

# Análisis de diversidad: Diversidad $\alpha$

## Parte I: Fundamentos teóricos

Juan D. Vásquez-Restrepo

Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México  
(mailto:#)juanda037@outlook.com (mailto:juanda037@outlook.com)

- 1. Índices de diversidad tradicionales
- 2. Estimadores de riqueza
- 3. Curvas de acumulación de especies
- 4. Rarefacción basada en tamaño y cobertura
- 5. Índices de diversidad verdadera
- 6. Perfiles de diversidad
- 7. Referencias

En esta sección abordaremos los fundamentos teóricos detrás de los análisis de diversidad, los índices, estimadores, curvas de acumulación y rarefacción, diversidades verdaderas, y perfiles de diversidad. Posteriormente, en la parte dos (Análisis de diversidad: Parte II (<https://vr-daniel.github.io/Diversidad-en-R/U1P2.html>)) veremos cómo aplicar la teoría usando R.

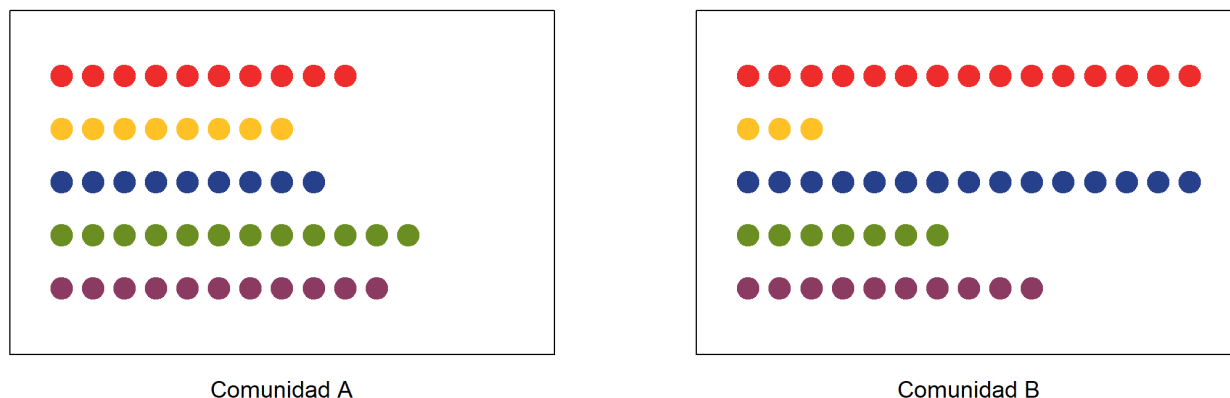
## 1. Índices de diversidad tradicionales

Los índices de diversidad son una aproximación numérica que nos permite cuantificar y comparar la diversidad en comunidades biológicas, entendiendo como diversidad no solo el número de especies o taxones diferentes sino también sus abundancias. Entre estos índices se encuentran aquellos que miden la diversidad a escalas locales o regionales, a lo que llamaremos diversidad alfa ( $\alpha$ ) y gamma ( $\gamma$ ), respectivamente, y los índices de diversidad beta ( $\beta$ ), que tienen que ver con las diferencias o el recambio de especies entre comunidades.

En esta unidad abordaremos los conceptos teóricos detrás de los índices de diversidad  $\alpha$ , tanto los tradicionales como las diversidades verdaderas. Vale la pena mencionar que, los índices que veremos a continuación son índices de diversidad taxonómica, la cual es solo una de las dimensiones de la biodiversidad. Por ejemplo, también existen índices de diversidad genética, filogenética o funcional, que de igual forma pueden medirse a diferentes escalas.

### Riqueza de especies

El índice más simple de diversidad  $\alpha$  que podemos encontrar es la riqueza de especies, es decir, el número de especies diferentes presentes en un sitio o una comunidad. Es importante tener presente que este valor no considera la abundancia, por lo que solo es informativo respecto a la composición y no a la distribución. Adicionalmente, la riqueza está influenciada por el esfuerzo de muestreo, de modo que, menores esfuerzos suelen capturar menor número de especies, sin que esto sea un indicativo real de la diversidad presente en un sitio.



**Figura 1.** Descripción gráfica de diversidad a partir de dos comunidades hipotéticas. Ambas comunidades tienen el mismo número de especies (5) y el mismo número de individuos (50), sin embargo, la Comunidad A tiene menos dominancia que la Comunidad B, por ende es más diversa.

Recordemos que una comunidad es más diversa a medida que acumula más especies diferentes y estas a su vez tienden a estar representadas por el mismo número de individuos, en otras palabras, son más equitativas y menos dominadas (**Fig. 1**). Sin embargo, es cierto que rara vez vamos a encontrar comunidades donde todas las especies estén igualmente representadas, sea porque las metodologías de muestreo tienen sesgos implícitos que no nos permiten detectar el 100% de los individuos por más rigurosas que sean, o porque la propia biología de las especies, sus interacciones con otras y las condiciones ambientales influyen de manera diferencial sus abundancias.

## Índices de Margalef y Menhinick

Una opción para considerar la abundancia de las especies dentro de un indicador de riqueza es utilizar los índices de Margalef (1958) o Menhinick (1964), los cuales relacionan el número de especies y el total de individuos en la muestra analizada. Si bien estos índices sí incluyen un valor de abundancia, tampoco son informativos respecto a cómo estas están distribuidas, dado que su intención es principalmente compensar el efecto que tiene el esfuerzo de muestreo sobre la riqueza.

$$D_{Ma} = \frac{S - 1}{\ln(N)}$$

**Índice de Margalef**

$$D_{Me} = \frac{S}{\sqrt{N}}$$

**Índice de Menhinick**

Donde: **S** es la riqueza de especies y **N** el número total de individuos.

La interpretación de ambos índices es directa, a mayor valor mayor es la diversidad. No obstante, a pesar de que el índice de Margalef y el de Menhinick ayudan a reducir la inflación de los valores de diversidad, siguen siendo altamente sensibles al número de individuos, por lo que deben ser utilizados con precaución ya que pueden llevarnos a conclusiones erradas cuando comparamos comunidades con diferentes esfuerzos de muestreo. Más adelante veremos otras aproximaciones que pueden ser más útiles en condiciones de esfuerzos diferenciales, pero por lo pronto, abordaremos la teoría detrás de los índices comúnmente más usados para entender sus pros y sus contras, porque solemos usarlos frecuentemente pero sin entender qué nos quieren decir y si son útiles para interpretar nuestros datos.

# Índices de dominancia y equitatividad

Como se mencionó anteriormente, cuando hablamos de diversidad debemos considerar no solo el número de especies diferentes sino también cómo sus abundancias están distribuidas. Para esto, existen algunos índices que involucran de manera directa las abundancias relativas de las especies, por ejemplo, los índices de Simpson, de Shannon o de Pielou. Estos, si bien no son los únicos, están entre los más ampliamente utilizados, y entender su trasfondo teórico nos será útil para posteriormente entrar a los índices de diversidad verdadera.

En cualquier caso, a la hora de aplicar estos índices no podemos ignorar la importancia que tiene el tamaño de muestra, porque es bien sabido que las inferencias basadas en muestras pequeñas tienen mayor sesgo (Smith & Grassle, 1977). Y aquí es donde tenemos que comenzar a ser críticos respecto a qué podemos y qué no podemos hacer a partir de inventarios rápidos o exprés.

## Índice de Simpson

El índice de Simpson o índice de dominancia de Simpson (Simpson, 1949), es un índice que representa la probabilidad de que dos individuos tomados al azar de una muestra correspondan a la misma entidad o especie, de modo que, es una medida de la dominancia o la inequitatividad. Este es quizás uno de índices que genera mayor confusión, no porque su cálculo o interpretación sean difíciles, sino porque tiene diferentes transformaciones que reciben diferentes nombres, las cuales se aplican e interpretan de manera diferente.

$$\lambda = \sum_{i=1}^S p_i^2$$

### Índice de Simpson

Donde:  $p_i$  es la abundancia relativa de cada especie.

Es importante mencionar que, la probabilidad calculada utilizando el índice de Simpson asume reemplazo, es decir, una vez que tomamos el primer individuo este es reemplazado inmediatamente antes de tomar el segundo. Por otro lado, para calcular la probabilidad sin reemplazo se utiliza un índice conocido como índice de Simpson corregido, el cual se deriva de una modificación del índice de original propuesta por Hunter y Gaston (1988)<sup>1</sup>.

$$D = \frac{\sum_{i=1}^S n_i(n_i - 1)}{N(N - 1)} = \frac{\sum_{i=1}^S n_i^2 - N}{N^2 - N}$$

### Índice de Simpson corregido

Donde:  $n_i$  corresponde a la abundancia de cada especie y  $N$  al total de individuos.

Sea que utilicemos el índice de Simpson original o el corregido, ambos expresan la dominancia dentro de una comunidad, siendo la dominancia lo opuesto a la diversidad. En estos casos, la interpretación es contraintuitiva, porque valores altos significan menor diversidad, y es por esto que existe un índice adicional conocido como índice de Gini-Simpson o índice de diversidad de Simpson (nótese que la diferencia está en el uso del término diversidad), el cual se expresa como el complemento de la dominancia.

$$D_{GS} = 1 - D$$

### Índice de Gini-Simpson

Donde:  $D$  corresponde a la dominancia de Simpson (original o corregida).

Si tenemos en cuenta que la dominancia es una proporción que varía entre cero y uno, lo que resta de la dominancia corresponde a la probabilidad de que dos individuos tomados al azar de nuestra muestra **NO** pertenezcan a la misma entidad, es decir, que sean diferentes. Parece algo trivial, pero son este tipo de cosas las que generan confusión, especialmente si utilizamos un software para hacer los análisis y no sabemos cuál forma del índice es la que se está calculando (por eso siempre es importante leer el manual).

## Índice de Shannon

El índice de Shannon o índice de entropía de Shannon, fue originalmente propuesto como una forma de medir la entropía en cadenas de texto (Shannon, 1948), pero posteriormente fue incorporado a la ecología ya que su utilidad no está limitada al contexto específico para el cual fue desarrollado. El problema de este índice tiene que ver con su interpretación, ya que su unidad de medida son bits<sup>2</sup> por especie. Sin embargo, para efectos prácticos tendemos a asumir que entre mayor sea su valor mayor es la diversidad.

$$H' = - \sum_{i=1}^S p_i \cdot \log(p_i)$$

### Índice de Shannon

Donde:  $p_i$  corresponde a la abundancia relativa de cada especie.

La notación original de este índice utiliza el  $\log_2$  (logaritmo base 2), pero también es común verlo expresado como  $\ln$  (logaritmo natural), lo cual no tiene efecto alguno en términos de la interpretación de los resultados, pero que sí debemos considerar a la hora de comparar valores si estos están calculados con bases diferentes.

Adicionalmente, el índice de Shannon se fundamenta en el supuesto que todas las especies están representadas en la muestra y que fueron muestreadas aleatoriamente. Por lo que debemos tener presente que este índice tiene un sesgo implícito, no solo cuando tenemos esfuerzos de muestreo que no son suficientes para capturar la totalidad de especies, sino porque muchas metodologías no son completamente aleatorias y favorecen la detectabilidad de algunos individuos. El índice de Shannon no tiene límite superior, puede tomar un valor mínimo de cero cuando solo hay una especie, o máximo como el logaritmo de la riqueza.

## Índice de Pielou

Finalmente tenemos el índice de equitatividad de Pielou, el cual es solo una transformación del índice de Shannon para que tome valores entre cero y uno, relacionando la diversidad observada y el máximo de la diversidad esperada.

$$J' = \frac{H'}{H'_{max}} = \frac{H'}{\log(S)}$$

### Índice de Pielou

Donde:  $H'$  es el índice de Shannon y  $S$  la riqueza.

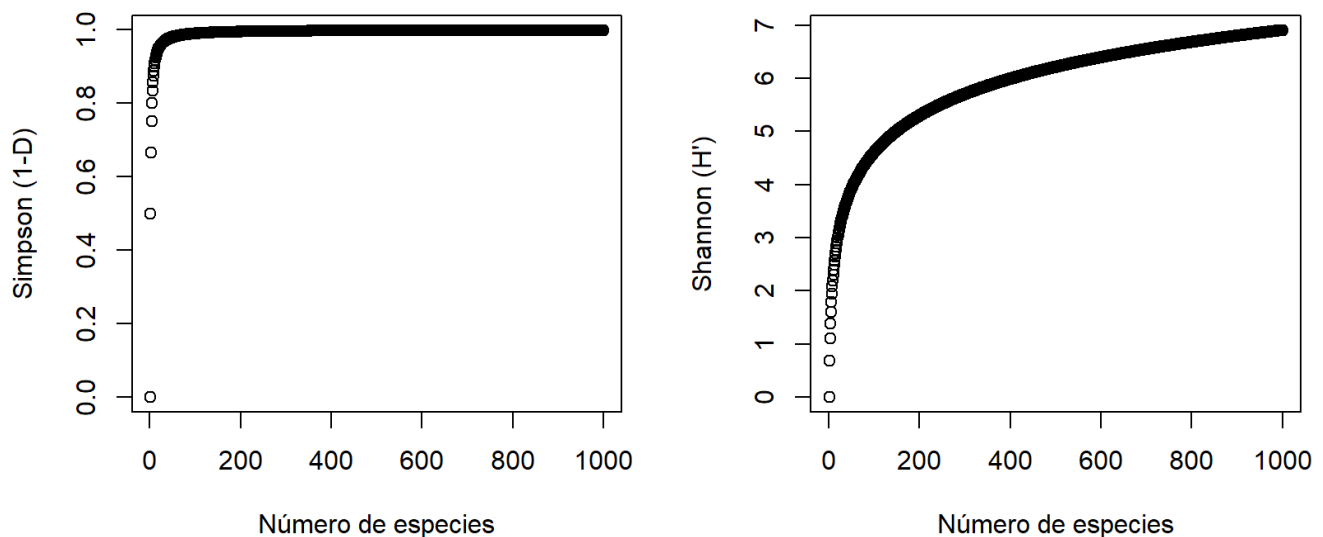
Esta transformación facilita un poco la interpretación, ya que el índice adquiere un valor de cero cuando hay una única especie, o de uno cuando todas las especies tienen la misma abundancia, por lo que nos da información sobre qué tan equitativa es una comunidad. Para el cálculo de este índice también es importante considerar que la base del logaritmo utilizada en el índice de Shannon sea la misma que utilizamos en el denominador.

# El problema de los índices tradicionales

Como ya hemos visto, los diferentes índices utilizados tradicionalmente para cuantificar y comparar la diversidad biológica se fundamentan en el uso de la riqueza y la abundancia. No obstante, algunos autores han resaltado varios problemas fundamentales subyacentes a estos (Jost, 2006; Jost, 2010; Daly *et al.*, 2018), problemas a los cuales durante mucho tiempo se les restó importancia, pero que tienen enormes implicaciones a la hora de realizar inferencias y sacar conclusiones.

Por ejemplo, además de que su interpretación puede ser difícil por tener unidades diferentes o complejas (número de especies vs. bits por especie vs. probabilidad), no varían linealmente, a lo que a partir de ahora llamaremos propiedad de duplicación. Tal como lo plantea Jost (2006), intuitivamente podemos llegar a la conclusión que una comunidad con 16 especies igualmente comunes es dos veces más diversa que una comunidad con 8 especies igualmente comunes, cosa que no sucede porque la mayoría de los índices de diversidad que usamos no varían proporcionalmente según los cambios en la riqueza.

Para demostrarlo podemos calcular algunos índices simulando una cantidad ridícula de sitios o comunidades, donde su riqueza aumente gradualmente de uno en uno y donde todas las especies estén igualmente representadas (**Fig. 2**).



**Figura 2.** Relación entre los índices de Simpson y Shannon con la riqueza.

Para ambos índices es evidente que el incremento se da de manera logarítmica antes de alcanzar una asíntota, con grandes cambios a riquezas bajas. Si para este ejemplo donde tenemos una comunidad con 1000 especies imaginásemos una catástrofe que eliminara el 80% de ellas, vemos cómo la variación en el índice de Simpson es apenas perceptible, lo que podría llevarnos a concluir que pese a la pérdida de la mayor parte de la diversidad, la comunidad no sufrió grandes cambios. Para el índice de Shannon pasa lo mismo, pero el cambio es un poco más marcado. Es por esto que el uso de estos índices ha sido cuestionado, pues no solo no son comparables entre sí, sino que nos pueden llevar a conclusiones erradas cuando queremos hacer inferencias basadas en extrapolaciones o interpolaciones. Más adelante, cuando entremos al tema de las diversidades verdaderas veremos de nuevo cómo cambia esta relación.

## 2. Estimadores de riqueza

Los estimadores de riqueza como su nombre lo dice, son métodos que nos permiten hacer aproximaciones teóricas sobre la riqueza real esperada en un sitio o comunidad, entendiéndose como riqueza esperada a la suma de la riqueza observada + la riqueza no observada. Son una forma mediante la cual podemos evaluar el esfuerzo

de muestreo.

## Estimadores paramétricos

Los estimadores paramétricos, son en principio funciones o distribuciones matemáticas conocidas que se pueden aplicar a nuestros datos con el objetivo de modelarlos teóricamente. Por el momento no vamos a detallar en la formulación matemática de los estimadores paramétricos, ya que son los menos usados por depender de modelos que por lo general no se ajustan a datos reales (asumen la distribución de las abundancias).

## Estimadores no paramétricos

A diferencia de los anteriores, los estimadores no paramétricos no asumen *a priori* ningún modelo conocido, en otras palabras, no ajustan una función existente a nuestros datos. A continuación veremos solo dos de los varios tipos de estimadores no paramétricos que existen, dado que son los más ampliamente utilizados y funcionan relativamente bien para la mayoría de conjuntos de datos.

### Chao

Los estimadores de Chao (Chao, 1984; 1987) son estimadores de riqueza que se basan en el uso del número de especies observadas y el número de especies raras o infrecuentes para calcular el número de especies no observadas. Entre estos se encuentran el Chao 1, que se utiliza cuando tenemos datos de abundancia, y el Chao 2, para datos de presencia-ausencia (también llamados datos de incidencia).

$$S_{Chao1} = S_{Obs} + \frac{f_1^2}{2f_2}$$

**Estimador de Chao 1**

$$S_{Chao2} = S_{Obs} + \frac{Q_1^2}{2Q_2}$$

**Estimador de Chao 2**

Donde:  $S_{Obs}$  es la riqueza observada y  $f_1/Q_1$  y  $f_2/Q_2$  el número singletons (especies con abundancia de uno o una única ocurrencia) y doubletons (especies con abundancias de dos o dos únicas ocurrencias).

Tanto Chao 1 como Chao 2 tienen una versión corregida (Chao, 2005), la cual ayuda a reducir el sesgo de la estimación en conjuntos de datos pequeños y permite el cálculo del estimador cuando los *doubletons* son cero (esto hace que el estimador no corregido sea indeterminado). Por defecto se recomienda utilizar las versiones corregidas, a menos que el coeficiente de variación de la abundancia o incidencia (que aprenderemos a calcular más adelante) sea mayor a 0.5, porque se ha visto que en estos casos el estimador se hace impreciso (Colwell, 2013).

$$S_{Chao1} = S_{Obs} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)}$$

**Estimador de Chao 1 corregido**

$$S_{Chao2} = S_{Obs} + \left( \frac{m-1}{m} \right) \left( \frac{Q_1(Q_1-1)}{2(Q_2+1)} \right)$$

### Estimador de Chao 2 corregido

Donde:  $S_{Obs}$  es la riqueza observada,  $f_1/Q_1$  y  $f_2/Q_2$  el número singletons (especies con abundancia de uno o una única ocurrencia) y doubletons (especies con abundancias de dos o dos únicas ocurrencias), y  $m$  el número total de muestras.

Estos estimadores suelen desempeñarse bien con la mayoría de conjuntos de datos, pero es importante tener presente a la hora de interpretarlos que Chao calcula riquezas mínimas, es decir el límite inferior del número de especies que esperaríamos encontrar.

## ACE e ICE

Los estimadores ACE (*Abundance-based Coverage Estimate*) e ICE (*Incidence-based Coverage Estimate*) (Colwell & Coddington, 1994) son una modificación de los estimadores basados en la cobertura de muestra propuestos por Chao y Lee (1992), pero que al igual que los anteriores, tratan de aproximar el número de especies esperadas a partir de los datos observados. De manera análoga a los estimadores de Chao, el ACE es la versión para datos de abundancia y el ICE para datos de presencia-ausencia. Matemáticamente el cálculo del ACE y el ICE es un poco más complejo, pero aunque existen diversos programas y paquetes que nos faciliten el trabajo, es importante saber qué es lo que están haciendo.

$$S_{ACE} = S_{Abund} + \frac{S_{Rare}}{C_{ACE}} + \frac{f_1}{C_{ACE}} \cdot \gamma_{ACE}^2$$

### Estimador ACE

$$C_{ACE} = 1 - \frac{f_1}{N_{Rare}}$$

### Estimador de cobertura de abundancia de muestra

$$\gamma_{ACE}^2 = \max \left\{ \frac{S_{Rare}}{C_{ACE}} \cdot \frac{\sum_{i=1}^{10} i(i-1)f_i}{(N_{Rare})(N_{Rare}-1)} - 1 \right\}$$

### Coefficiente de variación estimado para especies poco frecuentes

Donde:  $S_{Abund}$  es el número de especies abundantes (por defecto  $> 10$ ),  $S_{Rare}$  el número de especies raras (por defecto  $< 10$ ),  $f_1$  el número de singletons,  $N_{rare}$  el número total de individuos en las especies raras, y  $f_i$  el número de especies que tienen exactamente  $i$  individuos cuando se agrupan todas las muestras.

$$S_{ICE} = S_{Freq} + \frac{S_{Infr}}{C_{ICE}} + \frac{Q_1}{C_{ICE}} \cdot \gamma_{ICE}^2$$

### Estimador ICE

$$C_{ICE} = 1 - \frac{Q_1}{N_{Infr}}$$

### Estimador de cobertura de incidencia de muestra

$$\gamma_{ICE}^2 = \max \left\{ \frac{S_{Infr}}{C_{ICE}} \cdot \frac{N_{Infr}}{(N_{Infr} - 1)} \cdot \frac{\sum_{j=1}^{10} j(j-1)Q_j}{(N_{Infr})^2} - 1 \right\}$$

### **Coefficiente de variación estimado para especies poco frecuentes**

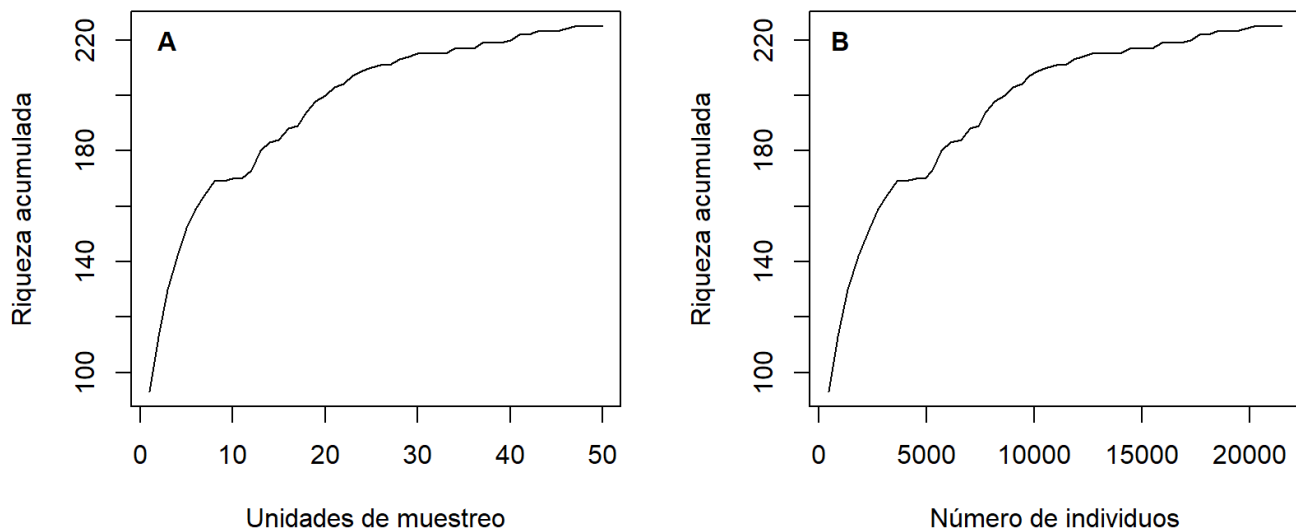
Donde:  $S_{Frec}$  es el número de especies frecuentes (por defecto > 10 muestras),  $S_{Infr}$  el número de especies infrecuentes (por defecto < 10 muestras),  $Q_1$  el número de singletons,  $N_{Infr}$  el número total de ocurrencias de las especies raras, y  $Q_j$  el número de especies que ocurren exactamente en  $j$  muestras.

Este grupo de estimadores asume como rareza o infrecuencia especies que tengan menos de 10 individuos u ocurrencias, por lo que su uso se recomienda para conjuntos de datos donde las especies suelen ser abundantes (aunque el valor por defecto se puede modificar). En este caso, los estimadores ACE e ICE también son estimadores de riquezas mínimas.

## 3. Curvas de acumulación de especies

La curva de acumulación de especies es un método que nos permite ver de manera gráfica la forma como varía la riqueza a medida que aumentamos el esfuerzo. El principal uso de las curvas de acumulación es evaluar qué tan completo es un muestreo a partir de la riqueza que un determinado esfuerzo logra capturar.

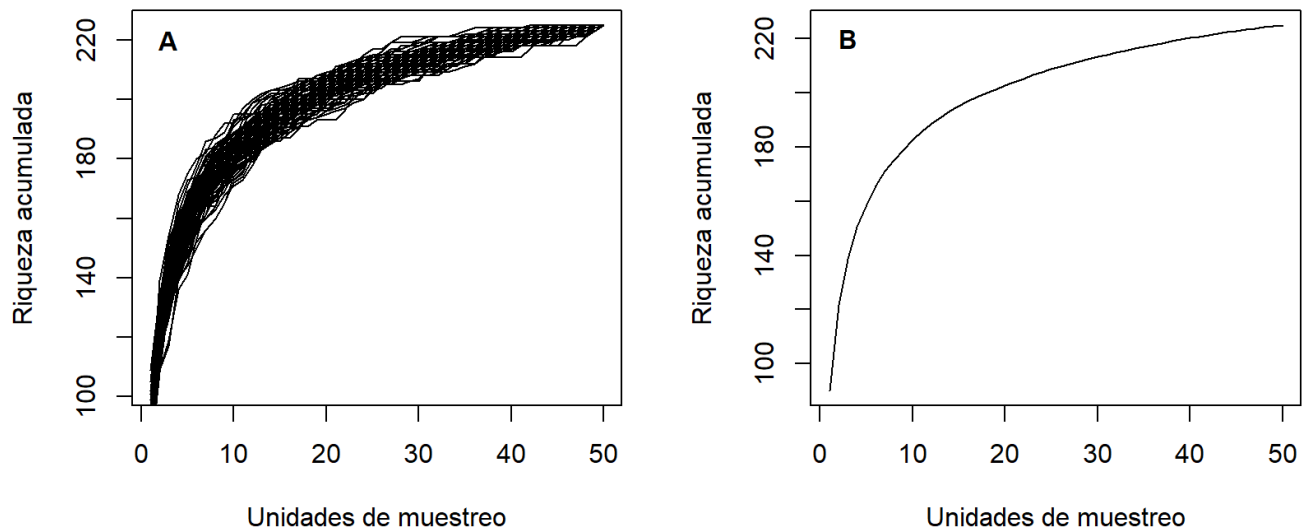
La forma más simple como podemos construir una curva de acumulación es utilizando el número acumulado de especies por unidad de muestreo (**Fig. 3A**) o el número acumulado de especies por el número de individuos (**Fig. 3B**), en el mismo orden en que estas se encuentren en nuestra matriz, lo que nos daría una curva que no está suavizada.



**Figura 3.** Curva de acumulación de especies no suavizada (sin permutación) utilizando unidades de muestreo (**A**) y número de individuos (**B**).

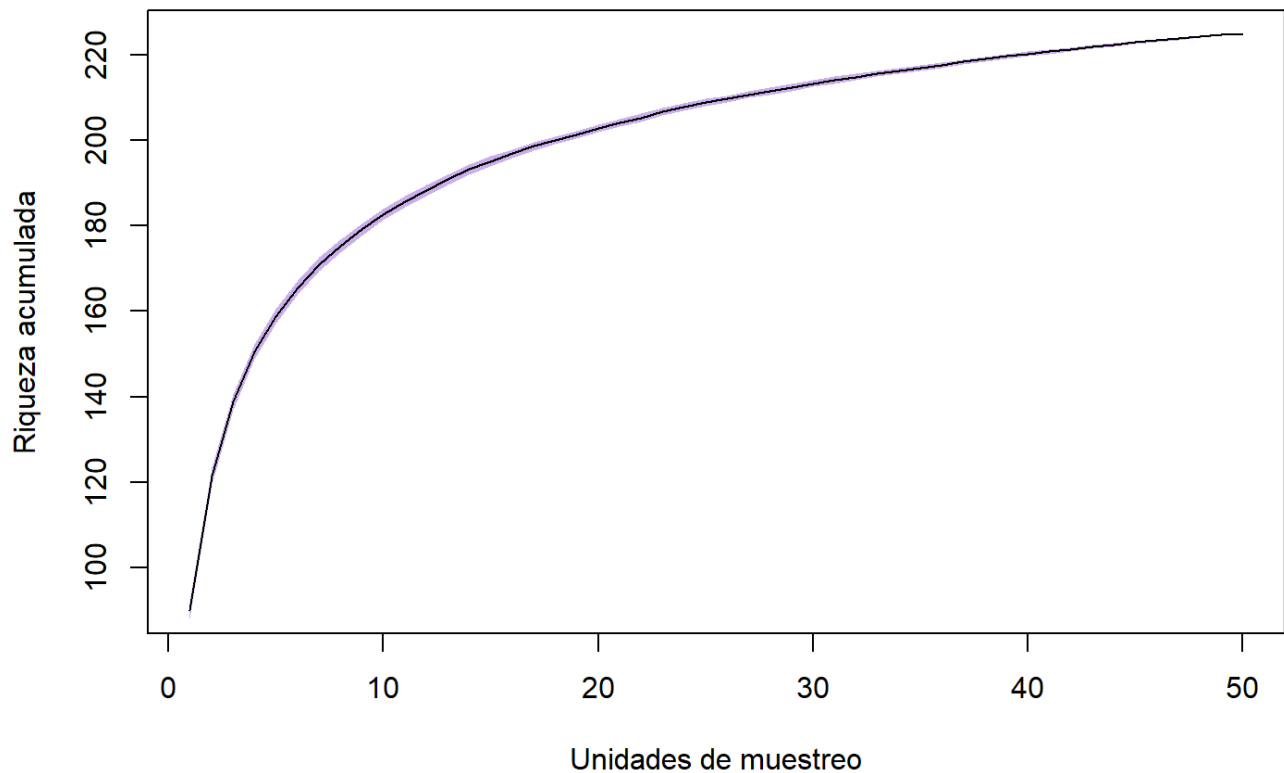
Según el orden en el que se organicen los datos, la forma de las curvas de acumulación puede variar significativamente. Es por esto que para suavizar las curvas realizamos permutaciones sobre nuestra matriz, con el objetivo de aleatorizar el orden de los sitios (**Fig. 4A**) y calcular el valor promedio de diversidad para cada unidad de muestreo (**Fig. 4B**). El número de permutaciones para construir una curva dependerá del criterio del investigador o la cantidad de datos, no hay un valor mínimo ni máximo, aunque es común utilizar 100 o 1000.





**Figura 4.** Posibles curvas de acumulación tras 100 permutaciones (A), curva de acumulación promedio de las 100 permutaciones (B).

Una de las ventajas de utilizar permutaciones es que también nos permiten calcular los intervalos de confianza para el promedio de la riqueza por sitio (**Fig. 5**), lo cual es no solo permite incorporar incertidumbre a nuestro análisis, sino que nos da información adicional sobre la comunidad. La amplitud de los intervalos está relacionada de manera indirecta con la equitatividad, y esto se da porque en comunidades más equitativas las permutaciones no generan matrices donde el número de especies por sitio al cambiar el orden varíe demasiado.

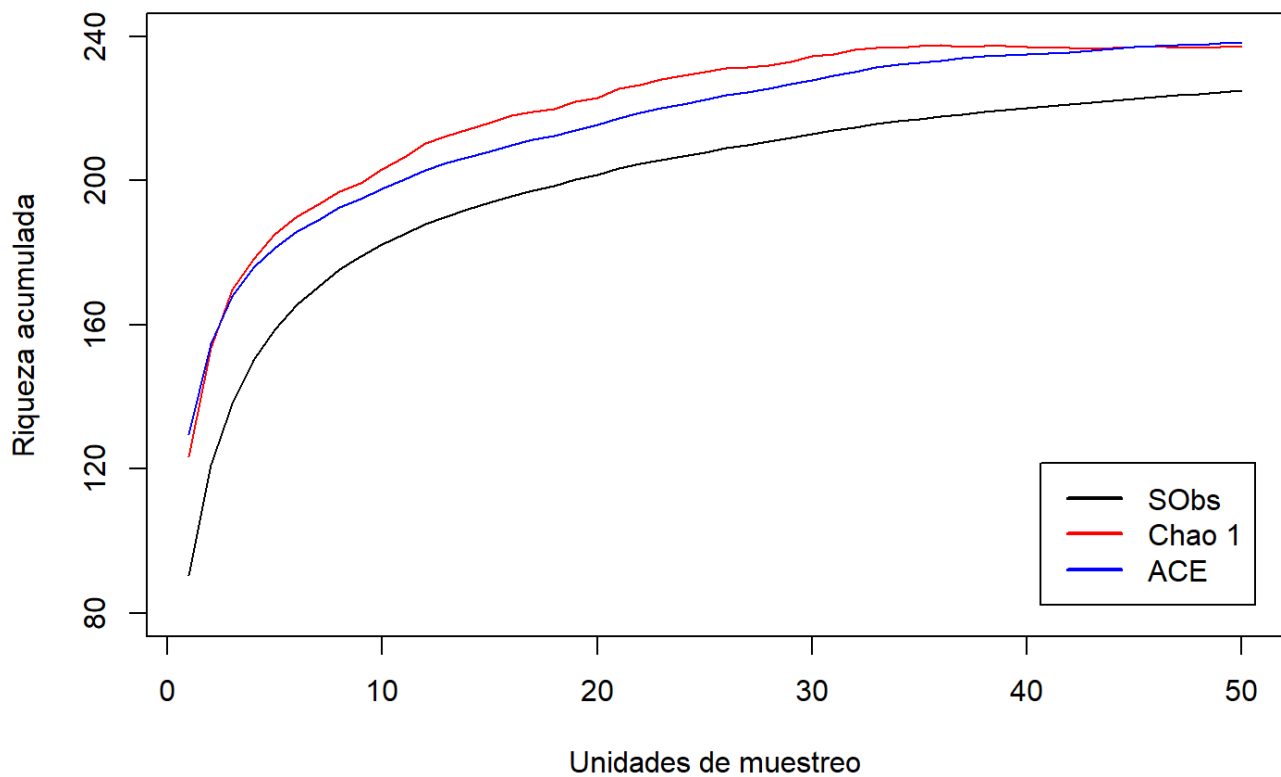


**Figura 5.** Curva de acumulación de especies suavizada con intervalos de confianza (región sombreada).

##	S	LCI	UCI
##	90.03	88.0718	91.9882
##	121.53	119.6387	123.4213
##	138.78	136.9064	140.6536
##	150.62	148.8604	152.3796
##	158.83	157.1396	160.5204

Respecto a la interpretación de las curvas de acumulación, lo que buscamos idealmente es una curva asintótica, que nos indica que la cantidad de nuevas especies que aparecen a medida que aumenta el esfuerzo de muestreo es muy baja. En otras palabras, que nuestro muestreo capturó la mayor parte de la diversidad. En caso contrario, una curva no asintótica nos dice que el esfuerzo de muestreo no ha sido suficiente. La velocidad a la que la curva alcanza la asíntota también nos dice de manera indirecta si nuestra comunidad es equitativa o no, por ejemplo, curvas que rápidamente se estabilizan son indicadoras de comunidades menos dominadas, y viceversa (Magurran, 2004). Vale la pena mencionar que, una curva asintótica no significa necesariamente que encontramos el 100% de las especies, sino que de haber más, serán especies raras.

Aquí, el uso combinado de los estimadores de riqueza y las curvas de acumulación nos permiten calcular la completitud del muestreo (en inglés *completeness*), es decir, qué tantas especies de las esperadas logramos encontrar (**Fig. 6**).



**Figura 6.** Curva de acumulación de especies observadas y estimadas.

##	N	S	Chao	ACE
##	50	225	237.2143	238.2177

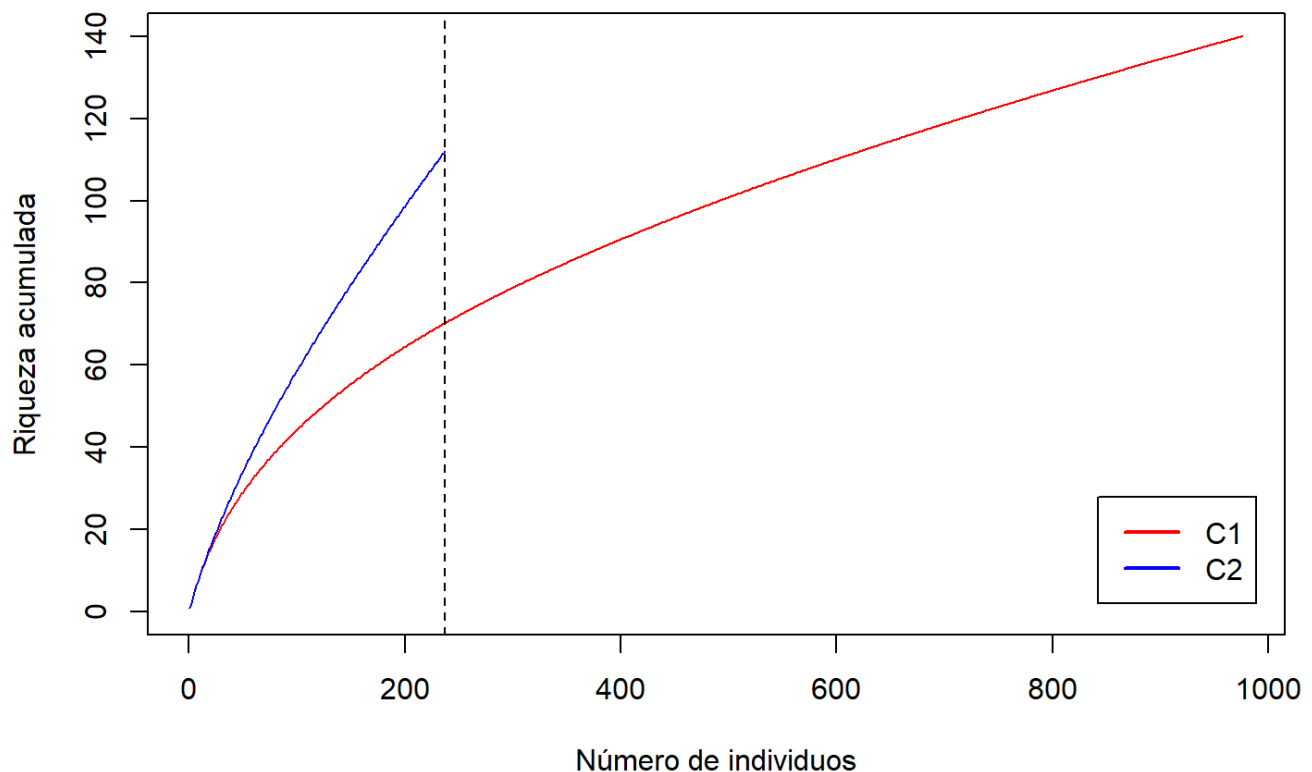
Por ejemplo, en este caso tenemos que la diversidad observada se mantiene por debajo de ambos estimadores, lo que significa que nuestro esfuerzo de muestreo aún no logra capturar el 100% de la diversidad que esperaríamos. Respecto a la diversidad acumulada observada tenemos que esta es de 225 especies, mientras

que la diversidad estimada usando Chao y ACE converge cerca de 238. Por lo tanto, podríamos decir que la completitud de nuestro muestreo es de cerca del 94.5%, en otras palabras, para el esfuerzo de muestreo realizado encontramos el 94.5% de las especies que esperaríamos encontrar. No existe un umbral mínimo para considerar si un esfuerzo de muestreo es bueno o malo, lo ideal es poder alcanzar una curva asintótica con diversidades observadas cercanas a las estimadas, pero que la completitud sea superior al 60%, 70%, 80% o 90% depende del criterio del investigador, la comunidad que estemos analizando y el area que estemos cubriendo en el espacio. Es cierto que valores demasiado bajos son un indicador de que necesitamos más esfuerzo de muestreo, pero comunidades extremadamente diversas pueden requerir demasiado esfuerzo para alcanzar la asintota.

## 4. Rarefacción basada en tamaño y cobertura

Como ya hemos visto, el esfuerzo de muestreo es un factor que debemos considerar a la hora de comparar la diversidad entre sitios o comunidades, y para esto se utiliza la rarefacción. La rarefacción es en principio una aplicación de la «Ley del Mínimo», busca estimar el número esperado de especies que tendría una comunidad respecto a aquella con el menor esfuerzo de muestreo.

Rarificar es equivalente a interpolar, es decir, utilizar los valores conocidos para obtener un valor dentro del rango de los datos. Por el contrario, extrapolar es utilizar los valores conocidos para tratar de estimar un valor fuera del rango de los datos. Existen diversos métodos para rarificar, pero los más comunes son aquellos que se basan en el uso de abundancias (en inglés *individual-based*) y que tratan de estimar el número de especies que se esperarían si todas las muestras tuvieran el mismo número mínimo común de individuos.



**Figura 7.** Curva de rarefacción basada en tamaño para dos comunidades hipotéticas con diferente esfuerzo de muestreo.

##	C1	C2
##	70.18935	112

Por ejemplo, para dos comunidades hipotéticas con diferente esfuerzo de muestreo no podemos hacer comparaciones directas entre ambas, por lo que tenemos rarificar (**Fig. 7**). En este caso, vemos que al interpolar al número mínimo de individuos común (línea punteada), la Comunidad 1 tendría cerca de 70 especies mientras que en la Comunidad 2 habrían 112, por lo que concluiríamos que la Comunidad 2 es más diversa. No obstante, aunque el esfuerzo de muestreo expresado como número de individuos es el mismo, la cobertura de la muestra no lo es, por lo que nuestra conclusión podría estar errada.

La cobertura de la muestra<sup>3</sup> (en inglés *sample coverage*) es una medida de qué tan completa es una muestra, pero a partir de la proporción del número total de individuos en una comunidad que pertenecen a las especies representadas en la muestra (Chao & Jost, 2012).

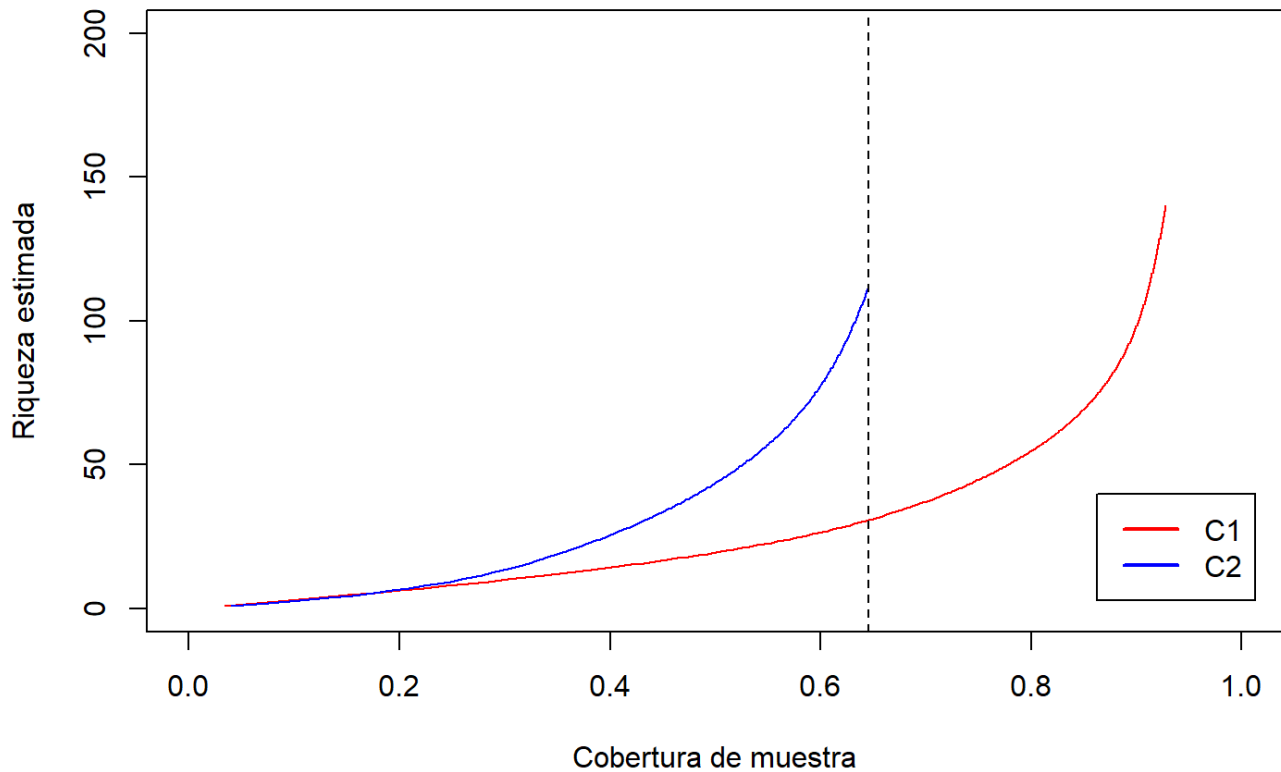
$$\hat{C}_n = 1 - \frac{f_1}{n} \left[ \frac{(n-1)f_1}{(n-1)f_1 + 2f_2} \right]$$

#### **Cobertura de la muestra**

Donde:  $n$  es el número de individuos, y  $f_1$  y  $f_2$  son el número *singletons* (especies con abundancia de uno) y *doubletons* (especies con abundancias de dos o dos únicas ocurrencias).

En palabras más simples, la cobertura nos indica qué tanto muestreamos una comunidad respecto a lo que esperaríamos. Es en parte similar a lo que hacemos al calcular la completitud con base en las especies observadas y los estimadores de riqueza, pero en este caso no decimos que encontramos un porcentaje determinado de las especies que esperábamos encontrar, sino que muestreamos un porcentaje determinado de la comunidad. La cobertura es una aproximación más adecuada para comparar entre comunidades, porque dependiendo del tipo de ecosistema o método de muestreo, el mismo número de individuos no necesariamente implica el mismo esfuerzo. Adicionalmente, si restamos la cobertura de muestra a uno, obtendremos el denominado déficit de muestreo, que es equivalente a la proporción de la comunidad que pertenece a especies no muestreadas. Este déficit también indica la probabilidad de que una nueva especie previamente no muestreada sea encontrada si la muestra aumentara en un individuo (Chao & Jost, 2012), por lo que el concepto de cobertura es mucho más informativo.

Si calculamos entonces la cobertura para nuestras dos comunidades tendríamos que, para la Comunidad 1 su cobertura es de 93% mientras que para la Comunidad 2 es de 65%. Por lo tanto, al tener coberturas diferentes no podemos asegurar cuál es más diversa, siendo necesario rarificar de nuevo pero esta vez con base en la cobertura y no en el tamaño (**Fig. 8**).



**Figura 8.** Curva de rarefacción basada en cobertura para dos comunidades hipotéticas con diferente esfuerzo de muestreo.

##	C1	C2
##	31.25367	116.0405

Una vez rarificamos con base en la cobertura de muestra podemos ver que, si bien la conclusión no cambió, los valores de riqueza sí. Usando la rarefacción basada en tamaño teníamos que para la Comunidad 1 se esperarían cerca de 70 especies y para la Comunidad 2 unas 112, pero al rarificar usando la cobertura tenemos que el número esperado si las hubiesemos muestreado en un 65% es de 31 para la Comunidad 1 y de 116 para la Comunidad 2. Dependiendo de nuestros datos, es posible que la conclusión cambie, de modo que lo ideal es siempre interpolar o extrapolar con base en coberturas. Para el caso de las extrapolaciones, se recomienda no hacerlo a más de dos veces el tamaño muestral, ya que más allá la incertidumbre se hace demasiado alta (Chao & Jost, 2012).

Para las curvas de rarefacción también es posible el cálculo de intervalos de confianza, los cuales cuando comparamos entre comunidades nos dan información adicional sobre la significancia estadística de la diferencia de medias, pero por simplicidad no abordaremos esa discusión en esta sección (lo veremos en la parte práctica).

Conceptual y operativamente la rarefacción puede generar algunas confusiones, dado que es un término que vemos aplicado en los estimadores de riqueza para una comunidad y como metodología para comparar la riqueza de varias comunidades. Lo que debemos tener presente es que, aunque en ambos casos se emplea el mismo método (la rarefacción), se utiliza con fines diferentes. En el primer caso, se utiliza para estimar el número de especies dentro de una comunidad si todas las muestras tuviesen un número mínimo de individuos común. Mientras que en el segundo, se utiliza para comparar la diversidad de dos o más comunidades diferentes, si todas tuvieran un mínimo esfuerzo de muestreo.

## 5. Índices de diversidad verdadera

Los índices de diversidad verdadera no son más que una transformación de los índices de diversidad tradicionales, pero que suponen algunas ventajas interpretativas y metodológicas. Si bien la matemática detrás de estos índices se conoce desde la década de los 70s (Hill, 1973), no tuvieron gran acogida hasta que Lou Jost y Anne Chao (Jost, 2006; Chao *et al.*, 2010; Jost, 2010; Chao & Jost, 2012; Chao *et al.*, 2014) los retomaron como parte de un marco teórico unificado para el análisis de diversidad en comunidades biológicas. Los números de Hill, como también se les conoce a los índices de diversidad verdadera, permiten expresar los índices tradicionales en las mismas unidades (número efectivo de especies), facilitando así su interpretación, además cumplen con la propiedad de duplicación, lo que hace que puedan ser extrapolables.

El concepto de número efectivo de especies (ENS, por sus siglas en inglés) es de suma importancia para los análisis de diversidad usando diversidades verdaderas, y no es más que el número de especies igualmente representadas que debería tener una comunidad para dar lugar a un valor particular de un índice, o el número de especies igualmente representadas que son necesarias para que la abundancia proporcional promedio en una comunidad de referencia sea equivalente a la del conjunto de datos original. En términos prácticos, el ENS puede interpretarse como el valor de diversidad verdadera de una comunidad (Jost, 2006). Es importante tener presente que el ENS varía en función de la distribución de las abundancias, por lo que tiende a ser menor a medida que aumenta la dominancia.

Por ejemplo, imaginemos dos comunidades hipotéticas con ocho especies cada una y no equitativas, es decir con diferentes abundancias por especie. Ambas comunidades tienen el mismo número total de individuos, con la única diferencia que la primera está menos dominada.

	A	B	C	D	E	F	G	H
Comunidad 1	40	35	25	60	12	20	45	25
Comunidad 2	88	12	5	32	15	2	46	62

Para la primera comunidad su ENS de orden uno (que aprenderemos a calcular más adelante) es de 7.25, lo que quiere decir que esta comunidad sería equivalente a una con 7.25 especies comunes igualmente abundantes. Por el contrario, para la segunda, que recordemos es menos equitativa, su ENS de orden uno es de 5.41.

## Números de Hill

El cálculo de la diversidad verdadera utiliza una fórmula generalizada para los diferentes índices, conocidos en este contexto como ordenes o números de Hill.

$${}^qD = \left( \sum_{i=1}^S p_i^q \right)^{1/1-q}$$

### **Formula general de los números de Hill**

Donde:  $q$  es el orden de diversidad que nos interesa y  $p_i$  es la abundancia relativa de cada especie.

En términos operacionales, es posible calcular el  ${}^qD$  para cualquier número entero o decimal, positivo o negativo. No obstante, los únicos ordenes con interpretación biológica son  $q = 0$ ,  $q = 1$  y  $q = 2$ . A continuación, veremos a partir de la simplificación matemática de la fórmula general a qué equivalen cada uno de los tres ordenes.

$${}^0D = \left( \sum_{i=1}^S p_i^0 \right)^{1/1-0} = \sum_{i=1}^S 1 = S$$

### ***Diversidad verdadera de orden 0 y su simplificación***

Para el caso de  $q = 0$ , el término  $p_i^0$  se hace uno, porque las propiedades de las potencias nos dicen que todo número diferente de cero elevado a cero es igual a uno. Así mismo, al resolver la operación del exponente general este también se hace uno, de modo que la expresión es equivalente a sumar uno tantas veces como el número de especies. Es por esta razón que este orden es equivalente a la riqueza.

$${}^1D = \left( \sum_{i=1}^S p_i^1 \right)^{1/1-1} = \exp \left( - \sum_{i=1}^S p_i \cdot \log(p_i) \right) = \exp(H')$$

### ***Diversidad verdadera de orden 1 y su simplificación***

Ahora, el primer orden es un caso especial, porque cuando  $q = 1$  el exponente general se hace indeterminado (división por cero) y por ende no es posible resolver la ecuación, sin embargo, su límite sí existe, es por esto que se calcula aproximando el orden a un valor muy cercano a uno. Esta aproximación no afecta realmente el valor del índice, ya que luego de simplificar la expresión obtenemos que el orden uno es equivalente al exponencial del índice de Shannon (véase la Eq. 1.6), el cual se puede calcular perfectamente sin el uso de la teoría de límites.

$${}^2D = \left( \sum_{i=1}^S p_i^2 \right)^{1/1-2} = \frac{1}{\left( \sum_{i=1}^S p_i^2 \right)} = \frac{1}{\lambda}$$

### ***Diversidad verdadera de orden 2 y su simplificación***

Finalmente, para  $q = 2$  el exponente general se hace uno negativo, si acudimos de nuevo a las propiedades de la potencia sabemos que, un número elevado a una potencia negativa es equivalente a su recíproco elevado al opuesto de la potencia, por lo que podemos simplificar la expresión como su inverso, obteniendo así lo equivalente a calcular el inverso del índice de Simpson (véase la Eq. 1.3).

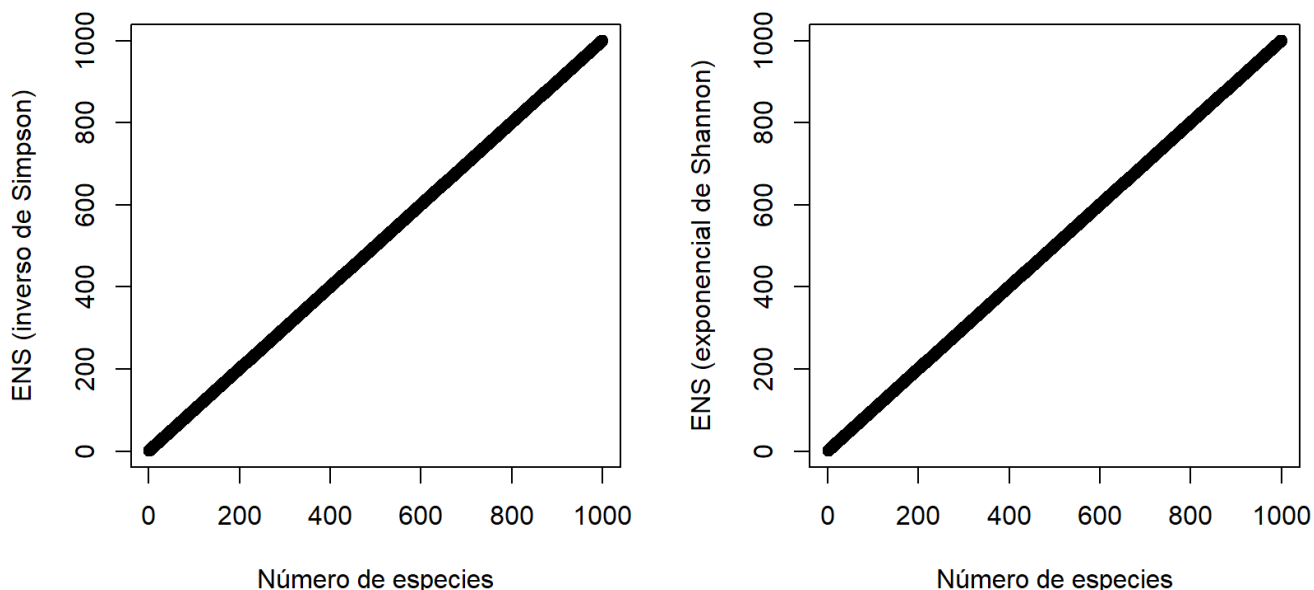
Dado que los índices de diversidad verdadera estandarizan las unidades en términos de especies, interpretarlos es mucho más simple. Por ejemplo, ya sabemos que el orden cero es equivalente a la riqueza, es decir, es el número de especies totales presentes en la muestra. El orden uno, por su parte, puede interpretarse como el número de especies comunes, y también nos da información indirecta sobre la equitatividad en la medida que se acerca al valor del orden cero. Además, la diferencia entre  ${}^0D$  y  ${}^1D$  nos da el número de especies raras. Por otro lado, el orden dos se puede interpretar como el número de especies dominantes, de modo que, entre más especies dominantes haya, la comunidad estará menos dominada, esta idea puede sonar un poco confusa, pero si todas las especies dominan por igual es porque la comunidad es totalmente equitativa. Finalmente, la relación entre  ${}^2D$  y  ${}^0D$  nos da una proporción de dominancia o factor de inequitatividad (Jost, 2010).

Es importante mencionar también que, los ordenes de diversidad indican además su sensibilidad a las especies raras o comunes (Jost, 2006; 2010), ya que el cálculo de las diversidades verdaderas da peso a las especies según sus abundancias relativas. En el caso de la diversidad de orden cero esta es insensible a la abundancia, mientras que las demás diversidades, menores o mayores que uno favorecen a las especies raras o comunes, respectivamente. Por lo tanto, el único orden que pesa a las especies según su frecuencia pero sin favorecer ni a las raras ni a las comunes, es la diversidad de orden uno, y por esta razón anteriormente utilizamos este orden para hablar de la equivalencia de nuestras comunidades hipotéticas en términos del ENS.

Como vemos y como veremos, las diversidades verdaderas no solo se interpretan de manera más simple y son matemáticamente más correctas, sino que además son sumamente informativas.

## Propiedad de duplicación

Si bien Jost (2006; 2010) no definió de manera explícita algo denominado propiedad de duplicación<sup>4</sup>, sí hizo referencia directa a la relación monotónica lineal que resulta de la transformación de los índices de diversidad tradicionales en índices de diversidad verdadera. Ahora, ya que en este punto tenemos las bases conceptuales del ENS, podemos graficar de nuevo la relación entre los índices de Simpson y Shannon (transformados a diversidades verdaderas) y la riqueza a medida que esta aumenta. Para ello, vamos a simular de nuevo 1000 comunidades cuya diversidad aumenta progresivamente de uno en uno (**Fig. 9**).



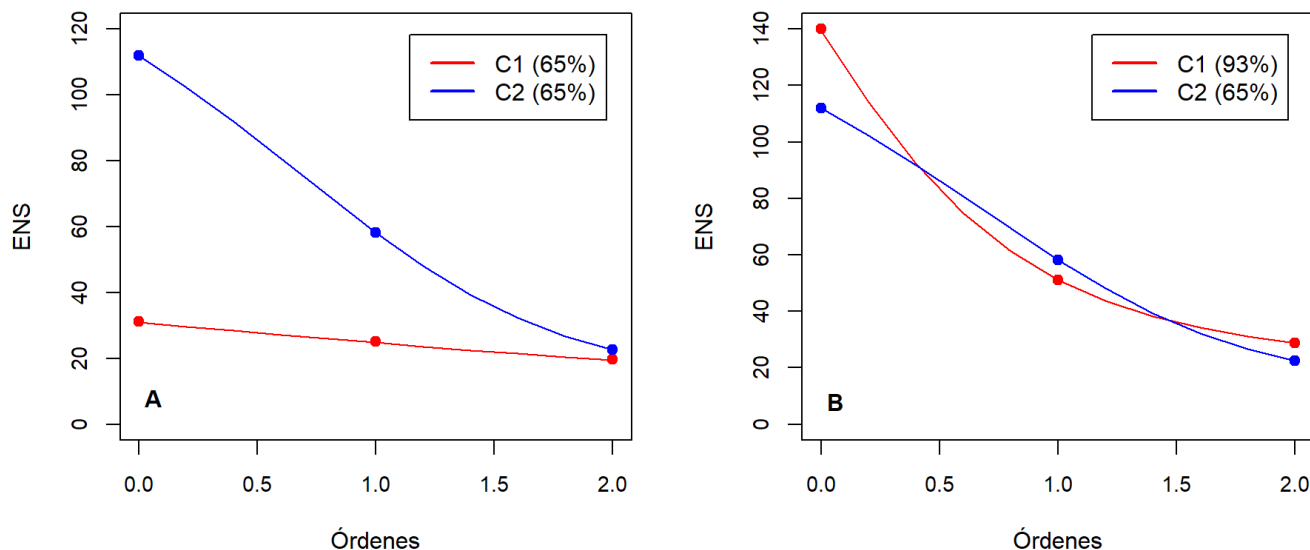
**Figura 9.** Relación entre el número efectivo de especies (ENS) y la riqueza para diversidades verdaderas.

Aquí podemos ver claramente como la relación que anteriormente no era lineal (**Fig. 2**) ahora sí lo es, lo que nos permite extrapolar de manera más simple. Por ejemplo, si el ENS de orden cero para una comunidad es 12, podemos concluir que otra con un ENS de 18 es 0.5 veces o 50% más diversa, o que una con 24 es dos veces más, y así sucesivamente. De igual forma, para los ordenes uno y dos podemos concluir cuántas veces una comunidad tiene más o menos especies comunes, raras o dominantes, así como cuántas veces son más o menos equitativas.

## 6. Perfiles de diversidad

Los perfiles de diversidad son una herramienta gráfica que nos ayuda a comparar la diversidad entre diferentes comunidades a partir de las diversidades verdaderas. A grandes rasgos, los perfiles nos permiten evaluar visualmente cosas como la riqueza y la equitatividad, siendo muy útiles si recordamos que diversidad no es solo cuántas especies sino también cuántos individuos y cómo están distribuidos.





**Figura 10.** Perfiles de diversidad para dos comunidades hipotéticas, con rarefacción basada en cobertura (A), sin rarefacción (B).

Construir un perfil de diversidad es bastante sencillo una vez tenemos calculadas las diversidades verdaderas, pero no podemos olvidar estandarizar el esfuerzo de muestreo (**Fig. 10A**), sea rareficanado o extrapolando, porque de lo contrario podríamos llegar a conclusiones erradas (**Fig. 10B**). La interpretación de los perfiles se basa en los órdenes de diversidad, de la misma manera que ya hemos mencionado, el orden cero indica la riqueza, el uno las especies comunes (en terminos de abundancia, no compartidas), y el dos las especies dominantes. Los órdenes intermedios no tienen una interpretación directa, pero en este contexto se calculan para ayudar a suavizar las curvas. Así, para estas dos comunidades hipotéticas tenemos que la Comunidad 2 tiene mayor riqueza que la Comunidad 1, pero la Comunidad 1 tiene un menor número de especies raras ( ${}^0D - {}^1D$ ) y menor dominancia ( ${}^2D / {}^0D$ ) al mismo esfuerzo.

## 7. Referencias

- Chao, A. (1984). Non-parametric estimation of the classes in a population. *Scandinavian Journal of Statistics*, 11(4), 265–270.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43(4), 783. <https://doi.org/10.2307/2531532> (<https://doi.org/10.2307/2531532>)
- Chao, A. (2005). Species estimation and applications. In: Kotz, S., Read, C. B., Balakrishnan, N. & Vidakovic, B. (Eds.). *Encyclopedia of statistical sciences*. 2nd. Vol. 12. pp. 7907–7916.
- Chao, A., Chiu, C. H. & Jost, L. (2010). Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3599–3609. <https://doi.org/10.1098/rstb.2010.0272> (<https://doi.org/10.1098/rstb.2010.0272>)
- Chao, A., Chiu, C. H. & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 297–324. <https://doi.org/10.1146/annurev-ecolsys-120213-091540> (<https://doi.org/10.1146/annurev-ecolsys-120213-091540>)
- Chao, A. & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, 93(12), 2533–2547. <https://doi.org/10.1890/11-1952.1> (<https://doi.org/10.1890/11-1952.1>)
- Chao, A. & Lee, S. M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, 87(417), 210–217. <https://doi.org/10.1080/01621459.1992.10475194> (<https://doi.org/10.1080/01621459.1992.10475194>)

- Colwell, R. K. (2013). *EstimateS 9.1.0 User's Guide*. EstimateS. Fecha de acceso: 24 de diciembre, 2021.  
Disponible en:  
<http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/EstimateSUsersGuide.htm>  
(<http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/EstimateSUsersGuide.htm>)
- Colwell, R. K. & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 345(1311), 101–118.  
<https://doi.org/10.1098/rstb.1994.0091> (<https://doi.org/10.1098/rstb.1994.0091>)
- Daly, A., Baetens, J. & Baets, B. D. (2018). Ecological diversity: Measuring the unmeasurable. *Mathematics*, 6(7), 119. <https://doi.org/10.3390/math6070119> (<https://doi.org/10.3390/math6070119>)
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432.  
<https://doi.org/10.2307/1934352> (<https://doi.org/10.2307/1934352>)
- Hunter, P. R. & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: An application of Simpsons index of diversity. *Journal of Clinical Microbiology*, 26(11), 2465–2466.  
<https://doi.org/10.1128/jcm.26.11.2465-2466.1988> (<https://doi.org/10.1128/jcm.26.11.2465-2466.1988>)
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x>  
(<https://doi.org/10.1111/j.2006.0030-1299.14714.x>)
- Jost, L. (2010). The relation between evenness and diversity. *Diversity*, 2(2), 207–232.  
<https://doi.org/10.3390/d2020207> (<https://doi.org/10.3390/d2020207>)
- Magurran, A. E. (2004). *Measuring biological diversity*. Blackwell: Oxford, United Kingdom.
- Margalef, R. (1958). Information theory in ecology. *General Systems*, 3, 36–71.
- Menhinick, E. F. (1964). A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology*, 45(4), 859–861. <https://doi.org/10.2307/1934933> (<https://doi.org/10.2307/1934933>)
- Rajaram, R., Castellani, B. & Wilson, A. N. (2017). Advancing Shannon entropy for measuring diversity in systems. *Complexity*, 2017, 1–10. <https://doi.org/10.1155/2017/8715605> (<https://doi.org/10.1155/2017/8715605>)
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.  
<https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (<https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>)
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(4148), 688–688. <https://doi.org/10.1038/163688a0>  
(<https://doi.org/10.1038/163688a0>)
- Smith, W. & Grassle, J. F. (1977). Sampling properties of a family of diversity measures. *Biometrics*, 33(2), 283.  
<https://doi.org/10.2307/2529778> (<https://doi.org/10.2307/2529778>)

1. El índice de Hunter y Gaston se propuso inicialmente para medir la capacidad de discriminación en sistemas de tipificado de bacterias. La notación original del índice incluye la substracción a uno, por lo que mide la probabilidad de que dos individuos tomados de manera aleatoria correspondan a entidades diferentes. Personalmente me fue imposible rastrear quién y por qué se comenzó a utilizar como índice solo el término que cuantifica la dominancia (sin la substracción).↵
2. Un bit es una unidad de medida ampliamente utilizada para cuantificar unidades de entropía en la información. En la Teoría de la Información, la entropía es el nivel promedio de incertidumbre que tienen los posibles resultados de un evento, o la información que un evento nos da sobre sus posibles resultados. Por ejemplo, la entropía de lanzar una moneda es de 1 bit, porque independientemente de la probabilidad de que caiga una cara o la otra, el resultado obtenido nos da información sobre el único otro posible resultado. En caso de lanzar dos monedas la entropía es de 2 bits, porque el resultado obtenido sea cual sea (p.e. XO, que para este caso es lo mismo que OX), nos da información sobre los únicos otros dos posibles resultados que se pueden obtener (OO o XX). Ahora, en términos biológicos esta unidad se hace mucho más difícil de interpretar, pero sistemas con mayor entropía indican mayor diversidad (Rajaram *et al.*, 2017).↵

3. El concepto de cobertura de muestra fue desarrollado en el ámbito de la Criptografía durante la Segunda Guerra Mundial, y se lo debemos a Alan Turing, el padre de la computación moderna.↵
4. Este término aparece posteriormente como un axioma de los índices de diversidad en el trabajo de Daly *et al.* (2018).↵