SOLUCIONS ALS PROBLEMES PROPOSATS AL VOLUM 20. N° 1

PROBLEMA N° 57

Consideremos una población finita de tamaño N, $U = \{1,2,\ldots,i,\ldots,N\}$. Sean «$y$» y «$x$» la variable de interés y auxiliar respectivamente. Tratamos de estimar la media poblacional de interés

$$\overline{y} = \sum_{i\in U} y_i/N,$$

haciendo uso del estadístico «media-de-productos»

$$\overline{p}_s = \frac{1}{n}\sum_{i\in s} y_1 x_i.$$

Para ello nos basamos en el trabajo de Murthy (1964). De (3.3) y (3.4), deducimos que bajo muestreo aleatorio simple sin reemplazamiento, tenemos los sesgos (siendo $\overline{y}_s$ y $\overline{x}_s$ las medias muestrales de las variables «y» y «x»),

$$B(\overline{y}_s\,\overline{x}_s) = E(\overline{y}_s\,\overline{x}_s) - \overline{y}\overline{x}$$

y

$$B(\overline{p}_s) = E(\overline{p}_s) - \overline{y}\overline{x}$$

donde $\overline{x}$ es la media poblacional de la variable auxiliar «x». Como $B(\overline{p}_s) = nB(\overline{y}_s\,\overline{x}_s)$, deducimos que

$$E(\overline{p}_s) - \overline{y}\overline{x} = nB(\overline{y}_s\,\overline{x}_s) - n\overline{y}\overline{x},$$

por lo que

$$(n-1)\overline{y}\overline{x} + E(\overline{p}_s) = nE(\overline{y}_s\,\overline{x}_s)$$

y por tanto

$$z = \overline{y}_s\,[(n-1)\overline{x} - n\overline{x}_s] + \overline{p}_s$$

es una variable aleatoria de esperanza nula. La clase de estimadores insesgados para $\overline{y}$ basados en el estadístico $\overline{p}_s$ será

$$t = \overline{y}_s + kz.$$

Esta metodología es análoga a la vista por Ruiz y Santos (1989), de manera que los últimos comentarios realizados en el párrafo final de estos autores, son válidos también en la clase aquí propuesta.

Referencias

Murthy, M.N. (1964). «Product method of estimation». *Sankhyā*. **Ser. A, 26**, 69–74.

Ruiz, M. y Santos, J. (1989). «Unbiased mean-of-the-ratios estimators». *Statistica*. **49**, 617–622.

M. Ruiz Espejo

UNED

PROBLEMA N° 58

a) La demostración de la condición suficiente, dada por Ruiz (1980, 1985), es básicamente la propuesta por Glasser (1962) hasta la ecuación (16). Pero en lugar de (16), señalamos que $m < n$, y entonces podemos deducir directamente que

$$k^2 > \frac{N}{n} - \frac{Nm-n}{n(N-1)} > \frac{N}{n} - 1,$$

y de aquí se deriva la nueva cota, ya que $k > 0$,

$$k > \sqrt{\frac{N}{n} - 1},$$

o en otras palabras

$$x > \mu + \sqrt{\frac{N}{n} - 1} \cdot \sigma.$$

b) Considerar una población finita de tamaño $N = 10$ cuyos valores de la variable de interés ordenados en orden creciente es: 1, 2, 2, 4, 4, 5, 7, 9, 12 y 14. Para una muestra de tamaño $n = 5$ tenemos la siguiente tabla de varianzas para varios procedimientos de muestreo

L = número de estratos	Diseño muestral	Varianza
1	Muestreo aleatorio simple sin reemplazo.	1.96
2	Estratificación especial con la cota de Glasser o Ruiz ($m = 2$).	0.98
2	Estratificación especial con $m = 3$.	0.75
2	Estratificación especial óptima ($m = 4$).	0.72

Este ejemplo garantiza que la cota inferior para el punto de estratificación especial con dos estratos no sea también cota inferior del punto de estratificación especial óptimo. En el ejemplo propuesto $\mu = 6$ y $\sigma^2 = 17.6$.

Referencias

Glasser, G.J. (1962). «On the complete coverage of large units in a statistical study». *Rev. Internat. Statist. Inst.* **30**, 28–32.

Ruiz, M. (1980). *Construcción de Estratos en el Diseño de Muestreo Estratificado Aleatorio.* Tesina de Licenciatura. Universidad Complutense de Madrid.

Ruiz, M. (1985). «Equiprecisional allocation and optimum stratification». *Statistics.* **16**, 559–562.

M. Ruiz Espejo

UNED

PROBLEMA N° 59

Prendrem la comparació del bucle `while` com a instrucció crítica. Així, podem observar que aquesta instrucció s'executa com a mínim 1 vegada i com a màxim i vegades en cada iteració del bucle `repeat`. D'altra banda, aquest bucle s'executarà de nou sempre que el bucle `while` s'hagi executat menys de i vegades (i, per tant, $j < i$).

Pel que fa al millor cas, doncs, podem veure fàcilment que correspon a l'execució del bucle `while` el màxim de vegades. Així, el bucle `repeat` s'executarà només una vegada. El cost en el millor cas serà, doncs:

$$T^m(n) = \sum_{i=1}^{n} i = \frac{n}{2}(n+1) = O(n^2).$$

Quant al pitjor cas, no es pot calcular estrictament ja que, per a un n donat, existeix una certa probabilitat que el cost sigui arbitràriament gran.

Si els valors aleatoris generats en cada iteració són independents, aleshores podem calcular el cost en el cas mitjà per a cada iteració separadament i després sumar per a totes les iteracions:

$$T(n) = \sum_{i=1}^{n} t(i) \tag{1}$$

Suposem que en la iteració i han estat generats $i-1$ valors no repetits que es troben en les $i-1$ primeres posicions del vector. És clar que el cost en la iteració i, $t(i)$, valdrà i amb probabilitat $\frac{n-i+1}{n}$ (la probabilitat de generar un valor diferent dels $i-1$ primers). Aquest valor és el mínim possible, i correspon a una única execució del bucle `repeat`.

Podem suposar, sense pèrdua de generalitat, que l'obtenció d'un dels $i-1$ primers valors en la iteració i és equiprobable. Per tant, tots els casos en què el bucle `repeat` s'execute dues vegades els podem resumir dient que el cost és igual a $\frac{i}{2}$ (cost mitjà si es genera qualsevol dels $i-1$ primers valors) més i (cost de l'última iteració) amb probabilitat $\frac{i-1}{n}\frac{n-i+1}{n}$ (probabilitat de generar un dels valors repetits per la probabilitat de generar-ne un no repetit).

De la mateixa manera, si el bucle `repetir` s'executa $j+1$ vegades (j valors repetits i un de no repetit) el cost corresponent seria $i + j\frac{i}{2}$ amb probabilitat $\left(\frac{i-1}{n}\right)^j \frac{n-i+1}{n}$. Per tant, el valor esperat del cost en cada iteració i el podem calcular com a

$$t(i) = \sum_{j=0}^{\infty}(i + j\frac{i}{2})\left(\frac{i-1}{n}\right)^j \frac{n-i+1}{n} = i\frac{n-i+1}{n}\left[\sum_{j=0}^{\infty}\left(\frac{i-1}{n}\right)^j + \frac{1}{2}\sum_{j=0}^{\infty} j\left(\frac{i-1}{n}\right)^j\right].$$

Les dues sèries que apareixen són convergents ja que $\frac{i-1}{n} < 1$.

A partir dels resultats $\sum_{j=0}^{\infty} r^j = \frac{1}{1-r}$, i $\sum_{j=0}^{\infty} jr^j = \frac{r}{(1-r)^2}$, obtenim

$$t(i) = \frac{i(2n-i+1)}{2(n-i+1)}.$$

Podem calcular ara el cost mitjà de l'algorisme fent servir l'Equació (1) com a

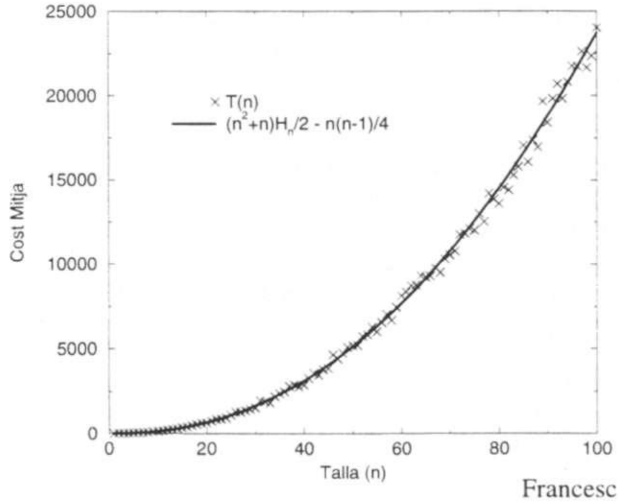
$$T(n) = \sum_{i=1}^{n} \frac{i(2n-i+1)}{2(n-i+1)} = \frac{1}{2}\sum_{k=1}^{n} \frac{(n-k+1)(n+k)}{k} = \frac{n^2}{2}\sum_{k=1}^{n}\frac{1}{k} - \frac{1}{2}\sum_{k=1}^{n}(k-1) + \frac{n}{2}\sum_{k=1}^{n}\frac{1}{k}.$$

Aleshores, el nombre mitjà de vegades que s'executa la instrucció crítica es pot escriure com a

$$T(n) = \frac{n^2+n}{2}H_n - \frac{n}{4}(n-1) = O(n^2 \log n),$$

on H_n és el nombre harmònic n-èsim.

Per tal de comprovar l'exactitud del càlcul que hem fet, en la següent figura es pot veure la funció $T(n)$ junt amb el cost mitjà mesurat (comptant les iteracions del bucle) després d'executar el corresponent programa 50 vegades per a valors de n entre 1 i 100. Encara que en la figura s'observa una relativa dispersió en els valors empíricament mesurats, es comprova que el càlcul del valor esperat de $T(n)$ és correcte.



Francesc F. Ferri

Universitat de València