

Capítulo 3

REVISION DE ALGUNOS PROBLEMAS ESTADISTICOS EN LA ESTIMACION DE LA BIODIVERSIDAD

CARLOS N. BOUZA HERRERA* y DANTE COVARRUBIAS MELGAR**

*Universidad de La Habana, Cuba

**Universidad Autónoma de Guerrero, México

RESUMEN

Se hace una presentación de los modelos que son de uso común en el estudio de la biodiversidad. Se hace hincapié en el problema que parece al hacer la estimación usando muestras aleatorias. Se intenta que este trabajo sirva de referencia para ulteriores desarrollos de modelos particulares para los estudios de la biodiversidad.

Palabras clave: estimación, errores de muestreo, índices de biodiversidad, índices de cubrimiento, rareza.

ABSTRACT

We present models of common use in biodiversity studies. We highlight the problems present in estimating using random samples. We intend that this paper will play the role of being a reference for posterior developments of particular models for studies of biodiversity.

Keywords: estimation, sampling errors , indexes of biodiversity, índices of coverage, rarity.

1. ENFOQUES Y CRITERIOS PARA LA PLANEACION Y DESARROLLO DEL LEVANTAMIENTO DE LOS DATOS

1.1. Enfoques para la experimentación

En el estudio de la biodiversidad podemos usar diversos enfoques. Ellos sea asocian a varios requerimientos básicos a tomar en cuenta al planear la recolección de datos. Estos son detallados en la tabla siguiente:

Inferencia basada en el diseño	Inferencia basada en el Modelo
Se requiere tamaños de muestra relativamente grandes	Se puede usar tamaños de muestra pequeños

aleatorias	representativas usando el modelo o un experto
La población objetivo está bien determinada	Se hacen consideraciones sobre clases de problemas, Fuentes o tendencias de cambios
Se puede hacer comparaciones entre poblaciones	Se hacen inferencias sobre relaciones bajo el modelo o se determinan modelos alternativos
Puede ser costosa para mantener niveles de precisión prefijados	Puede ser cara en términos de las dificultades de halla sitios representativos.

Tabla 1.1. Enfoques del proceso de toma de datos

El enfoque experimental por su parte es activo respecto a la adaptación del manejo. En el se hacen experiencias diferentes tratamientos para desarrollar comparaciones entre ellos. Uno de los resultados puede ser el mejor sistema de monitoreo fijados operacionalmente. Su uso permite evaluar modelos y usar los que sean convenientes para establecer las consecuencias de políticas de manejo y el impacto de ellas. Este enfoque ofrece muchas posibilidades para hacer comparaciones y puede ser costoso. Es claro que el uso de uno de estos enfoques requiere de un diseño minucioso y de considerar que métodos estadísticos serán utilizables para hacer inferencias.

Podemos caracterizar los enfoques y construir la tabla 2.2. El uso del principio inferencial basado en el diseño es adecuado al tomar en cuenta el uso de la teoría del muestreo lo que permite tipificar la toma de datos. El

procedimiento inferencial es soportado por la teoría del muestreo repetido que permite utilizar o desarrollar métodos de la llamada estadística clásica. Al tener tamaños de muestra suficientemente grandes se puede garantizar la validez del uso de la teoría normal y en todo caso algún método no paramétrico. Esto permite hacer comparaciones de las medidas de biodiversidad de varias poblaciones. La posible necesidad de tener tamaños de muestra grandes en modelos normales encarece su uso.

Por otra parte el uso de modelos para hacer la inferencia permite escapar de la necesidad de tomar muestras representativas viene dada por criterios asociados al modelo asumido como generador de las variables o de la biodiversidad. En las aplicaciones de biodiversidad se denominan tales sitios como 'centinelas'. Estos son en algunas ocasiones son utilizados para construir un modelo. Este es el caso al monitorear. Cada sitio representa un problema de interés para la biodiversidad y los estudios desarrollados en ellos son intensivos. Generalmente se estudian factores ecológicos, tales como la sensibilidad a ciertas tensiones o tendencias. De tales estudios se genera un modelo que se considera válido para otros sitios y se usa para hacer la inferencia. Su debilidad es que buscar los sitios puede ser muy difícil. Los centinelas pueden estar en regiones de difícil acceso y su identificación ya es de por sí complicada. En todos los casos hacer generalizaciones de experiencias basadas en modelos es muy riesgosa dada la complejidad de la biodiversidad en el mundo real.

Enfoque Operacional	Enfoque Experimental
Manejo Pasivo y Adaptativo	Manejo Activo y Adaptativo
Usa sitios disponibles y operacionales	Creación de sitios y tratamientos
Hace comparaciones entre los resultados de la operación común en curso (Cosecha, pesca, tala)	Hace comparaciones entre posibles formas de operar

etc.,)	
Limita el número de tratamientos a los que posibilita el manejo	Fija un número de tratamientos y se adapta el manejo a lo requerido por ellos
Es relativamente barato	Puede ser costoso
No hace propuestas con efecto profundo en el manejo	Hace propuestas que pueden conllevar profundos cambios en el manejo

Tabla 1.2 Caracterización de enfoques de los experimentos

El enfoque operacional es de carácter pasivo- adaptativo y maneja sitios operacionales de disponibilidad i en el caso de buscarse hacer comparaciones en cosechas, problemas forestales en lo que es importante monitorear el manejo. Este incluye estudios retrospectivos para conocer la historia de la biodiversidad a través de los datos acumulados. Por ello los métodos de monitoreo son muy usados, lo que conlleva que no sean útiles para comparar tratamientos. Estos llevan a cabo recomendaciones útiles para el manejo adaptativo de los sistemas.

Los estudios de biodiversidad buscan aprender o comprender los mecanismos que operan sobre ella en el hábitat escogido. Los datos por si mismos no dan conocimiento de este tipo. Al desear manejar el sistema se buscará explicar el mecanismo que le rige para poder operar sobre este. Para ello se pueden hacer comparaciones para aprender como opera el mecanismo o simplemente explicar el funcionamiento del mismo. Podemos decir que uno es de carácter indagatorio y el otro de carácter mecanicista .

1.2 Diseño estadístico de redes de monitoreo

El objetivo del monitoreo es detectar los cambios en los parámetros de interés en un periodo relativamente largo. Obtener una guía aceptable se vincula con la inferencia estadística. Adicionalmente es usual que se haga el diseño del estudio de monitoreo sin consultar a un estadístico agrava aun más las posibilidades reales de hacer inferencias confiables de los resultados. Claramente en ecología los experimentos y el monitoreo plantean problemas conceptuales a la Estadística que no siempre están resueltos satisfactoriamente. Por ejemplo diseñar una investigación en gran escala, las variaciones no controladas presentes (temperatura, incendios etc.). Dificultad de obtener réplicas, grandes variaciones los modelos de muestreo razonablemente. Por este motivo los programas de monitoreo están, generalmente, deficientemente diseñados desde el punto de vista estadístico y el método de medición es generalmente poco confiable. Es difícil hacer un diseño de monitoreo que satisfaga los requerimientos básicos para hacer inferencias confiables.

El ecólogo está generalmente interesado en que el estadístico le diseñe un programa que sea robusto desde vista inferencial. Esto es soportado por hacer un diseño tal que:

- Compare las diversas posibilidades de muestrear usando diversos tratamientos.
- Garantice un número suficiente de replicas para tratamientos tomando en cuenta la heterogeneidad de los ecosistemas y la incertidumbre.
- Haga una asignación usando criterios de estratificación para garantizar que haya alguna homogeneidad entre las unidades de las subzonas.

-Tome información suficiente (tiempo de monitoreo alto) para poder establecer el efecto de los tratamientos aislando las causas externas como cambios climáticos.

Son tareas del estadístico:

-Garantizar que la estratificación genere ganancias en precisión altas.-Comprobar las hipótesis correspondientes los métodos o desarrollar alternativas (no paramétricas por ejemplo).
-Establecer el efecto en la potencia de las pruebas de no satisfacerse los requerimientos de tamaño de muestra si no se acepta el fijado

2. MEDIDAS DE BIODIVERSIDAD

2.1 Generalidades

Existen diversas medidas para caracterizar la biodiversidad . Muchas de ellas no tiene un tratamiento estadístico debido al hecho mismo de que fueron propuestas sin tomar en cuenta el uso de estudios estadísticos , pensando en hacer inferencias.Los ecólogos usan medidas para caracterizar la biodiversidad y evaluar la salud de los ecosistemas. Su disminución indica la del ecosistema y llama la atención sobre el stress medio ambiental del mismo. La diversidad es evaluado mediante un sin número de medidas, propuestas puntualmente por ecólogos.

La pérdida en diversidad biológica por extinción es uno de los cambios más notables en la transformación global del medio ambiente hoy día. Los investigadores tratan de comprender el mecanismo de los cambio en la biodiversidad.

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

Al medir diversidad es necesario tomar en cuenta que se trata de describir la relación entre individuos de diferentes especies en una zona. El número de ellos es medido. Algunas asunciones deben ser hechas al considerar le evaluación de la biodiversidad

- **Las categorías son conocidas:** La clasificación de las especies es de universal aceptación **Todas las categorías son igualmente diferentes:** No hay error posible de clasificar un individuo en mas de una clase.

- **Usar una medida de la importancia de las especies:** Usualmente s es una medida que usa el cubrimiento y numero de individuos, densidad relativa.

- **La comunidad bajo estudio está bien definida:** Como la importancia relativa de una categoría variara dependiente de la definición de su tamaño y extensión de la zona.

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.

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.

2.2 Algunos índices

Los índices de biodiversidad mas frecuentemente utilizados utilizan medidas de riquezas o rareza y la distribución de las especies. Ejemplos de ellos son los de Simpson, (1949), Shannon (1948), Brillouin (1956), McIntosh (1967), Hurlbert (1971) y Fager (1972). Estos son re-escalados en diferentes trabajos por los ecólogos. Tendrá sentido hacerlo para garantizar propiedades estadísticas pero este no es el motivo usualmente,

Veremos algunas de ellas

2.2.1 La riqueza

La riqueza de las especies es la medida más simple. Es intuitivo considerar que este es la forma correcta de establecer la diversidad objetivamente. Esta es usada comúnmente en estudios conservacionistas para determinar la sensibilidad de ecosistemas en las especies que le habitan. Es usada como una primera medida de la diversidad en el hábitat. Fue llamado así por McIntosh (1967). El conteo puede ser complicado en muchas ocasiones. En general es útil al comparar dos hábitats. Ninguna inferencia puede ser hecha usándole.

2.2.2 Rareza

Esta es medida por la razón entre el valor de un índice de diversidad y su valor máximo. Este mide como se distribuye la abundancia entre las especies de una comunidad. El más usado es el referido al índice Shannon en el que se fija el parámetro como

$$E(\text{Shannon}) = \frac{-\sum_{i=1}^K \pi_i \ln(\pi_i)}{\ln(K)}$$

Por su definición, $E \in [0,1]$. Este es conocido muchas veces como índice de Pielou.

2.2.3 El índice de Brillouin

Para muestras no aleatorias es muy usado y es preferido a al índice de Shannon. Bajo el supuesto de que K es conocido sin error, este es

$$B = \frac{\ln(N!) - \ln(N_i!)}{N} B$$

Este fue usado por vez primera por Margalef (1958).

2.2.4 Métodos del tipo Monte Carlo

Se sabe que generalmente hay una correlación inversa entre riqueza y latitud: así como entre esta y la altitud. La riqueza de las especies no se relaciona con el tamaño del área estudiada y la observación de más especies puede deberse solo al uso de muestras más grandes. Áreas mayores por su parte comprenden un mayor número de hábitats dando más oportunidad a más especies a asentarse debido a la mayor diversidad. Esta medida es utilizada con frecuencia en estudios taxonómicos de comunidades o grupos funcionales. Tome como ejemplos la flora de un bosque (taxonómico) y la del zooplancton (funcional). Los principios del Jackknife y del Bootstrap son utilizados para examinar el proceso que refleja. Estos tratan de establecer el número de especies faltantes en la muestra. Para ello se evalúa en cada re-muestra el número de especies faltantes para estimar cuantas serían no detectadas.

El ajuste de una ecuación que describa la acumulación de las especies permite graficar una curva, la que se utiliza para predecir el número de especies. Los expertos utilizan a veces inopinadamente modelos como la ecuación de Michaelis-Menton. Esta describe la reacción de las enzimas-catalizadas y no es claro que sirva igualmente para describir la riqueza de animales o plantas, aunque se haga con frecuencia. Tiene más sentido hacer un ploteo con los datos obtenidos y alisar la curva usando algún método numérico. Un ejemplo de los trabajos de Rennolls-Laumonier (2006), quienes trabajaron con especies de árboles en Sumatra. Ellos asumieron que las especies raras en los sitios eran una muestra aleatoria de estas en la selva y estimaron el número de especies ocultas.

Un método novedoso es el aplicado por Royle-Dorazio (2008) quienes usaron los criterios Bayesianos implementados en el software comercial para fijar la probabilidad de detectar las especies y con ello estimar el total de ellas. La acumulación de especies permite juzgar como los estimadores de la riqueza trabajan para nuestros datos. La línea recta es el resultado estimado, en el que las especies se equilibran. Las líneas azul y roja describen como se comporta la riqueza al aumentar el número de individuos observados en dos estimadores. Su comparación permite valorar las diferencias entre dos zonas estudiadas. La azul refleja un comportamiento muy bueno. La negra describe el proceso teórico de nivelación de la riqueza.

En general es necesario responder si hay más especies y cuantas son. Para ello será necesario inferir por interpolación. Por ejemplo si en un sitio observamos 80 insectos de 10 y en otro 174 de 22 se desea establecer si hay más especies en el primer sitio. Tomamos k submuestras (i) del segundo sitio de tamaño 80 y se mide el número de especies observadas, $n(i)$.

$$\bar{n} = \sum_{i=1}^k \frac{n(i)}{k}$$

Este método es conocido como rarefacción (rarefaction), ver Kemton (2002).

En la figura siguiente tenemos las curvas alisadas de en los dos sitios. Podemos hacer la interpolación necesaria en la figura 3.2, usando este grafico.. Es confiable utilizar la extrapolación solo si se tienen tamaño muestral grande.

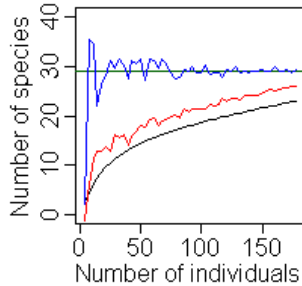


Figura 2.1. Ejemplo de un gráfico de la riqueza

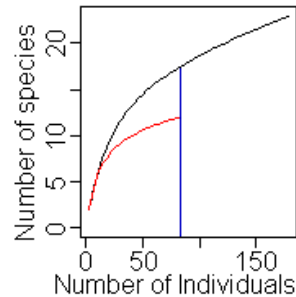


Figura Ejemplo de un gráfico de curvas alisadas

2.2.5 Estimación de la cobertura

El problema de estimar o predecir el número de clases en que se divide una población U está presente en muchas aplicaciones. Lo usual en la Teoría de Muestreo es que el decisor conoce las características que determinan las subpoblaciones y su cantidad. Este es el caso de la estratificación, la conglomeración y los dominios de estudio por ejemplo. Consideremos que U está formada por un número desconocido Δ de clases. Se toma una muestra $s \subset U$ usando muestreo simple aleatorio (MSA). Los objetos observados son clasificados en clases al ser identificados y considerando la similitud y diferenciación entre los grupos que determinan ellos. Este problema difiere del problema de la post estratificación sensiblemente. En la post estratificación conocemos los estratos y clasificamos los objetos después de tomar la muestra. Ahora lo que deseamos es conocer el número de clases. Podemos considerar que esta es una tarea de los métodos de aglomeración comunes en la paquetería estadística. Sin embargo el problema de fijar un valor Δ es de gran importancia en muchas aplicaciones. Si consideramos los problemas de la arqueología veremos que ese es el objetivo del investigador en ciertos momentos. Él analiza las piezas encontradas, determina grupos y piensa que el número de grupos formados es menor que los que existieron. En particular en los problemas medio ambientales es común una problemática similar pues se quiere conocer cuántos contaminantes están afectando el hábitat. Algunos ejemplos de problemas tipos son:

- Identificar el número de especies en una región o los contaminantes presentes en los desperdicios que una empresa bota en un vertedero natural.
- Determinar el léxico personal de un autor o las enfermedades causadas pro el nivel de toxicidad en el aire de una región de salud.
- Fijar los defectos presentes en las unidades de una línea de productos de una fabrica o las malformaciones en la vegetación circundante a una central átomo eléctrica.

-Establecer los tipos de plaga que están presentes en un cultivo o en las fuentes de los polutantes observados en un río.

El trabajo pionero en esta temática se remonta al trabajo del fundador de la estadística, Fisher et. Al. (1943) quien lo desarrolló estudiando un problema de Ecología aplicada. Posteriormente ese ha estudiado problemas diversos. Son muy conocidos los de Efron-Thisted (1976), quienes estudiaron el léxico de Shakespeare para establecer la autenticidad de ciertas obras, los de Engen (1978), Chao (1981) y Brained (1982) quienes continuaron la investigación de modelos para hacer una estimación de Δ en unos problemas biológicos. Otro enfoque es considerar la existencia de un proceso estocástico que explica la aparición de las clases en la muestra. Esto es más popular al considerar el arribo de especies a un punto de monitoreo, vea Chao-Bunge (2002) y Chao-Lee (1992). El problema teórico que este problema plantea varias interrogantes a la estadística. Muchas de ellas no están resueltas. Una línea es determinar cotas para Δ o estimarlas. Otra es estudiar el cubrimiento obtenido, por la muestra seleccionada, de las clases desconocidas a priori. En este trabajo se abordará este último problema. Los estimadores propuestos por Good (1953) y por Esty (1986) serán estudiados y se proponen alternativas. Tanto Good (1953) como Esty (1986) desarrollaron Teorema Centrales del Límite para determinar intervalos confidenciales basados en la aproximación normal. Sus hipótesis son algo fuerte al basarse en que la observación de las clases es descrita por un modelo de equiprobabilidad. En este trabajo proponemos el uso de un modelo superpoblacional del tipo regresión. Su aplicación en este caso conlleva un Modelo Lineal Generalizado como descriptor del comportamiento del cubrimiento. Esto no depende de la hipótesis de equiprobabilidad para garantizar la distribución normal asintótica del predictor.

Para analizar el comportamiento de las diversas alternativas para caracterizar el cubrimiento de la muestra utilizamos datos de dos investigaciones. Una son los datos obtenidos del estudio de la infestación de campos de caña por una plaga en Cuba. El otro es un estudio de biodiversidad forestal en México.

Para modelar este problema Bouza- Covarrubias () hicieron la partición de la población U como

$U = \cup_{i=1}^{\Delta} U_i$, $U_i \cap U_j = \emptyset$, si $i \neq j$. Al seleccionar una muestra s de tamaño n observamos D subpoblaciones. El modelo estocástico considera que hay Δ urnas y estas son ocupadas al azar observándose $D \leq \Delta$ ocupadas. Claramente, D es un estadígrafo suficiente para Δ . El cubrimiento de una muestra es definido por el parámetro $\theta = \sum I_t \pi_t$, donde $\pi_t = \text{Prob}(U_t \text{ aparece en la muestra})$ y

$$I_t = \begin{cases} 1 & \text{si el número de unidades de } U_t \text{ en } s \text{ es mayor que } 0 \\ 0 & \text{en otro caso} \end{cases}$$

Note que θ es un valor relacionado con la muestra y no con la población, pues es el cubrimiento de s . Este problema aparece también en los estudios de biodiversidad e lo que le índice está condicionado a la muestra observada.

Good (1953) propuso estimar θ mediante $\theta_1 = 1 - P_1$, $P_1 = n_1 / n$, siendo n_1 el número de clases observadas una vez. Esto parece sorprendente a primera vista pero es el índice de cubrimiento usado por Engen (1978), Starr (1979) y Chao (1981).

Otra solución fue propuesta por Esty (1986) al considerar que el número de clases observadas y la real se relaciona mediante la ecuación $D = \Delta^* [1 - \exp(-n/\Delta^*)]$ y usar como estimador $\theta_2 = D/\Delta^*$.

La conocida convergencia de ambos estimadores a θ se resume en el siguiente teorema:

Teorema 1. (Engen, Starr, Esty). Si $\pi_t = \pi$ para toda clase U_t , $n \rightarrow \infty$, $n/\Delta \rightarrow \infty$, $n/\Delta \rightarrow \alpha$ para $\alpha \in]0, 1[$ entonces se cumplen

1. $n^{1/2}(\theta_1 - \theta)$ tiene como distribución aproximada la normal $N\left(0, \frac{\alpha^2(1-\alpha)\ln(1-\alpha)}{\alpha + (1-\alpha)\ln(1-\alpha)}\right)$
2. $n^{1/2}(\theta_2 - \theta)$ tiene como distribución aproximada la normal $N\left(0, (1-\alpha)(\alpha - \ln(1-\alpha))\right)$.

Norris-Metter (1993) estudiaron el caso en que se seleccionan muestras con el fin de estimar el cubrimiento. Los resultados observados en s generan las variables Bernoulli

$$Z_{it} = \begin{cases} 1 & \text{si la clase } t\text{-ésima se observa en la selección } i \\ 0 & \text{en otro caso} \end{cases}$$

y podemos calcular $n_t = \sum_{i=1}^n Z_{it}$. Esta es un Binomial con parámetros n y π_t . Se N_j el número de clases observadas j veces en la muestra. Analizando N_0 , el número de clases no observadas, su esperanza es

$$E(N_0) = \sum_{t=1}^{\Delta} P(n_t = 0) = \sum_{t=1}^{\Delta} (1 - \pi_t)^n \geq \sum_{t=1}^D (1 - \pi_t)^n$$

Norris-Metter (1993) desarrollaron un análisis Bayesiano para derivar cotas inferiores para $E(N_0)$. Nosotros tomaremos $p_t = n_t / n$, el que estima insesgadamente a π_t , podemos usar el predictor derivado de los resultados anteriores $N^*_0 = \sum_{t=1}^D (1 - p_t)^n$. Es claro que $D + N^*_0 \cong \Delta$ por lo que un predictor ingenuo, que subestima el número de clases es $\hat{\Delta} = D + N^*_0$

Este corrige el sesgo de D y tiene uno menor.

Partiendo de este resultado, si seguimos las ideas del procedimiento de Esty (1986), podemos utilizar

$$\theta_3 = \frac{D}{\hat{\Delta}_0} = \left(1 + \frac{\hat{\Delta}_0}{D}\right)^{-1}$$

para predecir el cubrimiento. Note que este predictor es una variante del estimador de Esty (1986). Además si $n \rightarrow \infty$, $(1 - \pi_t)^n \rightarrow \varpi_t$, $\varpi > 0$, para al menos un $t \in \{1, \dots, \Delta\}$, entonces θ_3 tiene una distribución aproximadamente normal con media

$$[\sum_{t=1}^D (1 - \pi_t)^n]^{-1} \text{ y varianza } V_3 = V(D) + V(\hat{\Delta}_0) - \rho \sqrt{V(D)V(\hat{\Delta}_0)}$$

donde ρ es el coeficiente de correlación entre D y $\hat{\Delta}$. Las hipótesis utilizadas en el Teorema 1 son muy fuertes en la mayor parte de las aplicaciones. En los ejemplos utilizados en la introducción la equiprobabilidad del fenómeno que genera la observación de las distintas clases es poco realista. Ciertos contaminantes y polutantes aparecen con menor frecuencia que otros, algunas plagas son más prolíferas que otras. Algunas dolencias respiratorias son más comunes ante el enrarecimiento del aire por una cierta emisión de gases, etc.

Es poco realista pensar que podremos acudir a los expertos para que nos afirmen que la equiprobabilidad es válida, lo que difícilmente ocurre, que fijen un valor de α y nos aseguren la convergencia de los parámetros envueltos para grandes valores de n . Es más sencillo o esperable que los expertos puedan dar ideas sobre la relación existente entre una variable que identifique la clase y la probabilidad de ser detectada. Un modelo superpoblacional suficientemente sencillo es tomar

$$Y_{it} = \begin{cases} 1 & \text{si la observación } i \text{ pertenece a la clase } t \\ 0 & \text{en otro caso} \end{cases}$$

y valorar si es aceptable que esta se relacione con π_t mediante el modelo usado por Pothoof et al. (1992)

$Y_{it} = \pi_t + \varepsilon_{it}$, donde $E(Y_{it}) = \pi_t$ y

$$E(\varepsilon_{it} \varepsilon_{it'}) = \begin{cases} \sigma_t^2 & \text{si } Y_{it} = 1 \\ 0 & \text{en otro caso} \end{cases}$$

Ahora tenemos la variable Bernoulli

$$I(i, t | s) = \begin{cases} 1 & \text{si } Y_{it} = 1 \\ 0 & \text{en otro caso} \end{cases}$$

y $\hat{\pi}_t = \frac{\sum_{i=1}^n Y_{it}}{n}$ entonces tenemos que es modelo-insesgado el predictor $\theta_t = \frac{\sum_{i=1}^D \sum_{t=1}^n Y_{it}}{n}$ pues

$E(\hat{\pi}_t) = \pi_t$ y dada la independencia de las observaciones $V(\theta_t) = \sum_{t=1}^D \pi_t (1 - \pi_t) / n$. Note que el nuevo estimador es una función lineal de variables Binomiales independientes. Siguiendo las ideas de Bouza (1996) bajo el modelo superpoblacional se tiene que si $n \rightarrow \infty$ entonces $n^{1/2} (\theta_t - \pi_t) \rightarrow N(0, \sum_{t=1}^D \pi_t (1 - \pi_t) / n)$

Note que si $\pi_t = \pi$ para todo t la eficiencia de este predictor comparado con sus contrapartes y

$$E_{14} = \frac{-\alpha^2(1-\alpha)\ln(1-\alpha)n}{\alpha + (1-\alpha)\ln(1-\alpha)D\pi(1-\pi)}$$

y

$$E_{24} = \frac{(1-\alpha)(\alpha - \ln(1-\alpha))n}{D\pi(1-\pi)}$$

Ambas razones son mayores que uno por lo que el predictor es más exacto.

3. INDICES DE BIODIVERSIDAD MAS USADOS

3.1 El Índice de Simpson

El índice diversidad de Simpson es una de las medidas más utilizadas. Una búsqueda en Google arroja 28,600,000 de entradas bajo el descriptor 'Simpson indexdiversity' disminuyendo al hacerla en castellano a 53,000 al usar 'índice biodiversidad Simpson', de ellas 21,300 corresponden a aplicaciones. Esto da una idea de su popularidad.

Veamos los resultados existentes al considerar el problema estadístico. En este estudio el índice es un parámetro que se estima de su estudio como estimador de un parámetro. Este índice toma en consideración el número de especies presentes y su abundancia valorando la proporción de individuos en un y mide su importancia para la diversidad.

Este índice es usado fundamentalmente en cuantificar la biodiversidad de un hábitat por lo que es popular en estudios para fijar el perfil de biodiversidad. Este fue propuesto por Simpson en 1949. Desde entonces se ha hecho variaciones del mismo.

Sea N_i el número de individuos de la subespecie 'i' en el sitio analizado y N el total de los existentes.

$$\lambda_{Simp} = \sum_{i=1}^k \{1 - \pi_i\} \pi_i = 1 - \sum_{i=1}^k \pi_i^2$$

Se han propuesto varias variantes del índice como el índice de diversidad de Simpson

$$1 - \lambda_{Simp} = \sum_{i=1}^k \pi_i^2$$

Este puede verse como la probabilidad de que dos individuos seleccionados aleatoriamente en un área pertenezcan a dos diferentes subespecies. Este es conocido como índice de concentración de Herfindall que es usado comúnmente en economía en estudios descriptivos. Aquí tiene sentido hablar de dominancia de una especie. Ver Bruckman (1998 y 1999).

El índice recíproco de Simpson es $1/\lambda_{Simp}$ el que valora el número de subespecies igualmente comunes que producirían el valor observado del índice de Simpson. Este índice es influenciado por los parámetros por ciento de cada especies presente y su riqueza, Para una cierta riqueza el índice decrece cuando los porcentos de cada especies tiende a ser igualitaria ($\pi_i=1/K$). El investigador tiene que observar los patrones existentes con cuidado para interpretar el significado de los valores del índice,

Es usual encontrar en la literatura un estimador ingenuo para el índice de Simpson.

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

el sesgo de este estimador es $bias(\hat{\lambda}_{Simp}) = -\sum_{i=1}^k \frac{\pi_i(1-\pi_i)}{n}$

Note que el sesgo equivale a la suma de las $var\{n_i\}$, Bouza y Schubert (2003) hicieron la corrección del estimador y calcularon su varianza. Gimaret-Carpentier et al (1998) en un estudio de diversidad además de proponer un estimador insesgado y una varianza basada en métodos no paramétricos. Bouza-Covarrubias (200) analizaron el

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

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

Un estimador de esta es

$$\begin{aligned} & \hat{Var}_1\left\{\hat{\lambda}_{Simp(MAScr)}\right\} = \\ & = \frac{2}{n(n-1)^2(n-2)(n-3)} \left[\sum_{i=1}^k \left\{ 2n^2(n^2 + n - 3)\hat{\pi}_i^3 - n(5n^2 - 3n - 3)\hat{\pi}_i^2 + 3n(n-1)\hat{\pi}_i \right\} - \right. \\ & \left. - n^3(2n-3)\left(\sum_{i=1}^k \hat{\pi}_i^2\right)^2 + 2(2n-3)\sum_{i<j} n^2 \hat{\pi}_i^2 \hat{\pi}_j + n^2 \hat{\pi}_i \hat{\pi}_j^2 - n \hat{\pi}_i \hat{\pi}_j \right] \\ & = \frac{2}{n} \left[\sum \left\{ C_1 \hat{\pi}_i^3 - C_2 \hat{\pi}_i^2 + C_3 \hat{\pi}_i \right\} - C_4 \left(\sum_{i=1}^k \hat{\pi}_i^2\right)^2 + \sum_{i<j} C_5 \hat{\pi}_i^2 \hat{\pi}_j + C_5 \hat{\pi}_i \hat{\pi}_j^2 - C_6 \hat{\pi}_i \hat{\pi}_j \right] \end{aligned}$$

Podemos un estimador sesgado de la varianza que utiliza una fórmula más sencilla. La condición de insesgades se pierde pero este es consistente. Por su parte un estimador

consistente de la varianza del estimador de Simpson es $Var_2\left\{\hat{\lambda}_{Simp}\right\}$

$$= \frac{2}{n(n-1)} \left(\sum_{i=1}^k \hat{\pi}_i^2 + 2(n-1)\hat{\pi}_i^3 - (2n-3)\left(\sum_{i=1}^k \hat{\pi}_i^2\right)^2 \right)$$

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

Estos resultados se extienden al establecer que un estimador asintóticamente insesgado es:

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

Bouza-Covarrubias (2005) apuntaron 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.

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. 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, como usamos muestreo aleatorio simple con reemplazo, es valido usar la distribución Binomial de estos con parámetros n_j y π_i

El estimador ($\hat{\lambda}_{SimpSep(MAScr)}$) y el conjugado ($\hat{\lambda}_{SimpConj(MAScr)}$) fueron desarrollados tomando

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

Derivándose

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

Cuyo error es

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

donde los $\hat{\lambda}_j$ son los estimadores para un sitio j .

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 desarrolló el estimador $\hat{\lambda}_{SimpConj(MAScr)} = 1 - \sum_{i=1}^k \frac{n_i^*(n_i^* - 1)}{n^*(n^* - 1)}$

cuyo error es

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

El estimador de este error es

$$\begin{aligned} & \text{Var}\left\{\hat{\lambda}_{SimpConj(MAScr)}\right\} = \\ & = \frac{2}{n^*(n^* - 1)^2(n^* - 2)(n^* - 3)} \left[\sum_{i=1}^k \left\{ 2n^{*2}(n^{*2} + n^* - 3)\pi_i^3 - n^*(5n^{*2} - 3n^* - 3)\pi_i^2 + 3n^*(n^* - 1)\pi_i \right\} - \right. \end{aligned}$$

$$-n^{*3}(2n-3)\left(\sum_{i=1}^k \hat{\pi}_i^2\right)^2 + 2(2n^*-3)\sum_{i<j} n^{*2} \hat{\pi}_i^2 \hat{\pi}_j + n^{*2} \hat{\pi}_i \hat{\pi}_j^2 - n^* \hat{\pi}_i \hat{\pi}_j$$

Tomando el vector paramétrico $\pi(U)=(\pi_1, \dots, \pi_K)^T$ como aleatorio al considerar el MSP

$$\begin{aligned} \wp: E_{\wp}(\pi_i) &= \theta_i, \quad V_{\wp}(\pi_i) = \sigma_i^2, \quad \forall i, i' = 1, \dots, K; \\ \text{Cov}_{\wp}((\pi_i, \pi_{i'})) &= \rho \sigma_i \sigma_{i'} = \sigma_{ii'} \quad \forall i \neq i', i, i' = 1, \dots, K. \end{aligned}$$

Note que estos supuestos nos permite utilizar los resultados asociados a los momentos de una Multinomial. Después de hacer las substituciones correspondientes y agrupado convenientemente y definiendo los coeficientes

$$\psi(j) = \prod_{i=0}^{j-1} (N-j), \quad j=1, \dots, 4$$

El error esperado bajo este MSP es

$$\begin{aligned} E_{\wp} V\{L_S\} &= \frac{2}{n(n-1)} \left[\sum_{i=1}^k \sigma_i^2 + \left(\frac{2n(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] + \\ &+ \frac{2}{n(n-1)} \left[\left(1 + \frac{4n-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 \right] + \\ &+ \frac{2}{n(n-1)} \left[\frac{2n(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n-3)}{N^4} \Psi(4) \sum_{i=1}^k \theta_i^4 \right] - \\ &- \frac{2}{n(n-1)} \left[\frac{(2n-3)}{N^4} \left(\Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) + \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right) \right] \end{aligned}$$

A partir de este resultado son deducidos con facilidad los errores bajo al selección de una muestra simple aleatoria con reemplazo de m sitios de muestreo. Tenemos dos predictores alternativos.

Para el índice separado de Simpson el error de este bajo \wp es

$$\begin{aligned} E_{\wp} V\left\{ \hat{\lambda}_{SimpSep(MAScr)} \right\} &= \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[\sum_{i=1}^k \sigma_i^2 + \left(\frac{2n(j)(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] \right\} + \\ &+ \frac{1}{m^2} \sum_{j=1}^m \frac{2}{n(j)(n(j)-1)} \left(1 + \frac{4n(j)-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \\ &+ \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[\frac{2n(j)(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n(j)-3)}{N^4} \Psi(4) \sum_{i=1}^k \theta_i^4 \right] \right\} \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \frac{(2n(j)-3)}{N^4} \left(\Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right\} \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \frac{(2n(j)-3)}{N^4} \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right\} \end{aligned}$$

El resultado para el estimador combinado

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

y su esperanza bajo el modelo es

$$\begin{aligned} E_{\varphi} V\left\{\hat{\lambda}_{SimpConj(MAScr)}\right\} &= \frac{2}{n^*(n^*-1)} \left[\sum_{i=1}^k \sigma_i^2 + \left(\frac{2n^*(4N-3)}{N^4}\right) \Psi(1) \sum_{i=1}^k \theta_i + \right] + \\ &+ \frac{2}{n^*(n^*-1)} \left[\left(1 + \frac{4n^*-N-2}{N^4} \Psi(2)\right) \sum_{i=1}^k \theta_i^2 + \frac{2n^*(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 \right] - \\ &- \frac{2}{n^*(n^*-1)} \left[\frac{(2n^*-3)}{N^4} \left(\Psi(4) \left(\sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 - \sum_{i=1}^k \theta_i^4 \right) + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right] - \\ &- \frac{2}{n^*(n^*-1)} \left(\frac{(2n^*-3)}{N^4} \right) \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \end{aligned}$$

Consideremos el MSP caracterizado por las distribuciones utilizadas en la sección anterior

$$\wp(B): E_{\varphi}(\pi_i) = \theta_i, \quad V_{\varphi}(\pi_i) = \theta_i(1-\theta_i)/N, \quad \forall i, = 1, \dots, K.$$

Los errores esperados se expresan como

$$\begin{aligned} E_{\varphi} V\left\{\hat{\lambda}_{SimpConj(MAScr)}_{SS} | B\right\} &= \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[\sum_{i=1}^k \frac{\theta_i + (N-1)\theta_i^2}{N} \right] \right\} + \\ &+ \frac{1}{m^2} \sum_{j=1}^m \frac{2}{n(j)(n(j)-1)} \left[\left(\frac{4n(j)-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \left(\frac{2n(j)(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] \\ &+ \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[\frac{2n(j)(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 \right] \right\} - \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left(\frac{(2n(j)-3)}{N^4} \right) \left(\Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} - \Psi(4) \sum_{i=1}^k \theta_i^4 \right) \right\} - \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left(\frac{(2n(j)-3)}{N^4} \right) \left(\Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right\} \end{aligned}$$

y

$$\begin{aligned}
E_{\varphi} V \left\{ \hat{\lambda}_{SimpConj(MAScr)} | B \right\} &= \frac{2}{n^*(n^*-1)} \left[\sum_{i=1}^k \frac{\theta_i + (N-1)\theta_i^2}{N} \sigma_i^2 + \left(\frac{2n^*(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] + \\
&+ \frac{2}{n^*(n^*-1)} \left(\frac{4n^*-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \\
&+ \frac{2}{n^*(n^*-1)} \left[\frac{2n^*(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n^*-3)}{N^4} \Psi(4) \left(\sum_{i=1}^k \theta_i^4 - \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 \right) \right] \\
&\quad - \frac{2}{n^*(n^*-1)} \left[\frac{(2n^*-3)}{N^4} \left(\Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) + \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right) \right]
\end{aligned}$$

Cuando N es suficientemente grande para aceptar que $Hn(j)/N \cong 0$ para $H < 5$ estos errores son aproximados adecuadamente por

$$E_{\varphi} V \{ L_{SS} | B \} \cong \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[(2n(j)-4) \sum_{i=1}^k \theta_i^3 - (2n(j)-3) \left(\sum_{i=1}^k \theta_i^2 \right)^2 \right] \right\}$$

cuando usamos el estimador separado y por

$$E_{\varphi} V \{ L_{SC} | B \} \cong \left\{ \frac{2}{n^*(n^*-1)} \left[(2n(j)-4) \sum_{i=1}^k \theta_i^3 - (2n(j)-3) \left(\sum_{i=1}^k \theta_i^2 \right)^2 \right] \right\}$$

Para el caso del MSP

$$\wp(U): E_{\varphi}(\pi_i) = 1/2, \quad V_{\varphi}(\pi_i) = 1/12, \quad \forall i, = 1, \dots, K.$$

se obtiene

$$\begin{aligned}
E_{\varphi} V \{ L_{SS} | U \} &= \frac{k}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left(\frac{1}{2N} + \frac{(N-1)}{4N} \right) \left(\frac{2n(j)(4N-3)}{2N^4} \right) \Psi(1)_i \right\} + \\
&+ \frac{1}{m^2} \sum_{j=1}^m \left(\frac{2}{n(j)(n(j)-1)} \right) \left(\frac{2n(j)(N-1)-4N+3}{8N^4} \right) k \Psi(3) - \frac{(2n(j)-3)}{16N^4} k \Psi(4) - \\
&+ \frac{k}{m^2} \sum_{j=1}^m \left(\frac{4n(j)-N-2}{4N^4} \Psi(2) \right) - \left(\frac{2}{n(j)(n(j)-1)} \right) \left[\frac{(2n(j)-3)}{N^4} \left(\frac{\Psi(4)}{16} + \frac{\Psi(3)}{16} \right) \right] \cong \\
&\cong \frac{k}{m^2} \sum_{j=1}^m \left(\frac{n(j)-1}{2n(j)(n(j)-1)} \right) + \left(\frac{2n(j)-3}{8n(j)(n(j)-1)} \right)
\end{aligned}$$

y

$$\begin{aligned}
E_{\varphi} V \{ L_{SC} | U \} &= \frac{2\Psi(1)}{n^*(n^*-1)} \left(\frac{1}{2N} + \frac{N-1}{4N} \right) \frac{2kn^*(4N-3)}{2N^4} + \frac{k\Psi(2)}{4N^4} (4n^*-N-2) + \\
&+ \frac{2k}{n^*(n^*-1)} \left[\left(\frac{2n^*(N-1)-4N+3}{8N^4} \right) \Psi(3) - \left(\frac{(2n^*-3)k\Psi(4)}{16N^4} \right) \right] - \\
&- \frac{2}{n^*(n^*-1)} \left[\left(\frac{(2n^*-3)}{N^4} \right) \frac{\Psi(3) + \Psi(4)}{16} \right] \cong k \frac{n^*-2}{2n^*(n^*-1)} - \frac{2n^*-3}{8n^*(n^*-1)}
\end{aligned}$$

Mientras que para

$$\wp(Be): E_{\varphi}(\pi_i) = \alpha_i(\alpha_i+1)\beta_i, \quad V_{\varphi}(\pi_i) = \alpha_i\beta_i(\alpha_i+\beta_i)^{-2}(\alpha_i+\beta_i+1) \quad \forall i, = 1, \dots, K$$

estos errores esperados son aproximados por

$$E_{\phi} V \left\{ \hat{\lambda}_{\text{SimpConjMAScr}} | Be \right\} \cong \sum_{j=1}^m \frac{2}{n(j)(n(j)-1)} \left(\frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} + \left(\frac{\alpha_i}{(\alpha_i + \beta_i)} \right)^2 \right) + (2n(j)-4) \sum_{j=1}^m \left(\frac{\alpha_i}{(\alpha_i + \beta_i)} \right)_i^3 - (2n(j)-3) \left(\sum_{i=1}^k \left[\frac{\alpha_i}{(\alpha_i + \beta_i)} \right]^2 \right)_i^2$$

y

$$E_{\phi} V \left\{ \hat{\lambda}_{\text{SimpConjMAScr}} | Be \right\} \cong \frac{2}{n^*(n^*-1)} \left(\frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} + \left(\frac{\alpha_i}{(\alpha_i + \beta_i)} \right)^2 \right) + (2n^*-4) \left(\frac{\alpha_i}{(\alpha_i + \beta_i)} \right)_i^3 - (2n^*-3) \left(\sum_{i=1}^k \left[\frac{\alpha_i}{(\alpha_i + \beta_i)} \right]^2 \right)_i^2$$

3.2 El Índice de Shannon

El índice de Shannon es a veces citado como de Shannon-Wiener Io de Shannon-Weaver⁴ y es uno de los mas populares junto con el de Simpson. Usando en Google aparecen 157,000 citas bajo el descriptor 'indexbiodiversity Shannon', 15,700 para 'indexbiodiversity Shannon-Wiener' y 8,920 para 'indexbiodiversityshannon-weaver'. Claude Shannon publicó un par de trabajos seminales para la Teoría de Información. Por su parte Weaver en su rol de administrador de la Rockefeller Foundation fue un mecenas de esta nueva teoría dándole soporte a las investigaciones. Por su parte Wiener colaboró a la divulgación de esta teoría a través de su trabajo en cibernética. Ambos son considerados como personajes importantes al considerar el desarrollo de la Teoría de la Información pues su labor se jugó un rol decisivo después de la Segunda Guerra Mundial. Esta teoría es una rama de las, matemáticas aplicadas y es muy usada en ingeniería eléctrica. Una vez fijados los conceptos fundamentales esta teoría es usada en enfoques diversos. Uno de sus usos está en la estadística inferencial, el procesamiento del lenguaje natural, en criptografía, neurobiología análisis de datos y otras ramas de carácter aplicado. Su concepto clave es la entropía.

La matemática que soporta esta teoría fue desarrollada para la termodinámica. Por tanto se vincula a labor de J. Willard Gibbs y Ludwig Eduard Boltzmann fundadores de la estadística mecánica y termodinámica Estas conexiones fueron estudiadas por Rolf William Landauer el que demostró que cuando la información se pierde en un circuito irreversible la información se convierte en entropía y la energía asociada se disipa como calor.

En el caso de la biodiversidad la entropía de la información es asociada a la distribución. Las especies se consideran símbolos y su frecuencia relativa como probabilidades. La bondad dada a este índice es que toma en cuenta el numero de especies y la rareza.

Tomando :

N_i = numero de individuos de la especie i en la población (abundancia de la especie i).

K =numero de especies (riqueza)

$N = \sum_{i=1}^K N_i$ = Total de individuos en la población

$\pi_i = N_i / N$ = abundancia relativa de la especie (probabilidad de observar un individuo de la especie al hacer una selección aleatoria)

El índice de biodiversidad de Shannon es $S = -\sum_{i=1}^K \pi_i \ln(\pi_i)$. En muchas ocasiones se utiliza el índice $S' = S + [(K-1)/2N]$. Note que para cualquier número de especies fijado su máximo ocurre cuando el número en cada especie es igual en el que se tiene que $S_{\max} = \ln(K)$.

La estructura de este índice es no lineal en los π_i 's, lo que dificulta su estudio mediante las técnicas estadísticas clásicas. Sin embargo dada la base teórica, proveniente de la Teoría de la Información, que le soporta es muy usado.

Al tomar una muestra de tamaño n estimaríamos el índice de Shannon, bajo el supuesto de que K es conocido, en un sitio de muestreo al evaluar el número de individuos de cada especie i en la muestra, n_i , entre los n observados

$$\hat{S} = - \sum_{i=1}^K p_i \ln(p_i)$$

donde $p_i = n_i/n$. El vector $\underline{n} = (n_1, \dots, n_K)$ se distribuye Multinomial con parámetros n y $\underline{\pi} = (\pi_1, \dots, \pi_K)$ por lo que este estimador es Máximo Verosímil por serlo el vector $\underline{p} = (p_1, \dots, p_K)$ para $\underline{\pi}$. Este tiene un sesgo negativo y su esperanza es $E(\hat{S}) = S - [(K-1)/2N] + O(n^{-2})$.

Estos resultados fueron deducidos por Pielou (1975). Note que es necesario que $n > K$ y que usando

$$\hat{S}' = \hat{S} + [(K-1)/2N] + O(n^{-2}).$$

tenemos un estimador asintóticamente insesgado.

Si una especie no es observada $n_i=0$ y el modelo no es válido. Un estimador fue propuesto por Chao et al. (2002) basado en la estimación de Horvitz-Thomson. Este

$$\text{es } \hat{S}(HT) = - \sum_{i=1}^K \frac{p_i \ln(p_i)}{1 - (1 - p_i)^n} I(A(i))$$

donde

$$I(A(i)) = \begin{cases} 1 & \text{si } i < \text{especie } i \text{ está en la muestra} \\ 0 & \text{si no} \end{cases}$$

Estos propusieron usar el cubrimiento de la muestra

$$C = \sum_{i=1}^K \pi_i I(A(i))$$

Este mide la fracción del total de la abundancia de las especies en la muestra, por lo que es la probabilidad condicional de descubrir una nueva especie, ver Good (1953).

Entonces el modelo Multinomial tiene parámetros n y

$$\pi_i^* = \frac{\pi_i}{C}, i = 1, \dots, K$$

El estimador que proponen es

$$\hat{S}(HT) = - \sum_{i=1}^K \frac{\hat{C}p_i \ln(\hat{C}p_i)}{1 - (1 - \hat{C}p_i)^2} I(A(i))$$

Evidentemente en ningún caso los estimadores son insesgados. Se clama por la consistencia o la insesgadez asintótica.

En los modelos que le usan se hace lo que podríamos considerar una especie de estimador del tipo conjugado, pues al tomar la muestra de m sitios, que lo usual los conteos sean

$$n_i = \sum_{j=1}^m n_{ij}, \text{ donde}$$

n_{ij} es el numero de ejemplares de la especie i observados en la muestra j

Si decidimos usar el método separado al tomar m sitios de muestreo se computaría la media de los estimadores, digamos

$$\bar{S}(s) = \frac{1}{m} \sum_{j=1}^m \hat{S}(j), \quad \hat{S}(j) = \text{estimador de Sen la muestra } j\text{ésima}$$

$$p_i(j) = n_i(j)/n_j$$

En este caso $n_i(j)$ es el numero de individuos de la especie i en la muestra j y n_j el tamaño de esta.

Contemporáneamente a través de métodos de remuestreo como Bootstrap y Jackknife se hacen los estudios basados en este indice.

3.3 El Índice de Fager

El índice inicialmente propuesto por Fager (1972) es

$$:\lambda''_F = [N(K+1) - J(K-J)]/2 - \sum_{i=1}^K N_i R_i = \lambda^*_0 - \sum_{i=1}^K N_i R_i$$

Donde $J \in [0, K)$ es un entero y R_1, \dots, R_K son rangos asignados a las especies de interés en orden decreciente a su importancia. Esto es fijado de antemano por el ecólogo. Este llamó la atención por sus propiedades desde sus inicios. Referencias al mismo aparecen en la literatura con frecuencia en las aplicaciones, ver por ejemplo Oehl et al (2003),

En el orden más teórico no son abundantes las contribuciones. Esto ocurrió en los años cercanos a la propuesta del indice. Podemos ver que Hayes (1978) desarrolló un estudio de Monte Carlo para reproducir muestras de un grupo original de mediciones. Lo que llamó indice de Fager del grupo recurrente se calculó y se analizó el efecto del número de muestras. La existencia de un sesgo negativo fue reportado cuando el tamaño de la muestra era pequeño, sin hacer una referencia a propiedades teóricas del

índice como estadístico. Bouza-Schubert [2004] propusieron la simple transformación del índice:

$$\lambda_F = [N(K+1) - J(K-J)]/2N - \sum_{i=1}^K W_i R_i = \lambda_0 - \sum_{i=1}^K W_i R_i$$

donde

$$\lambda_0 = [N(K+1) - J(K-J)]/2N$$

Esto permite obviar el desconocimiento de N pues en particular esto nos permite considerar que

$$\lim_{N \rightarrow \infty} [N(K+1) - J(K-J)]/2N = [K+1]/2.$$

En este modelo tomamos un sitio aleatoriamente y observamos una estructura muestral (n_1, \dots, n_k) , donde n_j es el número de individuo de U_j en el sitio muestreado. Note que la distribución de (n_1, \dots, n_k) es una Multinomial y que el número de individuos en la muestra es $n = \sum_{i=1}^K n_i$. Por tanto $E(n_i) = nW_i$ y $p_i = n_i/n$ estima la proporción W_i de la especie i . Fager (1972) asumió que $R_i = \text{Rank}(N_i)$ es fijado por el ecólogo de antemano. Nosotros asumiremos, de ser necesario, además que $R_i \cong \text{Rank}(nW_i)$ en forma tal que la relación $R_i \cong \text{Rank}(n_i) = r_i$ sea aceptable. Entonces podemos escribir

$$\lambda^{\circ}_F = \lambda_0 - \sum_{i=1}^K p_i r_i$$

Entonces si usamos rangos basados en el rango asignado en función de información muestral $E(\lambda^{\circ}_F) \cong \lambda_F$ y su varianza es:

$$V(\lambda^{\circ}_F) = \sum_{i=1}^K R_i^2 W_i (1 - W_i) / n = V^2 V_2 / n$$

Cuyo estimador insesgado es, ver Bouza-Schubert (2003):

$$V^{\circ}(\lambda^{\circ}_F) = \sum_{i=1}^K R_i^2 p_i (1 - p_i) / n.$$

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. Esto determina la necesidad de utilizar muestras. Kepton (2002) señala la necesidad de dar una interpretación teórica a los índices como medidas de diversidad a partir de estudios empíricos. El problema de estimar índices se enfrenta a la necesidad de desarrollar una teoría de muestreo particular pues no se cuenta con un marco muestra.

En el marco usualmente utilizado en la teoría de encuestas se utiliza una sucesión de muestras aleatoriamente independientes $s_h, h=1, \dots, m$, de tamaño no necesariamente igual. Entonces tendremos una sucesión de variables Multinomiales $n_h = (n_{1h}, \dots, n_{Kh})$, $h=1, \dots, K$ y

$$\lambda_{F,h} = \sum_{i=1}^K p_{i,h} R_i = \sum_{i=1}^K R_i \sum_{h=1}^m n_{ih} / n_h, \quad h=1, \dots, m$$

y estima insesgradamente el índice de Fager utilizando a

$$\lambda^*_{F} = \lambda_0 - \sum_{i=1}^K \lambda_{F,h}/m.$$

Ya que $n_{ih}/n_h = p_{ih}$ estima insesgadamente a W_i . Estos resultados permiten dar como valida la siguiente Entonces al tomar m muestras independientes mediante muestreo simple aleatorio

$$\lambda^*_{F} = \lambda_0 - \sum_{i=1}^K \lambda_{F,h}/m.$$

es un estimador insesgado de λ_F donde λ_0 es fijado por [2.1] y $\lambda_{F,h}$ por [2.4]. Su varianza es

$$V[\lambda^*_{F}] = V^2 \sum_{h=1}^m n^{-1}_{,h}/m$$

Un estimador insesgado de ella es

$$V^\circ(\lambda^*_{F}) = \sum_{i=1}^K R_i^2 \sum_{h=1}^m p_{ih}(1-p_{ih})/n_h m$$

Si analizamos la estructura del índice de Fager es claro que esta muestra similitudes con los estadísticos lineales de rango. Sin embargo K no es el tamaño de una muestra sino un parámetro. Esto hace que los teoremas que aparecen en la literatura para establecer la convergencia de su distribución no sean utilizables directamente. Un expresión condensada de los teorías que pueden ser vinculados con tales estadísticos es el que brindamos a continuación y que se basa en Chernoff et al. [1967] usando la notación de Jurečková-Sen [1996] y Friedst-Gray [1997].

Teorema 3.1. [Convergencia de la distribución de estadísticos lineales de rango] . Sean

- X_1, \dots, X_m una sucesión de variables aleatorias de una función de distribución F .
- $X_{(1)}, \dots, X_{(m)}$ la correspondiente de los estadísticos de orden .
- Una función continua $J:]0, 1[\rightarrow \mathfrak{R}$ tal que $|J(u)|$ y $|F^{-1}(u)|$ están acotadas por $au^{2-r/2r}(1-u)^{1-2/2s}$ para $a, r, s > 0$ y $0 < u < 1$. Entonces $T_m = \sum_{i=1}^m J[u/(m+1)]X_{(i)}$ se distribuye asintóticamente $N\{\mu_m, \sigma^2(J, F)\}$ donde
- $\mu^m = \int_0^1 J(u) F^{-1}(u) du$
- $\sigma^2(J, F) = \int_0^1 [J(u)F^{-1}(u)]^2 du - \left[\int_0^1 J(u)F^{-1}(u) du \right]^2$ □

Ideas sobre su demostración puede ser obtenidas en Ferguson [1999] y Caaperà-Cutsem [1988] por ejemplo.

La estructura de la componente aleatoria de [2.5] es similar a los de los estadísticos T_m llamados de "tipo L-estimador" pues podemos escribirle como $\sum_{i=1}^K \lambda_{F,h}/m = \sum_{h=1}^m [1/n_{i,h}] [\sum_{i=1}^K n_{i,h} R_i] / m$, haciendo $J[h/(m+1)] = 1/n_h$. Para satisfacer las hipótesis 1 y 2 del teorema necesitaríamos que $\sum_{i=1}^K n_{i,h} R_i = X_{(h)}$.

lo que lograríamos al hacer una ordenación de los subíndices basados en la muestra. Esto es factible pero la función de pesos no es constante pro que la hipótesis 3 no se cumple. Sin embargo esto nos hace suponer que el estimador puede seguir una distribución asintótica normal.

Siguiendo otro camino como n_{ih} sigue una distribución la Binomial $B\{n_h, W_h\}$ por lo que si n_h es suficientemente grande p_{ih} sigue asintóticamente una distribución normal $N\{W_i, W_i[1-W_i]/n_h\}$. Esto es generalmente valido pues al trabajar sobre U las probabilidades W_i son constantes para cada especie y las n_h 's son suficientemente grandes para aceptar la aproximación normal de la Binomial paara cada i y h . Como los rangos son fijados usando información a priori sobre la variable $Y_{ih} = R_i p_{ih}$ se distribuye asintóticamente $N\{R_i W_i, R_i^2 W_h [1-W_i]/n_h\}$.

Sea $Y_{ih}^* = Y_{ih} - R_i W_i$ y $\sigma_i^2 = R_i^2 W_h [1-W_i]$ y la variable $Z_h = \sum_{i=1}^m Y_{ih}^*$ sigue una distribución

$N\{0, \sum_{i=1}^m \sigma_i^2 / n_h\}$. Al tomar una muestra aleatoria independiente mediante msa

$$Z_m = \sum_{h=1}^m r_{hm} Z_h = \sum_{h=1}^m r_{hm} [\sum_{i=1}^K Y_{ih} - E\{Y_{ih}\}]$$

tiene esperanza igual a cero y como varianza a $V[Z_m] = V^2 [\sum_{h=1}^m 1/r_h] / m^2 = V^2 / m^*$ al tomar $r_{hm} = 1/m$ para todo $h=1, \dots, m$. Como los tamaños de las muestras n_h 's en general son grandes la sumatoria en ese termino debe ser pequeña por lo que si $m \rightarrow \infty$ los conjuntos $c[h] = \{Z_h | Z_h \notin [-\varepsilon V \sqrt{m}, \varepsilon V \sqrt{m}], \forall \varepsilon \approx 0\}$

tienen medida de Lebesgue nula. Entonces

$$\lim_{m \rightarrow \infty} [V/\sqrt{m^*}]^{-1} \sum_{h=1}^m \{E(Z_h^2) I(|Z_h| < \varepsilon V \sqrt{m})\} = 0.$$

Esto no permite argüir, ver Ferguson [1999], que Z_m cumple con las condiciones del Teorema de Lindberg-Feller lo que garantiza la convergencia de $Z_m \sqrt{m^*} / V$ a una distribución normal standard. Estos razonamientos nos llevan al aceptar como válida la siguiente proposición.

Proposición 3.2. La distribución asintótica de $\lambda^*_F = \lambda_0 - \sum_{i=1}^K \lambda_{F,h}/m$ es $N\{\lambda_F = \lambda_0 - \sum_{i=1}^K R_i, W_i, V\{Z_m\}$, si $m \rightarrow \infty$. \square

Condicionando a que el número de individuos en la muestra s es fijo se hace el análisis usual pero es claro que al seleccionar un sitio el numero de elementos en el no es conocido, por lo que es una variable aleatoria. Al analizar el problema que nos plantea este hecho se debe buscar un modelo apropiado considerando la existencia de una fuente de variación debida a la aleatoriedad de n . Este fenómeno se modela al definir un cierto modelo probabilístico como generador del numero de individuos en la clase de interés. Un acercamiento al problema es considerar que el sitio de muestreo es generado por un modelo en el que se asignan v unidades a la población Q y que este es un valor conocido o conocible. Por ejemplo se puede saber que hay vv unidades de biomasa

usando un espectrómetro de masa al analizar fotos de un satélite que dio un mapa de la zona y usar la variable

$$Z_C(u) = 1 \text{ si } u \in C$$

$$Z_C(u) = 0 \text{ en otro caso}$$

Esta es una variable Bernoulli con parámetro P. Entonces tomado C como la población de interés U $n(v) = \sum_{h=1}^v Z_C(u)$ se distribuye binomialmente con parámetro v y P.

Bouza-Covarrubias tomaron un modelo superpoblacional en el se ligán estos puntos de vista en el análisis. Si consideramos que el espectrómetro evalúa d variables y que $Z_U[u] = g[q_1, \dots, q_d] + \varepsilon$

Donde ε es una variable aleatoria con esperanza cero.

-El problema de estimar la biodiversidad lo podemos enfocar en la forma siguiente:

-La naturaleza genera una población Q con v unidades.

-Es de nuestro interés solo las especies en $U \subset Q$.

-Se toma muestras usando un diseño muestral aleatorio y se observa en cada selección n(v) unidades de U con las que se computa un índice de biodiversidad.

-El error de muestreo del índice depende del numero de individuos observados en la muestra n(v) que es una variable aleatoria.

-Es de interés estimar el error de estimación pero también conocer su esperanza la que depende del modelo superpoblacional g a través del comportamiento de n(v) que es descrito por un modelo Binomial.

Es claro que dada la naturaleza del problema n(v) sigue una $B[v, P]$ por lo que $E[n(v)|g] = vP$ y $Var[n(v)|g] = vP[1-P]$. Como la población fija es generada por g, al observar una sucesión de realizaciones y dada la convergencia de la Binomial a la normal tenemos, ver Friedst-Gray [1997] que el proceso puede ser descrito asintóticamente por una distribución $N\{vP, vP[1-P]\}$ pues

$$[n(v) - vP] \sqrt{v} \longrightarrow_L N\{0, vP[1-P]\}$$

donde la convergencia es en ley. Para una realización n(v) el error de λ_F es

$$V[\lambda_F] = \sum_{i=1}^K R_i^2 W_i [1 - W_i] / n(v).$$

El resultado central del trabajo de referencia es que si v es suficientemente grande para aceptar que

$$[n(v) - vP] \sqrt{v} \longrightarrow_L N\{0, vP[1-P]\} \text{ entonces } E[V(\lambda_F|g)] \cong [vP]^{-1} [1 + [1-P]/vP] \sum_{i=1}^K R_i^2 W_i [1 - W_i]$$

De lo anterior tenemos que el índice de Fager posee las mejores condiciones desde el punto de vista estadístico para ser usado en las inferencias. Puede notarse que este es menos citado por los especialistas en sus investigaciones aplicadas. Otros se han impuesto por la tradición de su uso, como el de Shannon, pero su significado es confuso a la hora de hacer estudios, como la riqueza, basado en modelos estadísticos. Prefieren hacer análisis sin base probabilística por lo que las inferencias son dudablemente creíbles. Podemos citar algunas aplicaciones donde este es usado. Cuando que los especialistas poseen mayor control sobre el problema el índice Fager es más popular. Por ello podemos verle usado más frecuentemente en algunas aéreas, como la medicina, donde se puede aquilatar mejor la importancia de las especies en el problema bajo estudio.

Esto ha ocurrido por ejemplo: al investigar Albert et al. (2004) la variación de la diversidad clonal en la población de *V. myrtillus* en diferentes Hábitats de una área geográfica y al estudiar el particionamiento de la variabilidad genética del aislamiento de DNA ISOLATION extraído usando el modificaciones del procedimiento CTAB. En estos estudios la concentración de DNA se llevó a cabo mediante fluorometría usando bisbenzimidazol. Ver Labarca- Paigen, 1980). La diversidad genotípica fue estudiada usando 3 índices entre los que se incluye el propuesto por (Fager, 1972)

Bonet et al. (2002) consideraron la biodiversidad de microorganismos existentes en el intestino delgado humano. Hay un reconocimiento reciente de que, tener un cuadro de esta biodiversidad, ayuda a la biología molecular, al acceder directamente a la diversidad bacteriarDNA, en la comunidad natural del intestino. Esto ha generado estudios directos de DNA PCR sobre la comunidad. Estos son usados crecientemente para dilucidar incógnitas sobre la diversidad de la comunidad dentro de diversos ecosistemas. Tal es el caso de experiencias variadas como las Whitford et al., (1998) sobre el rumen, de los digestores anaerobios de Godonet et al., (1997), en fuentes termales de Barns et al., (1994), de suelos de Liesack-Stackebrandt, (1992) y en los sedimentos marinos de Devereux-Mundfrom, (1994). En su estudio Bonet et al (2002) examinaron el efecto del número de ciclos de PCR de la estructura inferida de una librería, considerando el comportamiento de métodos matemáticos que permitieron comparar estadísticamente la librería del clonaje de rDNA. Su objetivo era establecer si existían diferencias significativas entre las librerías ampliadas de rDNA, por usar diferentes números de ciclos en el PCR. Utilizaron Bootstrap y pruebas Chi-cuadrado en el análisis de la distribución de clones. En el análisis se incluyó el índice de Fager, considerando que se basa en el número de movidas, para caracterizar las movidas requeridos para convertir la distribución de clones, observada entre los OTUs, en una distribución adecuada, citando la experiencia de (Fager, 1972). Los intervalos de confianza derivados son más creíblemente aceptables, dada las características del índice como estadístico

Menares- Sepúlveda (2005) estudiaron simultáneamente 80 estaciones oceanográficas distribuidas en 24 transectos perpendiculares a la costa. Se estudió la estructura comunitaria de peces y crustáceos demersales, en la zona centro-sur de Chile. Como base del estudio se considera que las especies marinas presentan afinidad. Estos calcularon los índices de diversidad de Shannon y de Fager entre otros. Usando la idea presentada por Fager (1957) de que valores por encima del 0,5 con respecto al máximo número de afinidades posibles. Utilizaron Análisis de Correlación Canónica utilizando la abundancia relativa normalizada de las especies registradas y que estuvieron

presentes a lo menos en un 20% del total de los lances. La similitud faunística se determinó utilizando las especies dominantes determinadas según el método de ordenamiento propuesto por Fager (1957), donde las especies se ordenan (rankean) de crecientemente del 1 al 10 en términos de las abundancia dentro de cada muestra. Estos rangos se suman y se dividen por el número de lances. Según el método Fager, de un total de 96 lances y 38 especies presentes se determinaron las 10 especies más importantes. Esto permitió el ranqueo para el estudio de la biodiversidad marina.

REFERENCIAS

1. Adams, J. E. and E.D. McCune (1979): Application of the generalised Jackknife to Shannon's measure of information used as an index of diversity. En "Ecological Diversity in Theory and Practice" (J. F. Grassie et al., eds) International Co-operative Pub. House, Fairland, Maryland., 150-165.
2. Albert, Thierry, Olivier Raspé and Anne-Laure Jacquemart (2004): Clonal diversity and genetic structure in *vaccinium myrtillus* populations from different habitats, Belg. Journ. Bot. 137: 155-162.
3. Allende, S. C. Bouza, Amita D. Chakrabortya (2013): STUDY OF THE COVERAGE OF PLANTS AND THE EVALUATION OF DEFORESTATION, En DEFORESTATION, CONSERVATION POLICIES, ECONOMIC IMPLICATION AND ENVIRONMENTAL IMPACT. (C. Bouza, EDITOR), 101-114. NOVA PUBLISHERS, N. YORK. ISBN: 978-1-62948-241-5 • e-book ISBN: 978-1-62948-242-2.
4. Almazán A., Urbán G., González R., Tapia J., Villerías S., Beltrán E., Almazán M. (2003), "Ordenamiento Ecológico Territorial de la Subcuenca del Río San Juan del Estado de Guerrero," **Reporte Técnico** **Reporte Técnico de Investigación**, SBEJ-UAGroI
5. BanjShafiei, A., M. Akbarinia, S. GholamiJalili, P. Azizi and S. Mohsen Hosseini (2006): Effect of fire on herb layer biodiversity in a temperate forest of Northern Iran. *Pakistan J. of Biological Sciences*. 9. 2273-2277
6. Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* 91, 1609±1613.
7. Bonnet, R., A. Suau, J. l Doré, G. R. Gibson and Matthew D. Collins (2002): Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *International Journal of Systematic and Evolutionary Microbiology*, 52, 757-763
8. Bouza, C. (1996): Linear rank tests derived from a superpopulation model. *Biometrical J*, 38, 497-506.
9. Bouza, C. (2009): Estudio del cubrimiento vegetal: comparación de alternativas basadas en conjuntos ordenados. Estudio del cubrimiento vegetal: comparación de alternativas. Editorial Universitaria (Cuba): Ciudad de La Habana, ISBN 978-959-16-1083-6
10. Bouza C. y D. Covarrubias (2005). ESTIMACIÓN DEL ÍNDICE DE DIVERSIDAD DE SIMPSON EN m SITIOS DE MUESTREO. *Inv. Operacional*, 26, **187-195**
11. Bouza C. y D. Covarrubias (2007): .PROPUESTA DE UN MARCO SUPERPOBLACIONAL PARA EL ESTUDIO DE LA DIVERSIDAD MEDIANTE LOS INDICES DE FAGER Y SIMPSON. *REVISTA INVESTIGACIÓN OPERACIONAL*, . 28, 25-37,

12. Bouza C. N.y L. Schubert (2003). The Estimation of The Biodiversity and The Characterization Of The Dynamics : An Application To The Study Of A Pest. (2003) *Revista De Matematica E Estatistica*. 21, 85-98/
13. Beran, H. (1994): Eine system Konforme Maßzahlbildung für Klassifikatorische Merkmale. *Operations Res. Spektrum*. 16, 81-88.
14. Beran, H. (1999): A generalisation and optimisation of a measure of concentration. In "Modelling and Decisions in Economics" (U. Leopold-Wildburger, U. et. al. Eds.) Physika Verlag, Heidleberg.
15. BjörnNordén, Heidi Paltto, Frank Götmark, KjellWallin (2007): Indicators of biodiversity, what do they indicate? – Lessons for conservation of cryptogams in oak-rich forest. *Biological Conservation*, 135, 369-379.
16. Brained, B. (1982): On the relation between the type-token and species problem. *J. Applied Prob.* 19, 785-793.
17. Brown JH, Lomolino MV. (2004) *Biogeography*. 3rd. Sunderland, MA: Sinauer Associates,
18. Bruckman, G. (1969). Einige Bemerkungen zur statischen Messung der Konzentration. *Metrika*, 14, 183-213.
19. Buckland, S. T. , A.E. Magurran, R.E. Green and R.M. Fewster (2005): Monitoring change in biodiversity through composite indices. *Phil. Trans. R. Soc. B* 360, 243–254
20. Bunnell F.L., B.G. Dunsworth, D.J. Huggard, and L.L. Kremsater.(2003): . Learning to sustain biological diversity on Weyerhaeuser's coastal tenure. The Forest Project, Weyerhaeuser, Nanaimo, BC.
21. Burgmann MA, Lindenmayer DB. (1998) *Conservation Biology for the Australian Environment*. Chipping Norton NSW: Surrey Beatty and Sons Pty, Ltd,
22. Caaperáa, Ph. Y B. Van Cutsem (1988) *Asymptotique :Méthodes et Modèles en Statistique Non Paramétrique*. Exposition Fondamental. Bordas, Paris.
23. Chao, A. (1981): On estimating the probability of discovering a new species. *Ann. Statistics*, 9, 1339-1342.
24. Chao, A. and T. Bunge (2002): Estimating the number of species in a stochastic abundance model. *Biometrics*. 58, 530-539.
25. Chao, A. and S. M. Lee (1992): estimating the number of classes in sample coverage. *J. American Stat. Ass.* 87, 210-217.
26. Chao, A. and Tsung-Jen Shen (2002): Nonparametric estimation of Shannon's index of biodiversity when there are unseen species in simple. *Environmental and Ecological Statistics*, 10, 429-443.
27. Cook, A.,Marion, G., Butler, A. & Gibson, G. (2007). Bayesian inference for the spatio-temporal invasion of alien species. *Bulletin of Mathematical Biology*69(6), 2005-2025.
28. Chernoff H., J.L. Gaswirht Y M.V. John. (1967): Distribution Of Linear Combination Of Functions Or Order Statistics With Applications To Estimation. *Ann. Math. Statitis*. 38, 52-72.
29. Ducluzeau, R. (1988). Role of experimental microbial ecology in gastroenterology. In *Microbial Ecology and Intestinal Infections*, .(E. Bergone-Berezin.Editor). , 7-26Berlin: Springer.
30. Efron, B. and R. Thisted (1976): Estimating the number of unseen species: how many words did Shakespeare know? *Biometrika*, 63, 435-447.
31. Engen, S. (1996): *Stochastic Abundance Models*. Halted Press, N. York.
32. Esty, W.W. (1986): The efficiency of Good's nonparametric coverage estimator.

33. Fager, E.W. (1957). Determination and analyses of recurrent groups. *Ecology*, 38: 586-595.
34. Fager, E. W.(1963): The composition of sea-water comparative and descriptive oceanography. *Communities of organism.s*. In: M. N. Hill (Ed.), *Thea Sea. Ideas and Observations on Progress in the Study of the Seas. II.* 415-437 .Interscience Publ., Londres
35. Fager E.W. (1972): Diversity: a sampling study. *American Naturalist*, 106, 293-310.
36. Fager, E.W. & J.A. McGowan. (1963). Zooplankton species groups in the North Pacific. *Science*, 140: 453-460.
37. Feller, W. (1966): *An Introduction to Probability Theory and Its Applications*. Vol. 2. J. Wiley, N. York.
38. Ferguson T. S. (1996): *A Course in Large Sample Theory*. Chapman and Hall. N. York.
39. Friedst B. and Gray L. (1997) *A Modern Approach To Probability Theory*. Birhäuser, Boston.
40. Fisher, R.A, A. S. Corbett and S.C.B. Williams (1943): The relation between the number of species and the number of individuals in a random sample of animal population. *J. of Animal E.* 12, 42-58.
41. Gimaret-Carpentier C, Péliissier R., Pascal J. and Houllier F. (1998) "Sampling Strategies For The Assessment Of Tree Species Diversity," *Journal Of Vegetation Science*, 9, 161-172.
42. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. (1997). Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl Environ. Microbi*, 63, 2802-2813.
43. Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*40, 237-264.
44. Gotelli, and R.K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379-391.
45. Johnson N. L. and S. Kotz . (1969), *Distributions in Statistics: Discrete Distributions* John Wiley & Sons, N. York.
46. Jurecková, J. y P.K. Sen (19976): *Robust Statistical Procedures: Asymptotic And Interrelations*. J. Wiley.N. York.
47. Kempton R. A. (2002) "Species Diversity," *Encyclopedia of Environmetrics*, 4, 2086-2092. (edsAbbel H. El-Shaarawi and W. Piegorsh) John Wiley & Sons. N. York.
48. Kevin J. Gaston and John I. Spicer. 2004. *Biodiversity: an introduction*, Blackwell Publishing. 2nd Ed.
49. Keylock, C. J. (2005): Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy. *Oikos*, 109, 203-207.
50. Kirchhoff T, Trepl L (2001) Vom Wert der Biodiversität – Über konkurrierende politische Theorien in der Diskussion um Biodiversität. *Zeitschrift für angewandte Umweltforschung Sonderheft 13/2001*: 27-44. S
51. Labarca C. and Paigen K., 1980. — A simple, rapid, and sensitive DNA assay procedure. *Anal. Biochem.* 102: 344-352.
52. Liesack,W. and Stackebrandt, E. (1992). Occurrence of novel groups of the

- domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J Bacteriol* 174, 5072-5078.
53. Lindenmayer, D.B. and J.F. Franklin. (2002):. *Conserving forest biodiversity: A comprehensive multiscale approach*. Island Press, Washington, DC.
 54. Lindenmayer, D.B., C.R. Margules and D.B. Botkin(2000): Indicators of biodiversity for ecologically sustainable forest management. *Conservation Biology* 14(4): 941-950.
 - MacKay, David J. C. (2003): *Information Theory, Inference, and Learning Algorithms* Cambridge: Cambridge University Press.
 55. Margalef, D.R. 1958. Information theory in ecology. *Gen. Syst.* 3: 36-61
 56. McIntosh, R.P. (1967): An index of diversity and the relation to certain concepts of diversity. *Ecology*, 48, 392-404
 57. Menares, Brunilda y José I. Sepúlveda (2005): Grupos recurrentes de peces y crustáceos demersales en la zona centro-sur de Chile *Invest. Mar.*, Valparaíso, 33(1): 91-100G, r2u0p0o5s recurrentes
 58. Mulder, B.S., B.R. Noon, T.A. Spies, M.G. Raphael, C.J. Palmer, A.R. Olsen, G.H. Reeves and H.H. Welsh.(1999):. The strategy and design of the effectiveness monitoring program for the northwest forest plan. USDA Forest Service, Pacific Northwest Station. PNW-GTR-437.
 59. Nehring K, Puppe C (2002) A theory of diversity. *Econometrica* 70:1155-1198.
 60. Soriano, silvia v. ,ana m. Manacorda, nora b. Pierangeli, maria c. Navarro, alejandro l. Giayetto, liliana m. Barbieri, lorena e. Lazzarini, marta c. Minvielle, maría s. Grenovero y juan a. Basualdo (2005): intestinal parasitosis in relationtosocioeconomicfactors and habitatconditions in children of neuquén, Patagonia, Argentina . *ParasitolLatinoam* 60: 154 – 161
 61. Stirling, Gray and Brian Wilsey (2001): Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. *The American Naturalist*, . 158, 286–299
 62. Oehl, Fritz, EwaldSieverding, Kurt Ineichen, Paul Mäder, Thomas Boller, and Andres Wiemken (2003): Impact of Land Use Intensity on the Species Diversity of ArbuscularMycorrhizal Fungi in Agroecosystems of Central Europe.
 63. Patil G. P. and Taile C. (1982) “Diversity as a concept and its measurement,” *Journal of the American Statistical Association* 77, 548-567.
 - Pappert, Rebecca A. J. L. Hamrick and Lisa A. Donovan (2000): Genetic variation in *Puerarialobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States.*American Journal of Botany*, 7:1240-1245.
 - Pielou, E.C. (1977): *Mathematical Ecology*. Wiley, N.York.
 - Pla L. y S. D. Matteucci2 (2001): Intervalos de confianza bootstrap del índice de biodiversidad de Shannon. *Rev. Fac. Agron. (LUZ)*. 18: 222-234.
 64. Pothoff, R.R. , M.A. Woodbury and KI.G. Manton (1992): Equivalent sample size and equivalent degrees of freedom for refinements weights under superpopulation models. *J. Amer. Stat. Ass.* 87, 383-396.
 65. Qifa Zhang, and R. W. Allard (1986) Sampling variance of the genetic diversity index, *The Journal of Heredity* 77, 54-55
 66. Rennolls, K; Y Laumonier. 2006. A new local estimator of regional species diversity, in terms of ‘shadow species’, with a case study from Sumatra. *J Tropical Ecology*22:321-329. Renyi, A. 1961. On measures of entropy and information. In: *Proceedings of the Fourth*
 67. Berkeley Symposium on Mathematical Statistics and Probability. University of

68. California Press, Berkeley, CA. Vol. 1. Pages: 547-561.
69. Rico-G.;Alejandro, Juan P. Beltrán A.; Adriana Álvarez D.; Eduardo Flórez D. (2005): Diversidad de arañas (Arachnida: Araneae) en el Parque Nacional Natural Isla Gorgona, pacífico colombiano. *Biota Neotrop.* 5,
70. Royle, J. A. and Dorazio R. M. (2008): Hierarchical modeling and inference in ecology : the analysis of data from populations, metapopulations and communities. Academic Press, Amsterdam.
71. Sanders, H.L., 1968. Marine benthic diversity: a comparative study. *Amer. Nat.*, 102: 243–282.
72. Scheiner, S.M. and J. Gurevitch (eds.) (1993): Design and analysis of ecological experiments. Chapman & Hall Publishing, new York, NY
73. Shannon, C.E. (1948a): A mathematical theory of communication". *Bell System Technical Journal*27: 379–423
Shannon, C.E. (1948b): A mathematical theory of communication". *Bell System Technical Journal*27: 623–656.
74. Simberloff, D. (1978). Use of rarefaction and related methods in ecology. In *Biological Data in Water Pollution Assessment: Quantitative and Statistical Analyses*, (Edited by K. L. Dickson, J. Cairns, Jr & R. J. Livingston. West Conshohockeneditors) , 150-165. American Society for Testing and Materials. Pennsylvania
75. Simpson, E. M. (1949): Measurement of diversity. *Nature.* 163, 688.
76. Smith E.P. (2002), "Ecological Statistics," *Encyclopedia of Environmetrics*, 2, 589-602. (eds Abdel H. El-Shaarawi and Walter W. Piegorsch) John Wiley & Sons. N. YorkSom, R.K. (1996): Practical Sampling Techniques. Marcel Dekker, N. York.
77. Sørensen, T.A. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, 5, 1–34.
78. Scheiner, S.M. and J. Gurevitch (eds.) (1993): Design and analysis of ecological experiments. Chapman & Hall Publishing, new York, NY
79. Shannon, C.E. (1948a): A mathematical theory of communication". *Bell System Technical Journal*27: 379–423
Shannon, C.E. (1948b): A mathematical theory of communication". *Bell System Technical Journal*27: 623–656.
80. Simberloff, D. (1978). Use of rarefaction and related methods in ecology. In *Biological Data in Water Pollution Assessment: Quantitative and Statistical Analyses*, (Edited by K. L. Dickson, J. Cairns, Jr & R. J. Livingston. West Conshohockeneditors) , 150-165. American Society for Testing and Materials. Pennsylvania
81. Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688.
82. Starr, N. (1979): Linear estimation of the probability of discovering a new species. *Ann. Of Statistics.* 7, 644-652.
83. TortoleroLow, Leonardo José , Dalmiro José Cazorla Perfetti, Pedro Morales Moreno Y María Eugenia Acosta Quintero (2008), Prevalencia De Enteroparásitos En Perros
84. Domiciliarios De La Ciudad De La Vela, Estado Falcón, Venezuela. *Revista Científica, FCV-LUZ* , 18, 312 - 319, 2008.
85. Valle, A. F. Borges y C. Rincones (2000) Principales ,malezas en cultivos de caña de azúcar enb el municipio Unión del estado de Falcón , Venezuela.

REv.FAc.Agron. (LUZ) 17, 52-62.

86. Valliant, R.; A. H. Dorfmany R. M. Royall.; (2000); Finite Population Sampling And Inference; A Prediction Approach; John Wiley, New York.
87. Whitford, M. F., Forster, R. F., Beard, C. E., Gong, J. &Teather, R. M. (1998). Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* 4, 153-163.
88. Whittaker, R. H. (1972). "Evolution and Measurement of Species Diversity". *Taxon*, 21 (2/3), 213-251.
89. Yoshioka, Paul M. (2008): Misidentification of the Bray-Curtis similarity index, Marine Ecology Progress Series, Mar EcolProgSer, . 368, 309–310.
90. Zahl, S. (1977): Jackknifing an index of diversity. *Ecology*, 58, 907-913.

91. vonWintzingerode, F., Goss bel, U. B. &Stackebrandt, E. (1997).Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21, 213±229.
92. Weitzman M L (1992) On diversity. *Quarterly Journal of Economics* 107:363-405.
93. Weitzman M L (1998) The Noah's ark problem. *Econometrica* 66:1279-1298