Estudio de estimadores insesgados de algunas medidas de incertidumbre

POR RIGOBERTO PEREZ

Recibido: 10 de abril de 1985

Presentado por el académico numerario D. Francisco Azorín

Abstract

This paper is devoted to the estimation of the uncertainty, the useful uncertainty and the unquietness in finite populations when we consider the simple random sampligns with and without replacement.

The estimation is accomplished in four stages:

In the first stage, we examine the analogue estimators of the classical measures (or measures of the Shannon type) for a sample of size n, and conclude that there is not and exact relation between the expected values of those estimators and their corresponding population measures. This fact motivates us to study other measures for the considered problem. This study is developed in the second stage, in which we analyze the most relevant properties of the quadratic uncertaninty and introduce the quadratic useful uncertainty and unquietness.

In the third and fourth stages, we examine the last measures for the most elementary samplings in finite populations: simple random selection with replacement and unequal probabilities, and random selection without replacement and with graduated variable probabilities. Results in both cases show a satisfactory behavior.

Finally, and application of the preceding results to Economics is carrired out: the estimation of ıncome inequality in a population.

Resumen

El objetivo de este trabajo es la obtención de estimadores insesgados de la incertidumbre, la incertidumbre útil y la inquietud cuadrática en poblaciones finitas, a partir de una muestra de tamaño n cuando ésta se selecciona por un muestreo aleatorio con o sin reposición, así como estimaciones insesgadas de la varianza de estos estimadores. Existe, en cada caso, una ganancia de precisión cuando el muestreo es sin reemplazamiento.

Se desglosa esta investigación en cuatro apartados: en el primero, justificamos la inadecuación de las medidas tipo Shannon, en los procesos de estimación en poblaciones finitas. En el segundo, se estudian las propiedades de la incertidumbre de orden $\beta = 2$ y se introducen las medidas de incertidumbre útil e inquietud cuadráticas analizando las propiedades más notables de las mismas. En el tercer apartado examinamos el comportamiento de las medidas anteriores en las técnicas de muestreo mencionadas anteriormente.

En la última sección, se realiza una aplicación económica de la inquietud cuadrática como medida de desigualdad en la distribución personal de la renta.

1. INTRODUCCION

Consideremos una población finita E constituida por N individuos $\{w_1, ..., w_N\}$, sobre la cual está definida una variable X que asume unos valores $\{x_1, ..., x_M\}$ donde x_i se repite N_i veces

$$\left(\sum_i N_i = N\right),$$

así pues, este conjunto de valores lleva asociado un sistema de probabilidades $\{p_1, ..., p_M\}$, $(p_i = N_i/N)$. Supongamos además que cada valor x_i conlleva una utilidad positiva u_i $(u_i > 0)$ (no necesariamente en el sentido de Von Newmann), con lo que la variable X está dotada también de un sistema de un sistema de utilidades $\{u_1, .., u_M\}$.

Esta población (que podemos identificar con la variable X), entrañará un cierto grado de incertidumbre, incertidumbre útil e inquietud que tratamos de medir y estimar a partir de una muestra de tamaño $\boldsymbol{n}$, $\{w_1, ..., w_n\}$ sobre la que X *toma los valores* $\{x_1, ..., x_k\}$, con frecuencias absolutas $\{n_1, ..., n_k\}$,

$$\left(\sum_i n_i = n\right),$$

o frecuencias relativas $\{f_1, ..., f_k\}$, $(f_i = n_i/n)$. Esta muestra, que genéricamente denotamos por $\boldsymbol{x}$, lleva asociado un subsistema de utilidades $\{u_1, ..., u_k\}$.

Consideremos en primer lugar la estimación de la incertidumbre.

a) Estimación de la incertidumbre de Shannon

La medida de incertidumbre que comúnmente se utiliza (obteniéndose en muchos campos resultados satisfactorios), es la medida de entropía de Shannon. Es por ello que nos planteamos en primer lugar el comportamiento del estimador analógico de este indicador en los procesos de muestreo en poblaciones finitas.

Transcribiendo la definición de la incertidumbre de Shannon de una variable aleatoria $\boldsymbol{X}$, pueden establecerse las siguientes definiciones:

Definición 1.1. Llamamos *entropía poblacional de Shannon* relativa a la variable $\boldsymbol{X}$ (sobre la población $\boldsymbol{E}$), que denotamos por $H(\boldsymbol{X})$, al valor de la expresión:

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

Definición 1.2. Denominamos *entropía muestral de Shannon* de la muestra $\boldsymbol{x}$, que denotamos por $h(\boldsymbol{x})$, al valor de la expresión:

$$h(\boldsymbol{x}) = -\sum_{i=1}^{k} f_i \cdot \log f_i$$

El sesgo del estimador, dependerá de la técnica de selección empleada para obtener la muestra; en este trabajo se consideran los muestreos aleatorios con y sin reposición.

En un intento de obtener el sesgo de este estimador, calculemos su esperanza:

$$E[h(x)] = E\left[-\sum_{i=1}^{k} f_i \cdot \log f_i\right]$$

que introduciendo una variable auxiliar e_i, que nos indique el número de veces que el valor x_i es seleccionado en la muestra, resulta:

$$E[h(x)] = -\frac{1}{n} E\left[\sum_{i=1}^{M} e_i \log \frac{e_i}{n}\right] =$$

$$= H\left(\frac{1}{n}, ..., \frac{1}{n}\right) - \frac{1}{n} \sum_{i=1}^{M} E[e_i \log e_i]$$

Teniendo en cuenta la desigualdad de Jensen y la concavidad (hacia arriba) de la función $f(x) = x \cdot \log x$, se obtiene:

$$E[e_i \cdot \log e_i] \leqslant E(e_i \cdot \log E(e_i)$$

y no existe relación general exacta entre ambos miembros de la desigualdad; con lo cual no puede obtenerse un valor exacto de la esperanza de este estimador, ni por tanto el sesgo del mismo.

Nota 1. La distribución de la variable auxiliar e_i depende del proceso de selección. Cuando el muestreo es aleatorio con reposición, la distribución conjunta de la variable aleatoria $(e_1, ..., e_M)$ es polinomial, cuya función de cuantía es:

$$P(e_1, ..., e_M) = \frac{n!}{e_1! \ldots e_M!} p_1^{e_1} \ldots p_M^{e_M}; \qquad \left(\sum_{i=1}^{M} e_i = n\right) \qquad (*)$$

si el muestreo es aleatorio sin reemplazamiento, la distribución conjunta de las variables e_i es una hipergeométrica generalizada, cuya función de probabilidad es:

$$P(e_1, ..., e_M) = \frac{\binom{N_1}{e_1} \ldots \binom{N_M}{e_M}}{\binom{N}{n}}; \qquad \left(\sum_{i=1}^{M} e_i = n\right) \qquad (**)$$

(*) En el Apéndice I figuran los momentos de esta distribución.

(**) En el Apéndice II, figuran los momentos de esta distribución, también llamada poli-hipergeométrica o multi-hipergeométrica.

Nota 2. Si tratásemos de obtener directamente la esperanza del estimador, llegaríamos a obtener expresiones inmanejables que no nos permiten obtener el valor de tal esperanza. Para el muestreo con reposición, se obtiene:

$$E\Big[\sum_{i=1}^{k} n_i \cdot \log n_i\Big] =$$

$$= \sum_{\substack{[n_{1j}, \ldots, n_{Mj}],\ \sum_{i=1}^{M} n_{ij} = n \\ j,\ 0 \leqslant n_{ij} \leqslant \min\{N_i, n\},\ i = 1, \ldots, M}} \Big(\sum_{i=1}^{k} n_{ij} \log n_{ij}\Big) \cdot \sum_{\sigma \in [1, \ldots, M]} n! \prod_{i=1}^{M} \frac{p_i^{n_{\sigma(i)j}}}{n_{\sigma(i)j}!}$$

y para el muestreo sin reemplazamiento, se tiene:

$$E\Big[\sum_{i=1}^{k} n_i \cdot \log n_i\Big] = \sum_{\substack{[n_{1j}, \ldots, n_{Mj}],\ \sum_{i=1}^{M} n_{ij} = n \\ j,\ 0 \leqslant n_{ij} \leqslant \min\{M_i, n\},\ i = 1, \ldots, M}} \Big(\sum_{i=1}^{k} n_{ij} \log n_{ij}\Big) \cdot$$

$$\cdot \sum_{\sigma \in [1, \ldots, M]} \binom{N_1}{n_{\sigma(1)j}} \binom{N_M}{n_{\sigma(M)j}} \Big/ \binom{N}{n}$$

Nota 3. Los dos principales inconvenientes que no hacen operativa la medida de incertidumbre de Shannon en los procesos de muestreo son: la escasa manejabilidad de la función logarítmica en el muestreo finito (en el que invariablemente aparecen sumatorios), como se vio al calcular la esperanza del estimador, y el hecho de que muestras diferentes pueden proporcionar la misma incertidumbre muestral, motivo por el cual aparecen las complejas expresiones de la nota anterior.

b) Estimación de la incertidumbre útil (tipo Shannon)

Transcribiendo las definiciones de las medidas de incertidumbre útil dadas por P. Gil y M. A. Gil de un campo A, a una población finita $\boldsymbol{E}$ y a una muestra $\boldsymbol{x}$, se tiene:

Definición 1.3. Llamamos *incertidumbre útil poblacional*, relativa a la variable X, que denotamos por $HU(X)$ ó $HU(p_1, \ldots, p_M, u_1, \ldots, u_M)$ al valor de la expresión:

$$HU(X) = -\sum_{i=1}^{M} p_i \frac{u_i}{E(u)} \log p_i$$

donde

$$E(u) = \sum_{i=1}^{M} p_i u_i .$$

Definición 1.4. Denominamos *incertidumbre útil muestral*, relativa a la muestra x, y que denotamos por $hu(x)$ ó $hu(f_1, ..., f_k, u_1, ..., u_k)$, al valor de la expresión:

$$hu(x) = -\sum_{i=1}^{k} f_i \frac{u_i}{E_x(u)} \log f_i$$

donde por $E_x(u)$ denotamos la esperanza muestral de las utilidades.

Para la medida de M. A. Gil, se tiene:

Definición 1.5. Llamos *incertidumbre útil poblacional*, relativa a la variable X, y que denotamos por $HU(p_1, ..., p_M, u_1, ..., u_M)$ ó $HU(X)$, al valor de la expresión:

$$HU(X) = -\sum_{i=1}^{M} p_i \log \frac{u_i}{E(u)} p_i$$

Definición 1.6. Denominamos *incertidumbre útil muestral*, relativa a la muestra x, y que denotamos por $hu(f_1, ..., f_k, u_1, ..., u_k)$ ó $hu(x)$, al valor de la expresión:

$$hu(x) = -\sum_{i=1}^{k} f_i \log \frac{u_i}{E_x(u)} f_i$$

Si intentamos calcular la esperanza de estos estimadores, nos encontraremos con los mismos problemas que en la entropía de Shannon:

- Cuando se introduce la variable auxiliar, los inconvenientes son idénticos a los anteriores.
- Si empleamos el método directo, las expresiones finales resultan algo más sencillas, por cuanto muestras distintas, proporcionan una incertidumbre útil diferente, pero en cualquier caso se obtienen relaciones tan complejas que nos imposibilitan la obtención del valor esperado de estos estimadores.

c) Estimación de la inquietud

Se entiende por inquietud, la incertidumbre debida exclusivamente a las utilidades, que pueden obtenerse como la diferencia entre la incertidumbre útil y la indeterminación debida a las probabilidades (entropía de Shannon).

Formalmente, pueden establecerse las siguientes definiciones, basadas en la medida propuesto por M. A. Gil.

Definición 1.6. Llamamos *inquietud poblacional* de una variable X, que denotamos por $HU^*(p_1, ..., p_M, u_1, ..., u_M)$ ó $HU^*(X)$, al valor de la expresión:

$$HU^*(X) = -\sum_{i=1}^{M} p_i \log \frac{u_i}{E(\boldsymbol{u})}$$

Definición 1.7. Denominamos *inquietud muestral*, de una muestra $\boldsymbol{x}$ y que denotamos por $hu^*(f_1, ..., f_k, u_1, ..., u_k)$ ó $hU^*(\boldsymbol{x})$, al valor de la expresión:

$$hu^*(\boldsymbol{x}) = -\sum_{i=1}^{k} f_i \log \frac{u_i}{E_x(\boldsymbol{u})}$$

En un principio, podríamos pensar que este estimador tendría un mejor comportamiento en los procesos de muestreo en poblaciones finitas que los anteriores, dado que la probabilidad no figura explícitamente dentro de la función logarítmica, quedando así eliminado uno de los factores aleatorios. Sin embargo, al calcular la esperanza del estimador se observa que $E_x(u)$ depende de la muestra y por tanto, perdura una componente aleatoria en el argumento de la función logarítmica, lo que de nuevo nos imposibilita la obtención del sesgo de este estimador.

Nota 4. Podrían darse situaciones particulares que nos permitieran construir un estimador insesgado de HU^*, como:

- Caso de conocer $E(u)$ en la población.
- Si conociésemos el $\max_i \{u_i\}$ y definimos las utilidades relativas como:

$$v_i = u_i / \max_i \{u_i\} .$$

- Otras condiciones más generales, que con una elección adecuada del origen del referencial obtendríamos tal estimador insesgado.

Sin embargo, no se han desarrollado tales alternativas por considerarse como situaciones muy puntuales y que no nos permiten un tratamiento general del problema.

2. INCERTIDUMBRE, INCERTIDUMBRE UTIL E INQUIETUD DE ORDEN $\beta = 2$

Dado los inconvenientes que presentan las medidas de incertidumbre "tipo Shannon" en los procesos de estimación en poblaciones finitas, estudiamos en esta sección la incertidumbre, incertidumbre útil e inquietud de orden $\beta = 2$, que como veremos en el próximo apartado tienen un buen comportamiento en las técnicas de muestreo.

a) Incertidumbre cuadrática. Propiedades

Definición 2.1. Se denomina *incertidumbre cuadrática poblacional*, de la variable X, que denotamos por $H^2(p_1, \ldots, p_m)$, $H^2(X)$ ó $H^2(E)$ al valor de la expresión:

$$H^2(X) = H^2(p_1, \ldots, p_M) = -2 \sum_{i=1}^{M} (p_i^2 - p_i) =$$

$$= 2 \cdot \sum_{i=1}^{M} p_i (1 - p_i) = 2 \left(1 - \sum_{i=1}^{M} p_i^2 \right). \qquad (*)$$

Donde la última igualdad expresa la incertidumbre cuadrática en términos de la distancia de Bayes, o energía informacional de Onicescu.

Algunas de las propiedades más destacables son:

1) La entropía cuadrática es una función continua y simétrica de sus argumentos.
2) $H^2(X)$ es no negativa, anulándose si y sólo si la variable es degenerada.
3) La incertidumbre cuadrática de una variable es máxima cuando todos los resultados son equiprobables, en cuyo caso es una función creciente del número de valores distintos, siendo su cuantía:

$$2(1 - 1/M)$$

4) La medida de entropía cuadrática es una función cóncava (hacia arriba) de sus argumentos.
5) (Propiedad de ramificación). La incertidumbre de orden 2, satisface:

(*) Esta medida es una particularización de la entropía no aditiva de orden β, definidas por Havrda y Charvat como:

$$H^\beta(X) = \frac{1}{2^{1-\beta} - 1} \sum (p_i^\beta - p_i)$$

$$H^2(p_1, ..., p_M) = H^2(p_1 + p_2,\ p_3, ..., p_M) + (p_1 + p_2)^2 \cdot H^2\left(\frac{p_1}{p_1 + p_2},\ \frac{p_2}{p_1 + p_2}\right)$$

6) Sea $\{p_1, ..., p_M\}$ donde suponemos que $p_1 < p_2$, si construimos un nuevo sistema de probabilidades $\{p'_1, ..., p'_M\}$ con $p'_1 = p_1 + \varepsilon$, $p'_2 = = p_2 - \varepsilon$, $p'_l = p_l$, $l = 3, ..., M$; siendo $0 < \varepsilon < (p_2 - p_1)/2$, entonces se verifica:

$$H^2(p'_1, ..., p'_M) > H^2(p_1, ..., p_M)$$

7) Sea $A = (a_{ij})$ una matriz doblemente estocástica, si definimos un sistema de probabilidades $\{p'_1, ..., p'_M\}$ de forma que

$$p'_i = \sum_{j=1}^{M} a_{ij}\, p_j\,,$$

entonces se cumple:

$$H^2(p_1, ..., p_M) \leqslant H^2(p'_1, ..., p'_M)$$

y:

$$\sum_{j=1}^{M} p_j^2 H^2(a_{1j}, ..., a_{Mj}) \leqslant H^2(p'_1, ..., p'_M)$$

La entropía cuadrática cumple las principales propiedades que hacen a la entropía de Shannon adecuada como medida de incertidumbre, y su caracterización axiomática fue establecida por Forte and Ng. y Daróczy.

b) Incertidumbre útil cuadratica. Propiedades

Definición 2.2. Llamamos *incertidumbre útil cuadrática poblacional*, que denotamos por $HU^2(p_1, ..., p_M,\ u_1, ..., u_M)$ ó $HU^2(X)$, al valor de la expresión:

$$HU^2(X) = 2 \cdot \sum_{i=1}^{M} p_i \left(\frac{E(u)}{u_i} - p_i\right)$$

Sus propiedades más significativas son:

1) La incertidumbre útil cuadrática es una función continua y simétrica respecto a los pares (p_i, u_i).
2) La incertidumbre útil cuadrática es mayor o igual que la entropía cuadrática, alcanzándose la igualdad sólo y cuando las utilidades coinciden.

3) $HU^2(X)$ es invariante frente a homotecias positivas, con respecto a las utilidades.
4) (Propiedad de ramificación). La incertidumbre útil cuadrática satisface:

$$HU^2(p_1, ..., p_M, u_1, ..., u_M) =$$

$$= HU^2\left(p_1+p_2, p_3, ..., p_M, \frac{p_1u_1+p_2u_2}{p_1+p_2}, u_3, ..., u_M\right) +$$

$$+ \frac{(p_1+p_2)^2}{u_1^2+u_2^2} \cdot \left[\frac{(u_1-u_2)^2}{p_1u_1+p_2u_2} E(u) + 2u_1u_2\right] \cdot HU^2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}, u_1, u_2\right)$$

c) Inquietud cuadrática. Propiedades

Basándonos en la incertidumbre útil y la entropía cuadrática, podemos establecer la siguiente definición:

Definición 2.3. Llamamos *inquietud cuadrática poblacional*, relativa a la variable X, que denotamos por $HU^{*2}(p_1, ..., p_M, u_1, ..., u_M)$ ó $HU^{*2}(X)$, al valor de la expresión:

$$HU^{*2}(X) = 2 \cdot \sum_{i=1}^{M} p_i\left(\frac{E(u)}{u_i} - 1\right)$$

Sus propiedades más importantes son:

1) $HU^{*2}(p_1, ..., p_M, u_1, ..., u_M)$ es una función continua y simétrica respecto a los pares (p_i, u_i).
2) La inquietud cuadrática poblacional toma un valor no negativo, anulándose si y sólo si las utilidades coinciden para cualquier valor de la variable.
4) La inquietud cuadrática es invariante frente a homotecias positivas, con respecto a las utilidades.
5) (Propiedad de ramificación). Esta medida satisface:

$$HU^{*2}(p_1, ..., p_M, u_1, ..., u_M) =$$

$$= HU^{*2}\left(p_1+p_2, p_3, ..., p_M, \frac{p_1u_1+p_2u_2}{p_1+p_2}, u_1, ..., u_M\right) +$$

$$+ (p_1+p_2)^2\left(\frac{E(u)}{p_1u_1+p_2u_2}\right) HU^{*2}\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}, u_1, u_2\right)$$

6) Sea X una variable con sendos sistemas de probabilidades y utilidad $\{p_1, ..., p_M\}$, $\{u_1, ..., u_M\}$ respectivamente; donde $u_1 > u_2$, $p_1 > 0$, y $p_2 > 0$. Si definimos un nuevo sistema de utilidad $\{u'_1, ..., u'_M\}$ que comple: $u_2 < u'_1 < u_1$; $u_2 < u'_2 < u_1$, $u'_i = u_i$, $i = 3, ..., M$, y además:

$$E(u) = \sum_{i=1}^{M} p_i u_i = E(u') = \sum_{i=1}^{M} p_i u'_i$$

entonces, se verifica:

$$HU^{*2}(p_1, ..., p_M, u'_1, ..., u'_M) < HU^{*2}(p_1, ..., p_M, u_1, ..., u_M)$$

7) Puede obtenerse HU^{*2} como una función de los cocientes entre cada dos utilidades asignadas a los valores de X, esto es

$$HU^{*2}(X) = 2 \cdot \sum_{i=1}^{M} \mathrm{p}_i \left(\sum_{j=1}^{M} p_j \frac{u_j}{u_i} - 1 \right)$$

8) Sea $A = (a_{ij})$ una matriz cuadrada doblemente estocástica, si definimos un nuevo sistema de utilidades $\{u'_1, ..., u'_M\}$, de modo que:

$$u'_i = \sum_{j=1}^{M} a_{ij} u_j,$$

entonces se cumple que:

$$HU^{*2}(1/M, ..., 1/M, u'_1, ..., u'_M) \leqslant HU^{*2}(1/M, ..., 1/M, u_1, ..., u_M)$$

d) Medidas muestrales

Transcribimos ahora las definiciones anteriores para una muestra $\boldsymbol{x}$ de tamaño $\boldsymbol{n}$.

Definición 2.4. Llamamos *incertidumbre cuadrática muestral*, relativa a una muestra de tamaño n, que denotamos por $h_n^2(f_1, ..., f_k)$ o por $H_n^2(\boldsymbol{x})$, al valor de la expresión:

$$H_n^2(\boldsymbol{x}) = 2 \cdot \sum_{i=1}^{k} f_i (1 - f_i) = 2 \left(1 - \sum_{i=1}^{k} f_i^2 \right)$$

Definición 2.5. Denominamos *incertidumbre útil cuadrática muestral* de una muestra $\boldsymbol{x}$ de tamaño n, que denotamos por $HU_n^2(\boldsymbol{x})$ o por

$hu_n^2(f_1, ..., f_k, u_1, ..., u_k)$, al valor de la expresión:

$$HU_n^2(x) = 2 \cdot \sum_{i=1}^{k} f_i \left(\frac{E_x(u)}{u_i} - f_i \right) = 2 \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{u_j}{u_i} f_i f_j - \sum_{i=1}^{M} f_i^2 \right)$$

Definición 2.6. Llamamos *inquietud cuadrática muestral*, de una muestra de tamaño n, x, que denotamos por $hu^{*2}_n(f_1, ..., f_k, u_1, ..., u_k)$ ó por HU $^{*2}_n(x)$, al valor de la expresión:

$$HU^{*2}_n(x) = 2 \cdot \sum_{i=1}^{k} f_i \left(\frac{E_x(u)}{u_i} - 1 \right) = 2 \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{u_j}{u_i} f_i f_j - 1 \right)$$

Estas medidas muestrales serán consideradas como estimaciones de las correspondientes medidas poblacionales (obtenidas por el método de analogía); sin embargo, como se verá en la próxima sección tales estimadores resultarán sesgados.

3. ESTIMADORES DE LA INCERTIDUMBRE, LA INCERTIDUMBRE UTIL Y DE LA INQUIETUD

En este apartado, se obtienen estimadores insesgados para las anteriores medidas, así como una estimación de sus varianzas, para los métodos de selección con y sin reposición.

3.1. Estimador de la incertidumbre cuadrática poblacional(*)

a) Estimador insesgado

Teorema 3.1. En un proceso de selección aleatorio con reposición, el estimador:

$$\hat{H}_n^2(x) = \frac{n}{n-1} h_n^2(f_1, ..., f_n) = \frac{2n}{n-1} \left(1 - \sum_{i=1}^{k} f_i^2 \right)$$

es insesgado para estimar $H^2(X)$.

Teorema 3.2. Cuando el método de muestreo es aleatorio sin reemplazamiento, el estimador:

$$(\hat{H}_n^2(x))^c = \frac{N-1}{N} \frac{n}{n-1} h_n^2(f_1, ..., f_k) = 2 \frac{N-1}{N} \frac{n}{n-1} \left(1 - \sum_{i=1}^{k} f_i^2 \right)$$

es insesgado para estimar la incertidumbre cuadrática poblacional.

(*) Un tratamiento más extenso de este apartado, incluyendo algunas aplicaciones, puede verse en Pérez, R., Gil, M. A. and P. Gil (1985).

b) Varianza de los estimadores

Teorema 3.3. En un proceso de selección aleatorio con reposición, la varianza del estimador $\hat{H}_n^2(x)$, viene dada por:

$$\operatorname{Var}(\hat{H}_n^2(x)) = \frac{4}{n(n-1)} \left\{ (6-4n)\left(\sum_{i=1}^{M} p_i^2\right)^2 + \right.$$

$$\left. + 4(n-2)\left(\sum_{i=1}^{M} p_i^3\right) + 2\left(\sum_{i=1}^{M} p_i^2\right) \right\} = \qquad (*)$$

$$= [4(4n-7)H^2(X) + (6-4n)(H^2(X))^2 - 12(n-2)H^3(X)] / n(n-1)$$

Teorema 3.4. Si la muestra se obtiene de forma aleatoria sin reposición, entonces la varianza de $(H_n^2(x))^c$, es:

$$\operatorname{Var}((\hat{H}_n^2(x))^c) = \frac{4}{n(n-1)(N-2)(N-3)} \left\{ [6N(N-1) - 6n(n-1) - \right.$$

$$- 4nN(N-n)]\left(\sum_{i=1}^{M} p_i^2\right)^2 + 4(N-1)(N-n)(n-2)\left(\sum_{i=1}^{M} p_i^3\right) +$$

$$\left. + [2(N+1)(N-n)(N-n-1)/N] \cdot \left(\sum_{i=1}^{M} p_i^2\right) - 2(N-n)(N-n-1)/N \right\} =$$

$$= \frac{(N-n)}{n(n-1)(N-2)(N-3)} \left\{ \frac{4(N-1)}{N} [N(4n-7) - (n+1)] H^2(X) + \right.$$

$$\left. + [(6-4n)(N-1) + 2n](H^2(X))^2 - 12(n-2)(N-1)H^3(X) \right\}$$

Por lo general, estas varianzas no serán conocidas, por lo que nos planteamos su estimación a partir de la muestra de tamaño n, x.

c) Estimación de las varianzas

Teorema 3.5. En un proceso de selección aleatorio con resposición,

(*) Donde H^3, es la medida de entropía no aditiva de orden $\beta = 3$.

$$\widehat{\operatorname{Var}}\,(\hat{H}_n^2\,(\boldsymbol{x})) = \frac{2n}{(n-1)\,(n-2)\,(n-3)}\,\{n\,(3-2n)\,(H_n^2\,(\boldsymbol{x}))^2 +$$

$$+\,6n\,(n-1)\,H_n^3\,(\boldsymbol{x}) + 2\,(4n-1)\,(n-1)\,H_n^2\,(\boldsymbol{x})\}$$

es un estimador insesgado de $\operatorname{Var}\,(\hat{H}_n^2\,(\boldsymbol{x}))$.

Teorema 3.6. Cuando el método de muestreo es aleatorio sin reemplazamiento, el estimador:

$$\widehat{\operatorname{Var}}\,(\hat{H}_n^2\,(\boldsymbol{x}))^c) = n\,(N-n)\,(N-1)\,\{[(6-4n)\,(N-1)+2n]\,n\,(H_n^2\,(\boldsymbol{x}))^2 -$$

$$-\,12n\,(n-1)\,(N-2)\,H_n^3\,(\boldsymbol{x}) + 4\,[(4n-1)\,N-7n-1]\,(n-1)\,H_n^2\,(\boldsymbol{x})\}\,\cdot$$

$$\cdot\,(1/N^3\,(n-1)^2\,(n-2)\,(n-3))$$

es insesgado para estimar $\operatorname{Var}\,((\hat{H}_n^2\,(\boldsymbol{x}))^c)$.

3.2. Estimador de la incertidumbre útil cuadrática poblacional

a) Estimadores insesgados

Teorema 3.7. En un proceso de selección aleatorio con reposición, el estimador:

$$\widehat{HU}_n^2\,(\boldsymbol{x}) = \frac{n}{n-1}\,hu_n^2\,(f_1, \ldots, f_k,\ u_1, \ldots, u_k) =$$

$$= \frac{2n}{n-1}\left(\sum_{i=1}^{k}\sum_{j\neq i}^{k}\frac{u_j}{u_i}\,f_i f_j\right)$$

es insesgado para estimar la incertidumbre útil cuadrática de esa población a partir de una muestra de tamaño n.

Teorema 3.8. Cuando la muestra se selecciona de forma aleatoria sin reposición,

$$(\widehat{HU}_n^2\,(\boldsymbol{x}))^c = \frac{N-1}{N}\,\frac{n}{n-1}\,hu_n^2\,(\boldsymbol{x}) = 2\,\frac{N-1}{N}\,\frac{n}{n-1}\left(\sum_{i=1}^{k}\sum_{j\neq i}^{k}\frac{u_j}{u_i}\,f_i f_j\right)$$

es un estimador insesgado de la incertidumbre útil cuadrática poblacional.

b) Varianza de los estimadores

Teorema 3.9. Cuando el método de muestreo es con reemplazamiento, la varianza del estimador de la incertidumbre útil cuadrática, viene dada por la expresión:

$$\operatorname{Var}(\widehat{HU_n^2}(x)) = \frac{1}{n(n-1)} \Big\{ -2(2n-3)(HU^2(X))^2 + (n-2)[3HU^3(X) +$$

$$+ 3HU^3(1/X) + 4HU^2(X)] + 2HU^2(X^2) + 6(n-2)H^3(X) +$$

$$+ 2(2n-3)H^2(X) - 16(n-2)\Big[\sum_{i=1}^{M} p_i^2 \Big(\frac{E(u)}{u_i} - p_i\Big) +$$

$$+ \sum_{i=1}^{M} p_i^2 \Big(u_i E\Big(\frac{1}{u}\Big) - p_i\Big)\Big]\Big\} \qquad (*)$$

Teorema 3.10. En una técnica de muestreo aleatorio sin reposición, la varianza del estimador $(\widehat{HU_n^2}(x))^c$, es:

$$\operatorname{Var}((\widehat{HU_n^2}(x))^c) = \frac{4}{n(n-1)} \frac{N-1}{N} \Big\{ \Big[(n-2)(n-3) \frac{N}{N-2} \frac{N}{N-3} -$$

$$- n(n-1) \frac{N}{N-1} \Big] \frac{(HU^2(X))^2}{4} + \frac{N-n}{N-2} \frac{N}{N-3} \frac{3(n-2)}{4} \cdot$$

$$\cdot [HU^3(X) + HU^3(1/X)] + \frac{1}{2} \frac{N-n-1}{N-2} \frac{N-n}{N-3} [HU^2(X^2) + H^2(X)] +$$

$$+ \frac{N-n}{N-2} \frac{N}{N-3} (n-2)[HU^2(X) + (3/2)H^3(X) - H^2(X)] -$$

$$- 4(n-2) \frac{N-n}{N-2} \frac{N}{N-3} \Big[\sum_{i=1}^{M} p_i^2 \Big(\frac{E(u)}{u_i} - p_i\Big) +$$

$$+ \sum_{i=1}^{M} p_i^2 (u_i E(1/u) - p_i)\Big]\Big\}$$

(*) Por $1/X$, denotamos la variable que toma los mismos valores que X pero donde las utilidades asociadas a cada valor es el inverso de la utilidad original y por X^2 aquella que asocia utilidades cuadráticas.

c) Estimación de las varianzas

Una vez obtenidas las varianzas de estos estimadores, construimos ahora estimadores insesgados de las mismas a partir de la muestra x.

Teorema 3.11. En un proceso de selección aleatorio con reposición, el estimador:

$$\widehat{\mathrm{Var}}(\hat{H}_n^2(x)) = \frac{4}{n^2(n-1)^2(n-2)(n-3)} \left\{ \sum_{i=1}^{k} \sum_{j \neq i}^{k} \left(\frac{u_j^2}{u_i^2} + 1 \right) n_i n_j \cdot \right.$$

$$\cdot [-2(2n-3)n_i n_j + n(n-1)(n_i + n_j) - n(n-1)] +$$

$$+ \sum_{i=1}^{k} \sum_{j \neq i}^{k} \sum_{\substack{1 \neq i \\ 1 \neq j}}^{k} \left(2\frac{u_j}{u_i} + \frac{u_j u_1}{u_i^2} + \frac{u_i^2}{u_j u_1} \right) n_i n_j n_1 [-2(2n-3)n_i + n(n-1)] -$$

$$\left. - 2(2n-3) \sum_{i=1}^{k} \sum_{j \neq i}^{k} \sum_{\substack{1 \neq i \\ 1 \neq j}}^{k} \sum_{\substack{h \neq i \\ h \neq j \\ h \neq 1}}^{k} \frac{u_j u_h}{u_i u_1} n_i n_j n_1 n_h \right\}$$

es insesgado para estimar la varianza del estimador de la incertidumbre útil cuadrática poblacional.

Teorema 3.12. Cuando la muestra se selecciona de forma aleatoria sin reposición,

$$\widehat{\mathrm{Var}}(\hat{H}_n^2(x))^{c} = \frac{1}{n(n-1)(n-2)(n-3)} \frac{N-n}{N} \frac{N-1}{N} \cdot$$

$$\cdot \left\{ \sum_{i=1}^{k} \sum_{j \neq i}^{k} \left(\frac{u_j^2}{u_i^2} + 1 \right) n_i n_j \cdot \left[\frac{N-1}{N-n} \left[\frac{(n-2)(n-3)}{n(n-1)} - \right. \right. \right.$$

$$\left. \left. - \frac{N-2}{N-1} \frac{N-3}{N} \right] n_i n_j + \frac{N-2}{N}(n_i + n_j) - \frac{N-n+1}{N} \right] +$$

$$+ \sum_{i=1}^{k} \sum_{j \neq i}^{k} \sum_{\substack{1 \neq i \\ 1 \neq j}}^{k} \left(2\frac{u_j}{u_i} + \frac{u_j u_1}{u_i^2} + \frac{u_i^2}{u_j u_1} \right) n_i n_j n_1 \left[\frac{N-1}{N-n} \left[\frac{(n-2)(n-3)}{n(n-1)} + \right. \right.$$

$$- \frac{N-2}{N-1} \frac{N-3}{N} \Big] n_i + \frac{N-2}{N} \Big] + \sum_{i=1}^{k} \sum_{j \neq i}^{k} \sum_{\substack{1 \neq i \\ 1 \neq j}}^{k} \sum_{\substack{h \neq i \\ h \neq j \\ h \neq 1}}^{k} \frac{u_j u_h}{u_i u_1} \frac{N-1}{N-n} \cdot$$

$$\cdot \Big[\frac{(n-2)(n-3)}{n(n-1)} - \frac{N-2}{N-1} \frac{N-3}{N} \Big] n_i n_j n_1 n_h \Big\}$$

es una estimación insesgada de la varianza del estimador.

3.3. Estimador de la inquietud cuadrática poblacional

a) Estimadores insesgados

Teorema 3.13. En un proceso de selección aleatorio con reposición, el estimador:

$$\widehat{HU_n^{*2}}(x) = \frac{n}{n-1} hu_n^{*2}(x) = \frac{2n}{n-1} \Big(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{u_j}{u_i} f_i f_j - 1 \Big)$$

es insesgado para estimar la inquietud cuadrática poblacional, $HU^{*2}(X)$, a partir de una muestra de tamaño n.

Teorema 3.14. Cuando la muestra se selecciona de forma aleatoria con reposición,

$$(\widehat{HU_n^{*2}}(x))^c = \frac{N-1}{N} \frac{n}{n-1} hu_n^{*2}(x) =$$

$$= 2 \frac{N-1}{N} \frac{n}{n-1} \Big(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{u_j}{u_i} f_i f_j - 1 \Big)$$

es un estimador insesgado de la inquietud cuadrática poblacional.

b) Varianza de los estimadores

Teorema 3.15. Cuando el método de muestreo es con reemplazamiento, la varianza del estimador de la inquietud cuadrática, viene dada por:

$$\mathrm{Var}\,(\widehat{HU_n^2}(x)) = \frac{1}{n(n-1)} \{(6-4n)(HU^{*2}(X))^2 +$$

$$+ 4(4-3n) HU^{*2}(X) + 3(n-2)[HU^{*3}(X) + HU^{*3}(1/X)] + 2HU^{*2}(X^2)\}$$

Teorema 3.16. En una técnica de muestreo aleatorio sin reposición, la varianza del estimador $(HU^{*2}_n(x))^c$, es:

$$\mathrm{Var}\,((\widehat{HU}^2_n(x))^c) = \frac{(N-n)}{n(n-1)N(N-2)(N-3)}\{[(6-4n)N+$$

$$+6(n-1)]N(HU^{*2}(X))^2+4(N-1)[(4-3n)N+2(n+1)]HU^{*2}(X)+$$

$$+3(n-2)(N-1)N[HU^3(X)+$$

$$+HU^{*3}(1/X)]+2(N-1)(N-n-1)HU^{*2}(X^2)\}$$

c) Estimación de las varianzas

Teorema 3.17. En un proceso de selección aleatorio con reposición, el estimador:

$$\widehat{\mathrm{Var}}\,(\widehat{HU}^{*2}_n(x)) = \frac{1}{(n-1)^2(n-2)(n-3)}\{n^2(6-4n)(hu^{*2}_n(x))^2-$$

$$-4n(3n-2)(n-1)hu^{*2}_n(x)+3n^2(n-1)[hu^{*3}_n(x)+hu^{*3}_n(1/x)]- \qquad (*)$$

$$-2n(n-1)hu^{*2}_n(x^2)\}$$

es insesgado para estimar la varianza del estimador de la inquietud cuadrática poblacional.

Teorema 3.18. Cuando la muestra se selecciona de forma aleatoria sin reemplazamiento,

$$\widehat{\mathrm{Var}}\,((\widehat{HU}^{*2}_n(x))^c) = \frac{n(N-1)(N-n)}{N^3(n-1)^2(n-2)(n-3)}\{n[(6-4n)N+$$

$$+6(n-1)](hu^{*2}_n(x))^2+3n(n-1)(N-2)[hu^{*3}_n(x)+hu^{*3}_n(1/x)]-$$

$$-4(n-1)[(3n-2)N+4n+2]hu^{*2}_n(x)-2(n-1)(N-n-1)hu^{*2}_n(x^2)\}$$

es un estimador insesgado de $\mathrm{Var}\,((\widehat{HU}^{*2}_n(x))^c)$.

Nota 5. Como se observa, los estimadores insesgados mantienen la misma estructura en cada método de muestreo. Para el muestreo con reposición, los estimadores que se obtienen, se denominan: *cuasi-incertidumbre, cuasi-incertidumbre útil* y *cuasi-inquietud cuadrática* puesto que mantienen un esquema análogo a la estimación de la varianza.

(*) Por $1/x$ y x^2, dnotamos la misma muestra, donde se consideran las utilidades inversas y cuadráticas respectivamente.

3.4. Ganancia de precisión

Teorema 3.19. Los estimadores de la incertidumbre, incertidumbre útil e inquietud cuadrática poblacionales en el muestreo con reposición tienen una mayor precisión que en el muestreo con reemplazamiento.

Nota 6. El incremento de la varianza, en todos los casos, es una función de la varianza del estimador en el muestreo con reposición, de la misma varianza cuando la muestra es de tamaño 2 y del cuadrado de la medida poblacional; manteniéndose constantes los parámetros de la función.

Nota 7. Basándose en la entropía condicionada de orden β (Darozcy, 1970) se ha definido (Gil, M. A., Pérez, R. and I. Martínez, 1986) la *información mutua de orden* $\beta = 2$, que presenta una mejor adecuación a los procesos de estimación que su homóloga la cantidad de información de tipo Shannon.

En el trabajo citado, se obtiene un estimador insesgado de la información mutua así como una estimación insesgada de su varianza a partir de una muestra de tamaño n seleccionada por un método aleatorio con reposición. La simulación de muestras aleatorias de una población normal por este método de selección corroboran la idoneidad de la cantidad de información cuadrática.

4. LA INQUIETUD CUADRATICA COMO MEDIDA DE DESIGUALDAD

En esta sección se plantea una aplicación de la inquietud cuadrática como medida de desigualdad (de renta), tanto en su aspecto teórico como en su estimación a partir de una muestra de tamaño n.

Toda medida de desigualdad descomponible y satisfaciendo las propiedades de: *Normalización, independencia del tamaño poblacional, invariante por homotecias*, la condición de Pigou-Dalton y continua es de la forma(*):

$$I^N(X) = \sum_{i=1}^{M} \psi\left(\frac{x_i}{E(X)}\right) p_i$$

siendo $\psi(x)$ una función definida para cada número β real como:

$$\psi(\mathrm{x}) = \psi_\beta(x) = \begin{cases} x^\beta - 1 & \text{si } \beta < 0 \\ -\log x & \text{si } \beta = 0 \\ 1 - x^\beta & \text{si } 0 < \beta < 1 \\ x \cdot \log x & \text{si } \beta = 1 \\ x^\beta - 1 & \text{si } \beta > 1 \end{cases}$$

(*) Siguiendo la axiomática de D. Zagier. Otras caracterizaciones fueron dadas por: F. Bourguignon, F. A. Cowell y A. F. Shorrocks.

Si denotamos por X la variable renta, y consideramos la utilidad asociada a un individuo como la renta del mismo ($u_i = x_i$, suponemos que toda renta es positiva), entonces la inquietud cuadrática poblacional puede expresarse como:

$$HU^{*2}(X) = 2 \cdot \sum_{i=1}^{M} \mathrm{p}_i \left(\frac{E(X)}{\mathrm{x}_i} - 1 \right)$$

El indicador:

$$D_{-1}^{N}(X) = HU^{*2}(X)/2$$

puede ser interpretado como una medida de desigualdad, siendo la particularización $\beta = -1$ de las medidas generalizadas aditivamente descomponibles.

Además de los axiomas exigidos en su caracterización, otras propiedades importantes de esta medida son: *simetría o imparcialidad, el principio de transferencia regresivo, maximalidad o acotación* y *S-convexidad.*

Nota 8. Cumple también las características de comparar (por cociente) todas las rentas individuales y, confrontar la renta per-cápita de toda la población respecto a las rentas per-cápita de cada grupo.

En cuanto a la estimación de la desigualdad existente en una población finita E, a partir de una muestra de tamaño n, transcribiendo los teoremas 3.13 y 3.14, se tiene:

Teorema 4.1. En un proceso de selección aleatorio con reposición, el estimador:

$$d_{-1}^{n}(x) = \frac{n}{n-1} D_{-1}^{n}(x) = \frac{n}{n-1} \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{x_j}{x_i} f_i f_j - 1 \right)$$

es insesgado para estimar la desigualdad de renta en la población E, $D_{-1}^{N}(X)$.

Teorema 4.2. Cuando la muestra se obtiene por un método aleatorio con reemplazamiento, el estimador:

$$(d_{-1}^{n}(x))^{c} = \frac{N-1}{N} \frac{n}{n-1} D_{-1}^{n}(x) = \frac{N-1}{N} \frac{n}{n-1} \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{x_j}{x_i} f_i f_j - 1 \right)$$

es insesgado para estimar $D_{-1}^{N}(X)$.

Los siguientes teoremas nos proporcionan una estimación de la precisión de estos estimadores.

Teorema 4.3. Cuando el método de muestreo es aleatorio con reposición,

$$\widehat{\operatorname{Var}}\,(d_{-1}^{n}(x)) = \{n^2(6-4n)(D_{-1}^{n}(x))^2 - 2n(3n-2)(n-1)D_{-1}^{n}(x) + \\ + (9/16)\,n^2(n-1)\,[D_{-2}^{n}(x) + D_{-2}^{n}(1/x)] - \\ - n(n-1)D_{-1}^{n}(x^2)\} \,/\, (n-1)^2(n-2)(n-3) \qquad (*)$$

es un estimador insesgado de $\operatorname{Var}\,(d_{-1}^{n}(x))$.

Teorema 4.4. En un método de selección aleatorio simple sin reemplazamiento, el estimador:

$$\widehat{\operatorname{Var}}\,(d_{-1}^{n}(x))^{c}) = n(N-1)(N-2)\,\{n\,[(6-4n)\,N + \\ + 6(n-1)]\,(D_{-1}^{n}(x))^2 + (9/16)\,n(n-1)(N-2)\,[D_{-2}^{n}(x) + D_{-2}^{n}(1/x)] - \\ - 2(n-1)\,[(3n-2)\,N + 4n + 2]\,D_{-1}^{n}(x) - \\ - (n-1)(N-n+1)\,D_{-1}^{n}(x^2)\} \,/\, N^3(n-1)^2(n-2)(n-3)$$

es insesgado para estimar $\operatorname{Var}\,((d_{-1}^{n}(x))^{c})$.

Estos últimos teoremas ponen de manifiesto el buen comportamiento de las *medidas de desigualdad aditivamente descomponibles en orden* -1 en los procesos de estimación en poblaciones finitas.

BIBLIOGRAFIA

[1] ACZEL, J. AND Z. DAROCZY: *On Measures of Information and Their Characterizations.* Academic Press, New York 1975.

[2] AZORIN, F.: *Curso de muestreo y aplicaciones.* Ed. Aguilar, Madrid 1972.

[3] BHARGAVA, T. N. AND V. R. R. UPPULURI: "Sampling Distribution of Gini's Index of Diversity", *Applied Mathematics and Computation, 3* (1977), 1-24.

[4] BOURGUIGNON, F.: "Descomposable income inequality measures", *Econometrica, 4* (1979), 901-920.

[5] COCHRAN, W. G.: *Técnicas de Muestreo.* Ed. C.E.C.S.A., México 1980.

[6] COWELL, F. A.: "On the structure of additive inequality measures", *Review of Economic Studies, 47* (1980), 521-531.

[7] DAROCZY, Z.: "Generalized Information Functions", *Inf. and Contr., 16* (1970), 36-51.

[8] EICHHORN, W. AND W. GEHERIG: "Measurement of inequality in Economics", *Discussion Paper, Nr. 141,* Inst. Für Wirtschaftstheorie und operations research (1980).

[9] EMPTOZ, H.: *Informations Utiles et Pseudoquestionnaires.* These pour obtenir le Diplome de Docteur. Lyon 1976.

[10] GIL, M. A.: *Incertidumbre y utilidad.* Tesis Doctoral, Universidad de Oviedo, 1979.

(*) Donde $1/x$ y x^2 denotan la distribución de rentas inversas y cuadrática en la muestra y D, la medida aditivamente descomponible de orden -2.

[11] GIL, M. A.: "Criterion of maximizing the expected quietness (invariant by homotheties in relation to the utilities)", *R.A.I.R.O.-Rech. Opér., 16* (1982), 319-331.

[12] GIL, M. A., PEREZ, R. AND I. MARTINEZ: "The Mutual Information. Estimation in the Sampling with Replacement", *R.A.I.R.O-Rech. Opér.* (1986), accepted for publication).

[13] GIL, P.: *Teoría Matemática de la Información.* Ed. ICE, Madrid 1981.

[14] HAVRDA, J. AND F. CHARVAT: "Quantification method of classification processes", *Kybernetica, 3* (1967), 30-35.

[15] KENDALL, SIR M. AND A. STUART: *The Advanced Theory of Statistics.* Ed. Charles Griffin Co. Ltd., London, 1977-83.

[16] MATHAI, A. M. AND P. N. RATHIE: *Basic concepts in Information Theory and Statistics.* Wiley Eastern Limited, New Delhi 1975.

[17] ONICESCU, O.: "Energie Informationnelle", *C.R. Acad. Si., Paris, 263A,* (1966) 841-842.

[18] PEREZ, R., GIL, M. A. AND P. GIL: "Estimating the Uncertainty associated with a Variable in a Finite Population", *Kybernetes* 1986. Vol 15, 251-256.

[19] SANCHEZ-CRESPO, J. L.: *Curso Intensivo de Muestreo en Poblaciones Finitas. Ed. INE, Madrid 1976.*

[20] SHANNON, C. E.: "A mathematical Theory of communication", *Bell. System. Tech. J.27* (1948), 379-423.

[21] SHORROCKS, A. F.: "The class of additively decomposable inequality measures", *Econometrica, 48* (1980), 613-625.

[22] THEIL, H.: *Economics and Information Theory.* Nort Holland, Amsterdam 1967.

[23] ZAGIER, D.: "On the decomposability of the Gini coefficient and other indices of inequality", *Discussion paper, No. 108,* Projectgruppe Theoretische modelle, Universität Bonn, 1983.

APENDICE I

A.1. MOMENTOS DE LA DISTRIBUCION POLINOMIAL

Sea $(e_1, ..., e_M)$ un vector aleatorio con distribución polinomial de parámetros:

$$n, \quad p_1, ..., p_M, \quad \left(\sum_{i=1}^{M} p_i = 1 \right)$$

entonces:

(A.1.1) $E(e_i) = \mathrm{np}_i$

(A.1.2) $E(e_i^2) = np_i [(n-1) p_i + 1]$

(A.1.3) $E(e_i^3) = np_i [(n-1)(n-2) p_i^2 + 3(n-1) p + 1]$

(A.1.4) $E(e_i^4) = np_i [(n-1)(n-2)(n-3) p_i^3 + 6(n-1)(n-2) p_i^2 +$
$+ 7(n-1) p_i + 1]$

(A.1.5) $\mathrm{Var}(e_i) = np_i (1-p_i)$

(A.1.6) $E(e_i e_j) = n(n-1) p_i p_j$

(A.1.7) $\mathrm{Cov}(e_i e_j) = -np_i p_j$

(A.1.8) $E(e_i^2 e_j) = n(n-1) p_i p_j [(n-2) p_j + 1]$

(A.1.9) $E(e_i^3 e_j) = n(n-1) p_i p_j [(n-2)(n-3) p_i^2 + 3(n-2) p_i + 1]$

(A.1.10) $E(e_i^2 e_j^2) = n(n-1) p_i p_j [(n-2)(n-3) p_i p_j + (n-2)(p_i + p_j) + 1]$

(A.1.11) $E(e_i e_j e_1) = n(n-1)(n-2) p_i p_j p_1$

(A.1.12) $E(e_i^2 e_j e_1) = n(n-1)(n-2) p_i p_j p_1 [(n-3) p_i + 1]$

(A.1.13) $E(e_i e_{j\,1} e_h) = n(n-1)(n-2)(n-3) p_i p_j p_1 p_h$

APENDICE II

A.2. MOMENTOS DE LA DISTRIBUCION POLI-HIPERGEOMETRICA

Sea $(e_1, ..., e_M)$ un vector aleatorio con distribución poli-hipergeométrica de parámetros N, $Np_1, ..., Np_M$ y n. Entonces:

(A.2.1) $$E(e_i) = np_i$$

(A.2.2) $$E(e_i^2) = \frac{N-n}{N-1}\, np_i \left[\frac{N}{N-n}\,(n-1)\,p_i + 1 \right]$$

(A.2.3) $$E(e_i^3) = \frac{N-n}{N-1}\, np_i \left[\frac{N}{N-n}\,\frac{N}{N-2}\,(n-1)\,(n-2)\,p_i^2 + \right.$$

$$\left. + 3\,\frac{N}{N-2}\,(n-1)\,p_i + \frac{N-2n}{N-2} \right]$$

(A.2.4) $$E(e_i^4) = \frac{N-n}{N-1}\, np_i \left[\frac{N}{N-n}\,\frac{N}{N-2}\,\frac{N}{N-3}\,(n-1)\,(n-2) \cdot \right.$$

$$\cdot (n-3)\,p_i^3 + 6\,\frac{N}{N-2}\,\frac{N}{N-3}\,(n-1)\,(n-2)\,p_i^2 +$$

$$\left. + \frac{N}{N-2}\,(n-1)\,\frac{7N-11n+1}{N-3}\,p_i + \frac{N(N+1) - 6n(N-n)}{(N-2)(N-3)} \right]$$

(A.2.5) $$E(e_i\, e_j) = \frac{N}{N-1}\, n\,(n-1)\,p_i\,p_j$$

(A.2.6) $$E(e_i^2\, e_j) = \frac{N-n}{N-1}\,\frac{N}{N-2}\, n\,(n-1)\,p_i\,p_j \left[\frac{N}{N-n}\,(n-2)\,p_i + 1 \right]$$

(A.2.7) $$E(e_i^2\, e_j^2) = \frac{N-n}{N-1}\,\frac{N}{N-2}\,\frac{N}{N-3}\, n\,(n-1)\,p_i\,p_j \left[\frac{N}{N-n}\,(n-2) \cdot \right.$$

$$\left. \cdot (n-3)\,p_i\,p_j + (n-2)\,(p_i + p_j) + \frac{N-n-1}{N} \right]$$

$$E(e_i^3\, e_j) = \frac{N-n}{N-1}\,\frac{N}{N-2}\,\frac{N}{N-3}\, n(n-1)\, p_i\, p_j \Big[\frac{N}{N-n}\,(n-2)\cdot (n-3)\, p_i^2 + 3\,(n-2)\, p_i + \frac{N-2n+1}{N}\Big] \tag{A.2.8}$$

$$E(e_i\, e_j\, e_1) = \frac{N}{N-1}\,\frac{N}{N-2}\, n(n-1)(n-2)\, p_i\, p_j\, p_1 \tag{A.2.9}$$

$$E(e_i^2\, e_j\, e_1) = \frac{N-n}{N-1}\,\frac{N}{N-2}\,\frac{N}{N-3}\, n(n-1)(n-2)\, p_i\, p_j\, p_1 \cdot \Big[\frac{N}{N-n}\,(n-3)\, p_i + 1\Big] \tag{A.2.10}$$

$$E(e_i\, e_j\, e_1\, e_h) = \frac{N}{N-1}\,\frac{N}{N-2}\,\frac{N}{N-3}\, n(n-1)(n-2)\cdot (n-3)\, p_i\, p_j\, p_1\, p_h \tag{A.2.11}$$