

L'estimateur Chao1

Eric Marcon

14 février 2020

Section 1

Introduction

Problématique

Le Programme
Classé

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Estimer la richesse (le nombre d'espèces) d'une communauté en forêt tropicale par exemple est difficile.

Beaucoup d'espèces sont rares donc un échantillonnage aléatoire (inventaire) de taille raisonnable ne permet pas de les observer.

Des estimateurs de la richesse ont été développés pour estimer la richesse réelle à partir d'un inventaire incomplet.

Illustration

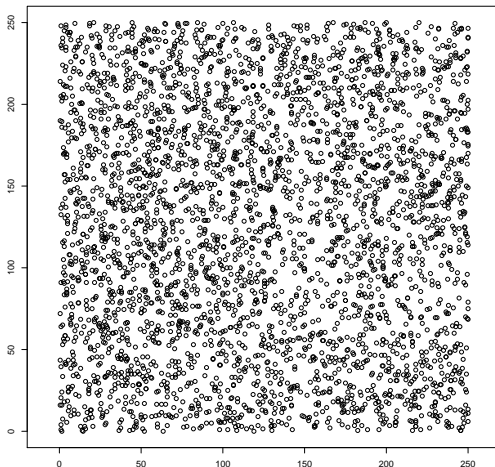
Eric Marcon

Introduction

Construction
 de l'estimateur

Application

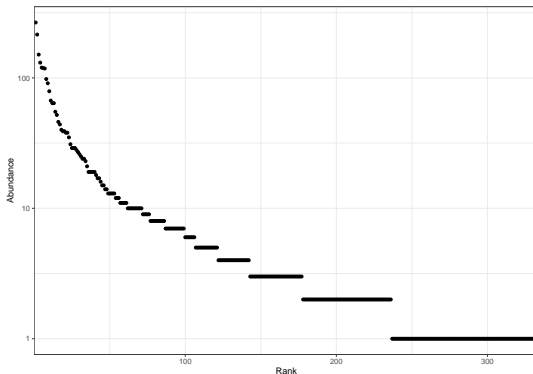
Inventaire d'une
 parcelle de
 Paracou, Sinamary,
 Guyane
 Nombre d'espèces
 observées : 334.
 Espèce la plus
 abondante (wapa :
Eperua falcata) :
 266 individus.



<https://paracou.cirad.fr>

Illustration

La parcelle est un échantillon de la communauté forestière locale.



Question : combien y a-t-il d'espèces d'arbres dans cette communauté ?

Estimateur Chao1

Estimateur
Chao1

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Développé par Anne Chao (Chao 2004).

Premier estimateur utilisé largement par les écologues, bon support mathématique.

Intuition :

- les espèces observées une fois auraient pu ne pas l'être.
- lien (à établir) entre les espèces observées un petit nombre de fois et les espèces manquées.

Section 2

Construction de l'estimateur

Notations

Introduction
Chapitre 1

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Un inventaire de n individus tirés indépendamment et aléatoirement est réalisé dans une communauté.

Les individus appartiennent à l'espèce s avec la probabilité p_s ,
$$\sum_1^S p_s = 1.$$

L'inventaire manque quelques espèces parmi les moins fréquentes : seules s_{obs} espèces sont observées.

s_n^ν est le nombre d'espèces observées ν fois dans un échantillon de taille n . C'est une réalisation de la variable aléatoire S_n^ν .

Observer une espèce

La probabilité qu'un individu inventorié ne soit pas de l'espèce s est

$$1 - p_s$$

La probabilité de ne pas inclure l'espèce s dans l'inventaire est

$$(1 - p_s)^n$$

La probabilité d'inclure l'espèce est donc

$$1 - (1 - p_s)^n$$

Observer une espèce ν fois

La probabilité d'observer l'espèce ν fois avant de ne plus l'observer dans le reste de l'inventaire est $p_s^\nu (1 - p_s)^{n-\nu}$.

La probabilité d'observer l'espèce ν fois dans l'inventaire est obtenue en prenant en compte toutes l'ordre des observations (combinaisons) :

$$\binom{n}{\nu} p_s^\nu (1 - p_s)^{n-\nu}$$

L'espérance du nombre d'espèces observées ν fois est obtenue en sommant cette probabilité sur toutes les espèces

$$\mathbb{E}(S_n^\nu) = \binom{n}{\nu} \sum_s p_s^\nu (1 - p_s)^{n-\nu}$$

Représentation vectorielle

Introduction
 Cours

Eric Marcon

Introduction

Construction
 de l'estimateur

Application

Soit le vecteur \mathbf{v}_ν dans \mathbb{R}^S dont les coordonnées sont

$$p_s^{\nu/2} (1 - p_s)^{(n-\nu)/2}$$

Le carré de la norme du vecteur \mathbf{v}_0 est

$$\sum_s (1 - p_s)^n,$$

c'est-à-dire $\mathbb{E}(S_n^0)$, l'espérance du nombre d'espèces non observées.

(Attention : on ne connaît pas les p_s !).

Représentation vectorielle

Éric Marcon
 Université de Guyane

Eric Marcon

Introduction

Construction
 de l'estimateur

Application

Le carré de la norme du vecteur \mathbf{v}_2 est

$$\sum_s p_s^2 (1 - p_s)^{n-2} = \frac{2}{n(n-1)} \mathbb{E}(S_n^2)$$

Enfin, le produit scalaire $\langle \mathbf{v}_0, \mathbf{v}_2 \rangle$ vaut

$$\sum_s p_s (1 - p_s)^{n-1} = \frac{1}{n} \mathbb{E}(S_n^1).$$

Représentation graphique

Introduction
Classification

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Soient deux espèces telles que $p_1 = 0,4$ et $p_2 = 0,6$, et $n = 6$.

Le vecteur \mathbf{v}_0 a pour coordonnées

$$([1 - 0,4]^3; [1 - 0,6]^3) = (0.216; 0.064)$$

.

Le vecteur \mathbf{v}_2 a pour coordonnées

$$(0,4 \times [1 - 0,4]^2; 0,6 \times [1 - 0,6]^2) = (0.144; 0.096)$$

.

Représentation graphique

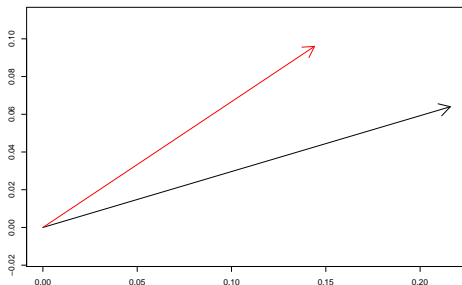
Présentation
 Class

Eric Marcon

Introduction

Construction
 de l'estimateur

Application



Le vecteur \mathbf{v}_0 dont le carré de la norme est $\mathbb{E}(S_n^0)$ est en noir.

Le vecteur \mathbf{v}_2 dont le carré de la norme est $\frac{2}{n(n-1)}\mathbb{E}(S_n^2)$ est en rouge.

Cauchy-Schwartz

Le produit scalaire est inférieur au produit des normes des vecteurs. La relation reste valide au carré:

$$\left[\sum_s p_s (1 - p_s)^{n-1} \right]^2 \leq \left[\sum_s (1 - p_s)^n \right] \left[\sum_s p_s^2 (1 - p_s)^{n-2} \right]$$

En substituant les espérances et en réarrangeant:

$$\mathbb{E}(S_n^0) \geq \frac{n-1}{n} \frac{[\mathbb{E}(S_n^1)]^2}{2\mathbb{E}(S_n^2)}$$

Estimateur

Estimateur
Classé

Eric Marcon

Introduction

Construction
de l'estimateur

Application

L'estimateur est obtenu en remplaçant les espérances par les valeurs observées:

$$\hat{S}_{Chao1} = s_n^{\neq 0} + \frac{(n-1)(s_n^1)^2}{2ns_n^2}$$

Usage

Introduction
Classification

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Il s'agit d'un estimateur minimum : l'espérance du nombre d'espèces est supérieure ou égale au nombre estimé.

L'estimation est bonne tant que l'inventaire n'est pas trop sous-échantillonné.

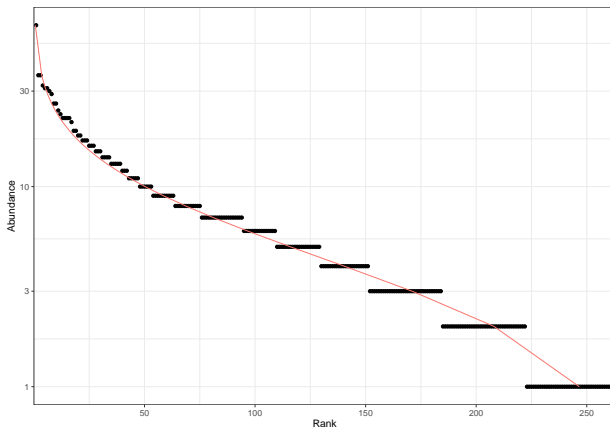
Règle empirique (Brose, Martinez, and Williams 2003) : pas plus d'un tiers des espèces observées une seule fois. Au-delà: sous estimation importante.

Section 3

Application

Simulation d'un inventaire

Communauté log-normale de 300 espèces, comparable à la forêt de Barro Colorado Island, Panama. Echantillon de 1800 arbres (3 ha de forêt).



Estimation

Estimateur
Chao1

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Nombre d'espèces observées : 263,

dont singletons : 41,

et doubletons : 38.

Estimateur Chao1 : 285 espèces.

Sous-échantillonnage

Le sous-échantillonnage

Eric Marcon

Introduction

Construction
de l'estimateur

Application

En limitant l'inventaire 600 arbres, environ 1 ha :

Estimateur Chao1 : 255 espèces, très sous-estimé.

Nombre d'espèces observées : 197, dont 75 singletons.

Il existe des estimateurs plus versatiles, comme le Jackknife
(Burnham and Overton 1978) :

```
## Jackknife 2
```

```
##           299
```

Et Paracou ?

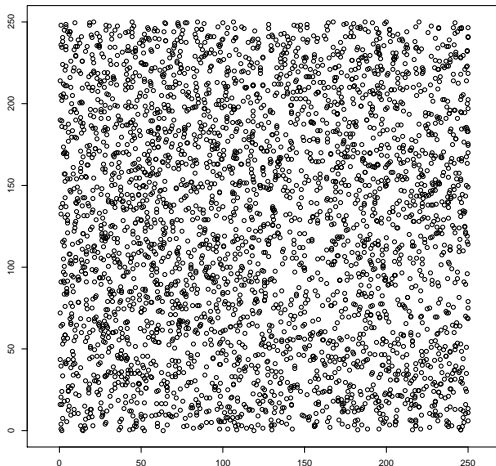
Eric Marcon

Introduction

Construction
 de l'estimateur

Application

6,25 ha inventoriés,
 environ 4000 arbres
 (mais la diversité
 est plus grande
 qu'à Panama !).
 Le nombre
 d'espèces observées
 est 334, dont 98
 singletons.
 L'estimateur Chao1
 donne 415 espèces.



Conclusion

Introduction
Chao

Eric Marcon

Introduction

Construction
de l'estimateur

Application

L'estimation de la richesse à partir d'un échantillon est possible sans faire aucune supposition sur la distribution des probabilités.

Les estimateurs de ce type sont dits "non-paramétriques". Ils sont bien supérieurs aux autres approches (estimateurs paramétriques ou extrapolation de la courbe aire-espèce).

L'estimateur de Chao est le plus connu. Il est très efficace quand l'échantillonnage est suffisant (moins d'un tiers de singletons).

Pour en savoir plus : Mesures de la biodiversité
(<https://hal-agroparistech.archives-ouvertes.fr/cel-01205813>)

References

Présentation
Chap 1

Eric Marcon

Introduction

Construction
de l'estimateur

Application

Brose, Ulrich, Neo D. Martinez, and Richard J. Williams. 2003. "Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns." *Ecology* 84 (9): 2364–77.
<https://doi.org/10.1890/02-0558>.

Burnham, K. P., and W. S. Overton. 1978. "Estimation of the Size of a Closed Population when Capture Probabilities vary Among Animals." *Biometrika* 65 (3): 625–33.

Chao, Anne. 2004. "Species richness estimation." In *Encyclopedia of Statistical Sciences*, edited by N Balakrishnan, C B Read, and B Vidakovic, 2nd ed. New York: Wiley.