

Similitud de Comunidades biológicas

Carlos Iván Espinosa

Octubre 2016

Contents

Prefacio

La comunidad biológica se refiere a una agrupación de poblaciones de especies que se presentan juntas en el espacio y el tiempo (Begon et al. 1999). Este concepto plantea que las comunidades tienen unos límites en el espacio y el tiempo, y que estos límites están dados por la distribución de las poblaciones. Sin embargo, la distribución de las poblaciones no es homogénea y cada población responde diferente en el espacio y el tiempo. Estas respuestas diferenciadas de cada población y de la comunidad en general genera la zonación y la sucesión respectivamente.

La identificación de las diferentes formaciones biológicas en el espacio (Zonación) o definir las etapas seriales a lo largo del tiempo (sucesión) implica que tenemos la capacidad de establecer en que momento una comunidad cambia. Parece una tarea sencilla, pero realmente no lo es, ¿cuanto debería cambiar una comunidad para poder hablar de etapas seriales o zonas distintas? y ¿cómo podemos calcular ese cambio? Una de las formas de responder estas preguntas puede ser intentar cuantificar las similitudes entre localidades.

Objetivos

En este ejercicio mostramos las bases del cálculo de similitudes entre comunidades, el cual se convierte en la base de los análisis multivariantes de la comunidad. Específicamente nos interesa;

- Comprender las bases teóricas para el cálculo de similitudes de la estructura de la comunidad entre localidades.
- Utilizar herramientas de análisis para calcular índices de similitud y distancias entre comunidades



Figure 1: *Stenocercus iridicens*

Chapter 1

Introducción

1.1 Similitud basada en la distancia

“La abundancia se refiere al número de individuos de una especie en una determinada área”
— (Smith and Smith 2010)

Cuando hablamos de la composición de especies de una comunidad nos referimos al conjunto de especies que habitan una determinada localidad. Típicamente, esto incluye cierto grado de abundancia de cada especie, pero puede también ser simplemente un listado de especies en esa localidad, donde se registra la presencia o ausencia de cada especie. Ahora, imaginemos que tenemos cuatro localidades (A, B, C, D) donde recogemos los datos de densidad de dos especies; *Tabebuia billbergii* y *Geofroea spinosa*, especies características de bosques secos tropicales. Podemos introducir datos hipotéticos de abundancia para cada especie en cada una de las localidades.

```
dens <- data.frame(T.bil = c(1, 1, 2, 3), G.spi = c(21, 8, 13, 5))
row.names(dens) <- LETTERS[1:4]
dens
```

```
##   T.bil G.spi
## A     1    21
## B     1     8
## C     2    13
## D     3     5
```

Generamos un gráfico para ver cuánto se parece cada sitio (Figura ??)

```
par(mar=c(4,4,1,1), mgp=c(1,0.3,0), tcl= -0.2)
plot(dens, type = "n", cex.axis=0.8, xlim=c(0,20))
text(dens, row.names(dens), col = "blue")
```

En la figura ?? vemos que la composición de especies en el sitio A es diferente de la composición del sitio D. Es decir, la distancia entre el sitio A y D es mayor que entre los otros sitios. Lo siguiente que nos deberíamos preguntar es; ¿qué tan distantes están los dos sitios? Claramente, esto depende de la escala de medición (los valores de los ejes), y sobre cómo medimos la distancia a través del espacio multivariado (?).

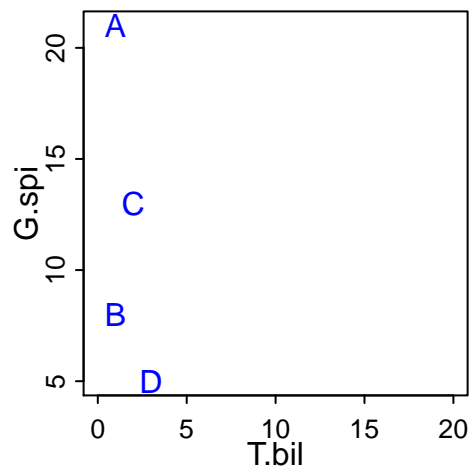


Figure 1.1: Distancias de cuatro localidades hipotéticas

Estas diferencias entre sitios son dependientes de la abundancia de cada especie. En el caso de *G. spinosa* su eje varía entre 5 y 21, mientras que para *T. billbergii* varía entre 1 y 3. Veamos ahora que sucede con las similitud si incremento la abundancia de *T. billbergii*.

```
par(mar=c(4,4,1,1), mgp=c(1,0.3,0), tcl= -0.2)
dens1 <- dens
dens1$T.bil <- dens1$T.bil*100
plot(dens1, type = "n", cex.axis=0.8, ylim=c(0,300))
text(dens1, row.names(dens1), col = "blue")
```

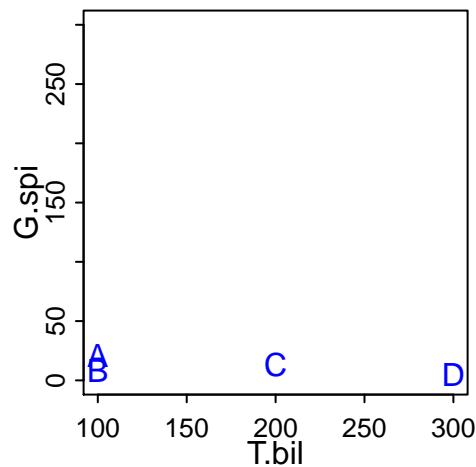


Figure 1.2: Distancias de cuatro localidades hipotéticas

Como vemos en la figura ?? las distancias entre cada uno de los sitios cambio, aunque la comunidad se

mantuvo igual. Una forma de corregir esta distorsión es calcular la densidad relativa de cada especie, de esta forma cada especie variará entre 0 y 1 (?). Cuando nos referimos a densidad relativa hablamos de la densidad de una especie con referencia a algo, en relación a la abundancia de individuos de la misma especie en otros sitios.

Para calcular la densidad relativa dividimos la abundancia de cada especie para la suma total de los individuos de las especies en esa muestra.

```
dens[,1]/sum(dens[,1])
```

```
## [1] 0.1428571 0.1428571 0.2857143 0.4285714
```

```
dens1[,1]/sum(dens1[,1])
```

```
## [1] 0.1428571 0.1428571 0.2857143 0.4285714
```

Ahora podemos ver cómo *T. billbergii* varía en su abundancia en los cuatro sitios. El sitio A y B tienen el 14% de individuos mientras que el D tiene el 42% de los individuos de esta especie. Interesantemente, no hay diferencias en las proporciones entre las dos medidas que tenemos. ¿Qué pasó con las distancias?

```
dens3 <- dens

for(i in 1:2){
  dens3[,i] <- dens[,i]/sum(dens[,i])
}
dens4 <- dens1

for(i in 1:2){
  dens4[,i] <- dens1[,i]/sum(dens1[,i])
}

par(mfcol=c(2,2), mar=c(4,4,1,1), mgp=c(1,0.3,0), tcl= -0.2)
plot(dens, type = "n", cex.axis=0.8, xlim=c(0,20), main = "Densidad")
text(dens, row.names(dens), col = "blue")

plot(dens1, type = "n", cex.axis=0.8, ylim=c(0,300), main = "Densidad 2")
text(dens1, row.names(dens1), col = "blue")

plot(dens3, type = "n", cex.axis=0.8, main = "Densidad relativa")
text(dens3, row.names(dens3), col = "blue")

plot(dens4, type = "n", cex.axis=0.8, main = "Densidad relativa2")
text(dens4, row.names(dens4), col = "blue")
```

En la figura ?? podemos apreciar que no hay diferencias entre las dos densidades cuando estoy usando la densidad relativa. Pero ¿Qué implicaciones biológicas tiene el usar las densidades relativas para calcular la distancia entre sitios?

Cuando usamos las densidades relativas lo que estamos haciendo es darles el mismo peso a todas las especies, de esta manera si yo tengo un ecosistema con una especie dominante y varias subordinadas, al usar la densidad relativa estoy eliminando esa dominancia. Es importante que tener claro este punto ya que las interpretaciones que puedo hacer con los datos de densidad y densidad relativa son distintos.

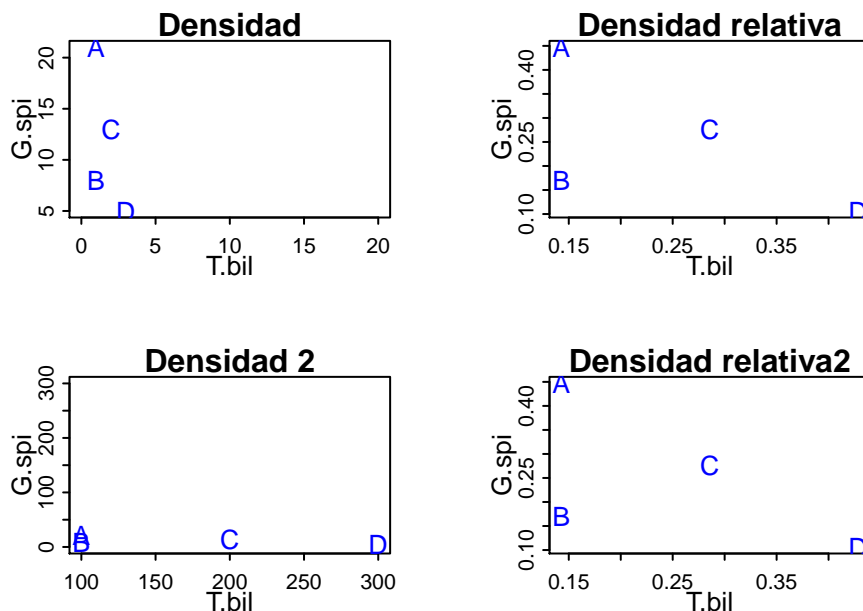


Figure 1.3: Distancias de cuatro localidades hipotéticas

Ya sea que nuestras medidas de abundancia son absoluta o relativa, nos interesa conocer cuan diferente es la comunidad de una muestra (o sitio) con relación a la otra. En el ejemplo ha sido fácil entender la diferencia entre las dos comunidades debido a que teníamos únicamente dos especies, pero con más de tres especies es complicado observar estas diferencias gráficamente. Tal vez la forma más sencilla de describir la diferencia entre los sitios es calcular las *distancias* entre cada par de sitios.

1.1.1 Distancias entre sitios

La *distancia* entre dos muestras está dada por la diferencia entre la abundancia y la composición de especies, como lo hemos visto esto genera una distancia, en el caso del ejemplo la comunidad A esta más alejada de la comunidad D que de las otras dos.

Existen muchas formas de poder calcular las distancias entre estos puntos una de las más sencillas es la distancia *Euclidiana*. La distancia euclidiana entre dos sitios es simplemente la longitud del vector que conecta los sitios y la podemos obtener como $\sqrt{x^2 + y^2}$, donde “*x*” y “*y*” son las coordenadas (x, y) de distancia entre un par de sitios.

En nuestro caso si queremos comparar B y C tenemos que la distancia en el eje *x* es la diferencia de la abundancia de *T. bilbergii* entre el sitio B y C.

```
x <- dens[2, 1] - dens[3, 1]
```

Mientras que la distancia en el eje *y* es la diferencia en la abundancia de *G. spinosa* entre el sitio B y C.

```
y <- dens[2, 2] - dens[3, 2]
```

Ahora obtenemos las distancias entre los dos sitios

```
sqrt(x^2 + y^2)
```

```
## [1] 5.09902
```

Pero como en *R* todo es sencillo podemos utilizar la función *dist*

```
dist(dens)
```

```
##           A           B           C
## B 13.000000
## C  8.062258  5.099020
## D 16.124515  3.605551  8.062258
```

Si bien este cálculo es sencillo con dos especies, si tenemos que calcular la distancia para una comunidad con más de tres especies los cálculos son tediosos y largos. Para calcular la distancia *Euclidiana* entre pares de sitios con *R* especies utilizamos la siguiente ecuación:

$$D_E = \sqrt{\sum_{i=1}^R (x_{ai} - x_{bi})^2}$$

Distancia Euclidiana

Existen otras formas de medir distancias entre dos localidades. En ecología una de las distancias más utilizada es la distancia de *Bray-Curtis*, conocida también como *Sorensen*. Esta distancia es calculada como:

$$D_{BC} = \sum_{i=1}^R \frac{(x_{ai} - x_{bi})}{(x_{ai} + x_{bi})}$$

Distancia de Bray-Curtis

La distancia *Bray-Curtis* no es más que la diferencia total en la abundancia de especies entre dos sitios, dividido para la abundancia total en cada sitio. La distancia Bray-Curtis tiende a resultar más intuitiva debido a que las especies comunes y raras tienen pesos relativamente similares, mientras que la distancia euclidiana depende en mayor medida de las especies más abundantes. Esto sucede porque las distancias euclidianas se basan en diferencias al cuadrado, mientras que Bray-Curtis utiliza diferencias absolutas. El elevar un número al cuadrado siempre amplifica la importancia de los valores más grandes. En la figura ?? se compara gráficos basados en distancias euclidianas y Bray-Curtis de los mismos datos.

Como se había comentado es virtualmente imposible representar una distancia en más de tres dimensiones (cada especie es una dimensión). Una forma sencilla de mostrar distancias para tres o más especies es crear un gráfico de dos dimensiones, intentando organizar todos los sitios para que las distancias sean aproximadamente las correctas. Está claro que esto es una aproximación nunca estas serán exactas. Una técnica que intenta crear un arreglo aproximado es escalamiento multidimensional no métrico (NMDS). Vamos a calcular las distancias para nuestra comunidad, primero vamos a añadir dos especies más a nuestra comunidad, *Ceiba trichistandra* y *Colicodendron scabridum*.

```
dens$C.tri<- c(11, 3, 7, 5)
dens$C.sca<- c(16, 0, 9, 4)
```

La función de escalamiento multidimensional no-métrico está en el paquete *vegan*. Aquí mostramos las distancias euclidianas entre sitios (Figura ??a) y las distancias de Bray-Curtis (Figura ??b).

```
library(vegan)

#Distancia Euclidiana
mdsE <- metaMDS(dens, distance = "euc", autotransform = FALSE, trace = 0)
#Distancia de Bray-Curtis
mdsB <- metaMDS(dens, distance = "bray", autotransform = FALSE, trace = 0)

par(mfcol=c(1,2), oma=c(1,1,1,1), mar=c(4,4,1,1),
    mgp=c(1,0.3,0), tcl= -0.2)

plot(mdsE, display = "sites",
     type = "text", main="a)Euclidiana",
     cex.axis= 0.7, cex.main=0.75, cex.lab=0.7)

plot(mdsB, display = "sites", type = "text",
     main="b)Bray-Curtis",
     cex.axis= 0.7, cex.main=0.75, cex.lab=0.7)
```

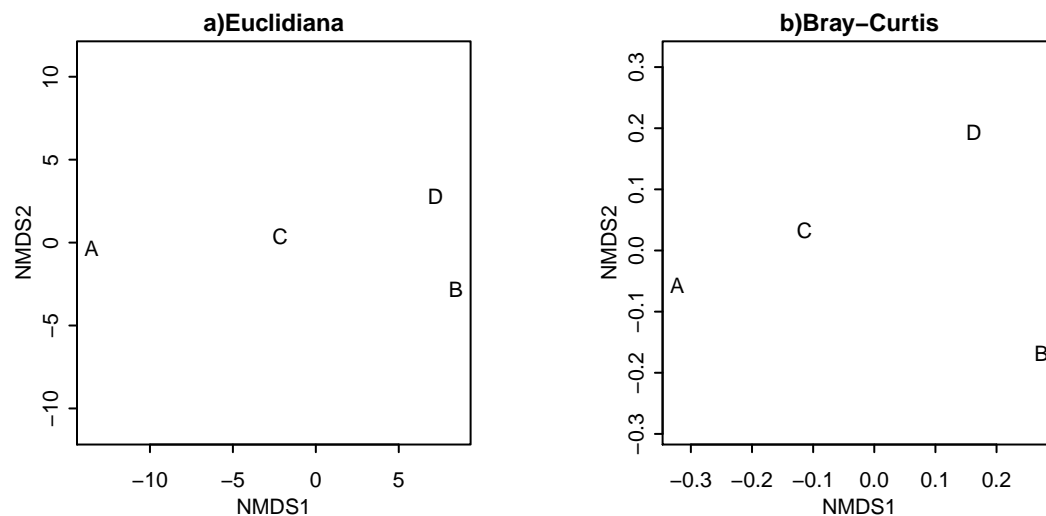


Figure 1.4: Arreglo de las parcelas en distancias multidimensionales no métricas (NMDS). Estas dos figuras muestran los mismos datos en bruto, pero las distancias euclidianas tienden a enfatizar las diferencias debidas a las especies más abundantes, mientras que Bray-Curtis no lo hace.

1.2 Similitud

Ahora que sabemos cuan distantes son los diferentes sitios, muchas veces nos podría interesar cuan similares son cada uno de los sitios a continuación se describen dos medidas de similitud; *Porcentaje de Similitud* e *Índice de Sorensen*.

El *porcentaje de similitud* puede ser simplemente la suma de los porcentajes mínimos de cada especie en la comunidad. Lo primero que debemos hacer es convertir la abundancia de cada especie a su abundancia relativa dentro de cada sitio. Para ello dividimos la abundancia de cada especie por la suma de las abundancias en cada sitio.

```
dens.RA <- t(apply(dens, 1, function(sp.abun) sp.abun/sum(sp.abun)))
dens.RA
```

```
##      T.bil      G.spi      C.tri      C.sca
## A 0.02040816 0.4285714 0.2244898 0.3265306
## B 0.08333333 0.6666667 0.2500000 0.0000000
## C 0.06451613 0.4193548 0.2258065 0.2903226
## D 0.17647059 0.2941176 0.2941176 0.2352941
```

El siguiente paso para comparar entre sitios, es encontrar el valor mínimo para cada especie entre los sitios que debemos comparar. Vamos a comparar los sitios A y B, para esto utilizamos la función `apply`, la cual nos permite encontrar el valor mínimo entre las filas 1 y 2 (sitio A y B respectivamente). Para *T. billbergi* en el sitio A la abundancia relativa es 0.02 que es menor a la abundancia en el sitio B que es de 0.08.

```
mins <- apply(dens.RA[1:2, ], 2, min)
mins
```

```
##      T.bil      G.spi      C.tri      C.sca
## 0.02040816 0.42857143 0.22448980 0.00000000
```

Finalmente para conocer el porcentaje de similitud entre los dos sitios sumamos estos valores y multiplicamos por 100.

```
sum(mins)*100
```

```
## [1] 67.34694
```

Esto significa que la comunidad A y B tienen un porcentaje de similitud del 67%.

El índice de Sorensen es la segunda medida de similitud que vamos a estudiar, este índice es medido como:

$$S_s = \frac{(2C)}{(A + B)}$$

Índice de Sorensen

Donde C es el número de especies en común entre los dos sitios, y A y B son el número de especies en cada sitio. Esto es equivalente a dividir las especies compartidas por la riqueza media.

Para calcular el índice de Sorensen entre los sitios A y B necesitamos definir el número de especies compartidas y luego la riqueza de cada uno de los dos sitios.

Definimos si alguna de las especies en uno de los sitios la abundancia no es igual a cero, eso nos dirá en qué casos se comparten especies. Finalmente, sumamos todas las especies que su abundancia es mayor a cero.