



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Proprietà di scala della biodiversità in comunità microbiche

Relatore

Prof. Samir Suweis

Correlatore

Dott.ssa Anna Tovo

Laureanda

Eleonora Manoli

Anno Accademico 2017/2018

Abstract

Indice

Abstract	iii
Indice	v
Introduzione	1
1 Master Equation	3
1.1 Distribuzione binomiale negativa	3
1.2 La distribuzione logaritmica di Fisher	4
1.2.1 La distribuzione di Fisher come caso particolare della binomiale ne- gativa	4
2 Il metodo di upscaling	5
2.1 Metodo della binomiale negativa	5
2.1.1 Proprietà di auto-somiglianza della distribuzione binomiale negativa	6
2.1.2 Il numero di specie a scala totale	6
2.2 Metodo della distribuzione di Fisher	7
2.2.1 Proprietà di auto-somiglianza della distribuzione logaritmica di Fisher	7
2.2.2 Il numero di specie a scala totale	8
3 Conclusioni	9
Elenco delle figure	11
Elenco delle tabelle	13

Introduzione

Uno dei più importanti obiettivi in ecologia è quello di dedurre le proprietà generali di un ecosistema campionando solo una frazione di esso.

Per ragioni pratiche la biodiversità viene misurata o monitorata tipicamente a piccole scale, ma è importante poter conoscere la biodiversità dell'intero ecosistema

....

Uno strumento statistico comunemente usato per descrivere la normalità o rarità della presenza delle specie in una comunità ecologica è la **relative species abundance distribution (SAD o RSA)** cioè il numero di individui per specie presenti all'interno di una regione. Normalmente la SAD è misurata a scale locali

1 Master Equation

In questo capitolo vediamo come si possono ricavare la distribuzione binomiale negativa e la Log Series modellizzando la dinamica dell'abbondanza delle specie attraverso un'equazione che descrive la nascita e la morte degli individui : la *birth-death master equation*.

1.1 Distribuzione binomiale negativa

L'equazione che regola l'evoluzione di $P_{n,s}(t)$ per $n \geq 0$ è la seguente:

$$\frac{\partial P_{n,s}(t)}{\partial t} = P_{n-1,s}(t)b_{n-1,s} + P_{n+1,s}(t)d_{n+1,s} - P_{n,s}(t)b_{n,s} - P_{n,s}(t)d_{n,s}. \quad (1.1)$$

Imponendo condizioni al contorno riflettenti, $b_{-1,s} = d_{0,s} = 0$, la (1.1) è valida anche per $n = 0$ e $n=1$. Per $n > 0$ la soluzione stazionaria è:

$$P_{n,s} = P_{0,s} \prod_{i=0}^{n-1} \frac{b_{i,s}}{d_{i+1,s}} \quad (1.2)$$

dove il termine $P_{0,s}$ è il fattore di normalizzazione che può essere trovato imponendo la condizione $\sum_{n=1}^{\infty} P_{n,s} = 1$. Notiamo che la somma inizia da $n=1$ in quanto non si considerano specie con abbondanza nulla.

Assumiamo ora che $b_{n,s}$ dipenda da un termine indipendente dal numero di individui b_s , cioè il tasso di nascita pro capite, e dal termine r_s , che tiene conto di eventi di immigrazione o di interazioni intraspecifiche:

$$b_{n,s} = b_s(n + r_s). \quad (1.3)$$

Analogamente assumiamo che il termine $d_{n,s}$ dipenda da d_s , cioè dal tasso di morte pro capite:

$$d_{n,s} = d_s n. \quad (1.4)$$

Queste supposizioni sono ragionevoli in ecologia.

Sostituendo questi ultimi termini nella (1.2) e denotando con $\xi_s = b_s/d_s$, si ottiene:

$$P_{n,s} = P_{0,s} \binom{n + r_s - 1}{n} \xi_s^n.$$

La costante di normalizzazione può essere calcolata imponendo:

$$1 = \sum_{n=1}^{\infty} P_{n,s} = P_{0,s} \sum_{n=0}^{\infty} \binom{n + r_s - 1}{n} \xi_s^n = P_{0,s} [1 - (1 - \xi_s)^{r_s}] (1 - \xi_s)^{-r_s}$$

Dunque la probabilità che una specie s abbia s individui all'equilibrio è data da una binomiale negativa di parametri (r_s, ξ_s) e normalizzata per abbondanze non nulle ($n \geq 1$):

$$P_{n,s}^{NB} = \frac{1}{1 - (1 - \xi_s)^{r_s}} \binom{n + r_s - 1}{n} \xi_s^n (1 - \xi_s)^{r_s}. \quad (1.5)$$

Sotto l'ipotesi della teoria neutrale, secondo la quale le specie sono considerate demograficamente equivalenti (cioè ogni individuo ha la stessa probabilità di procreare, morire e migrare), possiamo rimuovere l'indice s di specie dall'equazione sopra, ottenendo così una RSA per l'ecosistema in esame.

1.2 La distribuzione logaritmica di Fisher

Notiamo che, scegliendo in modo diverso i termini $b_{n,s}$ e $d_{n,s}$, si può ottenere, partendo dalla *birth death master equation* (1.1), un'altra importante distribuzione: la Fisher Log Series. Assumiamo che la dinamica della popolazione di una comunità sia governata dal corso ecologico e dalla speciazione casuale invece che dalla migrazione da comunità esterne (?). Allora possiamo porre:

$$b_{n,s} = b_s n + \delta_{n,0} \nu \quad (1.6)$$

Aggiungendo la condizione al contorno riflettente $b_{0,s} = \nu$ si ha che il tasso di nascita tiene conto della riproduzione e della speciazione. In particolare, il parametro ν assicura che, se le specie si estinguono, la comunità rimane sempre popolata da un individuo. Dunque sostituendo la (1.4) e la (1.6) nella (1.2) e definendo $x_{emphs} = b_s/d_s$, si trova la seguente soluzione stazionaria:

$$P_{n,s} = P_{0,s} \frac{\nu}{b_s} \frac{x_s^n}{n}. \quad (1.7)$$

La costante di normalizzazione $P_{0,s}$ si determina imponendo:

$$1 = \sum_{n=1}^{\infty} P_{n,s} = P_{0,s} \frac{\nu}{b_s} \sum_{n=0}^{\infty} \frac{x_s^n}{n} = P_{0,s} \frac{\nu}{b_s} [-\log(1 - x_s)].$$

Dunque abbiamo: Anche in questo caso assumiamo che le specie siano equivalenti e possiamo dunque omettere l'indice s .

1.2.1 La distribuzione di Fisher come caso particolare della binomiale negativa

Osserviamo che la distribuzione binomiale negativa converge ad una distribuzione logaritmica nel limite di r che tende a zero:

$$\lim_{r \rightarrow 0} P_n^{NB} = \lim_{r \rightarrow 0} \frac{(1 - \xi)^r}{1 - (1 - \xi)^r} \binom{n + r - 1}{n} \xi^n = \frac{\xi^n}{-n \ln(1 - \xi)}, \quad (1.8)$$

dove si è usato il fatto che:

$$\binom{n + r - 1}{n} = \frac{\Gamma(n + r)}{\Gamma(n + 1)\Gamma(r)} \approx \frac{r}{n + 1},$$

per $r \approx 0$.

Notiamo dunque che la (1.8) coincide con la (??) ponendo $x = \xi$.

2 Il metodo di upscaling

In questa sezione vediamo come è possibile ricostruire la biodiversità di un intero ecosistema a partire da un campione ridotto di SAD.

....

2.1 Metodo della binomiale negativa

Il quadro analitico all'interno del quale si svolge questo lavoro è bastato sui seguenti passaggi:

- Campionare una frazione p^* dell'intera foresta e ottenere il vettore delle abbondanze delle S^* specie campionate, $n_{p^*} = n_1, n_2, \dots, n_{S^*}$
- Usare una combinazione lineare di binomiali negative con lo stesso $\hat{\xi}_{p^*}$ e diversi valori di r per fittare la SAD sperimentale al desiderato grado di accuratezza.

Di seguito analizzeremo in dettaglio il metodo, le proprietà e i passaggi che ci permettono di ottenere le informazioni desiderate.

Quando facciamo upscaling siamo interessati alla SAD ed al numero totale di specie presenti a scala totale, cioè in tutta l'area della foresta A . Denotiamo con $P(n|1)$ la probabilità che una specie abbia esattamente n individui a scala totale (qui con il numero 1 si denota l'intera foresta), anche nota come *abbondanza relativa delle specie* RSA. Notiamo che $P(n|1)$ deve essere definita solamente per $n \geq 1$ poiché, a scala totale, una data specie deve avere almeno un individuo. In questo contesto si ipotizza che la SAD segua una distribuzione binomiale negativa, $P(n|r, \xi)$ di parametri (r, ξ) :

$$P(n|1) = c(r, \xi) P(n|r, \xi) \quad (2.1)$$

con

$$P_n = \binom{n+r-1}{n} \xi^n (1-\xi)^r, \quad c(r, \xi) = \frac{1}{1 - (1-\xi)^r} \quad (2.2)$$

dove c è la costante di normalizzazione. Quest'ultima è stata calcolata imponendo $\sum_{n=1}^{\infty} P(n|1)$, dove la somma parte da 1 poiché le specie con abbondanza nulla a scala totale saranno assenti anche a scale ridotte. Notiamo che $p(n|r, \xi)$ è normalizzata per $n \geq 0$: questo perché, nei sotto campioni, esiste una probabilità non nulla di trovare una specie, presente nell'intera foresta, avente $n=0$ individui. In questo modo si tiene conto del numero di specie mancanti nei sotto campioni.

Consideriamo ora un campione di foresta di area a e definiamo $p=a/A$ la scala del campione, cioè la frazione di foresta osservata. Come primo passaggio calcoliamo la RSA del campione assumendo che quest'ultima non sia influenzata da correlazioni spaziali. Quest'ipotesi è ben soddisfatta ed è stata verificata usando dati di foreste generati *in silico* a vari gradi di correlazione spaziale. (citare qualcosa, ci devo tornare sopra??)

Sotto queste ipotesi la probabilità che una specie presenti k individui in un'area $a=pA$, condizionata dal fatto che presenta n individui nella regione totale A è data dalla distribuzione binomiale:

$$P_{binom}(k/p, n) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{se } k = 0, \dots, n \\ 0, & \text{se } nk > n \end{cases} \quad (2.3)$$

Infatti, in assenza di correlazioni spaziali, la probabilità che uno degli individui di una specie si trovi in una regione di area a è esattamente p . (?controllare?)

Mostriamo ora un risultato chiave per il metodo di upscaling:

2.1.1 Proprietà di auto-somiglianza della distribuzione binomiale negativa

Sia $P(n|1) = c(r, \xi)P(n|r, \xi)$ la RSA della foresta a scala totale e denotiamo con $P(k|r, \xi)$ la probabilità che una specie abbia abbondanza k alla scala $p \in (0, 1)$, condizionata dal fatto che alla scala totale A sono presenti n individui di quella specie. Se $P(k|n, p) = P_{\text{binom}}(n|r, \xi)$ segue una distribuzione binomiale, allora la RSA $P_{\text{sub}}(k|p)$ alla scala di campionamento p è ancora una binomiale negativa per $k \geq 1$ con il parametro ξ riscalato e lo stesso r :

$$P_{\text{sub}}(k|p) = \begin{cases} c(r, \xi)P(k|r, \xi), & k \geq 1 \\ 1 - c(r, \xi)/c(r, \hat{\xi}_p), & k=0 \end{cases} \quad (2.4)$$

con

$$\hat{\xi}_p = \frac{p\xi}{1 - \xi(1 - p)} \quad (2.5)$$

DIMOSTRAZIONE?

Ricordiamo che questo metodo fa uso solamente delle informazioni che si possono ottenere da un campione ad una certa scala p^* , infatti noi abbiamo informazioni solo sulle abbondanze delle $S^* \leq S$ specie presenti nel campione esaminato. Denotando il numero di specie di abbondanza k alla scala p^* con $S^*(k)$, otteniamo, per $k \geq 1$:

$$\frac{S^*(k)}{S^*} \equiv P(k|p^*) = \frac{P_{\text{sub}}(k|p^*)}{\sum_{k' \geq 1} P_{\text{sub}}(k'|p^*)} = \frac{P(k|r, \hat{\xi}_{p^*})}{\sum_{k' \geq 1} P(k'|r, \hat{\xi}_{p^*})} = c(r, \hat{\xi}_{p^*})P(k|r, \hat{\xi}_{p^*}) \quad (2.6)$$

che, dalla (2.1), è una binomiale negativa normalizzata per $k \geq 1$, mentre $P(k|r, \hat{\xi}_{p^*})$ è normalizzata per $k \geq 0$. Per quanto detto sopra otteniamo dunque il seguente risultato: partendo da una distribuzione binomiale negativa per la RSA a scala globale, anche la RSA a scala ridotta risulta distribuita secondo una binomiale negativa di parametri lo stesso r e $\hat{\xi}_p^*$ riscalato. Una RSA avente la proprietà di avere la stessa forma funzionale a scale differenti è detta essere *invariante per forma*.

2.1.2 Il numero di specie a scala totale

Fittando la RSA dei dati alla scala p^* possiamo dunque trovare i parametri r e $\hat{\xi}_p^*$, invertendo l'equazione (2.5), troviamo:

$$\xi = \frac{\hat{\xi}_{p^*}}{p^* + \hat{\xi}_{p^*}(1 - p^*)} \quad (2.7)$$

Usando ancora la (2.5) per eliminare ξ dall'ultima equazione, otteniamo la seguente relazione per il parametro ξ alle due scale p e p^* :

$$\hat{\xi}_p = \frac{p\hat{\xi}_{p^*}}{p^* + \hat{\xi}_{p^*}(p - p^*)} \equiv U(p, p^*|\hat{\xi}_{p^*}) \quad (2.8)$$

dalla quale, ovviamente, è possibile riottenere sia la (2.5) che la (2.7) ponendo $\xi \equiv \hat{\xi}_{p=1}$.

Vogliamo ora determinare la relazione tra il numero totale di specie S alla scala totale $p=1$ e il numero totale di specie osservate localmente S_p alla scala p . D'ora in avanti per

denotare il numero di specie alla scala locale useremo la notazione $S^* \equiv S_{p^*}$. Notiamo che:

$$P_{sub}(k=0|p^*) = \frac{S - S^*}{S} \quad (2.9)$$

$$P_{sub}(k=0|p^*) = \frac{S^*(k)}{S}. \quad (2.10)$$

Usando la seconda delle (2.4), il numero di specie presenti nell'intera foresta è dato, in termini dei dati del campione osservato, da:

$$S = \frac{S^*}{1 - P_{sub}(k=0|p^*)} = S^* \frac{1 - (1 - \xi)^r}{1 - (1 - \hat{\xi}_p^*)^r} \quad (2.11)$$

Notiamo che, se si assume che la RSA segua una distribuzione binomiale negativa a scala globale, il valor medio dell'abbondanza totale riscalda linearmente con l'area, infatti: (AGGIUNGERE EQ S26)

2.2 Metodo della distribuzione di Fisher

Ora mostreremo che è possibile risalire al numero di specie anche quando si suppone che la SAD a scala globale sia distribuita secondo una log-series.

Supponiamo che la RSA a scala globale sia distribuita secondo una distribuzione logaritmica con parametro x :

$$P(n|1) = P^{LS}(n|x) = \alpha(x) \frac{x^n}{n}, \alpha(x) = -(\log(1 - x))^{-1} \quad (2.12)$$

dove $\alpha(x)$ è la costante di normalizzazione. Assumendo anche questa volta che la RSA del campione non sia affetta da correlazioni spaziali si trova che anche la log-series soddisfa la proprietà di auto somiglianza.

2.2.1 Proprietà di auto-somiglianza della distribuzione logaritmica di Fisher

Sia $P(n|1) = \alpha(x)P^{LS}(n|x)$ la RSA alla scala globale e denotiamo con $P(k|n, p)$ la probabilità che una specie abbia abbondanza k nel campione alla scala $p \in (0,1)$ condizionata dal fatto alla scala totale A la specie possiede n individui.

Se $P(k|n, p) = P_{binom}(k|n, p)$ è distribuita secondo una binomiale, allora la RSA alla scala del campione, $P_{sub}^{LS}(k|p)$, è ancora una log-series per $k \geq 1$ con il parametro x riscalato:

$$P_{sub}^{LS}(k|p) = \begin{cases} n/2, & \text{se } n \text{ pari} \\ 3n + 1, & \text{se } n \text{ dispari} \end{cases} \quad (2.13)$$

con

$$\hat{x}_p = \frac{px}{1 - x(1 - p)} \quad (2.14)$$

DIMOSTRAZIONE??

Notiamo che (2.14) è analoga a (2.5). Dunque l'analogo di (2.7) è

$$x = \frac{\hat{x}_p}{p + \hat{x}_p(1 - p)} \quad (2.15)$$

e l'equazione (2.8) vale anche in questo caso.

La RSA può essere ottenuta come nell'equazione (2.6) ed è data da:

$$P(k|p) = \frac{P_{sub}^{LS}}{\sum_{k' \geq 1} P_{sub}^{LS}(k'|p)} = \alpha(\hat{x}_p) \frac{\hat{x}_p^k}{k} = P^{LS}(n|\hat{x}_p) \quad (2.16)$$

Poiché la distribuzione logaritmica di Fisher è un caso particolare della binomiale negativa, è anch'essa invariante per scala.

2.2.2 Il numero di specie a scala totale

Il numero di specie con popolazione $k \geq 1$ presenti nel sotto-campione di area $a=pA$ è dato da:

$$S_p(k) \equiv SP_{sub}(k|p) = S\alpha(x) \frac{\hat{x}_p^k}{k} = \hat{\alpha} \frac{\hat{x}_p^k}{k} \quad (2.17)$$

dove abbiamo unito le costanti S e $\alpha(x)$ in un unico termine $\hat{\alpha}$ che non dipende dalla scala p del campione. Quando ci riferiremo alla scala p^* useremo, per brevità di notazione, $S^*(k) \equiv S_{p^*}(k)$.

Allora il numero totale di specie S^* e l'abbondanza totale N^* (?) alla scala p^* sono date rispettivamente da:

$$S^* = \sum_{k=1}^{\infty} S^*(k) = -\hat{\alpha} \log(1 - \hat{x}_{p^*}) \quad (2.18)$$

$$N^* = k \sum_{k=1}^{\infty} S^*(k) = \hat{\alpha} \frac{\hat{x}_{p^*}}{1 - \hat{x}_{p^*}} \quad (2.19)$$

Poiché S^* e N^* sono note dal campione, possiamo trovare $\hat{\alpha}$ risolvendo la seguente equazione:

$$N^* - \hat{\alpha}(\exp(\frac{S^*}{\hat{\alpha}}) - 1) = 0 \quad (2.20)$$

che si ottiene inserendo l'espressione di \hat{x}_{p^*} da (2.18) nella (2.19).

Vogliamo ora dedurre le informazioni a scala globale $p=1$ dai dati disponibili alla scala $p=p^*$. Dalle considerazioni precedenti sappiamo che $\hat{\alpha}$ è un parametro indipendente dalla scala, dunque abbiamo le seguenti relazioni per S e N :

$$S = -\hat{\alpha} \log(1 - x) \quad (2.21)$$

$$N = \hat{\alpha} \frac{x}{1 - x} \quad (2.22)$$

dalle quali otteniamo:

$$S = \hat{\alpha} \log(1 + \frac{N}{\hat{\alpha}}), \hat{\alpha} = S\alpha(x). \quad (2.23)$$

Dunque per dedurre la biodiversità a scala globale, S , è necessaria una stima dell'abbondanza totale N . Prendiamo $N=N^*/p^*$. Notiamo che questo è consistente con il nostro quadro teorico nel quale assumiamo che la RSA sia "form-invariant(?)": infatti si può dimostrare che, se si assume che la RSA segua una distribuzione di Fisher a scala globale, il valor medio dell'abbondanza totale riscalda linearmente con l'area:

$$\mathbb{E}(N^*) = \sum_{k=1}^{\infty} k S^*(k) = \sum_{k=1}^{\infty} k \hat{\alpha} \frac{\hat{x}_{p^*}^k}{k} = \alpha \frac{\hat{x}_{p^*}}{1 - \hat{x}_{p^*}} = \hat{\alpha} \frac{px}{1 - x} = p^* \mathbb{E}(N), \quad (2.24)$$

dove abbiamo usato la (2.14).

3 Conclusioni

Elenco delle figure

Elenco delle tabelle
