PROPUESTA DE UN MARCO SUPERPOBLACIONAL PARA EL ESTUDIO DE LA DIVERSIDAD MEDIANTE LOS INDICES DE FAGER Y SIMPSON

Carlos N. Bouza* y Dante Covarrubias **
*Universidad de La Habana, **Universidad Autónoma de Guerrero
Email: bouza@matcom.uh.cu

RESUMEN:
El problema de estimar un índice de diversidad es analizado utilizando el marco Bayesiano conformado por la teoría de los modelos superpoblacionales. Se estudian las propiedades fundamentales de la teoría correspondiente al utilizar los índices de Fager y Simpson. Se utilizan varias distribuciones teóricas para ilustrar el comportamiento de los errores esperados bajo los modelos analizados.

ABSTRACT:
The estimation of a diversity index is studied using the Bayesian framework provided by the superpopulation model theory. The main theoretical properties are studied when Fager and Simpson indexes are used. Different distributions are considered for illustrating the behaviour of the expected errors under the analysed models.
Key Words: Model unbiasedness, expected error, beta distribution, uniform distribution

Key Words: a priori distributions, Bayesian, expected sampling error

MSC 62D05

1. INTRODUCCIÓN

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 es igualmente desconocido.

Este índice propuesto por Simpson (1949) es uno de los más utilizados por los investigadores. Menos conocido es el de Fager que se basa en rangos. Ver Patil y Taille (1982) . Bouza-Covarrubias (2005a y 2005b) estudian estos índices utilizado el marco usual de muestreo. En este trabajo hacemos un estudio superpoblacional de los mismos utilizando clases de modelos populares en la literatura. En la sección 2 se presentan estos modelos. En la sección 3 se desarrolla la teoría ara el índice de Fager y en la 4 para el de Simpson. Los resultados obtenidos permiten que el experto valore cuan conveniente es el uso de uno de los estimadores alternativos propuestos por Bouza-Covarrubias (2005a y 2005b). Esto facilitaría la valoración de la conveniencia de utilizar uno de los índices en forma teórica al asignar valores a los parámetros de los modelos superpoblacionales

2. MODELOS SUPERPOBLACIONALES

El modelo básico de la teoría del muestreo se asocia a la búsqueda de una medida de probabilidad llamada diseño muestral. Esta permite seleccionar una muestra aleatoria s de un población finita {1,..,N} al asignarle la probabilidad d(s). Tomando a S como el conjunto de todas las muestras (espacio muestral), como el cardinal $|S<\infty$ podemos tomar como σ-álgebra el conjunto potencia 2^S. Así el problema probabílístico asociado a la selección de muestras aleatorias esta definido sobre $\{S, 2^S, d\}$ que es un espacio de probabilidad. El aspecto de la estimación parte de la existencia de una variable de interés Y la que determina, al evaluarse en todo $i\in U$, un vector $Y(U)=(Y_1 ,...,Y_N)^T$, parámetro universal o vector paramétrico. Cada Y es mensurable pero es desconocido su valor antes de tomar la muestra. Por un problema de formalidad utilizaremos un intervalo de definición para cada i lo que nos hace trabajar $¥\subseteq\Re^N$.
$¥=\{Y(U)\,|\,Y_i\in]a_i, b_i[, i=1,...,N\}$

El problema de estimar una función paramétrica $\theta:¥\longrightarrow\Re^K$, a partir de la selección de un muestra aleatoria $s\subset U$, es un problema de estimación estadística con características particulares dadas por la estructura del espacio de probabilidad con que se trabaja. Al obtener s se puede identificar los individuos de U que será evaluada Y . Denotemos $s=\{i_1,..,i_n\}$ medimos Y en cada $i_j\in s$. Así se obtiene el vector aleatorio $Y(s)=\left(Y_{i_1},...,Y_{i_n}\right)^T=\left(y_1,...,y_n\right)^T$ cuyas componentes son variables aleatorias. Un estimador de θ es una función de Y(s), $\theta^*(Y(s))$. Las propiedades de este se deducen a partir de las especificidades del problema probabílístico que se define en la teoría del muestreo evaluándose su precisión usando el llamado error de muestreo

$E((\theta^*-\theta)^2\ =ECM(\theta^*)$.

Este enfoque constituye la inferencia basada en el diseño, la que fue elaborada a partir de los trabajos d Godambe, Basu, Joshi y otros a partir de la segunda mitad el siglo pasado. El criterio de d-insesgadez es clave en este enfoque . Este parte de considerar que

$E_d(\theta^*)=\sum_{s\in S}\theta^*(s)d(s)=\theta$

es una propiedad deseable y que el error de muestreo , d-error, es definido usando este mismo marco como

$E_d(\theta^*-\theta)^2=\sum_{s\in S}(\theta^*(s)-\theta)^2 d(s)$

Esta es la medida de precisión popularmente utilizada para evaluar una estrategia muestral
{d, θ*}.. La decisión de preferir una estrategia a otra se asociará a la comparación de los errores de muestreo de ellas. Este es motivado por el hecho de que no existen estimadores óptimos, bajo condiciones suficientemente generales como para que sean de interés práctico. Este hecho es ampliamente discutido en los libros de texto modernos ver Brewer (2002), Chaudhuri- Stenger (1992) y Vaillant et. Al. (2000).

Otro enfoque muy utilizado para evaluar estrategias es el concepto de Superpoblación. Este considera Y(U) como generado por un mecanismo aleatorio ה que es caracterizado por una distribución a priori perteneciente a una clase $\Phi=\{\phi\}$. En general estos se caracterizan por una serie de relaciones estructurales de las esperanzas de funciones de las variables . Utilizándoles se comparan las estrategias al calcular las esperanzas de los errores de muestreo de las estrategias., $E_ה[\ ECM_d\ (\theta^*)]$.

Godambe (1955) estudió este problema dando inicio al desarrollo unificado de la teoría del muestreo. Sin embargo este fue usado anteriormente por diversos estadísticos como Farfield- Smith en 1938 y Cochran en 1939 entre otros, vea Vaillant et.al (2000). Este enfoque es de naturaleza Bayesiana por ser una clase de distribuciones apriorística para determinar la esperanza a priori del error de muestreo de una estrategia muestral. Como este no trata de controlar el riesgo a posteriori es catalogado como Bayes –Parcial o Bayes-Ecléctico. Esto es evidente cuando se nota que el objetivo de un estudio basado en el enfoque superpoblacional no tiene por finalidad derivar un estimador Bayesiano.

Un modelo superpoblacional suficientemente general, ver Chaudhuri-Stenger (1992) es

n: $E_n(Y_i)=\mu_i$, $V_n(Y_i)=\sigma_i^2$ y $C_n(Y_i Y_j)=\rho\sigma_i \sigma_j$

Los parámetros envueltos son desconocidos en general. Se considera en ocasiones la existencia de una variable auxiliar X, conocida para todo $i\in U$. Los parámetros superpoblacionales son descritos por un modelo que depende de X. Es muy utilizado considerar que $\mu_i=\alpha_i+\beta X_i+\varepsilon_i$, donde α_i y β pueden ser conocidos o no pero $X_1,.., X_N$ si lo es. Denotamos este como m(1) .

La línea usual de análisis es asumir que un modelo superpoblacional general Y y comparar los errores de muestreo esperados bajo el modelo para evaluar el comportamiento de varias estrategias alternativas. Esto permite seleccionar una estrategia o mejorar la seleccionada en el sentido de que su error esperado bajo el modelo superpoblacional es disminuido e idealmente minimizado. El teorema siguiente brinda una serie de resultados clásicos al estimar el total

$T=\sum_{i=1}^{N} Y_i$

Teorema 1. Sean α , b(s) y b(s) funciones independientes de Y(U) y d(n) un diseño que le asigna probabilidades cero a las muestras de tamaño diferente de n y $P_i=fX_i$ la probabilidad de observar un individuo $i\in U$.

.$t_1=b(s)+\sum_{i\in s} b(s_i)y_i$

con $E_d(t_1)$ caracteriza la clase de los estimadores lineales insesgados de T.

Si m(1) es el modelo que genera Y

- $t'_1=\sum_{i\in s}(y_i-\alpha_i-\beta X_i)/nX_i+\sum_{i=1}^{N}\alpha_I+-\beta X_i$

pertenece a esta clase (Basu , 1978) , si f=n/N

- .$t_2=N\sum_{i\in s}(y_i-\alpha_i)/nX_i+\sum_{i=1}^{N}\alpha_i$

Entonces

- $E_{m(1)}V_{d(n)}(t_1)\geq E_{m(1)}V_{d(n)}(t_2)$

Sea

- $t_3=\sum_{i\in s} b(s_i)y_i$

(Godambe- Thompson , 1977), el estimador que caracteriza la clase de los estimadores lineales homogéneos insesgados . Si m(2) es el MSP en el que $\sigma_{ij}=0$ para todo $i\neq j$ y $\mu_i>0$,

- $E_{m(2)}V_{d(n)}(t_3)\geq V$

Cuando los Y_i son independientes el MSP m(3) es asumido Definiendo

- $t_4=\sum_{i\in s}(y_i-\mu_i)/P_i+\mu$

si $P_i=n\mu_i/\mu$

- $E_{m(3)} V_{d(n)} (t_3) \geq E_{m(3)} V_{d(n)} (t_4)$

(Ho, 1980 y Godambe , 1955) donde

$$V = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1-\pi_i}{\pi_i} \right) + E_d E_{\mathbf{m(2)}} (t_3 - T)^2$$

Si el MSP m(4) identifica las relaciones

$E_{m(4)}(Y_i)=\mu_i=\alpha_i+\beta X_i$ $V_{m(4)}(Y_i)=\sigma g_i$,$g_i>0$, $P_i=ng_i/g$,
.$g=\sum_{j=1}^{N} g_j$

es insesgado

- $t_5= \sum_{i\in s} (y_i -\alpha_i-\beta X_i)/P_i + \sum_{i=1}^{N}\alpha_i -\beta X_i$

Si α=0 y β no se conoce, Tam (1984)

- $E_{m(4)}V_{d(n)} (t_1) \geq E_{m(4)}V_{d(n)} (t_5)$

  Y

- $t´_5= \sum_{i\in s} y_i/P_i + B(\sum_{i=1}^{N}X_i- \sum_{i\in s} X_i/P_i)$

donde B es una estimación de β, entonces $t´_5$ es robusto. □

Otras modificaciones de las relaciones expresadas en los MSP´s vistos anteriormente han generado diversas posibilidades de comparación de estrategias muestrales.

3. ESTUDIO DEL ÍNDICE DE FAGER BAJO EL ENFOQUE SUPERPOBLACIONAL

Bouza y Schubert (2004) y Bouza –Covarrubias (2005 a)] han estudiado el índice de biodiversidad debido a Fager (1972). Este es definido como

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

donde $J\in[0,K)$ es un entero y $R_1,...,R_K$ son rangos asignados a las especies de interés en orden decreciente a su importancia .

Considerando que la población está estratificada de acuerdo a $U=U_1+....+U_K$, donde U_i es el conjunto de individuos de la especie i en la comunidad. N_i denota el número individuos de la especie i. Por lo que:

$N=\sum_{i=j}^{K} N_j$

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

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

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

El modelo usual considera que se selecciona un sitio aleatoriamente y se mide $\mathbf{n}=(n_1,...,n_k)^T$, donde

$n_j = \sum_{j\in s} I(j|i)$

Sean las variables aleatorias independientes con distribución Bernoulli de parámetro π_i :

$$I(j|i) = \begin{cases} 1 & si \quad j \in U_i \\ 0 & si \quad j \notin U_i \end{cases}$$

Su suma en j., n_j , es el número de individuos de U_i en el sitio muestreado. La distribución de
$\mathbf{n} = (n_1, ..., n_K)^T$ es una Multinomial y observamos en los sitios seleccionados, o sea en la muestra s,

$n = \sum_{i=1}^{K} n_i$.

individuos. Por tanto $E(n_i)=n\pi_i$ y $p_i=n_i/n$ estima la proporción poblacional de la especie i. Fager (1972) asumió que R_i =Rank (N_i) es fijado por el ecólogo de antemano. Entonces λ_F es estimado insesgadamente por

$$L_F = \lambda_0 - \sum_{i=1}^{K} p_i R_i \qquad (3.1)$$

Cuyo error de muestreo es:

$$V(L_F) = [\sum_{i=1}^{K} R_i^2 \pi_i - (\sum_{i=1}^{K} R_i \pi_i)^2]/n. \qquad (3.2)$$

De hecho al considerar un sitio de muestro usamos un modelo probabílístico caracterizado por que

$E(I(j|i)=\pi_i$. Para todo i=1,..,K y j=1,..,n.

Podemos considerar que la biodiversidad es generada por un MSP.

Consideremos que $\pi(U)=(\pi_1, ..., \pi_K)^T$ es generado por el modelo

$\wp$: $E_\wp(\pi_i)=\theta_i$, $V_\wp(\pi_i)=\sigma_i^2$, $\forall i, i'=1,\ldots,K$;
$Cov_\wp((\pi_{I,}(\pi_i)=\rho\sigma_i\sigma_{i'}=\sigma_{ii'} \ \forall\ i\neq i', i,i'=1,\ldots,K$.

Veamos cual es el error esperado de L_F bajo este MSP. Como

$$V(L_F) = [\sum_{i=1}^{K} R_i^2 \pi_i - (\sum_{i=1}^{K} R_i^2 \pi_i^2 + \sum_{i\neq=i'} R_i R_{i'} \pi_i \pi_{i'}).]/n \qquad (3.3)$$

Bajo $\wp$ tenemos que

$$E_\wp V(L_F) = [\sum_{i=1}^{K} R_i^2 \theta_i - (\sum_{i=1}^{K} R_i^2(\theta_i^2 + \sigma_i^2) + \sum_{i\neq=i'} R_i R_{i'} (\sigma_{ii'} + \pi_i \pi_{i'}).]/n$$

El que puede ser reducido a:

$$E_\wp V(L_F) = [\sum_{i=1}^{K} R_i^2 \theta_i (1-\theta_i) - (\sum_{i\neq=i'} R_i R_{i'} \sigma_{ii'} + \sum_{i=1}^{K} R_i^2 \sigma_i^2]/n \quad (3.4)$$

Al tomar un m sitios de muestreo usando el diseño muestreo simple aleatorio con reemplazo evaluamos una sucesión Ç={s(1),..,s(m) }de muestras y por eso tenemos dos formas de conformar un estimador del índice de Fager. Bouza-Covarrubias (2005) propusieron el uso del estimador separado

$L_F(S)=\sum_{s(t)\in Ç} L_F\,(s(t))=\lambda_0\text{-}m^{-1} \sum_{s(t)\in Ç} [\sum_{i=1}^{K} R_i\, n_i(t)/n(t)]$ (3.5)

Tomando $n_i(t)$ como el número de individuos de U_i que aparecieron en s(t) y

$\sum_{i=1}^{K} n_i(t)=n(t)$

Cuyo error de muestreo es

$V(L_F(S)) = m^{-2} \sum_{s(t)\in Ç} [[\sum_{i=1}^{K} R_i^2\, \pi_i - (\sum_{i=1}^{K} R_i^{\,2}\pi_i^{\,2} + \sum_{i\neq=i'} R_i\, R_{i'}\, \pi_i\, \pi_{i'}\,).]/n(t)]$ (3.6)

La otra alternativa de usar la información brindada por S es combinar los resultados y estimar cada π_i mediante

$.p^*_i = \sum_{s(t)\in Ç} n_i\,(t)/ \sum_{s(t)\in Ç} n(t)= \sum_{s(t)\in Ç} n_i\,(t)/\, n^*$

Como $E(n_i\,(t))= \pi_i$, $\forall i=1,..,K$, $t=1,..,m$ es insesgado el estimador

$L_F(C)=\lambda_0\text{-}m^{-1} \sum_{i=1}^{K} R_i\, p^*_i$

Cuya varianza es

$V(L_F(C))= [\sum_{i=1}^{K} R_i^2\, \pi_i - (\sum_{i=1}^{K} R_i^{\,2}\pi_i^{\,2} + \sum_{i\neq=i'} R_i\, R_{i'}\, \pi_i\, \pi_{i'}\,).]/n^*$ (3.7)

Esta estructura ayuda a establecer que los errores esperados bajo $\wp$ son

$E_{\wp} V(L_F(S))=m^{-2}\sum_{s(t)\in Ç}[\sum_{i=1}^{K} R_i^2\, \theta_i\,(1-\theta_i\,)\text{-}(\,\sum_{i\neq=i'} R_i\, R_{i'}\, \sigma_{ii'} + \sum_{i=1}^{K} R_i^2\, \sigma_i^2]/n(t)]$ (3.8)
$E_{\wp}\, V(L_F(C\,)=[\sum_{i=1}^{K} R_i^2\, \theta_i\,(1-\theta_i\,)\text{-}(\,\sum_{i\neq=i'} R_i\, R_{i'}\, \sigma_{ii'} + \sum_{i=1}^{K} R_i^2\, \sigma_i^2)]/n^*$ (3.9)

Pudiéramos considerar que las regularidades usadas en la fijación de los rangos por un especialista aseguran la independencia estadística de ellos. El modelo superpoblacional genera los rangos seria del tipo m(3). Sea este modelo

$\wp^*$: $E(R_i\,)=\tau_i$. $V(R_i\,)=\delta_i^2$,$\forall i=1,..,K$.
$E_{\wp^*}\, V(L_F) = [\sum_{i=1}^{K} \pi_i\,(\tau_i^2 + \delta_i^2\,)\text{–}(\sum_{i=1}^{K} \pi_i^2(\tau_i^2 + \delta_i^2\,) + \sum_{i\neq=i'} \pi_i\, \pi_{i'}\, R_iR_{i'}).]/n=V^*(L_F)$

Y obtenemos la expresión

$V^*(L_F)=[\sum_{i=1}^{K} \tau_i^2\, \pi_i - (\sum_{i=1}^{K} \tau_i\, \pi_i\,)^2\,]+ \sum_{I=1}^{K} \delta_i^2\, \pi_i\,(1-\pi_i)]]/n$

Analizando nuevamente los dos estimadores alternativos del índice de Fager al seleccionar m sitios de muestreo podemos calcular las esperanzas de los errores bajo el MSP. Estos son

$E_{\wp^*}V(L_F(S))=m^{-2}\sum_{s(t)\in Ç}[\sum_{i=1}^{K} \tau_i^2\, \pi_i - (\sum_{i=1}^{K} \tau_i\, \pi_i\,)^2\,]+ \sum_{I=1}^{K} \delta_i^2\, \pi_i\,(1-\pi_i)]]/n(t)]$ (3.10)
$E_{\wp^*}V(L_F(C))=[\sum_{i=1}^{K} \tau_i^2\, \pi_i - (\sum_{i=1}^{K} \tau_i\, \pi_i\,)^2\,]+ \sum_{I=1}^{K} \delta_i^2\, \pi_i\,(1-\pi_i)]]/n^*$ (3.11)

Note que para ambos MSP´s $L_F(S)$ es más preciso que $L_F(C)$ si

$m^{-2} \sum_{s(t)\in Ç} n(t)^{-1} \leq 1/n^*$

Para ejemplificar el comportamiento de los errores esperados consideremos que $\rho=0$ y que el MSP es descrito por una distribución particular.

Si una distribución Binomial esta generando poblaciones con $(\pi_1\,,..,\,\pi_K\,)$ y

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

(3.10) y (3.11) quedan como

$E_{\wp}V(L_F(S))=m^{-2}\sum_{s(t)\in Ç}[\sum_{i=1}^{K}R_i^2\,\theta_i\,(1-\theta_i\,)(N-1)N^{-1}-\sum_{i\neq=i'}R_i\,R_{i'}\,\theta_i\,\theta_{I'}\,]/n(t)]\cong V(L_F(S)\mid B)$
(3.12) .

Donde

$V(L_F(S)\mid B)= m^{-2}\sum_{s(t)\in Ç}[\sum_{i=1}^{K}R_i^2\,\theta_i\,(1-\theta_i\,)-\sum_{i\neq=i'}R_i\,R_{i'}\,\theta_i\,\theta_{I'}\,]/n(t)]$ (3.13)

Y para el combinado el resultado es

$E_{\wp}\,V(L_F(C\,)= [\sum_{i=1}^{K}R_i^2\,\theta_i\,(1-\theta_i\,)(N-1)N^{-1}-\sum_{i\neq=i'}R_i\,R_{i'}\,\theta_i\,\theta_{I'}\}]/n^*\cong V(L_F(C)\mid B)$

Siendo esta aproximación, cuando N≅N-1,

$V(L_F(C\,)\mid B)= [\sum_{i=1}^{K}R_i^2\,\theta_i\,(1-\theta_i\,)-\sum_{i\neq=i'}R_i\,R_{i'}\,\theta_i\,\theta_{I'}\}]/n^*.$ (3.14)

Si una distribución uniforme en el intervalo [0,1] es la generadora de los parámetros superpoblacionales

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

y los errores esperados son:

$$V\left(L_F(S)\middle|U\right)=\frac{\sum_{s(t)\in Ç}\frac{1}{n(t)}\left[\sum_{i=1}^{K}\frac{R_i^2}{6}-\sum_{i\neq i'}\frac{R_iR_{i'}}{4}\right]}{m^2}=$$
$$=\left(\frac{5K(K+1)(2K+1)}{72m^2}-\frac{K^2(K+1)^2}{16m^2}\right)\sum_{s(t)\in Ç}\frac{1}{n(t)} \qquad (3.15)$$

$$V\left(L_F(c)\middle|U\right)=\frac{\sum_{i=1}^{K}\frac{R_i^2}{6}-\sum_{i\neq i'}\frac{R_iR_{i'}}{4}}{n^*}=\frac{5K(K+1)(2K+1)}{72n^*}-\frac{K^2(K+1)^2}{16n^*} \qquad (3.16)$$

Bajo este MSP queda más clara la relación que establece cuando el estimador separado es más preciso que el combinado.

Considerando que en la Teoría Bayesiana la distribución Beta es tomada usualmente como la conjugada de la distribución Binomial la tomaremos en consideración . Para ella el modelo queda como

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

quedando las expresiones de los errores esperados bajo este nuevo MSP como

$$V\left(L_F(S)\middle|Be(\alpha_i,\beta_i)\right)=\frac{\sum_{s(t)\in Ç}\frac{1}{n(t)}\left[\sum_{i=1}^{K}\frac{R_i^2\left(\alpha_i\beta_i(\beta_i+1)\right)}{(\alpha_i+\beta_i)^2(\alpha_i+\beta_i+1)}-\sum_{i\neq i'}\frac{R_iR_{i'}\alpha_i\alpha_{'i}\beta_i\beta_{i'}}{(\alpha_i+\beta_i)(\alpha_{i'}+\beta_{i'})}\right]}{m^2}$$
(3.17)

$$V\left(L_F(C)\middle|Be(\alpha_i,\beta_i),i=1,..,K\right)=\frac{\sum_{i=1}^{K}\frac{R_i^2\left(\alpha_i\beta_i(\beta_i+1)\right)}{(\alpha_i+\beta_i)^2(\alpha_i+\beta_i+1)}-\sum_{i\neq i'}\frac{R_iR_{i'}\alpha_i\alpha_{'i}\beta_i\beta_{i'}}{(\alpha_i+\beta_i)(\alpha_{i'}+\beta_{i'})}}{n^*}$$
(3.18)

Si $\alpha_i=\alpha$ y $\beta_i=\beta$, $\forall i=1\ldots,K$ estas expresiones quedan simplificadas quedando

$$V\left(L_F(S)\middle| Be(\alpha,\beta)\right)=\left(\frac{5K(K+1)(2K+1)\alpha\beta(\beta+1)}{72m^2(\alpha+\beta)^2(\alpha+\beta+1)}-\frac{K^2(K+1)^2\alpha^2\beta^2}{16m^2(\alpha+\beta)^2}\right)\sum_{s(t)\in Ç}\frac{1}{n(t)} \quad (3.19)$$

$$V\left(L_F(C)\middle\| Be(\alpha,\beta)\right)=\left(\frac{5K(K+1)(2K+1)\alpha\beta(\beta+1)}{72n^*(\alpha+\beta)^2(\alpha+\beta+1)}-\frac{K^2(K+1)^2\alpha^2\beta^2}{16n^*(\alpha+\beta)^2}\right) \quad (3.20)$$

Los rangos son variables discretas y en la práctica estos son fijados por el especialista en forma subjetiva. El desconocimiento de estas nos puede llevar a suponer que ellos son generados por un modelo probabílístico en el que

$\wp$(R): $E_{\wp(R)}(R_i)=\xi_i$, $V_{\wp(R)}(R_i))=\varpi^2_i\ \forall$i, =1,…,K.

Bajo este MSP

$E_{\wp(R)}V(L_F) = [\sum_{i=1}^{K} (\xi_i+\varpi_i^2)\pi_i -(\sum_{i=1}^{K}(\xi_i{}^2+\varpi_i^2)\pi_i{}^2 + \sum_{i\neq i'}\xi_i\,\xi_{i'}\,\pi_i\,\pi_{i'})]/n$

Lo que nos queda explícitamente como

$E_{\wp(R)}V(L_F) = [\sum_{i=1}^{K} \xi_i\,\pi_i\,(1-\pi_i{}^2)+ \sum_{i=1}^{K}\varpi_i^2\ \pi_i\,(1-\pi_i{}^2)-[\sum_{i=1}^{N}\xi_i\ \pi_i]^2]/n$

Por tanto los errores esperados de los dos estimadores alternativos del índice de Fager

$E_{\wp(R)}V(L_F(S))=m^{-2}\sum_{s(t)\in Ç}[\sum_{i=1}^{K}\xi_i\,\pi_i(1-\pi_i{}^2)+\sum_{i=1}^{K}\varpi_i^2\,\pi_i(1-\pi_i{}^2)-[\sum_{i=1}^{N}\xi_i\,\pi_i]^2]/n(t)]$ . . . (3.21)

$E_{\wp(R)}V(L_F(C))=[\sum_{i=1}^{K}\xi_i\,\pi_i(1-\pi_i{}^2)+\sum_{i=1}^{K}\varpi_i^2\,\pi_i(1-\pi_i{}^2)-[\sum_{i=1}^{N}\xi_i\,\pi_i]^2]/n^*$ (3.22)

Si la probabilidad de observar un valor del rango esta asociado a la equiprobabilidad

$P(R_i=r)=1/K$ para todo i=1,..,K. Entonces

$\wp$(R | Equip.): $E_{\wp(R|Equip)}(R_i)=(K+1)/2$, $V_{\wp(R|Equip)}(R_i))=(K^2-1)/12\ \forall$i, =1,…,K.

establecen el valor de los parámetros superpoblacionales. Sustituyendo en (3.21) y (3.22) tenemos que

$$E_{\wp(R|Equip)}V\left(L_F(S)\right)=\frac{\sum_{s(t)\in Ç}\left(\frac{(K+1)(K+5)}{12}\sum_{i=1}^{K}\pi_i(1-\pi_i)-\frac{(K-1)^2}{4}\left(\sum_{i=1}^{K}\pi_i\right)^2\right)/n(t)}{m^2}$$

(3.23)

y

$$E_{\wp(R|Equip)}V\left(L_F(C)\right)=\frac{\frac{(K+1)(K+5)}{12}\sum_{i=1}^{K}\pi_i(1-\pi_i)-\frac{(K-1)^2}{4}\left(\sum_{i=1}^{K}\pi_i\right)^2}{n^*} \quad (3.24)$$

4. ESTUDIO DEL ÍNDICE DE SIMPSON BAJO EL ENFOQUE SUPERPOBLACIONAL

Otro índice de biodiversidad es el debido a Simpson. Este es

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

El que puede considerarse como un caso particular del de Fager . Note sus similitudes al tomar el mismo rango para todo i y usar como peso a 1-□ᵢ. De utilizar el método de la substitución tendremos como su estimador a

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

pero este posee como sesgo

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

Al hacer una corrección del estimador tenemos el estimador insesgado

$$L_S = 1 - \sum_{i=1}^{k} \frac{n_i (n_i - 1)}{n(n-1)}$$

cuya varianza es

$$V\{L_S\} = \frac{2}{n(n-1)} \left[\sum_{i=1}^{k} \pi_i^2 + 2(n-2)\pi_i^3 - (2n-3)\left(\sum_{i=1}^{k} \pi_i^2\right)^2 \right]$$

Ver Bouza-Covarrubias (2005b)

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

$\wp$: $E_\wp(\pi_i)=\theta_i$, $V_\wp(\pi_i)=\sigma_i^2$, $\forall i, i'=1,\ldots,K$;
$Cov_\wp((\pi_i, \pi_{i'}) = \rho\sigma_i \sigma_{i'} = \sigma_{ii'}$ $\forall$ $i \neq i'$, $i,i'=1,\ldots,K$.

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

$\psi(j)=\Pi_{j=0}^{j-1}(N-j)$, j=1,..,4

El error esperado bajo este MSP es

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

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. Como en el caso del Fager tenemos dos predictores alternativos.

El separado para el índice de Simpson está defino como

$$L_{SS} = \sum_{j=1}^{m} \frac{L_{Sj}}{m}$$

Siendo, su varianza considerando el mismo modelo Multinomial donde las n(j) en cada muestra son fijas,

$$V\{L_S S\} = \frac{1}{m^2}\sum_{j=1}^{m} V\{L_{Sj}\}$$

donde L_{Sj} es el índice de Simpson computado en la muestra j-ésima. A partir del resultado obtenido con la esperanza del error de L_S para un sitio y tomando en cuenta lo señalado en la sección anterior tenemos que el error de este bajo $\wp$ es

$$E_\wp V\{L_{SS}\} = \frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left[\sum_{i=1}^{k}\sigma_i^2 + \left(\frac{2n(j)(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\theta_i + \left(1+\frac{4n(j)-N-2}{N^4}\Psi(2)\right)\sum_{i=1}^{k}\right.\right.$$

$$+\frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left[\frac{2n(j)(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\theta_i^3 - \frac{(2n(j)-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\theta_i^4\right]\right\}+$$

$$-\frac{1}{m^2}\sum_{j=1}^{m}\{\frac{2}{n(j)(n(j)-1)}\left[\frac{(2n(j)-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\theta_i^2\theta_{i'}^2 + \Psi(3)\sum_{i\neq i'}\theta_i\,\theta_{i'}\left(\theta_i+\theta_{i'}\right) + \Psi(2)\sum_{i\neq i'}\right.\right.$$

El estimador combinado

$$L_{SC} = 1-\sum_{i=1}^{k}\frac{n_i^*(n_i^*-1)}{n^*(n^*-1)}$$

también es insesgado, ver Bouza-Covarrubias (2005b), la varianza es

$$V\{L_{SC)}\} = \frac{2}{n^*(n^*-1)}\sum_{i=1}^{k}\pi_i^2 + 2(n^*-2)\pi_i^3 - (2n^*-3)\left(\sum_{i=1}^{k}\pi_i^2\right)^2$$

y su esperanza bajo el modelo es

$$E_\wp V\{L_{SC}\} = \frac{2}{n*(n*-1)}\left[\sum_{i=1}^{k}\sigma_i^2 + \left(\frac{2n*(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\theta_i + \left(1+\frac{4n*-N-2}{N^4}\Psi(2)\right)\sum_{i=1}^{k}\theta_i^2\right]+$$

$$+\frac{2}{n*(n*-1)}\left[\frac{2n*(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\theta_i^3 - \frac{(2n*-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\theta_i^4\right]+$$

$$-\frac{2}{n*(n*-1)}\left[\frac{(2n*-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\theta_i^2\theta_{i'}^2 + \Psi(3)\sum_{i\neq i'}\theta_i\,\theta_{i'}\left(\theta_i+\theta_{i'}\right) + \Psi(2)\sum_{i\neq i'}\theta_i\,\theta_{i'}\right)\right]$$

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

$\wp$(B): $E_\wp(\pi_i)=\theta_i$, $V_\wp(\pi_i)=\theta_i(1-\theta_i)/N$, $\forall$i, =1,…,K .

Los errores esperados se pueden expresar como

$$E_{\wp}V\{L_{SS}|B\}=\frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left[\sum_{i=1}^{k}\frac{\theta_i+(N-1)\theta_i^2}{N}+\left(\frac{2n(j)(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\theta_i+\left(\frac{4n(j)-N-}{N^4}\right.\right.\right.$$

$$+\frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left[\frac{2n(j)(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\theta_i^3-\frac{(2n(j)-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\theta_i^4\right]\right\}+$$

$$-\frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left[\frac{(2n(j)-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\theta_i^2\theta_{i'}^2+\Psi(3)\sum_{i\neq i'}\theta_i\,\theta_{i'}\left(\theta_i+\theta_{i'}\right)+\Psi(2)\sum_{i\neq i'}\right.\right.\right.$$

y

$$E_{\wp}V\{L_{SC}|B\}=\frac{2}{n*(n*-1)}\left[\sum_{i=1}^{k}\frac{\theta_i+(N-1)\theta_i^2}{N}\sigma_i^2+\left(\frac{2n*(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\theta_i+\left(\frac{4n*-N-2}{N^4}\Psi(2)\right)\sum_{i=1}^{k}\theta\right.$$

$$+\frac{2}{n*(n*-1)}\left[\frac{2n*(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\theta_i^3-\frac{(2n*-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\theta_i^4\right]+$$

$$-\frac{2}{n*(n*-1)}\left[\frac{(2n*-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\theta_i^2\theta_{i'}^2+\Psi(3)\sum_{i\neq i'}\theta_i\,\theta_{i'}\left(\theta_i+\theta_{i'}\right)+\Psi(2)\sum_{i\neq i'}\theta_i\,\theta_{i'}\right)\right]$$

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

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

cuando usamos el estimador separado y por

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

Para el caso del MSP

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

se obtiene

$$E_{\wp}V\{L_{SS}|U\} = \frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{2}{n(j)(n(j)-1)}\left(\frac{1}{2N}+\frac{(N-1)}{4N}\right)\left(\frac{2n(j)(4N-3)}{2N^4}\right)k\Psi(1)_i + k\left(\frac{4n(j)-N-2}{4N^4}\Psi(2)\right)\right\}$$

$$+\frac{1}{m^2}\sum_{j=1}^{m}\left(\frac{2}{n(j)(n(j)-1)}\right)\left(\frac{2n(j)(N-1)-4N+3}{8N^4}\right)k\Psi(3) - \frac{(2n(j)-3)}{16N^4}k\Psi(4) +$$

$$-\frac{1}{m^2}\sum_{j=1}^{m}\ \{\frac{2}{n(j)(n(j)-1)}\ \left[\frac{(2n(j)-3)}{N^4}\left(\frac{\Psi(4)}{16}+\frac{\Psi(3)}{16}_{i'}\right)\right]k \quad \cong$$

$$\cong \frac{k}{m^2}\sum_{j=1}^{m}\left(\frac{n(j)-1}{2n(j)(n(j)-1)}\right)+\left(\frac{2n(j)-3}{8n(j)(n(j)-1)}\right)$$

y

$$E_{\wp}V\{L_{SC}|U\} = \frac{2}{n^*(n^*-1)}\left(\frac{1}{2N}+\frac{(N-1)}{4N}\right)\left(\frac{k2n^*(4N-3)}{2N^4}\right)\Psi(1)_i + k\left(\frac{4n^*-N-2}{4N^4}\Psi(2)\right)+$$

$$+\left(\frac{2}{n^*(n^*-1)}\right)\left[\left(\frac{2n^*(N-1)-4N+3}{8N^4}\right)k\Psi(3) - \frac{(2n^*-3)}{16N^4}k\Psi(4)\right]+$$

$$-\frac{2}{n^*(n^*-1)}\left[\frac{(2n^*-3)}{N^4}\left(\frac{\Psi(4)}{16}+\frac{\Psi(3)}{16}_{i'}\right)\right] \qquad \cong k\,\frac{n^*-2}{2n^*(n^*-1)} - \frac{2n^*-3}{8n^*(n^*-1)}$$

Mientras que para

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

estos errores esperados son

$$E_{\wp}V\{L_{SS}|Be\} = \sum_{j=1}^{m}\ \frac{2}{m^2 n(j)(n(j)-1)}\left[\sum_{i=1}^{k}\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i)(\alpha_i+\beta_i+1)}+\left(\frac{2n(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\ \frac{\alpha_i^{\ 2}}{(\alpha_i+\beta_i)^2}\right]+$$

$$+\left(1+\frac{4n-N-2}{N^4}\Psi(2)\right)\sum_{i=1}^{k}\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^2$$

$$+\sum_{j=1}^{m}\ \frac{2}{n(j)m^2(n(j)-1)}\left[\frac{2n(j)(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\ \left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^3 - \frac{(2n(j)-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\left[\frac{\alpha}{(\alpha_i}\right.\right.$$

$$-\sum_{j=1}^{m}\ \frac{2}{m^2 n(j)(n(j)-1)}\ \left[\frac{(2n(j)-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\ \left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)^2\left(\frac{\alpha_{i'}}{(\alpha_{i'}+\beta_{i'})}\right)^2_{i\,i'}\right)\right] \quad \cong$$

$$\cong \sum_{j=1}^{m}\ \frac{2}{n(j)(n(j)-1)}\left(\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i)(\alpha_i+\beta_i+1)}+\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)^2 + (2n(j)-4)\sum_{j=1}^{m}\ \left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^3 - (2n(j)-3)\left(\sum_{i=1}^{k}\right.\right.$$

y

$$E_{\wp}V\{L_{SC}|Be\} = \frac{2}{n^*(n^*-1)}\left[\sum\nolimits_{i=1}^{k}\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i)(\alpha_i+\beta_i+1)} + \left(\frac{2n^*(4N-3)}{N^4}\right)\Psi(1)\sum_{i=1}^{k}\frac{\alpha_i{}^2}{(\alpha_i+\beta_i)^2}\right] +$$

$$+\left(1+\frac{4n^*-N-2}{N^4}\Psi(2)\right)\sum_{i=1}^{k}\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^2$$

$$+\frac{2}{n^*(n^*-1)}\left[\frac{2n^*(N-1)-4N+3}{N^4}\Psi(3)\sum_{i=1}^{k}\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^3 - \frac{(2n^*-3)}{N^4}\Psi(4)\sum_{i=1}^{k}\left[\frac{\alpha_i}{(\alpha_i+\beta_i)}\right]_i^4\right]+$$

$$-\frac{2}{n^*(n^*-1)}\left[\frac{(2n^*-3)}{N^4}\left(\Psi(4)\sum_{i\neq i'}\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)^2\left(\frac{\alpha_{'i}}{(\alpha_{i'}+\beta_{'i})}\right)_{i\,i'}^2\right)\right] \cong$$

$$\cong \frac{2}{n^*(n^*-1)}\left(\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i)(\alpha_i+\beta_i+1)} + \left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)^2 + (2n^*-4)\left(\frac{\alpha_i}{(\alpha_i+\beta_i)}\right)_i^3 - (2n^*-3)\left(\sum_{i=1}^{k}\left[\frac{\alpha_i}{(\alpha_i+\beta_i)}\right]^2\right)_i^2\right.$$

REFERENCIAS

ARNOLD.D, B.C., N. BALAKRISHNAN AND H. N. NAGARAJA (1992): A First Course in Order Statistics. **Wiley,** N.York.

BASU , D. (1978): Relevance of randomization in data analysis. En **"Survey Sampling and Measurement",** 267-292. . (N.K. Naambodiri, Editor**).Academic Press**, N. York.

BOUZA C.N. Y D. COVARRUBIAS. (2005a): Estudio estadístico del índice de biodiversidad de Fager. Presentado en **el "6th International Workshop on Operations Research**", La Habana.

BOUZA, C.N. Y COVARRUBIAS, D.(2005B): Estimación del índice de diversidad de Simpson en m sitios de muestreo **Inv. Operacional. 26**, 187-195.

BOUZA C.N. AND L. SCHUBERT (2004): The estimation of bio-diversity and the characterization of the dynamics: an application to the study of a pest. **Rev. Matemática e Estatística. 21**, 85-98.

BREWER, K.(2002): Combined survey sampling inference. Weighing Basu's elephants**. Arnold**, London.

CHAUDHURY A. AND H. STENGER (1992): Sampling Survey. **M. Dekker**, N. York.

FAGER, E.W. (1972): Diversity: a sampling study**. American Naturalist. 106**, 293-310.

GODAMBE, V.P. (1955): A unified theory of sampling. J**. Royall Statistical. Soc. B, 17**, 269-278.

GODAMBE, V.P. Y M.E. THOMPSON (1976): Philosophy of survey-sample practice. . **Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science. Vol II**, 103-123 (W.L. Harper y C.A. Hooker Editores). D. Reidel Pub. Comp., Dordrecht .

GODAMBE, V.P. Y M.E. THOMPSON (1977): Robust near optimal estimation in survey practice. **Bull. International Statistical Institute. 47**, 129-146.

HO, E.W.H. (1980): Model-unbiasedness and the Horvitz-Thompson estimator in finite population sampling**. Australian Journal of Stat. 22**, 218-225.

KEMPTON R. A. (2002) “Species Diversity,” **Encyclopedia Of Environmetrics, 4**, 2086-2092. (Eds Abbel H. El-Shaarawi and W. Piegorsh**) John Wiley & Sons**. N. York

PATIL G. P. AND TAILLIE C. (1982) “Diversity As A Concept And Its Measurement,” Journal Of **The American Statistical Association 77**, 548-567.

TAM, S.M. (1984): Optimal estimation in survey sampling a under regression superpopulation model**. Biometrika, 71**, 645-664.

VALLIANT, R., DORFMAN, A. H. AND ROYALL, R. M. :(2000).: Finite population sampling and inference. A prediction approach. **Wiley Series in Probability and Statistics**., New York.

Received March 2006
Revised October 2006