研究はありそうな random effects は十分に考慮している.
- ただし random effects をもっと複雑にしたり, また別の明るさ依存性パラメーターを導入できる.

7. Conclusion

- この研究は trade-off の存在自体を否定するものではない. しかしこれまでの研究で「見つかった」 trade-off は (環境・個体の) ばらつきによって強調されすぎている.
- モデルはむやみに複雑にするな— 複雑な “error” 構造は複雑モデルの推定を誤らせるものだから.
- かかる複雑な状況から単純なモデルを推定するには本研究で使った階層 Bayes モデルが役にたつ.