

ESTIMACIÓN DEL ÍNDICE DE DIVERSIDAD DE SIMPSON EN m SITIOS DE MUESTREO*

Carlos N. Bouza¹, Universidad de La Habana, Cuba

Daniel Covarrubias², Universidad Autónoma de Guerrero, México

RESUMEN

Se estudia la estimación del índice de Simpson para un sitio de muestreo. Las propiedades del estimador son desarrolladas condicionadas al número de individuos observados. Se construyen estimadores para m muestras considerando dos posibles estructuras. El diseño utilizado es el muestreo simple aleatorio con reemplazo. Los estimadores desarrollados son insesgados. Sus varianzas así como sus correspondientes estimadores son obtenidos. El comportamiento de los estimadores propuestos es estudiado utilizando los datos de una investigación sobre Ordenamiento Ecológico Territorial llevado a cabo en Estado de Guerrero, México. Experimentos de Monte Carlo son desarrollados.

ABSTRACT

The estimation of the index of Simpson is studied for a sampling plot. The properties of the estimator are developed for a fixed number of observed individuals. Estimators are built for m samples considering two possible structures. The design used is Simple Random Sampling with replacement. The developed estimators is unbiased. Their variances as well as their corresponding estimators are obtained. The behavior of the proposed estimators is studied using the data of an investigation Ordering of Territorial Ecological Classification carried out in the State of Guerrero, Mexico. Experiments of Monte Carlo are developed.

Key words: Biodiversity, unbiasedness, simple random sampling, Monte Carlo experiments.

MSC: 62D05

1. INTRODUCCIÓN

Una visión hacia las políticas actuales de los países permite medir el nivel de importancia que a cobrado la ecología en la actualidad. El creciente interés por el tema en el mundo y el apoyo de diferentes organismos internacionales manifiestan una verdadera preocupación por conocer y preservar nuestros recursos naturales.

De acuerdo con (Smith, 2002) a lo largo de las últimas cuatro décadas se han registrado avances importantes en los estudios ecológicos y particularmente de estadística en la ecología, ejemplos de ello son las contribuciones al estudio de las distribuciones espaciales de organismos y los diseños de estudios ecológicos.

*Esta investigación ha recibido el soporte de PROMEP (México) y TWAS (UNESCO).

¹E-mail: bouza@matcom.uh.cu

²dte_fer@yahoo.com.mx

La implementación de políticas que incorporan caracterizaciones del estado que guarda una comunidad ecológica, hacen uso de indicadores propuestos varias décadas antes, los *índices de riqueza, equilibrio y diversidad*.

El concepto de diversidad ha sido ampliamente discutido como se puede ver en la literatura especializada. Un aporte a esta discusión la hacen Patil y Taille (1982) quienes consideran a la *diversidad* como una propiedad medible intrínseca de la comunidad, definida como el promedio de la rareza de especie.

El trabajo se encuentra dividido en cuatro secciones, la segunda describe el uso de los índices de diversidad además de presentar una caracterización de los mismos. La tercera sección muestra la metodología propuesta para encontrar los estimadores de los índices de diversidad para la clase aditiva, se presenta el caso Simpson (1949). Por último el comportamiento de la aplicación de los índices de Simpson es estudiado haciendo experimentos de Monte Carlo con los datos obtenidos por el Centro de Investigación Área de Ciencias Naturales de la Universidad Autónoma de Guerrero, México, en una investigación Financiada por el SIBEJ México, en el Estado de Guerrero.

2. ÍNDICES DE DIVERSIDAD

Distintos autores (Krebs 1999, Gove **et al.**, 1994, Ludwig y Reynolds 1988, Pielou 1977) coinciden en señalar que el índice de diversidad está formado por dos componentes: el *número de especies* o riqueza de especie y la *abundancia* o equilibrio de especie.

De acuerdo con (Patil y Taille 1982) podemos considerar a una comunidad ecológica como el par $C(k, \pi)$ donde k es el número finito conocido de especies que están físicamente presentes, y $\pi = \{\pi_1, \pi_2, \dots, \pi_k\}$ el vector de abundancia de especie. Se dice que una comunidad está completamente equilibrada si

$$\pi_1 = \pi_2 = \dots = \pi_k = 1/k. \text{ En general en una población con } N \text{ individuos } \pi_i = N_i/N \text{ con } N = \sum_{i=1}^k N_i.$$

Por lo que definen el índice de diversidad $\lambda(C)$ asociado a una comunidad $C(k, \pi)$ como el promedio de la rareza y está dado por:

$$\lambda(C) = \sum_{i=1}^k \pi_i R(i; \pi) \quad (2.1)$$

A partir de esta definición podemos ver que en la construcción de índices de diversidad sólo se difiere en la forma en que es medida $R(i, \pi)$, rareza de especie.

2.1. Clases de índices de diversidad

La literatura apunta la existencia de una gran cantidad de índices de diversidad. Patil y Taille (1982) proponen clasificar los índices de acuerdo a la medida de rareza utilizada, entonces tenemos índices dicotómicos y de rango. Para el caso dicotómico la rareza de especie depende del valor de π_i , por otro lado los índices basados en rangos dependen del vector de rangos de especie, definido este como $\pi^\# = \{\pi_1^\#, \pi_2^\#, \dots, \pi_k^\#\}$. Sin perder en generalidad tomemos $\pi_1^\# \geq \pi_2^\# \geq \dots \geq \pi_k^\#$. Los datos provienen del *vector de abundancias relativas*.

El problema teórico puede analizarse tomando en cuenta criterios matemáticos. Proponemos dos clases generales para los índices a partir de su estructura: la clase aditiva Λ^+ y la clase multiplicativa Λ^\times . Esto nos permite hacer un tratamiento unificado del problema de la estimación.

La clase aditiva la definimos como:

$$\Lambda^+ = \left\{ \lambda = \sum_{l=1}^k \{ \tau_l R(l, \pi_l) \} \pi_l : \pi_l = \frac{N_l}{N}, N > 0, l = 1, 2, \dots, k \right\} \quad (2.2)$$

y la clase multiplicativa por:

$$\Lambda^\times = \left\{ \lambda = \prod_{l=1}^k \{ \tau_l R(l, \pi_l) \} \pi_l : \pi_l = \frac{N_l}{N}, N > 0, l = 1, 2, \dots, k \right\} \quad (2.3)$$

τ_l es el rango que se le asigna a la especie l , $R(l, \pi_l)$ la medida de rareza de especie antes definida y π_l la abundancia relativa de la especie l .

Son dos las diferencias básicas en ésta definición de clases: 1) Al definir las clases aditiva y multiplicativa se hace hincapié en el uso de el *vector de abundancia relativa* para la determinación de la rareza de especie, más adelante se expondrán los argumentos. 2) Por otra parte el definir la clase aditiva de esta manera permite hacer la búsqueda de estimadores usando procedimientos propios de la teoría de estimación .

2.2. Aplicación de los índices, un problema de muestreo

Muchas de las investigaciones en ecología tienden a caracterizar una comunidad y en la mayoría de las ocasiones se desea obtener información de áreas o superficies muy grandes, dada la usual imposibilidad de hacer un censo se utilizan muestras. Kepton (2002) menciona que si bien es cierto se ha llegado a dar una interpretación teórica a los índices como medidas de diversidad se necesitan estudios empíricos del desempeño de sus estimadores muestrales.

Analizando con detenimiento el problema de la estimación de los índices, nos encontraremos con situaciones no muy cómodas para aplicar la teoría de muestreo, en principio no se cuenta con un marco muestral o su construcción sería muy costosa y el vector de abundancia de especie π es igualmente desconocido.

Sabemos que n no es un buen estimador de N , por lo tanto se propone que se estime la abundancia relativa de especies en la comunidad $\pi_i = N_i/N$ en lugar de estimar la absoluta N_i dado que π_i puede ser estimado mediante $\hat{\pi}_i = n_i/n$.

Definimos a n_i como el número de individuos de la especie i que están contenidos en la muestra aleatoria seleccionada. En ella hay n individuos, entonces la definimos como

$$n = \sum_{i=1}^k n_i \quad (2.4)$$

note que n es una variable aleatoria.

En nuestro estudio consideraremos que el tamaño de muestra está fijo, por lo que las estimaciones serán condicionadas a n . Kempton (2002) menciona este mecanismo de fijar n es utilizado en técnicas como la rarefration, técnica sugerida en la literatura. Las consideraciones que incluyan el caso más general en el que n sigue una cierta distribución será tratado en trabajos posteriores.

3. ESTIMACIÓN DE UN ÍNDICE DE DIVERSIDAD BAJO UN PROCEDIMIENTO EQUIVALENTE AL MAS, SUPONIENDO n FIJA

Supondremos que se selecciona un sitio de muestreo y que las variables aleatorias son producidas por un mecanismo aleatoria que es considerado equivalente al MAS. Note que en el Muestreo aleatorio simple con reemplazo [MAScr] para n fija tenemos que las n_i siguen una distribución Binomial con parámetros n y π_i , entonces se tiene que $E\{n_i\} = n\pi_i$ y $Var\{n_i\} = n\pi_i(1 - \pi_i)$.

Si los datos se consideran generados por un mecanismo equivalente al Muestreo aleatorio simple sin reemplazo [MASsr] para n fija las n_i siguen una distribución Hipergeométrica con parámetros N , N_i , n . Lo que presenta un nuevo problema, pues en la definición de la densidad tenemos a N , que es un parámetro desconocido. Por ello es recomendable trabajar con una distribución que no contenga a N como parámetro. La Hipergeométrica converge rápidamente a la Binomial. Por ejemplo Johnson y Kotz (1969) proponen como aproximación de la distribución Binomial a la Hipergeométrica si $n/N < 0.1$ y $N > 30$ condiciones que no son difíciles de cumplir en nuestro caso.

Para derivar los estimadores de los índices de diversidad así como las varianzas, utilizaremos el método de momentos. Lo que hace necesario el uso de los momentos de la Binomial y los momentos de la Multinomial, dado que se presentan productos cruzados cuando se estima la varianza de los estimadores.

3.1. Índice de Simpson para un sitio de muestreo

Este índice propuesto por Simpson (1949) es uno de los más utilizados por los investigadores. Patil y Taille (1982) mencionan a este índice como uno de los tres más populares. De acuerdo a la clasificación de los índices propuesta anteriormente el índice de Simpson pertenece a la clase aditiva (2.8) si hacemos que $\tau_i = 1$, es decir todas las especies tienen el mismo rango y $R(\pi) = 1 - \pi_i$. Entonces

$$\lambda_{\text{Simp}} = \sum_{i=1}^k \{1 - \pi_i\} \pi_i = 1 - \sum_{i=1}^k \pi_i^2 \quad (3.5)$$

Se puede notar que el índice (3.5) es $\lambda_{\text{Simp}} = 1 - H$, donde H es el índice de concentración de Herfindall, comúnmente usado como un estadístico descriptivo en la economía.

Es usual encontrar en la literatura un estimador ingenuo para el índice de Simpson. Este está dado por (3.6).

$$\hat{\lambda}_{\text{Simp}} = 1 - \sum_{i=1}^k \left(\frac{n_i}{n} \right)^2 \quad (3.6)$$

el sesgo de este estimador es

$$\text{bias}(\hat{\lambda}_{\text{Simp}}) = - \sum_{i=1}^k \frac{\pi_i(1 - \pi_i)}{n} \quad (3.7)$$

Note que el sesgo equivale a la suma de las $\text{var}\{n_i\}$, Bouza y Schubert (2003) hicieron la corrección del estimador y calcularon su varianza. Gimaret-Carpentier et al (1998) en un estudio de diversidad además se proponen un estimador insesgado y una varianza basada en métodos no paramétricos.

En este trabajo además de demostrar que el estimador se Simpson efectivamente es insesgado para el caso de un sitio de muestreo, lo que daría un marco teórico a las propuestas de la literatura, se tiene interés en obtener una varianza teórica del estimador por una vía analítica. Esto se puede ver en la siguiente proposición.

Proposición 1

Un estimador insesgado para el índice de Simpson en un sitio de muestreo, su varianza y la varianza estimada en el MAScr con n fija están dados por:

$$\hat{\lambda}_{\text{Simp(MAScr)}} = 1 - \sum_{i=1}^k \frac{n_i(n_i - 1)}{n(n-1)} \quad (3.8)$$

$$\text{Var}\left\{\hat{\lambda}_{\text{Simp(MAScr)}}\right\} = \frac{2}{n(n-1)} \left[\sum_{i=1}^k \pi_i^2 + 2(n-2)\pi_i^3 - (2n-3) \left(\sum_{i=1}^k \pi_i^2 \right)^2 \right] \quad (3.9)$$

$$\text{Var}\left\{\hat{\lambda}_{\text{Simp(MAScr)}}\right\} = \frac{2}{n(n-1)^2(n-2)(n-3)} \left[\sum_{i=1}^k \left\{ 2n^2(n^2 + n - 3)\pi_i^3 - n(5n^2 - 3n - 3)\pi_i^2 + 3n(n-1)\pi_i \right\} - n^3(2n-3) \left(\sum_{i=1}^k \pi_i^2 \right)^2 + 2(2n-3) \sum_{i < j} n^2 \pi_i^2 \pi_j + n^2 \pi_i \pi_j^2 - n \pi_i \pi_j \right] \quad (3.10)$$

también la podemos escribir como:

$$= \frac{2}{n} \left[\sum \left\{ C_1 \hat{\pi}_i^3 - C_2 \hat{\pi}_i^2 + C_3 \hat{\pi}_i \right\} - C_4 \left(\sum_{i=1}^k \hat{\pi}_i^2 \right)^2 + \sum_{i < j} C_5 \hat{\pi}_i^2 \hat{\pi}_j + C_5 \hat{\pi}_i \hat{\pi}_j^2 - C_6 \hat{\pi}_i \hat{\pi}_j \right] \quad (3.10a)$$

donde k es el número de especies conocidas y $\hat{\pi}_i = \frac{n_i}{n}$.

Prueba:

La prueba de que el estimador (3.8) es insesgado es sencilla. Para probar que es válida (3.9) desarrollamos:

$$\begin{aligned} \text{Var}\left\{\hat{\lambda}_{\text{Simp(MAScr)}}\right\} &= \frac{1}{n^2(n-1)^2} \sum_{i=1}^k \text{Var}\{n_i^2 - n_i\} + \frac{2}{n^2(n-1)^2} \sum_{i \neq j}^k \text{Cov}\{(n_i^2 - n_i)(n_j^2 - n_j)\} \\ &= \sum_{i=1}^k \frac{2\pi_i^2 + 4(n-2)\pi_i^3 - 2(2n-3)\pi_i^4}{n(n-1)} - \frac{2}{n^2(n-1)^2} \sum_{i < j} 2n(n-1)(2n-3)\pi_i^2 \pi_j^2 \\ &= \frac{2}{n(n-1)} \left(\sum_{i=1}^k \pi_i^2 + 2(n-1)\pi_i^3 - (2n-3) \left(\sum_{i=1}^k \pi_i^2 \right)^2 \right) \end{aligned}$$

que es la varianza (3.9) dada en la proposición.

Siguiendo el mismo procedimiento para demostrar que el estimador de la varianza (3.10) es insesgado, tenemos que:

$$\begin{aligned}
E\left\{\widehat{\text{Var}}(\widehat{\lambda}_{\text{Simp}})\right\} &= \frac{2}{n(n-1)} \left[\sum_{i=1}^k \left(\frac{2(n^2+n-3)}{n-3} \pi_i^3 + \frac{n^2+9n-15}{(n-2)(n-3)} \pi_i^2 + \frac{2n-3}{(n-1)(n-2)(n-3)} \pi_i \right) - (2n-3) \sum_{i=1}^k \pi_i^4 \right. \\
&\quad \left. - \sum_{i=1}^k \left(\frac{6(2n-3)}{n-3} \pi_i^3 + \frac{7(2n-3)}{(n-2)(n-3)} \pi_i^2 + \frac{2n-3}{(n-1)(n-2)(n-3)} \pi_i \right) - 2(2n-3) \sum_{i<j} \pi_i^2 \pi_j^2 \right] \\
&= \frac{2}{n(n-1)} \left[\sum_{i=1}^k \left(\frac{n^2-5n+6}{n-3} \pi_i^3 + \frac{n^2-5n+6}{(n-2)(n-3)} \pi_i^2 \right) - (2n-3) \left(\sum_{i=1}^k \pi_i^4 + 2 \sum_{i<j} \pi_i^2 \pi_j^2 \right) \right] \\
&= \frac{2}{n(n-1)} \left[\sum_{i=1}^k 2(n-2) \pi_i^3 + \pi_i^2 - (2n-3) \left(\sum_{i=1}^k \pi_i^2 \right)^2 \right]
\end{aligned}$$

es decir que la estimación es insesgada, con lo que la prueba termina. ■

Tenemos un estimador sesgado de la varianza que utiliza una fórmula más sencilla. La condición de insesgader se pierde pero este es consistente.

Corolario 1.1

Un estimador consistente de la varianza del estimador de Simpson puede estar dado por:

$$\widehat{\text{Var}}_2\left\{\widehat{\lambda}_{\text{Simp}}\right\} = \frac{2}{n(n-1)} \left(\sum_{i=1}^k \widehat{\pi}_i^2 + 2(n-1) \widehat{\pi}_i^3 - (2n-3) \left(\sum_{i=1}^k \widehat{\pi}_i^2 \right)^2 \right) \quad (3.11)$$

en donde el número de especies k es conocida, las $\widehat{\pi}_i = \frac{n_i}{n}$ dentro del MAScr.

Es importante analizar el efecto del tamaño de la muestra en los coeficientes involucrados en (3.10a). El corolario siguiente nos proporciona un estimador que es asintóticamente insesgado.

Corolario 1.2

Un estimador asintóticamente insesgado de la varianza del estimador del índice de Simpson está dado por:

$$\widehat{\text{Var}}_3(\widehat{\lambda}_{\text{Simp}}) = \frac{2}{n} \left[\sum_{i=1}^k \left(C_1 \widehat{\pi}_i^3 - C_2 \widehat{\pi}_i^2 + C_3 \widehat{\pi}_i \right) - C_4 \left(\sum_{i=1}^k \widehat{\pi}_i^2 \right)^2 + \sum_{i<j} C_5 \widehat{\pi}_i^2 \widehat{\pi}_j + C_5 \widehat{\pi}_i \widehat{\pi}_j^2 - C_6 \widehat{\pi}_i \widehat{\pi}_j \right] \quad (3.12)$$

Estudiamos el efecto de los posibles valores de n en los valores de los coeficientes C_1, C_2, \dots, C_6 donde los C_1, C_2, \dots, C_6 son valores aproximados de los coeficientes dados en (3.16a) se tiene la Tabla 3.1.

Tabla 3.1. Coeficientes de la varianza estimada del estimador del índice de Simpson.

n	C_1	C_2	C_3	C_4	C_5	C_6
30	2.6244	0.2079	0.004 1	2.4206	0.0807	0.0028
60	2.2880	0.0929	0.000 9	2.1960	0.0366	0.0006
80	2.2118	0.0678	0.000	2.1455	0.0269	0.0003

			5			
100	2.1675	0.0533	0.000	2.1145	0.0211	0.0002
			3			
200	2.0818	0.0258	0	2.0561	0.0103	0.0000
250	2.0652	0.0205	0	2.0447	0.0082	0
300	2.0541	0.0170	0	2.0372	0.0068	0
400	2.0405	0.0127	0	2.0278	0.0051	0

Note que a medida que crece n los productos cruzados se acercan a cero, lo que podría sugerir que la importancia de la interacción entre las especies no tiene un peso importante en la estimación de la varianza si el número de observaciones es grande.

3.2. Índice de Simpson para más de un sitio de muestreo

Es usual abordar la estimación de los índices de diversidad para el caso de m sitios de muestreo donde m > 1 de un total de M en la población ya que como se apuntó antes los investigadores en ecología realizan muestreos en sus trabajos.

Nuevamente retomaremos las estimaciones condicionadas a un tamaño fijo, es decir, las n_j serán consideradas como fijas ($j = 1, 2, \dots, m$) por lo que la suma de los individuos en la muestra n^* será fijo. La distribución de las especies en cada sitio de muestreo cuando estamos en el muestreo aleatorio simple con reemplazo sigue una distribución Binomial con parámetros $n_{\cdot j}$ y π_j .

Para el índice definido en (3.5) se proponen dos estimadores: el separado ($\hat{\lambda}_{\text{SimpSep(MAScr)}}$) y el conjugado ($\hat{\lambda}_{\text{SimpConj(MAScr)}}$), para lo cual definimos, $n_{\cdot j}$ y n^*

$$n_{\cdot j} = \sum_{i=1}^m n_{ij} \quad \text{y} \quad n^* = \sum_{j=1}^m n_{\cdot j}$$

Obtener un índice por sitio y calcular un promedio como estimación tiene sentido considerar que tenga una distribución Normal independientemente de la distribución original de los datos en los sitios de muestreo.

Proposición 2

El estimador separado para el índice de Simpson, su varianza y varianza estimada en el MAScr en más de un sitio de muestreo $m > 1$ considerando $n_{\cdot j}$ fijas están dados por:

$$\hat{\lambda}_{\text{SimpSep(MAScr)}} = \sum_{j=1}^m \frac{\hat{\lambda}_j}{m} \quad (3.13)$$

$$\text{Var}\{\hat{\lambda}_{\text{SimpSep(MAScr)}}\} = \frac{1}{m^2} \sum_{j=1}^m \text{Var}\{\hat{\lambda}_j\} \quad (3.14)$$

donde los $\hat{\lambda}_j$ son los estimadores de la proposición 1.

Prueba:

Es evidente la demostración, dado que la construcción del estimador 3.13 sigue una estructura lineal y $E\{\hat{\lambda}\} = \sum_j^m E\{\hat{\lambda}_j\}$ conserva la propiedad de insesgadez. Además se observa que existe independencia de los $\hat{\lambda}_j$ por lo que $\text{Var}\left\{\sum_j^m \hat{\lambda}_j\right\} = \sum_j^m \text{Var}\{\hat{\lambda}_j\}$. ■

Por otra parte el estimador Conjugado es el más utilizado por los investigadores en ecología cuando desean caracterizar una población ecológica, siguiendo la idea de conformar una muestra grande con pequeñas muestra, lo que puede representar un problema si no se toma en cuenta los supuestos del MAS. Para estimar de manera conjugada el índice de Simpson se tiene la siguiente proposición.

Proposición 3

El estimador conjugado para el índice de Simpson, su varianza y varianza estimada en el MAScr en más de un sitio de muestreo $m > 1$ con $n_{.j}$ fijas están dados por:

$$\hat{\lambda}_{\text{SimpConj(MAScr)}} = 1 - \sum_{i=1}^k \frac{n_i^*(n_i^* - 1)}{n^*(n^* - 1)} \quad (3.15)$$

$$\text{Var}\left\{\hat{\lambda}_{\text{SimpConj(MAScr)}}\right\} = \frac{2}{n^*(n^* - 1)} \sum_{i=1}^k \pi_i^2 + 2(n^* - 2)\pi_i^3 - (2n^* - 3) \left(\sum_{i=1}^k \pi_i^2\right)^2 \quad (3.16)$$

$$\begin{aligned} \text{Var}\left\{\hat{\lambda}_{\text{SimpConj(MAScr)}}\right\} &= \\ &= \frac{2}{n^*(n^* - 1)^2(n^* - 2)(n^* - 3)} \left[\sum_{i=1}^k \left\{ 2n^{*2}(n^{*2} + n^* - 3)\hat{\pi}_i^3 - n^*(5n^{*2} - 3n^* - 3)\hat{\pi}_i^2 + 3n^*(n^* - 1)\hat{\pi}_i \right\} - \right. \\ &\quad \left. - n^{*3}(2n^* - 3) \left(\sum_{i=1}^k \hat{\pi}_i^2\right)^2 + 2(2n^* - 3) \sum_{i < j}^k n^{*2}\hat{\pi}_i^2\hat{\pi}_j + n^{*2}\hat{\pi}_i\hat{\pi}_j^2 - n^*\hat{\pi}_i\hat{\pi}_j \right] \quad (3.17) \end{aligned}$$

Prueba:

Para probar la proposición se sigue los pasos utilizados para demostrar en la proposición 1 con los nuevos parámetros. ■

Note que también podemos extender los resultados a el Corolario 1.1 y el Corolario 1.2. Lo que permitiría alcanzar los valores de n de la Tabla 1.1 con mayor facilidad.

4. EXPERIMENTOS DE MONTE CARLO

Se realiza un estudio comparativo de los estimadores separado (3.13) y combinado (3.15) mediante experimentos de Monte Carlo. La base de datos proporcionada forma parte de la investigación Ordenamiento Ecológico Territorial desarrollada en la subcuenca del río San Juan al norte del Estado de Guerrero, México. Investigación financiada por el SIBEJ Sistema de Investigación Benito Juárez, ver Almazán **et al.** (2003).

Para generar la base de datos de este trabajo, se registraron las especies con un número de individuos mayor de 9, de esta manera la población artificial quedó conformada por los 60 sitios y 820 individuos clasificados en 66 especies.

Para la simulación se utilizaron $H = 50, 100, 200$ y 500 como el número de corridas. El error absoluto de estimación es calculado a partir de las muestras que se seleccionaron mediante un procedimiento aleatorio simple con reemplazo. De forma similar se hace con las estimaciones de la varianza, donde es comparado el valor de la estimación dada la muestra.

Consideramos cada corrida como una muestra independiente de la misma población. Comparamos los resultados obtenidos y obtuvimos que el p-value de la prueba no-paramétrica de Kruskal-Wallis es 0.00001 para el caso del índice combinado. Esto nos indica que n influye en la precisión significativamente. Por su parte para el separado es de 0.3275 lo que nos sugiere una conclusión contraria.

Esta comparación de ambos índices se puede visualizar en la Tabla 4.1. Como se puede ver los índices estimados son sensiblemente diferentes y el error del separado es considerablemente menor, lo que le hace mas recomendable. Esto es muy claro al analizar el Gráfico 4.3.

Tabla 4.1. Comparativo de los estimadores del índice de Simpson.

f	$\hat{\lambda}_{\text{SimpConj}}$	$\hat{\lambda}_{\text{SimpSep}}$	Error $\{\hat{\lambda}_{\text{sc}}\}$	Error $\{\hat{\lambda}_{\text{ss}}\}$	V $\{\hat{\lambda}_{\text{sc}}\}$	V $\{\hat{\lambda}_{\text{ss}}\}$
0.03	0.80893	0.66789	0.18201	0.32305	0.000003	0.000001
0.10	0.90253	0.67039	0.08842	0.32054	0.000018	0.000001
0.17	0.92552	0.66665	0.06542	0.32429	0.000043	0.000001
0.25	0.93370	0.67327	0.05725	0.31707	0.000084	0.000001
0.51	0.94610	0.67179	0.04485	0.31915	0.000218	0.000001
0.75	0.94942	0.67132	0.04152	0.31962	0.000231	0.000001
0.80	0.95238	0.67219	0.03857	0.31876	0.000207	0.000001
0.90	0.95159	0.67220	0.03936	0.31875	0.000108	0.000001
0.95	0.95164	0.67142	0.03930	0.31952	0.000048	0.000001

En la Tabla 4.2 se tienen las varianzas obtenidas en el experimento para el índice conjugado. Note que éstas son muy similares para todas las fracciones de muestreo. Esto soporta que el uso de la fórmula es menos costosa computacionalmente. El Gráfico 4.4 nos da un visión del comportamiento de las varianzas.

Tabla 4.2. Comparativo de las varianzas del estimador conjugado del índice de Simpson.

F	$\hat{\text{Vâr}}_1\{\hat{\lambda}_{\text{SimpConj}}\}$	$\hat{\text{Vâr}}_2\{\hat{\lambda}_{\text{SimpConj}}\}$	$ \hat{\text{V}}_1\{\hat{\lambda}\} - \text{V}\{\hat{\lambda}\} $	$ \hat{\text{V}}_2\{\hat{\lambda}\} - \text{V}\{\hat{\lambda}\} $
0.03	0.001317	0.001350	0.001314	0.001347
0.10	0.000126	0.000129	0.000109	0.000123
0.17	0.000049	0.000050	0.000033	0.000033
0.25	0.000029	0.000029	0.000060	0.000060

0.51	0.000009	0.000010	0.000209	0.000209
0.75	0.000006	0.000006	0.000226	0.000226
0.80	0.000005	0.000005	0.000206	0.000206
0.90	0.000004	0.000004	0.000144	0.000144
0.95	0.000004	0.000004	0.000126	0.000126

Gráfico 4.3. Errores de los estimadores del índice de Simpson

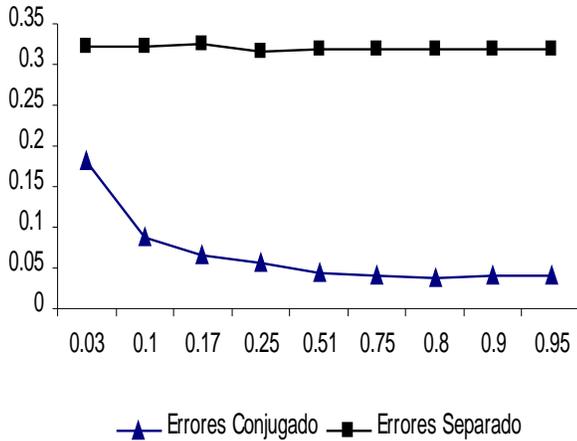
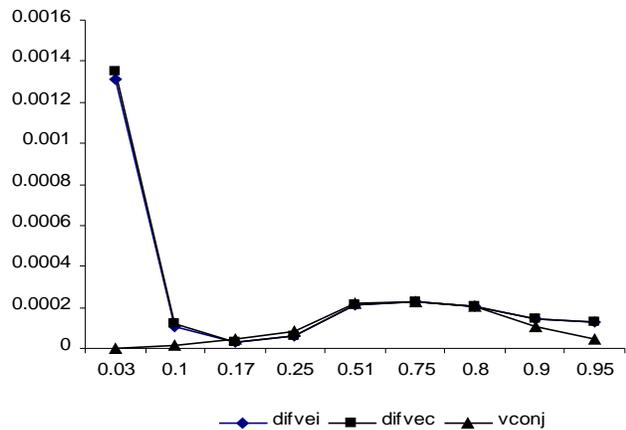


Gráfico 4.4. Varianzas de estimado conjugado del índice de Simpson.



REFERENCIAS

ALMAZÁN, A. y otros (2003): "Ordenamiento Ecológico Territorial de la Subcuenca del Río San Juan del Estado de Guerrero", **Reporte Técnico de Investigación**, SIBEJ-UAGro.

BOUZA, C. and L. SCHUBERT (2002): "A bio-diversity study: index and dynamics", **Agrociencia**. Aceptado y por publicar.

COCHRAN, W.G. (1977): **Sampling Techniques**, Wiley, New York.

GIMARET-CARPENTIER, C.; R. PÉLISSIER; J. PASCAL and F. HOULLIER(1998): "Sampling Strategies for the Assessment of Tree Species Diversity," **Journal of Vegetation Science**, 9, 161-172.

GOVE, J.H.; G.P. PATIL; B.F. SWINDEL and C. TAILLE (1994): "Ecological diversity and Forest management", **Handbook of statistics**, 12, (eds G. P. Patil and C. R. Rao, Elsevier Science B. V.).

HELTSHE, J.F. and D.W. BITZ (1979): "Comparing diversity measures in sampling communities", **Ecological Diversity in Theory and Practice**, 133-144. International Co-operative Publishing House, Fairland Maryland.

HOLLANDER, M. and D. WOLFE (1999): **Nonparametric Methods**, Wiley & Sons. New York.

- JOHNSON, N.L. and S. KOTZ (1969): **Distributions in Statistics: Discrete Distributions**, John Wiley & Sons, New York.
- KEMPTON, R.A. (2002): "Species diversity", **Encyclopedia of Environmetrics**, 4, 2086-2092. (eds Abbel H. El-Shaarawi and W. Piegorsch) John Wiley & Sons. N. York.
- KREBS, C. (1999): **Ecological Methodology**, 2da. Edition, Adison Wesley, London.
- LUDWIG, J.A and J.F. REYNOLDS (1988): **Statistical ecology: A primer in methods and computing**, John Wiley, New York.
- PATIL, G.P. and C. TAILE (1982): "Diversity as a concept and its measurement," **Journal of the American Statistical Association** 77, 548-567.
- PIELOU, E.C. (1974): **Population and Community Ecology, Principles and Methods**.
- SMITH, E.P. (2002): "Ecological Statistics," **Encyclopedia of Environmetrics** 2, 589-602. (eds Abdel H. El-Shaarawi and Walter W. Piegorsch) John Wiley & Sons. New York.