Shimatani

Données :

* Paracou15.csv : les arbres de la parcelle 15, avec coordonnées
* Taxonomie.csv : les données pour faire une taxonomie
* DataLifeTraits.csv, ind\_traits\_MICE.csv : de quoi faire un arbre fonctionnel.

Le début du code calcule deux arbres : une taxonomie ***Taxonomy*** et un arbre fonctionnel ***Functional***. Il lit le jeu de points ***Paracou15*** et lui fabrique un petit frère poissonnien ***X*** avec 5 espèces pour des exemples rapides.

Le code permet de calculer la fonction de Shimatani :

# Version originale

Selon l’équation 2 de l’article.

Shimatani <- function(X, R, correction = "isotropic", Biased = TRUE)

X est un ppp (planar point process de spatstat), R un vecteur de distances dont le premier terme est 0. Le calcul utilise Kest() de spatstat en lui passant la correction des effets de bord choisie.

J’ai ajouté la correction de biais.

L’estimation de l’équation 2 donne exactement l’équation 9 si on estime le carré de l’intensité par n(n-1)/A² (Stoyan et Stoyan, 1994). L’équation 9 s’interprète facilement : l’indice de Simpson vaut 1 moins la probabilité que deux individus tirés au hasard soient de même espèce ; cette probabilité peut être estimée en calculant la proportion de paires de points monospécifiques. Shimatani restreint les paires de points aux paires de voisins (à distance inférieure à un seuil choisi). L’utilisation de K accélère les calculs (et permet la correction des effets de bord si on veut).

La définition de Simpson est avec remise, c'est-à-dire en comptant les paires d’un seul point tiré deux fois. L’option Biased = FALSE le fait, et Shimatani débiaisé égale Simpson quand r est plus grand que la fenêtre.

# Version phylogénétique

La définition peut être étendue : on prend en compte toutes les paires de voisins, et on calcule la distance moyenne entre les deux points.

A un facteur de surface près (incluant pi.r² et A) qui se simplifie :

* Le nombre total de paires de points à distance r est , comme l’a vu Shimatani dans son eq.9
* Le nombre de paires de points d’espèces et est . Pour une seule espèce (), s’entend comme la fonction K univariée appliquée au jeu de points réduit à l’espèce i.

Pour calculer la distance moyenne entre deux points tirés au hasard, il suffit de calculer la moyenne des distances (dans l’arbre) pondérée par le nombre de paires de points concernées :

Les estimateurs de densité comptent le nombre de points, quand i et j sont différents, quand i=j et au dénominateur. Les surfaces A se simplifient.

L’équation 9 de Shimatani est un cas particulier : en version neutre, vaut 1 si , 0 sinon :

Comme , il est plus simple de calculer seulement les K univariés : , ce qui donne l’équation 9 après simplification.

PhyloShimataniK <- function(X, r, Tree = NULL, Normalized = FALSE, correction = "isotropic", Biased = TRUE, CalculateRao = TRUE, ShowProgressBar = TRUE)

Tree est un arbre au format hclust qui sert à calculer une matrice de distances. Si NULL, la matrice est 1-I et on revient à la version neutre. Normalized décide de la normalisation de l’arbre pour que sa hauteur soit 1. Si FALSE, on ne touche à rien.

La valeur de PhyloShimataniK() est exactement égale à celle de PhyloShimatani() en absence de correction d’effets de bords (*correction="none"*), mais elle est beaucoup plus rapide. Elle est exactement égale aussi à Shimatani neutre si on utilise un arbre où toutes les distances valent 1 et si on utilise la même correction des effets de bord. Elle atteint exactement la valeur de Rao avec *correction="none"* et *Biased=FALSE* pour r >= la distance entre les deux points les plus éloignés.

On pourrait gagner un peu de temps de calcul en considérant que mais les estimateurs ne sont pas exactement identiques à cause des effets de bord justement, donc autant faire simple.

# Exemples

## Comparaison des fonctions

Tirage d’un semis de points de 5 espèces, le carré fait 250m de côté :

> X <- rmpoispp(lambda = c(.5, .3, .1, .05, .05)/1000, win= owin(c(0,250), c(0,250)), types=sample(names(TaxoPhy$leaves), 5))

> plot(X)

Melastomataceae\_Miconia\_tschudyoides Celastraceae\_Maytenus\_oblongata

1 2

Fabaceae\_Swartzia\_polyphylla Sapotaceae\_Pouteria\_torta

3 4

Chrysobalanaceae\_Parinari\_campestris

5



Comparaison de Shimatani sans correction d’effet de bord, débiaisé, à longue distance (1000m), et Simpson :

> Ns <- tapply(X$marks, X$marks, length)

> Ps <- Ns/sum(Ns)

> Nall <- sum(Ns)

> Simpson(Ps)

[1] 0.6010101

> Shimatani(X, c(0,1000), "none", Biased=FALSE)

[1] NaN 0.6010101

A distance 0, Shimatani vaut toujours NaN s’il n’y a pas de points superposés, mais le vecteur R doit obligatoirement commencer à 0, sinon Kest() renvoie une erreur.

Comparaison de Shimatani phylogénétique à Shimatani non débiaisé :

> Shimatani(X, c(0,1000), "none", Biased=TRUE)

[1] NaN 0.6102564

> PhyloShimatanir(1000, X)

[1] 0.6102564

> PhyloShimatanir(1000, X, Biased=FALSE)

[1] 0.6010101

La correction du biais de Shimatani phylogénétique se fait en comptant les paires monospécifiques.

Effet de la correction sur un semis de point sans structure (X, plus haut) :

> pShi <- PhyloShimatani(X, R, Taxonomy, Normalized=FALSE)

==========================================================================

> plot(x=R, y=pShi, type="l")

> pShi2 <- PhyloShimatani(X, R, Taxonomy, Biased=FALSE)

==========================================================================

> lines(x=R, y=pShi2, lty=2)

> Ns <- tapply(X$marks, X$marks, length)

> Ps <- Ns/sum(Ns)

> Nall <- sum(Ns)

> abline(h = Rao(Ps, Taxonomy), lty=3)



La version débiaisée est en pointillés. À faible distance, la diversité est plus faible parce que le point central de chaque voisinage est compté. Le trait horizontal est Rao.

Comparaison des temps de calcul (en secondes pour un jeu de 4128 points) entre la version neutre (optimisée, utilise Kest) et la version phylogénétique (utilise appynbd) :

> system.time(Shimatani(Paracou15, c(0,1000), "none", Biased=TRUE))

utilisateur système écoulé

1.45 0.00 1.45

> system.time(PhyloShimatanir(1000, Paracou15))

utilisateur système écoulé

85.32 0.70 86.22

## Paracou 15, taxonomie

Les fonctions ShimataniEnvelope et PhyloShimataniEnvelope permettent de calculer un intervalle de confiance sous l’hypothèse d’étiquetage aléatoire. C’est long : 36h pour chaque exemple.

> R <- c(seq(0, 10, 1), seq(15, 50, 5), seq(60, 120, 10))

> pShEnvTaxo <- PhyloShimataniEnvelope(Paracou15, R, NumberOfSimulations = 100, Alpha = 0.05, Tree=Taxonomy)



3 est la distance entre familles différentes, on est près de la situation où toutes les espèces seraient de familles différentes.

Il y a un excès significatif de diversité à courte distance (Janzen–Connell ?) et ensuite un déficit. Le trait rouge est la moyenne des simulations, qui converge très vite vers Rao.

## Paracou 15, arbre fonctionnel

Idem, sauf l’excès de diversité fonctionnelle à courte distance.

> pShEnvFun <- PhyloShimataniEnvelope(Paracou15, R, NumberOfSimulations = 100, Alpha = 0.05, Tree=Functional)

> plot(pShEnvFun)

