# Muestreo Estadístico: Diseño muestral de Poisson [*]

García Prado, Sergio
sergio@garciparedes.me

25 de septiembre de 2017

## 1. Definición

El muestreo probabilístico de *Poisson* se caracteriza por ser un diseño muestral con probabilidades desiguales. Es decir, si $\pi_k$ es la probabilidad de añadir al individuo $k \in \{1, ..., N\}$ en la muestra, en este caso no se cumple que $\forall\ i, j\ \pi_i = \pi_j,\ i \neq j$. De esta manera, la estimación de los estadísticos se hace más complicada, sin embargo se añade un mayor grado de versatilidad al muestreo.

$$p(s) = \prod_{k \in s} \pi_k \prod_{k \in U \setminus s} (1 - \pi_k) \tag{1}$$

Si se define la variable $I_k \sim B(\pi_k),\ k \in \{1, ..., N\}$, es decir, como una distribución de *Bernoulli* de parámetro $\pi_k$, entonces la probabilidad de seleccionar la muestra $s$ de entre todo el conjunto de posibles muestras $S$ de una población $U$ se define tal y como se indica en la ecuación (1). En este diseño muestral se cumple la propiedad de que $\pi_{kl} = \pi_k \pi_l \quad k \neq l$.

Este diseño muestral se puede llevar a cabo de manera sencilla, generando $n$ valores aleatorios a partir de una distribución uniforme en el intervalo $[0, 1]$, de tal manera que $\epsilon_k$ sea el $k$-ésimo valor aleatorio. Entonces se añade el elemento $k$ a la muestra $s$ si se cumple que $\epsilon_k < \pi_k$ y se deja fuera en caso contrario.

Nótese por tanto, que este diseño muestral no tiene un un tamaño de muestra fijo, sin embargo es posible estimarlo: El tamaño $n_s$ de la muestra obtenida tendrá una esperanza de $E[n_s] = \sum_U \pi_k$ y una varianza $Var[n_s] = \sum_U \pi_k(1 - \pi_k)$.

Sea $Y$ la variable de estudio, entonces el $\pi$-estimador (insesgado) de la suma total es:

$$t = \sum_U y_k \qquad \widehat{t}_\pi = \sum_s \frac{y_k}{\pi_k} \tag{2}$$

Cuya varianza es:

$$Var[\widehat{t}_\pi] = \sum_U (\frac{1}{\pi_k} - 1) y_k^2 \qquad \widehat{Var}[\widehat{t}_\pi] = \sum_s (\frac{1}{\pi_k} - 1) \frac{y_k}{\pi_k} \tag{3}$$

Dicha estimación de la suma total tiene una varianza muy elevada, para lo cual se propone la elección apropiada de los valores $\pi_k$ a partir de un determinado ratio relacionado con los valores $y_k$, lo cual suele ser inaccesible en la mayoría de casos. Sin embargo, en algunas ocasiones se puede obtener los valores de otra variable $X$ relacionada con la variable de estudio $Y$, de tal manera que los valores $\pi_k$ sean proporcionales a $y_k$.

[*] URL: https://github.com/garciparedes/statistical-sampling-poisson-design

Otra alternativa es la elección de un estimador diferente para la suma total. Por contra, dicho estimador no es insesgado, pero su varianza es mucho menor. Este se indica a continuación:

$$\widehat{t}_{alt} = N \cdot \frac{\sum_s \frac{y_k}{\pi_k}}{\sum_s \frac{1}{\pi_k}} \tag{4}$$

## Referencias


[SSW03] Carl-Erik Särndal, Bengt Swensson, and Jan Wretman. *Model assisted survey sampling.* Springer Science & Business Media, 2003.

[TG18] Jesús Alberto Tapia García. Muestreo Estadístico 1, 2017/18.