FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

Inferencia estadística con datos sesgados por longitud

Irene Sanmartín Dopico

2018/2019

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GRAO DE MATEMÁTICAS

Traballo Fin de Grao

Inferencia estadística con datos sesgados por longitud

Irene Sanmartín Dopico

Xullo 2019

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

Área de Coñecemento: Estatística e Investigación Operativa
Título: Inferencia estadística con datos sesgados por longitud
Breve descrición do contido
Los datos sesgados por longitud surgen cuando los individuos no tienen la misma probabilidad de forma parte de la muestra. Esto es frecuente en múltiples situaciones de muestreo, por ejemplo si se toman turistas al azar de entre los que se encuentran en el lugar de destino, pues los turistas con estancias más largas tendrán más probabilidad de ser encuestados. Para que los estimadores (por ejemplo, de la duración media de la estancia) no estén sesgados, se han diseñado estimadores específicos con este tipo de datos. Cabe destacar que estos estimadores están relacionados con la media armónica. En este trabajo se revisarán estos estimadores y se ilustrarán con datos simulados y datos reales.
Recomendacións
Outras observacións

Índice general

Resumen

En este trabajo se hace un recorrido a lo largo de varias situaciones que nos llevan a la aparición de sesgo por longitud en una muestra. Presentaremos el modelo de sesgo por longitud, y a partir de él veremos que el estimador adecuado de la media poblacional no es la media aritmética simple, sino que es la media armónica. Mediante un estudio de simulación analizaremos las propiedades de la media armónica y comprobaremos que la media aritmética no es un buen estimador de la media poblacional, en el caso de tener muestras sesgadas.

También trabajaremos los métodos de estimación paramétrica más habituales (mínimos cuadrados ponderados) adaptados al sesgo por longitud. Y empleando el programa R generaremos muestras sesgadas de un modelo de regresión y estimaremos los coeficientes con y sin ponderación, para hacer una comparación de los sesgos y desviaciones típicas aproximadas.

Abstract

Several situations that exhibit biased sampling are presented in this paper. The length bias model is introduced and we will see that the appropriate estimator of the population mean is not the simple arithmetic mean, but is the harmonic mean. Through a simulation study, we will analyze the properties of the harmonic mean and we will verify that the arithmetic mean is not a good estimator of the population mean in the case of biased samples.

We will also work with the most usual parametric estimation methods (weighted least squares) adapted to the length bias. And using the R program, we will generate biased samples of a regression model, and we will estimate the coefficients with and without weighting, to make a comparison of approximate biases and standard deviations.

Introducción

En este trabajo vamos a tratar con datos sesgados por longitud. Aunque normalmente cuando hablamos de sesgo solemos referirnos al estimador y al problema de estimación, los datos sesgados no se refieren a esta cuestión, sino al proceso de registro de los propios datos.

Recordemos que el sesgo de un estimador $\hat{\theta}$ para un parámetro poblacional θ es: $Sesgo(\hat{\theta}) = E(\theta) - \theta$, y decimos que el estimador es insesgado si su sesgo vale cero. Sin embargo, los datos sesgados son aquellos que no tienen la misma probabilidad de aparecer en la muestra. Si una muestra contiene datos sesgados, puede incluir preferencia o marginar cierto tipo de resultados. Como consecuencia del sesgo en la toma de datos, se puede producir un sesgo en los estimadores ordinarios, como por ejemplo en la media muestral, de modo que es necesaria alguna corrección en los estimadores, posteriores a la toma de la muestra.

Si $0 \leq w(x) \leq 1$ es la probabilidad de registrar una observación x, el problema que plantean los datos sesgados o las observaciones sesgadas es que provienen de una población X^w, que recoge el comportamiento estocástico de la población de interés X, y a este comportamiento le añade el efecto que el procedimiento de observación de dicha población pudiera tener. Ocurre entonces que aunque la distribución que sigue X^w está relacionada con la de X, ambas son diferentes.

En Patil (1984) podemos ver un claro ejemplo de lo que son los datos sesgados por longitud. En él se pretende estimar la duración media de la estancia de turistas en Marruecos. Para ello se contactó con turistas que se encontraban saliendo del país, y con turistas durante su estancia en hoteles. Las duraciones medias fueron 9 días y 17,8 días respectivamente. Claramente, vemos que la duración media de los turistas alojados en hoteles es casi el doble de los que se encuentran en la salida del país. Esto no supone ninguna contradicción, pues en cada encuesta se están midiendo variables diferentes. En el caso de los hoteles, los turistas encuestados forman una muestra sesgada, ya que al ser los instantes de visita

a los hoteles aleatorios, los turistas que más frecuentemente encontraban los encuestadores eran aquellos cuya estancia era más larga, y por ello las personas que tenían estancias más largas eran recogidos en la muestra con mayor probabilidad. Sin embargo, en el primer caso la muestra es simple, ya que los individuos tienen la misma probabilidad de ser encuestados.

Los datos sesgados por longitud aparecen en muchos problemas de muestreo, no solo de tiempo o duración, sino también en el ámbito tecnológico e industrial, como se puede ver en Cox (1969). También tienen gran importancia en problemas económico-sociales, asociados a tiempos de duración, como puede ser el estudio del tiempo de desempleo de una población.

Supongamos que en un instante al azar contactamos con una Oficina de Empleo, y extraemos una muestra de individuos que se encuentran desempleados. Dicha muestra está sesgada por longitud, ya que las personas que llevan más tiempo sin empleo son incluidas en ella con mayor probabilidad. Por otra parte, si de cada individuo sólo podemos medir el tiempo desde que empezó a estar sin empleo hasta el momento en el que se realiza el muestreo, las observaciones no miden toda la duración del intervalo de desempleo, luego van a ser lo que denominamos como datos "censurados", que requieren un tratamiento estadístico especial.

El origen y problemática de tratamiento de los datos sesgados por longitud está relacionado con la Paradoja del tiempo de espera que podemos ver en Feller (1971). En ella se supone que los instantes en los que llega un autobús a una parada siguen una Poisson de parámetro λ. Luego si un observador que se encuentra en la parada, registra el tiempo que transcurre entre la llegada de dos autobuses consecutivos, verá que dicho tiempo es aproximadamente λ. Sin embargo, si consideramos otro observador que llega a la parada en un instante aleatorio y uniformemente distribuido a lo largo del tiempo que transcurre entre la parada de dos autobuses consecutivos, éste también llegará a que el tiempo medio que hay que esperar hasta que pase el siguiente autobús es λ y no $\lambda/2$ como podríamos pensar. Esto es debido a que los tiempos medidos por ambos observadores tienen distribuciones distintas, ya que están observando el tiempo que transcurre de forma diferente.

Además, en dicha paradoja también podemos ver que este tipo de datos quedan caracterizados por el hecho de que la distribución F^w de la población observada, X^w es proporcional a x. Es decir,

$$dF^w(x) = \frac{x}{\mu_x} dF(x). \tag{1}$$

En general, la mayoría de fenómenos en los que se accede a los datos mediante el "descubrimiento" son candidatos a ser datos sesgados, luego necesitan métodos de estimación e inferencia diferentes de los habituales.

Capítulo 1

Estimación con datos sesgados por longitud

Los valores que obtenemos de una muestra nos permiten conocer los valores que encontraríamos en una determinada población. Luego cuestiones como la forma en la que se seleccionen los sujetos o el tamaño de la muestra, tienen una gran importancia en el momento de poder determinar en qué medida es representativa de la población a la cual se refiere.

Una muestra aleatoria o probabilística es aquella en la que todos los sujetos de la población tienen la misma probabilidad de ser escogidos. Las muestras aleatorias aseguran o garantizan mejor el poder extrapolar los resultados, y en ellas tenemos más seguridad de que se encuentren representadas las características importantes de la población, en la proporción que les corresponde. Si el 20 % de la población tiene la característica A, (una determinada edad, una determinada situación económica, etc.) podemos esperar que en la muestra también habrá en torno a un 20 % con esa característica.

Si la muestra no es aleatoria (no probabilística) puede suceder que esté sesgada y que por lo tanto no sea representativa de la población general, porque predominan más unos determinados tipos de sujetos que otros. Por ejemplo, si hacemos una pregunta a los conductores que se paran ante un semáforo, prescindimos de los que utilizan otro medio de transporte, y si hacemos la pregunta a la salida de una estación de metro, prescindimos de los que tienen y utilizan coche particular, etc.

En la primera sección de este capítulo veremos cómo estimar la función de distribución

con una muestra no sesgada, así como la media y la varianza poblacionales. También mencionaremos algunos resultados relacionados con propiedades de la función de distribución empírica, como el Teorema de Glivenko-Cantelli. En la segunda sección, vamos a volver a estimar la función de distribución, pero en este caso, considerando muestras sesgadas por longitud. Que una variable esté sesgada por longitud, quiere decir que en lugar de observarla a ella, observaremos otra que toma los mismos valores que la primera pero con probabilidades diferentes. Por lo tanto, veremos que para hacer esta estimación, tendremos que relacionar la función de distribución de una variable sesgada, con la de otra variable no sesgada. También veremos que la media armónica es un estimador de la media poblacional de la variable sesgada por longitud y aproximaremos su distribución haciendo uso del Teorema central del límite y del Método delta.

1.1. Estimación con una muestra no sesgada

Vamos a comenzar trabajando con una muestra no sesgada. En esta sección veremos quiénes son los estimadores de la función de distribución, de la media y de la varianza poblacional, así como algunas propiedades.

1.1.1. Estimación de la función de distribución

Definición 1.1. Sea $X_1, ..., X_n$ una muestra aleatoria de X con función de distribución F. Se define la **función de distribución empírica** como la función

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), x \in \mathbb{R}, \tag{1.1}$$

que a cada número real x le asigna la proporción de valores observados que son menores o iguales que x.

Es inmediato comprobar que F_n así definida es una función de distribución que cumple:

1. $F_n(x) \in [0, 1]$ para todo $x \in \mathbb{R}$.

2. F_n es continua por la derecha.

3. F_n es no decreciente.

4. $\lim_{x \to -\infty} F_n(x) = 0$.

5. $\lim_{x\to\infty} F_n(x) = 1$.

En general, F_n se puede considerar como una función de distribución de una variable aleatoria discreta (que podemos llamar X_e), que asigna probabilidad $\frac{1}{n}$ a cada uno de los n valores X_i, con $i = 1, 2, ..., n$. Además F_n es un estimador de F.

x_i	x_1	x_2	...	x_n
$p_i = P(X_e = x_i)$	1/n	1/n	...	1/n

De este modo, F_n asigna a un conjunto A del espacio muestral de X la probabilidad empírica

$$\frac{1}{n}\sum I(X_i \in A),$$

que es la proporción de datos X_i que pertenecen a A.

Si se fija el valor de x entonces la variable aleatoria $I(X_i \leq x)$ toma valores $0, 1$, con probabilidad $1 - p, p$ respectivamente. Es decir, es una Bernoulli de parámetro $p = F(x)$. De ahí se deduce que F_n es una variable aleatoria y que $nF_n(x)$ tiene distribución binomial con parámetros n y $p = F(x)$. Es decir, con media $nF(x)$ y varianza $nF(x)(1-F(x))$.

Así,

$$\mathbb{E}(nF_n(x)) = nF(x) \Longrightarrow \mathbb{E}(F_n(x)) = F(x) \tag{1.2}$$

$$Var(nF_n(x)) = nF(x)(1 - F(x)) \Longrightarrow Var(F_n(x)) = \frac{1}{n}F(x)(1 - F(x)). \tag{1.3}$$

Como hemos visto, la esperanza de la función de distribución empírica coincide con la función de distribución teórica, $\mathbb{E}(F_n(x)) = F(x)$. Por lo tanto $F_n(x)$ es un estimador insesgado de $F(x)$.

De lo anterior se sigue que la función de distribución empírica es un proceso estocástico: si consideramos un espacio probabilístico (Ω, A, P) donde están definidas las sucesiones de variables aleatorias $\{X_n\}_{n\geq 1}$ a partir de las cuales definiremos la función de distribución empírica, tenemos que

$$Fn : (\Omega, A, P) \times (\mathbb{R}, B) \longrightarrow [0, 1]$$

$$(\omega, x) \longrightarrow Fn(x)(\omega) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)(X_i(\omega)).$$

Fijado x, $F_n(x)(\cdot) : (\Omega, A, P) \longrightarrow [0,1]$ es una variable aleatoria.

Fijado ω, $F_n(\cdot)(\omega) : \mathbb{R} \longrightarrow [0,1]$ es una función de distribución.

Por lo tanto, la función de distribución empírica es una función de distribución aleatoria.

Podemos englobar lo dicho anteriormente en el siguiente teorema.

Teorema 1.2. *Sea $\{X_n\}_{n\geq 1}$, sucesión de variables aleatorias independientes e idénticamente distribuidas definidas en el espacio de probabilidad (Ω, A, P) con función de distribución común F. Se denota por F_n la función de distribución empírica obtenida de las n primeras variables aleatorias $X_1, ..., X_n$. Sea $x \in \mathbb{R}$. Se verifica lo siguiente:*

1. $P(F_n(x) = j/n) = \binom{n}{j} F(x)^j \left(1 - F(x)\right)^{n-j}$, $j = 0, ..., n$.

2. $\mathbb{E}(F_n(x)) = F(x)$, $Var(F_n(x)) = \dfrac{1}{n} F(x)\left(1 - F(x)\right)$.

3. $F_n(x) \to F(x)$ *casi seguro.*

4. *Se tiene*

$$\frac{\sqrt{n}\left(F_n(x) - F(x)\right)}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{d} Z \tag{1.4}$$

donde Z es una variable aleatoria con distribución normal estándar, y la convergencia es en distribución.

Demostración. Los apartados (1) y (2) son consecuencia inmediata del hecho de que $nFn(x) \sim B(n, p = F(x))$, como hemos visto.

Por otro lado, si definimos $Y_i = I(X_i \leq x)$, se tiene que $F_n(x)$ es igual a la media aritmética de las variables aleatorias $Y_1, ..., Y_n$. Así, el apartado (3) es una aplicación inmediata de la ley fuerte de los grandes números y el apartado (4) es consecuencia del teorema central de límite. □

Como hemos visto en la asignatura de *Probabilidad y Estatística*, la convergencia casi segura, implica convergencia en probabilidad, luego del teorema anterior deducimos que

$$F_n(x) \xrightarrow{p} F(x), x \in \mathbb{R}.$$

Pero se tiene un resultado más potente, que es el llamado Teorema de Glivenko-Cantelli. Éste nos dice que $F_n(x)$ converge a $F(x)$ de forma uniforme para todos los valores de x.

Teorema 1.3 (Glivenko-Cantelli). *Sea $\{X_n\}_{n\geq 1}$, sucesión de variables aleatorias independientes e idénticamente distribuidas definidas en el espacio de probabilidad (Ω, A, P) con función de distribución común F. Se denota por F_n la función de distribución empírica obtenida de las n primeras variables aleatorias $X_1, ..., X_n$. Entonces:*

$$Sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \xrightarrow{c.s.} 0.$$

La demostración del teorema anterior puede encontrarse en Vélez y García (1993), p.36.

Es importante resaltar que según el apartado (3) del Teorema 1.2, las distribuciones empíricas asociadas a muestras de tamaño n convergen débilmente a la distribución de probabilidad teórica identificada por F, para casi todas las muestras de tamaño infinito que se extraigan de F. Esta es una de las consecuencias más importantes de dicho teorema: la distribución empírica converge débilmente con probabilidad 1 a la poblacional cuando el tamaño de la muestra tiende a infinito. Esto garantiza la posibilidad de realizar inferencia estadística. Los aspectos probabilísticos de una característica X, medida en una población, se resumen de forma estilizada en una distribución de probabilidad F, la cual puede ser aproximada mediante las distribuciones empíricas F_n obtenidas por muestreo de la población en estudio. El Teorema de Glivenko-Cantelli afirma que esas aproximaciones son uniformes en x. Por esta razón el Teorema de Glivenko-Cantelli se llama a veces Teorema Fundamental de la Estadística Matemática: da una fundamentación de la inferencia estadística, cuyo objetivo principal consiste en extraer información sobre F a partir de las observaciones muestrales.

1.1.2. Estimación de la media y la varianza

Definición 1.4. Dada una muestra aleatoria simple $X_1, X_2, ..., X_n$ de X, con $\mathbb{E}(X) = \mu$ y $Var(X) = \sigma^2$. La **media muestral** es el estadístico obtenido tomando la media aritmética de los elementos de la muestra. La denotaremos mediante $\bar{X}$:

$$\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i$$

La media aritmética muestral es un estimador insesgado de la media poblacional, pues:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{n} X_i\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\left(\mathbb{E}[X_1] + ... + \mathbb{E}[X_n]\right) = \frac{1}{n}\, n\, \mu = \mu$$

Por otra parte, la varianza de la media muestral sería:

$$Var(\bar{X}) = Var\left[\sum_{i=1}^{n} \frac{1}{n} X_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\, n\, \sigma^2 = \frac{\sigma^2}{n}$$

Como podemos ver, la media muestral en general es diferente a la media poblacional. La magnitud del error que estamos cometiendo dependerá de la distribución de la media muestral, y en particular de su variabilidad. A medida que aumenta el tamaño muestral n, el valor de la varianza de la media muestral decrece, reduciéndose así el error.

El requisito mínimo deseable para un estimador es que a medida que el tamaño de la muestra crece, el valor del estimador tienda a ser el valor del parámetro poblacional, propiedad que se denomina consistencia.

Definición 1.5. Se define el **error cuadrático medio** de un estimador como

$$ECM(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right].$$

Y diremos que un estimador converge en media cuadrática al verdadero valor del parámetro que está estimando, si su ECM tiende a cero cuando el tamaño de la muestra se va a infinito.

Además también se cumple que

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + (Sesgo(\hat{\theta}))^2.$$

Definición 1.6. Se dice que un **estimador** es **consistente** si converge en probabilidad al valor verdadero del parámetro que pretende estimar, cuando el número de datos de la muestra tiende a infinito.

La convergencia en media cuadrática implica la convergencia en probabilidad, tal y como vimos en la asignatura de *Probabilidad y Estadística*. Luego si un estimador converge en media cuadrática es consistente.

Además, como el error cuadrático medio es la suma de la varianza y del cuadrado del sesgo y ambos sumandos son no negativos, es inmediato el siguiente resultado:

Proposición 1.7. *La sucesión* $\{\hat{\theta}_n\}_{n=1}^{\infty}$ *de estimadores de un parámetro* θ*, es consistente en media cuadrática, si y sólo si se cumplen las dos condiciones siguientes:*

1. *Es asintóticamente insesgado. Esto es,* $\lim_{n\to\infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

2. $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$.

Luego si el estimador es insesgado, la consistencia en media cuadrática equivale a $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$ (ya que el sesgo es igual a cero).

La media muestral es un estimador consistente de la media poblacional, pues:

$$\lim_{n\to\infty} \mathbb{E}(\hat{\theta}) = \lim_{n\to\infty} \mathbb{E}(\bar{X}) = \mu$$

$$\lim_{n\to\infty} Var(\hat{\theta}) = \lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Definición 1.8. Dada una muestra aleatoria simple $X_1, X_2, ..., X_n$ de X, con $\mathbb{E}(X) = \mu$ y $Var(X) = \sigma^2$. La **varianza muestral** se define como :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Vamos a calcular la esperanza de la varianza muestral y comprobar así que es un estimador sesgado de la varianza poblacional. Para ello vamos a realizar algunos cálculos previos sumando y restando la esperanza de la variable aleatoria poblacional.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X} + \mu - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \mu) - (\bar{X} - \mu)\right]^2.$$

Desarrollando el cuadrado:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)\right] =$$

$$= \frac{1}{n}\left[\sum_{i=0}^{n}(X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu)\right] =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(X_i-\mu)^2 + n(\bar{X}-\mu)^2 - 2(\bar{X}-\mu)(n\bar{X}-n\mu)\right] =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - (\bar{X}-\mu)^2\right]$$

Entonces:

$$\mathbb{E}(S^2) = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i-\mu)^2\right] - \mathbb{E}\left[(\bar{X}-\mu)^2\right] =$$

$$= \sigma^2 - \mathbb{E}[(\bar{X}-\mu)^2].$$

Pero por definición y empleando resultados anteriores, se tiene que

$$\mathbb{E}[(\bar{X}-\mu)^2] = ECM(\bar{X}) = Var(\bar{X}) + (Sesgo(\bar{X}))^2 = \frac{\sigma^2}{n}.$$

Luego,

$$\mathbb{E}(S^2) = \frac{n-1}{n}\sigma^2.$$

Por lo tanto concluimos que la varianza muestral es un estimador sesgado de la varianza poblacional, como ya habíamos dicho.

Ahora vamos a definir un estimador insesgado de la varianza poblacional y después probaremos que lo es.

Definición 1.9. Dada una muestra aleatoria simple $X_1, X_2, ..., X_n$ de X, con $\mathbb{E}(X) = \mu$ y $Var(X) = \sigma^2$. La **cuasi-varianza muestral** se define como :

$$Sc^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

La cuasi-varianza muestral es un estimador insesgado de la varianza poblacional.

Como $S_c^2 = \frac{n}{n-1}\cdot S^2$, podemos obtener de lo anterior que

$$\mathbb{E}\left[S_c^2\right] = \mathbb{E}\left[\frac{n}{n-1}S^2\right] = \frac{n}{n-1}\mathbb{E}\left[S^2\right] = \left(\frac{n}{n-1}\right)\left(\frac{n-1}{n}\right)\sigma^2 = \sigma^2.$$

Por lo tanto la cuasi-varianza muestral es un estimador insesgado de la varianza.

1.2. Estimación con una muestra sesgada

El sesgo muestral, a veces también llamado efecto de selección o error muestral es una distorsión de un análisis estadístico que surge debido al método de recolección de muestras. Es importante tenerlo en cuenta, pues sino algunas de las conclusiones podrían ser erróneas.

En esta sección vamos a estimar la función de distribución de variables sesgadas por longitud, tanto en el caso discreto, como en el continuo. Ahora bien, que una variable esté sesgada por longitud, quiere decir que en lugar de observarla a ella, observaremos otra que toma los mismos valores que la primera, pero con probabilidades diferentes. Por lo tanto, para hacer esta estimación, tendremos que relacionar la función de distribución de la variable sesgada con la de otra variable no sesgada.

Veremos que un estimador de la media poblacional de la variable sesgada por longitud, es la media armónica, y aproximaremos su distribución haciendo uso del Teorema central del límite y del Método delta. Aplicando propiedades llegaremos a una relación entre las esperanzas de las variables, sesgada y no sesgada, que vamos a denotar por X e Y respectivamente. Concretamente $\mathbb{E}(X) \leq \mathbb{E}(Y)$, lo cual resulta intuitivo, pues en una muestra sesgada por longitud los datos con mayor longitud tienen más probabilidad de ser seleccionados que el resto.

Finalmente, en el último apartado nos centraremos en la estimación de la función de distribución no ponderada a partir de una muestra sesgada por longitud. También calcularemos su distribución asintótica, haciendo uso nuevamente del Teorema central del límite y del Método delta, pero en el caso multivariante.

1.2.1. Distribución de la variable sesgada

Comencemos exponiendo un ejemplo sencillo para ver el significado de muestra sesgada por longitud. Vamos a considerar una urna con bolas de diferentes tipos y sus frecuencias, y veremos cómo varían dichas frecuencias, en el caso de suponer que la muestra estuviera sesgada.

Supongamos que tenemos una urna con 100 bolas: 25 marcadas con un uno, 50 marcadas con un dos, y otras 25 marcadas con un tres.

Valores	1	2	3
Frecuencias	25	50	25

Entonces, la probabilidad de observar una bola marcada con un uno es 1/4, con un dos es 1/2 y con un tres es 1/4.

Supongamos ahora que la muestra anterior está sesgada por longitud. Esto querría decir que mientras que las bolas marcadas con un 1 las observaríamos una única vez cada una, las bolas marcadas con un 2 las observaríamos dos veces cada una, y las marcadas con un 3, tres veces cada una. Por lo tanto es como si en total tuvieramos 200 bolas, 25 marcadas con un 1, 100 marcadas con un 2, y 75 marcadas con un 3.

Valores	1	2	3
Frecuencias	25	100	75

Y las probabilidades con las que se observa cada tipo de bola pasarían a ser:

$$P(Y = 1) = \frac{25}{200} = 1/8$$

$$P(Y = 2) = \frac{100}{200} = 1/2$$

$$P(Y = 2) = \frac{75}{200} = 3/8$$

Veámoslo ahora formalmente para el caso discreto. Consideremos una población, modelizada matemáticamente mediante una variable aleatoria discreta X que puede tomar n valores: $x_1, x_2, ..., x_n$, con probabilidades $p_1, p_2, ..., p_n$ respectivamente. El sesgo por longitud de la variable X supone que realmente vamos a observar una variable Y, cuyas probabilidades de observación se ven afectadas proporcionalmente por el valor de la variable original. Es decir, estamos observando una variable aleatoria Y, que toma los mismos valores que X, pero con probabilidades $q_1, q_2, ..., q_n$, siendo $q_i = C \cdot x_i \cdot p_i$, para todo $i = 1, ..., n$, donde C es una constante. Además podemos calcular C, pues:

$$1 = q_1 + ... + q_n = C(x_1 \cdot p_1 + ... + x_n \cdot p_n)$$

Luego,

$$C = \frac{1}{(x_1 \cdot p_1 + ... + x_n \cdot p_n)} = \frac{1}{E(X)}.$$

Por lo tanto,

$$P(Y = y) = \frac{yP(X = y)}{\mu_x}.$$

Extendiendo al caso continuo, sea $X_1, ..., X_n$ una muestra de variables aleatorias positivas, con función de densidad no ponderada $f(x)$. La densidad ponderada o sesgada por longitud de $f(x)$ es

$$g(x) = \frac{xf(x)}{\mu}, x > 0, \tag{1.5}$$

donde μ es la media de $f(x)$. Más aún, en términos generales podríamos decir que $G(y) = \int \frac{y}{\mu}\, dF(y)$.

Como decíamos, la variable X no puede observarse, luego para estimar su media poblacional, podemos tratar de relacionarla con características de la variable observable y sesgada Y. Así, se prueba fácilmente que:

$$\mathbb{E}\left(\frac{1}{Y}\right) = \left(\frac{1}{x_1}\right) \cdot q_1 + ... + \left(\frac{1}{x_n}\right) \cdot q_n = C(p_1 + ... + p_n) = C = \frac{1}{\mathbb{E}(X)}.$$

También es sencillo calcular la media empleando la función de densidad:

$$\mathbb{E}\left(\frac{1}{Y}\right) = \int \frac{1}{y}\, g(y)\, dy = \int \frac{1}{y}\, \frac{yf(y)}{\mu}\, dy = \frac{1}{\mathbb{E}(X)}.$$

Luego tanto en el caso discreto, como en el continuo, se tiene que $\mathbb{E}(X) = \frac{1}{\mathbb{E}(1/Y)}$. Es decir, la media de la variable de interés es la media armónica teórica de la variable sesgada por longitud.

Esta relación de las medias de las variables sesgadas y no sesgadas, nos permite construír un estimador de la media poblacional de X, a partir de una muestra de Y. Supongamos que observamos una muestra de tamaño n de la variable sesgada por longitud: $Y_1, Y_2, ..., Y_n$. Si nuestro interés es estimar la media de la variable original $\mathbb{E}(X)$, es lógico hacerlo usando la media armónica de la muestra de la variable Y:

$$\hat{\mu} = \frac{n}{\frac{1}{Y_1} + \cdots + \frac{1}{Y_n}}.$$

Definición 1.10. La **media armónica** de una cantidad finita de números se define como el recíproco o inverso de la media aritmética de los recíprocos de dichos valores.

Así, dados n números $y_1, y_2, ..., y_n$ la media armónica será igual a:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{y_i}} = \frac{n}{\frac{1}{y_1} + \cdots + \frac{1}{y_n}}.$$

Una propiedad de la media armónica es que siempre es menor o igual que la media aritmética, ya que para cualquier $Y > 0$:

$$\frac{1}{\mathbb{E}(1/Y)} \leq \mathbb{E}(Y).$$

Por lo tanto, podemos afirmar que también se cumple que:

$$\frac{n}{\frac{1}{y_1} + \cdots + \frac{1}{y_n}} \leq \frac{y_1 + \cdots + y_n}{n}$$

La propiedad anterior se obtiene directamente aplicando la llamada "Desigualdad de Jensen" para funciones convexas, la cual se puede escribir de la siguiente forma:

Desigualdad de Jensen

Sea φ una función convexa, entonces se cumple:

$$\varphi\left(\mathbb{E}\{Y\}\right) \leq \mathbb{E}\{\varphi(Y)\}.$$

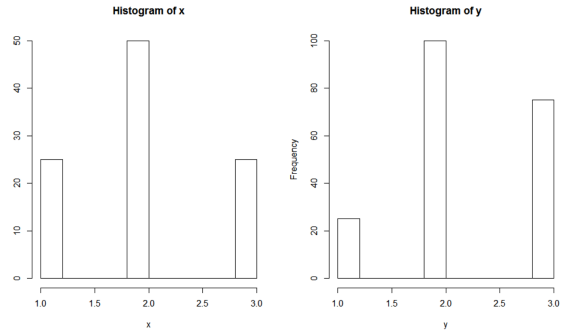
Tomando como función convexa $\varphi(Y) = \frac{1}{Y}$, se tiene que $\frac{1}{\mathbb{E}(Y)} \leq \mathbb{E}(1/Y)$. Y despejando, obtenemos $\frac{1}{\mathbb{E}(1/Y)} \leq \mathbb{E}(Y)$, que es lo que queríamos probar.

Como habíamos visto que $\mathbb{E}(X) = \frac{1}{\mathbb{E}(1/Y)}$, y empleando la propiedad de la media armónica que acabamos de ver, llegamos a:

$$\mathbb{E}(X) = \frac{1}{\mathbb{E}(1/Y)} \leq \mathbb{E}(Y).$$

Es decir, la media aritmética de la variable no sesgada X, siempre es menor o igual que la media aritmética de la variable sesgada Y. Esto resulta fácil de ver intuitivamente. En una muestra sesgada por longitud los datos con mayor longitud tienen más probabilidad de ser seleccionados que el resto, luego la media de esta muestra ha de ser mayor que la de otra no sesgada.

Si consideramos el ejemplo que vimos al principio de este apartado, se tiene que en la variable no sesgada la media es 2, mientras que en la variable sesgada es 2,25.



1.2.2. Estimación de la media con una muestra sesgada: media armónica

En el apartado anterior hemos construido un estimador de $\mathbb{E}(X)$. Ahora nos ocuparemos de calcular su distribución límite empleando el Teorema central del límite y el Método delta.

Consideremos un conjunto de variables aleatorias $Y_1, Y_2..., Y_n$, mutuamente independientes y con la misma distribución que Y, con media $\mathbb{E}(1/Y) = \dfrac{1}{\mathbb{E}(X)}$. Sea $\hat{\mu}$, el estimador de $\mu = \mathbb{E}(X) = \dfrac{1}{\mathbb{E}(1/Y)}$ calculado previamente.

Comencemos viendo que $\hat{\mu}$ es un estimador sesgado de la media poblacional, y sin embargo $\dfrac{1}{\hat{\mu}}$ es un estimador insesgado de la inversa de la media poblacional:

$$\mathbb{E}\left(\frac{1}{\hat{\mu}}\right) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\frac{1}{Y_i}\right] = \frac{1}{n}\left[\mathbb{E}\left(\frac{1}{Y_1}\right) + ... + \mathbb{E}\left(\frac{1}{Y_n}\right)\right] =$$

$$= \mathbb{E}\left(\frac{1}{Y}\right) = \frac{1}{\mathbb{E}(X)}.$$

Luego $\frac{1}{\hat{\mu}}$ es un estimador insesgado de la inversa de la media poblacional. Sin embargo,

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}\left(\frac{n}{\frac{1}{Y_1} + ... + \frac{1}{Y_n}}\right).$$

Pero como $\mathbb{E}\left(\frac{1}{\hat{\mu}}\right) \neq \frac{1}{\mathbb{E}(\hat{\mu})}$, se tiene que $\mathbb{E}(\hat{\mu}) \neq \mathbb{E}(X)$. Por lo tanto $\hat{\mu}$ es un estimador sesgado de μ.

Para ver la convergencia de $\hat{\mu}$, en primer lugar tenemos que aplicar la Ley de los grandes números para ver la de $\frac{1}{\hat{\mu}}$. Posteriormente, empleamos el Teorema de la función continua y obtenemos la primera. Por lo tanto vamos a comenzar enunciando la Ley de los grandes números para ver que estamos en las hipótesis del teorema.

Teorema 1.11. *(Teorema de Kolmogorov) Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad, mutuamente independientes y con la misma distribución que X, entonces*

$$\mathbb{E}(X) = \mu < \infty \Longleftrightarrow \bar{X}_n \xrightarrow{c.s.} \mu.$$

Como la sucesión de variables aleatorias que estamos considerando cumple las hipótesis del Teorema anterior, se tiene que

$$\frac{1}{\hat{\mu}} = \frac{1}{n}\left(\frac{1}{Y_1} + ... + \frac{1}{Y_n}\right) \xrightarrow{c.s.} \mathbb{E}\left(\frac{1}{Y}\right).$$

Y como la convergencia casi segura implica la convergencia en probabilidad,

$$\frac{1}{\hat{\mu}} \xrightarrow{P} \mathbb{E}\left(\frac{1}{Y}\right).$$

Empleando ahora el teorema de la función continua, podemos concluir que la media armónica también converge casi segura y en probabilidad a $\mathbb{E}(X)$, es decir:

$$\hat{\mu} = \left(\frac{n}{\frac{1}{Y_1} + ... + \frac{1}{Y_n}} \right) \xrightarrow[P]{c.s} \frac{1}{\mathbb{E}(1/Y)} = \mathbb{E}(X).$$

Ahora vamos a aproximar la distribución de $\hat{\mu}$ haciendo uso del Teorema central del límite y del Método delta. Comencemos recordando estos resultados.

Teorema 1.12. *(Teorema de Levy-Linderberg) Sea $\{X_n\}$ una sucesión de variables aleatorias mutuamente independientes y con la misma distribución que cierta variable X, de la que suponemos que tiene media y varianza, que denotaremos $\mu = \mathbb{E}(X)$ y $\sigma^2 = Var(X)$. Entonces se tiene que*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

siendo $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$.

Teorema 1.13 (Método delta)**.** *Si $\{T_n\}$ es una sucesión de variables aleatorias tales que*

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

y g es una función diferenciable en θ, entonces

$$\sqrt{n}\,(g(T_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2).$$

En nuestro caso tenemos un conjunto de variables aleatorias, independientes e idénticamente distribuidas $\frac{1}{Y_1}, \frac{1}{Y_2}, ..., \frac{1}{Y_n}$ con media $\mathbb{E}\left(\frac{1}{Y}\right)$ y varianza $Var\left(\frac{1}{Y}\right)$. Entonces, aplicando el Teorema central del límite:

$$\frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i} - \mathbb{E}\left(\frac{1}{Y}\right)}{\sqrt{\frac{Var\left(\frac{1}{Y}\right)}{n}}} \xrightarrow{d} N(0,1)$$

Por lo tanto,

$$\sqrt{n}\left(\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right) \xrightarrow{d} N(0, Var(1/Y)). \tag{1.6}$$

Si definimos $\mu_{-1} = \mathbb{E}(1/X)$, tenemos que:

$$\mathbb{E}\left(\frac{1}{Y^2}\right) = \int \frac{1}{y^2} \cdot g(y)dy = \int \frac{1}{y^2} \cdot y \cdot \frac{f(y)}{\mu}dy = \frac{1}{\mu}\int \frac{1}{y} \cdot f(y)dy = \frac{1}{\mu}\mathbb{E}\left(\frac{1}{X}\right) = \frac{\mu_{-1}}{\mu}$$

$$\left[\mathbb{E}\left(\frac{1}{Y}\right)\right]^2 = \left(\frac{1}{\mu}\right)^2$$

Así, la varianza es:

$$Var\left(\frac{1}{Y}\right) = \mathbb{E}\left(\frac{1}{Y^2}\right) - \left[\mathbb{E}\left(\frac{1}{Y}\right)\right]^2 = \frac{\mu_{-1}}{\mu} - \left(\frac{1}{\mu}\right)^2 = \frac{\mu\mu_{-1} - 1}{\mu^2}.$$

Y sustituyendo en (1.6):

$$\sqrt{n}\left(\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \frac{\mu\mu_{-1} - 1}{\mu^2}\right).$$

Pero como lo que queremos ver es la distribución de $\hat{\mu}$, tenemos que aplicar el método delta considerando $h(x) = 1/x$:

$$\sqrt{n}\left(h\left(\frac{1}{\hat{\mu}}\right) - h\left(\frac{1}{\mu}\right)\right) \xrightarrow{d} N\left(0, h'\left(\frac{1}{\mu}\right)^2 \sigma^2\right). \tag{1.7}$$

$$h\left(\frac{1}{\hat{\mu}}\right) = \hat{\mu}$$

$$h\left(\frac{1}{\mu}\right) = \mu$$

$$h'\left(\frac{1}{\mu}\right)^2 = \mu^4 \implies h'\left(\frac{1}{\mu}\right)^2 \sigma^2 = (\mu\mu_{-1} - 1)\mu^2$$

Por lo tanto, sustituyendo en (1.7):

$$\sqrt{n}\,(\hat{\mu} - \mu) \xrightarrow{d} N(0, (\mu\mu_{-1} - 1)\mu^2).$$

Es decir, la distribución límite de $\hat{\mu}$ es una normal con media μ y varianza $\dfrac{(\mu\mu_{-1} - 1)\mu^2}{n}$.

Así, si queremos calcular un intervalo de confianza asintótico de nivel $1 - \alpha$ para la media:

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{\sqrt{(\mu\mu_{-1} - 1)\mu^2}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

Y haciéndo cálculos llegamos a:

$$P\left(\hat{\mu} - z_{\alpha/2}\sqrt{(\mu\mu_{-1} - 1)\mu^2} \leq \mu \leq \hat{\mu} + z_{\alpha/2}\sqrt{(\mu\mu_{-1} - 1)\mu^2}\right) = 1 - \alpha$$

Luego un intervalo de confianza asintótico para la media, sería:

$$\left(\hat{\mu} - z_{\alpha/2}\sqrt{(\mu\mu_{-1} - 1)\mu^2}, \hat{\mu} + z_{\alpha/2}\sqrt{(\mu\mu_{-1} - 1)\mu^2}\right)$$

donde $z_{\alpha/2}$ es el valor de una distribución normal estándar que deja a su derecha una probabilidad de $\alpha/2$ para un intervalo de confianza de $(1 - \alpha)$.

1.2.3. Estimación de la función de distribución a partir de una muestra sesgada por longitud

En este apartado vamos a estimar la función de distribución no ponderada (*cdf*) evaluando en z. Posteriormente aproximaremos su distribución, empleando de nuevo el Teorema central del límite y el Método delta.

Se puede ver fácilmente que:

$$\frac{\mathbb{E}\left(\frac{1}{Y}\right) I\{Y \leq z\}}{\mathbb{E}\left(\frac{1}{Y}\right)} = \mu \int_0^z \frac{1}{y}\, g(y)\, dy = \mu \int_0^z \frac{1}{y}\, \frac{y f(y)}{\mu}\, dy = \int_0^z f(y)\, dy = F(z)$$

Es decir, $F(z) = \dfrac{\mathbb{E}\left(\frac{1}{Y} I\{Y \leq z\}\right)}{\mathbb{E}\left(\frac{1}{Y}\right)}$, lo cual se puede estimar mediante la media muestral:

$$\hat{F}(z) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i} P\{\frac{1}{Y_i} \leq z\}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i}}.$$

Para calcular la distribución límite de $\hat{F}(z)$, vamos a hacer uso del Teorema central del límite y del Método delta en el caso multidimensional. Comencemos recordando estos resultados.

Teorema 1.14 (Teorema de Levy-Lindeberg multivariante)**.** *Sea $\{X_n\}$ una sucesión de vectores aleatorios mutuamente independientes y con la misma distribución que cierto vector X, del que suponemos que tiene vector de medias y matriz de covarianzas, que denotaremos $\mu = \mathbb{E}(X)$ y $\Sigma = Cov(X, X)$. Entonces se tiene que*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} N(0, \Sigma).$$

Teorema 1.15 (Método delta multivariante)**.** *Si $\{T_n\}$ es una sucesión de vectores aleatorios de dimensión k tales que*

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{d} N_k(0, \Sigma)$$

y $g : R^k \longrightarrow R^m$ es una función diferenciable en θ, siendo $D_g(\theta)$ la matriz jacobiana de g en θ, entonces

$$\sqrt{n}\,(g(T_n) - g(\theta)) \xrightarrow{d} N(0, D_g(\theta)\, \Sigma\, D_g(\theta)^t)$$

Definamos $\hat{F}(z) = \frac{\sum_{i=1}^{n} U_i(z)}{\sum_{i=1}^{n} V_i}$, donde

$$U_i(z) = \begin{cases} 1/Y_i & \text{si } Y_i \leq z \\ 0 & \text{si } Y_i > z \end{cases}$$

$$V_i = \frac{1}{Y_i}.$$

La distribución asintótica de $\hat{F}(z) = \frac{\sum_{i=1}^{n} U_i(z)}{\sum_{i=1}^{n} V_i}$, se encuentra al examinar la distribución conjunta de numerador y denominador. Fijado un i, vamos a calcular la esperanza, la varianza y la covarianza de U_i y V_i.

En primer lugar, denotemos como $\mu_r(z) = \int_0^z x^r f(x)\, dx$. Luego, en particular se tiene que $\mu_0(z) = F(z)$. Entonces:

$$\mathbb{E}_g\{U_i(z)\} = \frac{\mu_0(z)}{\mu}$$

$$\mathbb{E}_g(V_i) = \mathbb{E}\left(\frac{1}{Y_i}\right) = \frac{1}{\mathbb{E}(X_i)} = \frac{1}{\mu}$$

$$Var_g\{U_i(z)\} = \mathbb{E}(U_i^2) - (\mathbb{E}(U_i))^2 = \frac{\mu_{-1}(z)}{\mu} - \frac{\mu_0^2(z)}{\mu^2} = \frac{\mu\mu_{-1}(z) - (\mu_0(z))^2}{\mu^2}$$

$$Var_g(V_i) = \mathbb{E}(V_i^2) - (\mathbb{E}(V_i))^2 = \mathbb{E}\left(\frac{1}{Y_i^2}\right) - \left[\mathbb{E}\left(\frac{1}{Y_i}\right)\right]^2 = \frac{\mu_{-1}(z)}{\mu}\,\frac{\mu_0^2(z)}{\mu^2} = \frac{\mu\mu_{-1}(z) - 1}{\mu^2}$$

$$Cov_g\{U_i(z), V_i\} = \mathbb{E}(U_i \cdot V_i) - \mathbb{E}(U_i) \cdot \mathbb{E}(V_i) = \frac{\mu_{-1}(z)}{\mu}\frac{\mu_0(z)}{\mu} = \frac{\mu_{-1}}{\mu} - \frac{1}{\mu} \cdot \frac{\mu_0}{\mu} =$$

$$= \frac{\mu\mu_{-1}(z) - \mu_0(z)}{\mu^2}$$

Consideremos ahora el vector (A_n, B_n), siendo:

$$A_n = \frac{1}{n}\sum_{i=1}^{n} U_i(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{Y_i} P\{\frac{1}{Y_i} \leq z\}$$

$$B_n = \frac{1}{n}\sum_{i=1}^{n} V_i(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{Y_i}$$

Tenemos $(U_1(z), V_1), ...(U_n(z), V_n), ...$ una sucesión de vectores aleatorios, mutuamente independientes y con la misma distribución, con vector de medias $\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}$, y matriz de covarianzas $\Sigma = \begin{pmatrix} \frac{\mu\mu_{-1}(z) - (\mu_0(z))^2}{\mu^2} & \frac{\mu\mu_{-1}(z) - \mu_0(z)}{\mu^2} \\ \frac{\mu\mu_{-1}(z) - \mu_0(z)}{\mu^2} & \frac{\mu\mu_{-1}(z) - 1}{\mu^2} \end{pmatrix}$.

Aplicando el Teorema central del límite para el caso multivariante, se tiene:

$$\sqrt{n}\left((A_n, B_n) - \begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}\right) \xrightarrow{d} N_2(0, \Sigma)$$

Pero como queríamos la distribución asintótica de $\hat{F}(z) = \frac{\sum_{i=1}^{n} U_i(z)}{\sum_{i=1}^{n} V_i}$, tenemos que aplicar el Método delta para el caso multivariante. Consideremos la función diferenciable

$$g : (0, +\infty) \times (0, +\infty) \longrightarrow \mathbb{R}.$$

$$(a, b) \longrightarrow a/b$$

Considerando dicho teorema, se tiene que:

$$\sqrt{n}\left(g(A_n, B_n) - g\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}\right) \xrightarrow{d} N_2\left(0, D_g\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix} \Sigma D_g\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}^t\right), \quad (1.8)$$

donde $D_g(\theta)$ es la matriz jacobiana de g en $\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}$.

Hacemos los cálculos necesarios:

$$g(A_n, B_n) = \frac{A_n}{B_n}$$

$$g\begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix} = \mu_0(z)$$

$$D_g(a, b) = \begin{pmatrix} \frac{1}{b} & \frac{-a}{b^2} \end{pmatrix}$$

$$D_g \begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix} \Sigma D_g \begin{pmatrix} \mu_0(z)/\mu \\ 1/\mu \end{pmatrix}^t =$$

$$= \begin{pmatrix} \mu & -\mu_0(z)\mu \end{pmatrix} \begin{pmatrix} \dfrac{\mu\mu_{-1}(z) - (\mu_0(z))^2}{\mu^2} & \dfrac{\mu\mu_{-1}(z) - \mu_0(z)}{\mu^2} \\ \dfrac{\mu\mu_{-1}(z) - \mu_0(z)}{\mu^2} & \dfrac{\mu\mu_{-1}(z) - 1}{\mu^2} \end{pmatrix} \begin{pmatrix} \mu \\ -\mu_0(z)\mu \end{pmatrix} =$$

$$= \begin{pmatrix} \mu_{-1}(z)(1-\mu_0(z)) & \mu_{-1}(z)(1-\mu_0(z)) \end{pmatrix} \begin{pmatrix} \mu \\ -\mu_0(z)\mu \end{pmatrix} =$$

$$= \mu\mu_{-1}(z) - 2\mu\mu_{-1}(z)\mu_0(z) + (\mu_0(z))^2 \mu\mu_{-1}(z).$$

Y sustituimos en (1.8). Así llegamos a:

$$\sqrt{n}\left(\frac{A_n}{B_n} - \mu_0(z)\right) \xrightarrow{d} N_2\left(0, \mu\mu_{-1}(z) - 2\mu\mu_{-1}(z)\mu_0(z) + (\mu_0(z))^2 \mu\mu_{-1}(z)\right)$$

Luego siempre que todos los términos sean finitos, el numerador y el denominador de $\hat{F}(z)$ tienen asintóticamente una distribución normal bivariada. Por lo tanto $\hat{F}(z)$ es asintóticamente normalmente distribuido con media

$$\mu_0(z) = F(z) \tag{1.9}$$

y varianza

$$\frac{\mu\mu_{-1}(z) - 2\mu\mu_{-1}(z)\mu_0(z) + \mu\mu_{-1}(\mu_0(z))^2}{n}. \tag{1.10}$$

Capítulo 2

Estudio de simulación sobre la media armónica

El presente estudio de simulación tiene por objetivo realizar un análisis de las propiedades de la media armónica. En él consideramos distribuciones continuas y discretas, para cada una de las cuales se simulan mil muestras sesgadas. Se calcula la media armónica de cada muestra sesgada, y luego se hace una comparación de diferentes propiedades de ésta, como son el sesgo, la varianza, el ECM y la varianza asintótica.

2.1. Modelos considerados en la simulación y sus versiones sesgadas

En nuestro estudio vamos a emplear los modelos recogidos en el Cuadro 2.1 y sus formas modificadas bajo sesgo por longitud. Es sencillo ver cómo se obtienen algunas de estas formas sesgadas,si empleamos la densidad ponderada de la que hemos hablado en el capítulo anterior: $g(y) = \dfrac{yf(y)}{\mu}$, donde f es la densidad no sesgada y μ su media.

Para simular tenemos dos opciones:

- En una muestra establecer un sesgo de muestreo. Es decir, un sesgo en la que se recoge una muestra de tal manera que algunos miembros de la población destinada tienen menos probabilidades de ser incluidos que otros. El resultado es una muestra sesgada, una muestra no aleatoria de una población en la que todas las personas o instancias, no tienen las mismas probabilidades de ser seleccionados.

- Emplear distribuciones sesgadas por longitud. Las más básicas se encuentran en el

Cuadro 2.1. Este es el método que emplearemos en nuestro estudio.

Variable aleatoria (V.a.)	Función de densidad ó probabilidad	V.a. sesgada por longitud
Uniforme(0,1)	1	$2x$
Gamma, $G(p,a)$	$\frac{1}{\Gamma(p)}a^p x^{p-1}e^{-ax}$	$G(p+1,a)$
Beta, $B(p,q)$	$\frac{1}{\beta(p,q)}x^{p-1}(1-x)^{q-1}$	$B(p+1,q)$
Binomial, $B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	$1+B(n-1,p)$
Binomial negativa, $BN(k,p)$	$\binom{x-1}{k-1}q^k(1-p)^{x-k}$	$1+BN(k+1,p)$
Poisson, $Po(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$1+Po(\lambda)$
Hipergeométrica, $H(n,M,N)$	$\binom{n}{x}M^{(x)}(N-M)^{(n-x)}N^{(n)}$	$1+H(n-1,M-1,N-1)$
Equiprobable	$\frac{1}{k}$, $\forall j \in \{1,...,k\}$	$\frac{j}{\frac{k(k+1)}{2}}$, $\forall j \in \{1,...,k\}$

Cuadro 2.1: Algunas distribuciones básicas y sus formas sesgadas por longitud.

En primer lugar consideremos las distribuciones continuas Gamma y Beta y las distribuciones discretas Equiprobable y Poisson. Veamos cómo se obtienen sus formas sesgadas.

Gamma

Este modelo depende de dos parámetros positivos: a y p. Su función de densidad es de la forma

$$f(x) = \frac{1}{\Gamma(p)}\, a^p x^{p-1} e^{-ax}, x > 0,$$

y su esperanza es p/a.

La $\Gamma(p)$ es la denominada función Gamma de Euler que representa la siguiente integral: $\Gamma(p) = \int_0^\infty x^{p-1}\, e^{-x}\, dx$. Además es fácil comprobar que verifica que $\Gamma(p) = (p-1)\,\Gamma(p-1)$.

En efecto, si hacemos integración por partes considerando

$$u = x^{p-1} \qquad du = (p-1)x^{p-2}$$

$$dv = -e^{-x}dx \qquad v = -e^{-x}$$

llegamos a

$$\Gamma(p) = \int_0^\infty x^{p-1} e^x dy = (p-1) \int_0^\infty e^{-x} x^{p-2} = (p-1)\Gamma(p-1).$$

Repitiendo el proceso sucesivamente, se llega a que $\Gamma(p) = (p-1)(p-2)...\Gamma(1)$, para p entero positivo. Por otra parte, por integración directa vemos que $\Gamma(1) = 1$. Entonces, deducimos que $\Gamma(p) = (p-1)!$ para p entero positivo.

Comprobemos que su forma sesgada por longitud coincide con la escrita en el Cuadro 2.1. Sea $f(x)$ la función de densidad de una $Gamma(p, a)$. Empleando la distribución no ponderada calculada previamente se tiene:

$$g(x) = \frac{x f(x)}{\mu} = \frac{x\, a^p\, x^{p-1}\, e^{-ax}}{\Gamma(p)\mu} \overset{(a)}{=} \frac{a^{p+1}\, x^p\, e^{-ax}}{\Gamma(p)\, p} \overset{(b)}{=} \frac{a^{p+1}\, x^p\, e^{-ax}}{\Gamma(p+1)}.$$

En la igualdad (a) se sustituye $\mu = p/a$ y en la igualdad (b) se aplica que $\Gamma(p+1) = p\,\Gamma(p)$.

Luego, como queríamos probar, $g(x)$ es la función de densidad de una $Gamma(p+1, a)$.

Beta

Su función de densidad para valores $0 < x < 1$ está representada por la expresión

$$f(x) = \frac{1}{\beta(p,q)}\, x^{p-1}\, (1-x)^{q-1},$$

donde $\beta(p, q)$ es la función beta, definida por

$$\beta(p,q) = \int_0^1 x^{p-1}\, (1-x)^{q-1}\, dy = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}.$$

Veamos cuál es la forma sesgada por longitud de esta distribución:

$$g(x) = \frac{x f(x)}{\mu} = \frac{x^p\,(1-x)^{q-1}}{\beta(p,q)\,\mu} \overset{(b)}{=} \frac{x^p\,(1-x)^{q-1}}{\beta(p,q)\,\frac{p}{p+q}} \overset{(b)}{=} \frac{x^p\,(1-x)^{q-1}}{\beta(p+1,q)}.$$

Donde en la igualdad (a) únicamente sustituimos $\mu = \dfrac{p}{p+q}$, y en la igualdad (b) aplicamos la relación entre las funciones Γ y β, así como la propiedad $\Gamma(p+1) = p\,\Gamma(p)$:

$$\beta(p+1,q) = \frac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q+1)} = \frac{p\Gamma(p)\Gamma(q)}{(p+q)\Gamma(p+q)} = \frac{\Gamma(p)\Gamma(q)}{\frac{p+q}{p}\Gamma(p+q)} = \beta(p,q)\frac{p}{p+1}.$$

Por lo tanto $g(x)$ es la función de densidad de una $Beta(p+1,q)$, como queríamos demostrar.

Equiprobable

La equiprobable es una distribución que asigna probabilidades iguales a un conjunto finito de puntos del espacio. Es decir, si la variable X puede tomar los valores $1, 2, ..., k$ con igual probabilidad, entonces,

$$P(X = j) = \frac{1}{k}, \;\; \forall j \in \{1, ..., k\}.$$

Por lo tanto podemos ver cuál es su forma sesgada por longitud,

$$P(Y = j) = \frac{j\,P(X = j)}{\mu_x} \overset{(a)}{=} \frac{j\,\frac{1}{k}}{\frac{k+1}{2}} = \frac{j}{\frac{k(k+1)}{2}}.$$

Donde en la igualdad (a) estamos usando que

$$\mathbb{E}(X) = \sum_{j=1}^{k} j\,\frac{1}{k} = \frac{1}{k}\sum_{j=1}^{k} j = \frac{1}{k}\,\frac{k(k+1)}{2} = \frac{k+1}{2}.$$

Debemos destacar un problema que surge en las distribuciones discretas que presentan probabilidad no nula en cero. Esto ocurre en distribuciones tan conocidas como la Binomial o la Poisson.

El hecho de tomar el valor cero con probabilidad no nula hace que en la distribución sesgada ese valor tenga probabilidad cero, que a efectos prácticos implica que se va a observar con probabilidad cero.

Consideremos una variable X que toma tres valores posibles, 0, 1 y 2, todos con la misma probabilidad $p = 1/3$. Por ejemplo, supongamos que tenemos 90 bolas, 30 marcadas con un 0, otras 30 marcadas con un 1, y las 30 restantes marcadas con un 2.

Valores de X	0	1	2
Probabilidad	$1/3$	$1/3$	$1/3$

La función de densidad de esta variable será:

$$f(k) = P(X = k) = 1/3, \quad k \in \{0, 1, 2\}.$$

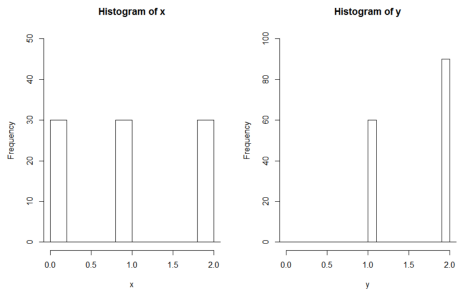
Luego si la muestra estuviera sesgada por longitud, las probabilidades serían:

$$P(Y = 0) = 0, \quad P(Y = 1) = \frac{1}{3}, \quad P(Y = 2) = \frac{2}{3}.$$

Es decir, pasaríamos a tener:

Valores de Y	0	1	2
Probabilidad	0	$1/3$	$2/3$

Si hacemos la representación gráfica, al pasar a la forma sesgada observamos la misma cantidad de bolas marcadas con un 1, y el doble de bolas marcadas con un 2; sin embargo ya no observaríamos ninguna bola marcada con un 0.



Obviamente se tiene que la $E(X) = 1$, pero sin embargo vamos a ver que $E(X) \neq \frac{1}{E(1/Y)}$. Para ello consideremos la siguiente tabla:

Valores de $1/Y$	1	1/2
Probabilidad	1/3	2/3

y calculemos la media armónica de la muestra de la variable Y:

$$E(1/Y) = 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{2}{3}$$

Por lo tanto, $E(X) \neq \frac{1}{E(1/Y)}$, es decir, la pérdida del cero es imposible reconstruirla con ponderaciones. Sin embargo, vamos a considerar la variable X^t resultante de suprimirle a X el valor $x_0 = 0$. Si calculamos su esperanza vemos que obtenemos que $E(X^t) = \frac{1}{E(1/Y)}$, pues

$$E(X^t) = \frac{\left(1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}\right)}{\frac{2}{3}} = \frac{3}{2}.$$

Es decir, en los casos en los que el valor cero tome probabilidad no nula, se tiene que $E(X) \neq \frac{1}{E(1/Y)} = E(X^t)$.

Ahora vamos a calcular $E(X^t)$ para una variable X en general. Supongamos que tenemos una variable X que toma $k+1$ valores posibles, $x_0 = 0, x_1, ..., x_k$, con probabilidades $p_0, p_1, ..., p_k$. Como vimos en el capítulo anterior, el sesgo por longitud de una variable X supone que realmente vamos a observar una variable Y, cuyas probabilidades se ven afectadas proporcionalmente por el valor de la variable original. Es decir, observamos una variable aleatoria Y que toma los mismos valores que X, pero con probabilidades $q_0, q_1, ..., q_k$, donde $q_i = Cx_ip_i$.

Sea X^t la variable aleatoria resultante de suprimirle a X el valor $x_0 = 0$. Es sencillo calcular la esperanza de X^t:

$$\mathbb{E}(X^t) = \frac{\sum_{i=1}^{k} x_i p_i}{\sum_{i=1}^{k} p_i} = \frac{\sum_{i=1}^{k} x_i p_i}{1 - p_0}.$$

Es decir,

$$E(X^t) = \frac{E(X)}{1-p_0}.$$

Además,

$$1 = \sum_{i=1}^{k} q_i = C\sum_{i=1}^{k} x_i p_i = C(1-p_0)\mathbb{E}(X^t),$$

por lo tanto se tiene que,

$$C = \frac{1}{\mathbb{E}(X^t)(1-p_0)}.$$

Y sustituyendo,

$$q_i = Cx_i p_i = \frac{x_i p_i}{(1-p_0)\mathbb{E}(X^t)}.$$

Llegados a este punto ya estamos en las condiciones de poder obtener la forma sesgada por longitud de una Poisson(λ).

Poisson

Sea una variable aleatoria X que sigue la distribución de Poisson. Entonces,

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \{0,1,2,...\}.$$

Vamos a obtener la forma sesgada de la distribución,

$$P(Y=k) = \frac{kP(X=k)}{\mu_x} = \frac{kP(X=k)}{(1-P(X=0))\mathbb{E}(X^t)} = \frac{ke^{-\lambda}\frac{\lambda^k}{k!}}{(1-P(X_0))\mathbb{E}(X^t)} \overset{(a)}{=} \frac{k}{\lambda}e^{-\lambda}\frac{\lambda^k}{k!} =$$

$$= \frac{1}{x} \quad e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}, \quad k \in \{1, 2, ...\}.$$

Donde en la igualdad (a) estamos utilizando que

$$(1 - P(X = 0))\mathbb{E}(X^t) = \sum_{k=1}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} kf(X = k) - 0 \cdot P(X = 0) = \mathbb{E}(X) = \lambda.$$

Por lo tanto concluimos que la forma sesgada de una Poisson(λ) es 1+Poisson(λ), tal y como habíamos visto en el Cuadro 2.1 .

2.2. Procedimiento general de simulación

El Cuadro 2.1 tiene importancia en nuestro estudio para generar distribuciones ponderadas, pues algunas de las distribuciones que aparecen en él son las que vamos a emplear en la simulación.

Para cada una de esas distribuciones hemos simulado mil muestras sesgadas. Y como de cada muestra se obtiene una media armónica, hemos creado un vector con las mil medias armónicas.

$$\begin{array}{ccc} Y_{1,1} & ... & Y_{1,n} \longrightarrow \hat{\mu}_1 \\ & ... & \\ & ... & \\ Y_{1000,1} & ... & Y_{1000,n} \longrightarrow \hat{\mu}_{1000} \end{array}$$

Una vez obtenido, nuestro interés se centrará en la aproximación de características de ella, como el sesgo, la varianza, el ECM ó la varianza asintótica. Para cada distribución, haremos unas tablas que nos permitan hacer una comparación de las aproximaciones de las características de la media armónica, y de la varianza de la media aritmética simple de la variable no sesgada X, variando el tamaño muestral, al que llamaremos n.

Aproximamos la esperanza de la media armónica mediante la media aritmética de todas las armónicas obtenidas de cada una de las muestras. Es decir, la esperanza de la media

armónica será,

$$\hat{\mu} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\mu}_i.$$

Sabemos que el sesgo de un estimador $\hat{\theta}$ para un parámetro poblacional θ se define como $Sesgo(\hat{\theta}) = \mathbb{E}(\theta) - \theta$. Por lo tanto, para dar una aproximación del sesgo, simplemente le tenemos que restar a la esperanza de la media armónica calculada anteriormente el verdadero valor de la media de la distribución. Es decir,

$$\hat{Sesgo}(\hat{\mu}) = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\mu}_i - \mu.$$

Por otra parte aproximaremos la varianza de la media armónica por

$$\hat{Var}(\hat{\mu}) = \frac{1}{1000-1}\sum_{i=1}^{1000}(\hat{\mu}_i - \bar{\hat{\mu}})^2,$$

siendo

$$\bar{\hat{\mu}} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\mu}_i.$$

Si calculamos el error típico aproximado por simulación, y hacemos la raíz cuadrada de la media armónica, obtenemos una aproximación de la varianza de la media armónica. Ésta será parecida a la varianza asintótica, que habíamos visto en el capítulo anterior que era

$$\frac{(\mu\mu_{-1} - 1)\mu^2}{n}.$$

También hemos visto que el error cuadrático medio de un estimador θ, se define como $ECM(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2$, y cumple que $ECM(\hat{\theta}) = Var(\hat{\theta}) + (Sesgo(\hat{\theta}))^2$. Luego en este caso, una aproximación del ECM sería

$$\hat{ECM}(\hat{\mu}) = \hat{Var}(\mu) + (\hat{Sesgo}(\hat{\mu}))^2.$$

Por último, como hemos dicho, en las tablas también va a aparecer la varianza de la media aritmética simple de la variable no sesgada X, que no es ninguna característica de la media armónica. El motivo de añadir esta columna es hacer una comparación de las muestras anteriores con otras no sesgadas.

Nótese que la varianza de la media aritmética simple no debería diferir en gran cantidad de la varianza de la media armónica. Luego tampoco lo debería hacer de la varianza asintótica, ni del ECM.

2.3. Distribuciones continuas

Vamos a comenzar simulando muestras sesgadas por longitud de distribuciones continuas. Concretamente una Uniforme, una Gamma y una Beta. Para cada una de ellas consideraremos varios casos, y haremos comparaciones mediante representaciones gráficas y tablas.

En cada tabla vamos a aproximar las propiedades de la media armónica explicadas en la sección anterior, así como su varianza asintótica y la varianza de la media aritmética simple. Una vez que hallemos estas aproximaciones, haremos una comparación de cómo varían los resultados obtenidos dependiendo del tamaño muestral. También compararemos las diferentes tablas de las distribuciones.

Por otra parte, como se ha explicado en el capítulo anterior, en una muestra sesgada por longitud los datos con mayor longitud tienen mayor probabilidad de ser seleccionados. Y claramente el sesgo por longitud afecta más cuando hay valores de x próximos al cero con probabilidades altas. Este es el motivo por el que algunas funciones de densidad, se deforman más que otras al pasar a su forma sesgada, tal y como vamos a ver en esta sección.

Uniforme(a,b)

La distribución Uniforme es el modelo (absolutamente) continuo más simple. Corresponde al caso de una variable aleatoria que sólo puede tomar valores comprendidos entre dos extremos a y b, de manera que todos los intervalos de una misma longitud (dentro de

(a, b)) tienen la misma probabilidad.

En nuestro caso trabajaremos con una $Uniforme(0, 1)$ y con una $Uniforme(9, 10)$. La primera es un caso particular de la distribución beta, que se corresponde con una beta de parámetros $a = 1$ y $b = 1$. Luego su forma sesgada por longitud la podemos deducir del Cuadro 2.1. Sin embargo, en dicha tabla no tenemos una forma sesgada para la $Uniforme(a, b)$, con $a \neq 0$ y $b \neq 1$. Por lo tanto para hacer simulaciones con esta distribución, tendremos que hallar la función de distribución sesgada $G(y)$, y calcular $Y = G^{-1}(U)$, siendo U una variable con distribución $Uniforme(a, b)$.

Como conocemos la función de densidad no ponderada de una Uniforme, a partir de ella podemos calcular la ponderada o sesgada,

$$g(y) = \frac{yf(y)}{\mu_y} = \frac{y}{\frac{a+b}{2}(b-a)} = \frac{2y}{b^2 - a^2}.$$

Por lo tanto, la función de distribución es

$$G(y) = \int_a^y \frac{2u}{b^2 - a^2}\, du = \frac{1}{b^2 - a^2}\left[u^2\right]_a^y = \frac{y^2 - a^2}{b^2 - a^2}, y \in [a, b].$$

Sea ahora $U \in Uniforme(a, b)$, pongamos $Y = G^{-1}(U)$, entonces:

$$\frac{y^2 - a^2}{b^2 - a^2} = u \Longrightarrow y^2 = a^2 + u(b^2 - a^2) \Longrightarrow Y = \sqrt{a^2 + U(b^2 - a^2)}$$

Para calcular la varianza asintótica, vamos a obtener μ_{-1} para las dos Uniformes que hemos considerado.

En el caso de $Uniforme(0, 1)$:

$$\mu_{-1} = \mathbb{E}\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x} f(x)\, dx = \int_0^1 \frac{1}{x}\frac{1}{1-0}\, dx = [ln(x)]_0^1 = +\infty,$$

y en el caso de la $Uniforme(9,10)$:

$$\mu_{-1} = \mathbb{E}\left(\frac{1}{X}\right) = \int_9^{10} \frac{1}{x} f(x)\, dx = \int_9^{10} \frac{1}{x} \frac{1}{10-9}\, dx = [ln(x)]_9^{10} = ln(10) - ln(9).$$

En el Cuadro 2.2 se presentan para una distribución Uniforme(0,1) los valores de sesgo, varianza, ECM y varianza asintótica de la media armónica como estimador de la media de X. Se añade una columna con σ^2/n, que es la varianza de $\bar{X}$, que sería el estimador natural de la media con datos no sesgados. Así podremos considerar a la media armónica como el estimador óptimo de la media en condiciones ideales. Se muestran los resultados para tamaños muestrales crecientes $n = 20, 50, 100, 500$. En el Cuadro 2.3 se presentan los resultados para la Uniforme(9,10).

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	2,24	1,11	1,16	Inf	0,03
50	1,19	0,53	0,53	Inf	0,013
100	0,85	0,36	0,37	Inf	0,000694
500	0,40	0,09	0,09	Inf	0,0000139

Cuadro 2.2: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Uniforme(0,1). (Los resultados vienen en centésimas). Se añade el valor σ^2/n, también en centésimas.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	1,742	3,932	3,935	4,173	4,16
50	0,120	1,72	1,72	1,66	1,66
100	−0,85	0,83	0,83	0,83	0,83
500	−0,12	0,17	0,17	0,16	0,16

Cuadro 2.3: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Uniforme(9,10). (Los resultados vienen en milésimas). Se añade el valor σ^2/n, también en milésimas.

En el Cuadro 2.2 vemos que tanto el sesgo como la Varianza y el ECM convergen a cero cuando el tamaño muestral tiende a infinito. Esto muestra que la media armónica es un estimador consistente en media cuadrática. También presentamos la varianza asintótica que hemos obtenido en el capítulo anterior, y vemos que su valor es $+\infty$. Esto se debe a que $\mu_{-1} = +\infty$. Aunque en los demás casos que consideraremos la varianza asintótica se parecerá mucho a la aproximación de la varianza.

Si hacemos una comparación de los Cuadros 2.2 y 2.3, vemos que la media armónica para la $Uniforme(0,1)$ tiene mayor sesgo que para la $Uniforme(9,10)$. El hecho de que la $Uniforme(9,10)$ esté poco sesgada, se ve reflejado en que la $Var(\bar{X})$ y la varianza asintótica son muy parecidas. Por otra parte, que la varianza asintótica de la $Uniforme(0,1)$ sea infinito es consecuencia de que $\mu_{-1} = \infty$, como vimos anteriormente.

Un sesgo pequeño implica que la varianza y el ECM de la media armónica son muy parecidos. Vemos que en la Uniforme(9,10) el sesgo aproximado es más pequeño que en la Uniforme(0,1), y en consecuencia la varianza y el ECM son mucho más parecidos que en el primer caso.

En la Figura 2.1 se muestra una representación de la función de densidad de las dos Uniformes de las que hemos hablado y sus formas sesgadas por longitud. Además también añadimos la $Uniforme(100,101)$. Las funciones de densidad no sesgadas están representadas en rojo, y las sesgadas en verde. Por lo tanto, en la primera columna de la figura podemos ver una $Uniforme(0,1)$, una $Uniforme(9,10)$ y una $Uniforme(100,101)$, y en la segunda sus respectivas formas sesgadas.

La función de densidad de la $Uniforme(0,1)$ se deforma mucho más al pasar a su forma sesgada que la de la $Uniforme(9,10)$. Y a su vez, las dos anteriores se deforman más que la función de densidad de la $Uniforme(100,101)$, en la que a penas vemos cambio al pasar de una forma a otra.

Como las funciones de densidad de las Uniformes son rectas, podemos ver de forma numérica lo que se deforman al pasar de la forma no sesgada a la sesgada. Simplemente tenemos que hallar y comparar las pendientes de dichas rectas.

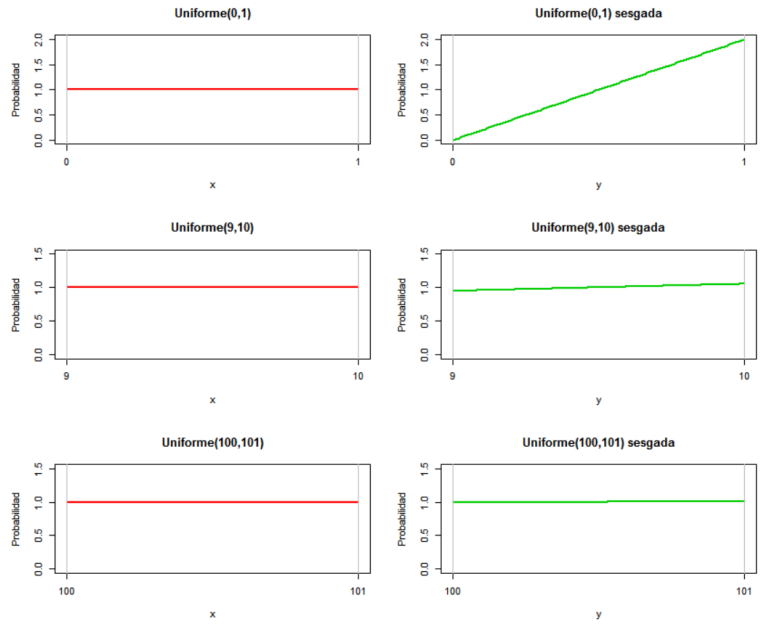


Figura 2.1: Uniforme no sesgada y sesgada

Es conocido que cualquier función de densidad de una Uniforme no sesgada es una recta horizontal con pendiente cero. Por otra parte, es sencillo ver que la pendiente de la recta asociada a la forma sesgada de la $Uniforme(0,1)$ es 2, mientras que las de la $Uniforme(9,10)$ y $Uniforme(100,101)$ son 0,1053 y 0,0099 respectivamente. Por lo tanto, la función de densidad de la $Uniforme(0,1)$ es la que más se deforma al pasar a su forma sesgada, seguida de la $Uniforme(9,10)$, y por último la $Uniforme(100,101)$. Esto es debido a que en la función de densidad de la $Uniforme(0,1)$ hay valores de x próximos al cero que toman probabilidades altas, mientras que en las $Uniforme(9,10)$ y $Uniforme(100,101)$ las x que toman probabilidades altas están muy alejadas del cero. Estamos hablando de sesgo por longitud, luego a medida que consideramos valores de x próximos al cero que toman probabilidades altas, afecta en mayor medida.

Es más, vimos que la función de densidad sesgada de una $Uniforme(a,b)$, es $g(y) = \dfrac{2y}{b^2-a^2}$, luego si a y b toman valores muy grandes, la función de densidad sesgada es muy parecida a la no sesgada, pues:

$$g(a) = \frac{2a}{b^2-a^2} = \frac{2a}{(a+b)(b-a)} \overset{(a)}{\simeq} \frac{1}{b-a}$$

$$g(b) = \frac{2b}{b^2 - a^2} = \frac{2b}{(a+b)(b-a)} \overset{(a)}{\simeq} \frac{1}{b-a}$$

Donde en (a) estamos usando que como a y b toman valores muy grandes, $2a$ y $2b$ se parecen mucho a $(a+b)$.

Así, a medida que x toma valores más grandes, la forma sesgada por longitud difiere menos de la no ponderada o no segada.

Gamma(p,a)

En una distribución Gamma, según los valores que tome el parámetro de forma, p, la función de densidad presenta perfiles muy diversos. Con valores de p menores o iguales que 1, la función de densidad muestra un perfil decreciente; en cambio, si p es mayor que la unidad, la función de densidad crece hasta el valor $x = \frac{p-1}{a}$ y decrece a partir de este valor. Vamos a considerar dos casos, uno con p=2 y a=1 y otro con p=8 y a=1. Como se manifiesta en el Cuadro 2.1, la forma sesgada por longitud de una $Gamma(2,1)$ es una $Gamma(3,1)$, y la de una $Gamma(8,1)$, es una $Gamma(9,1)$.

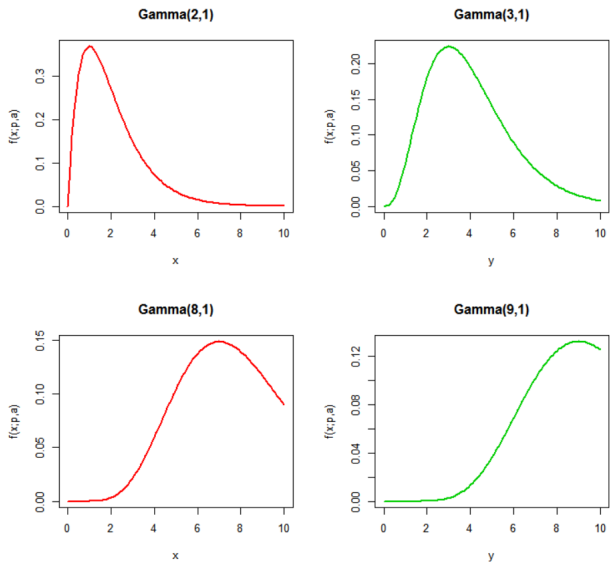


Figura 2.2: Distribuciones Gamma no sesgadas y sesgadas

En la Figura 2.2, podemos ver una $Gamma(2,1)$ y una $Gamma(8,1)$ con sus respectivas formas sesgadas por longitud.

Tal y como ocurría con las Uniformes, la $Gamma(2,1)$ se deforma más al pasar a su forma sesgada que la $Gamma(8,1)$. Esto es debido a que en el primer caso hay valores de x próximos al cero que toman probabilidades altas.

A continuación vamos a ver quién es $\mu_{-1} = \mathbb{E}(1/X)$, necesaria para el cálculo de la varianza asintótica.

$$\mu_{-1} = \mathbb{E}\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x} f(x)\, dx \overset{(a)}{=} \int_0^\infty \frac{1}{x}\frac{1}{\Gamma(p)}\, a^p\, x^{p-1}\, e^{-ax} \overset{(b)}{}\, dx =$$

$$= \int_0^\infty \frac{(ax)^{p-2}}{\Gamma(p)}\, a^2\, e^{-ax}\, dx \overset{(c)}{=} \frac{a^2}{p-1}\int_0^\infty \frac{(p-1)\,(ax)^{p-2}\, e^{-ax}}{\Gamma(p)} dx \overset{(d)}{=} \frac{a^2}{p-1}.$$

En la igualdad (a) sustituimos la función de densidad de una $Gamma(p,a)$, que es $f(x) = \dfrac{1}{\Gamma(p)} a^p x^{p-1} e^{-ax}$. En (b) simplemente hacemos operaciones. En (c) multiplicamos y dividimos por $(p-1)$ para obtener dentro una $\Gamma(p)$. Y finalmente en la igualdad (d) simplificamos las dos $\Gamma(p)$ empleando lo siguiente:

$$\Gamma(p) = \int_0^\infty x^{p-1} e^x dx,$$

luego si llamamos

$$u = x^{p-1} \qquad du = (p-1)x^{p-2}dx$$
$$dv = e^{-x}dx \qquad v = -e^{-x}$$

Integrando por partes, se obtiene lo siguiente:

$$\Gamma(p) = \left[-e^{-x}x^{p-1}\right]_0^\infty + \int_0^\infty e^{-x}(p-1)x^{p-2}dx = (p-1)\int_0^\infty x^{p-2}e^{-x}dx.$$

En los Cuadros 2.4 y 2.5 se presentan las aproximaciones de las características de la media armónica y la varianza de la media aritmética simple en el caso de considerar una $Gamma(2,1)$ y una $Gamma(8,1)$.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	0.0996	0.1451	0.1550	0.2	0.1
50	0.0332	0.0715	0.0726	0.08	0.04
100	0.0154	0.0375	0.0377	0.04	0.02
500	0.0006	0.0076	0.0076	0.008	0.004

Cuadro 2.4: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Gamma(2,1). Se añade el valor σ^2/n.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	0.0606	0.482	0.4864	0.4571	0.4
50	0.0157	0.1845	0.1847	0.1828	0.16
100	-0.0014	0.0932	0.0932	0.0914	0.08
500	-0.0085	0.0188	0.0188	0.0183	0.016

Cuadro 2.5: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Gamma(8,1). Se añade el valor σ^2/n.

La aproximación de la varianza de la media armónica a medida que aumenta el tamaño muestral se parece más al ECM en ambos casos. Esto es debido a que el sesgo tiende a cero. Así mismo la varianza asintótica también toma un valor similar a ambos.

Si nos fijamos en la columna del sesgo de las dos tablas, vemos que la $Gamma(2,1)$ está más sesgada que la $Gamma(8,1)$. En consecuencia, la varianza asintótica de la $Gamma(8,1)$ se parece más a $Var(\bar{X}) = \sigma^2/n$, que en el primer caso. Es decir, en una $Gamma(p,a)$ un incremento del valor de p produce una disminución del sesgo aproximado de la media armónica. Y en consecuencia la varianza asintótica de la $Gamma(p,a)$ se parece más a $Var(\bar{X}) = \sigma^2/n$ que en el caso de otra $Gamma(p,a)$ que tuviera un valor de p inferior.

Beta(p,q)

La distribución beta es adecuada para variables aleatorias continuas que toman valores en el intervalo $(0,1)$, lo que la hace muy apropiada para modelar proporciones. En nuestro caso vamos a considerar una $Beta(2,1)$ y una $Beta(9,1)$.

Comencemos viendo cómo obtener μ_{-1}, necesaria para el cálculo de la varianza asintótica.

$$\mathbb{E}\left(\frac{1}{X}\right) = \int_{-\infty}^{+\infty} \frac{1}{x}\, f(x) \overset{(a)}{=} \frac{1}{\beta(p,q)} \int_{-\infty}^{+\infty} \frac{1}{x}\, x^{p-1}\,(1-x)^{q-1}\, dx \overset{(b)}{=}$$

$$= \frac{1}{\beta(p,q)} \int_0^1 x^{(p-1)-1}\,(1-x)^{q-1} = \frac{\beta(p-1,q)}{\beta(p,q)} \overset{(c)}{=} \frac{\dfrac{\Gamma(p-1)\,\Gamma(p)}{\Gamma(p-1+q)}}{\dfrac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}} \overset{(d)}{=} \frac{p+q-1}{p-1}.$$

En la igualdad (a) sustituimos la función de densidad de una $Beta(p,q)$, que es,

$$f(x) = \frac{1}{\beta(p,q)}\, x^{p-1}\,(1-x)^{q-1}.$$

En (b) agrupamos términos para conseguir una función $\beta(p-1,q)$, sabiendo que

$$\beta(p,q) = \int_0^1 x^{p-1}\,(1-x)^{q-1} dy.$$

En (c) empleamos la igualdad demostrada anteriormente:

$$\beta(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}.$$

Y finalmente en (d) simplificamos empleando que $\Gamma(p-1) = \dfrac{\Gamma(p)}{p-1}$.

Consideremos los Cuadros 2.6 y 2.7. Nuevamente vemos que en ambos casos el sesgo aproximado se va a cero a medida que aumenta el tamaño muestral, y que la varianza de la media armónica y el ECM se parecen cada vez más.

Como en el caso de la *Gamma*, volvemos a tener que la $Beta(2,1)$ está más sesgada que la $Beta(9,1)$. Luego la varianza asintótica de la segunda se parece más a $Var(\bar{X})$ que en el primer caso.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	0.78	0.52	0.53	0.74	0.015
50	0.13	0.25	0.25	0.29	0.00617
100	0.19	0.13	0.13	0.14	0.00308
500	0.07	0.027	0.027	0.029	0.000617

Cuadro 2.6: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Beta(2,1). (Los resultados vienen en centésimas). Se añade el valor σ^2/n, también en centésimas.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	-0.74	0.48	0.48	0.5	0.00334
50	-0.7	2	2	0.2	0.00133
100	-0.4	0.1	0.1	0.1	0.000669
500	-0.0396	0.0209	0.209	0.02025	0.000133

Cuadro 2.7: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Beta(9,1). (Los resultados vienen en milésimas). Se añade el valor σ^2/n, también en milésimas.

En la Figura 2.3 se hace una representación de las funciones de densidad correspondientes a ambas distribuciones. En la primera fila vemos una $Beta(2,1)$ y su forma sesgada por longitud, y en la segunda una $Beta(9,1)$ y su forma sesgada por longitud.

La $Beta(2,1)$ se deforma más al pasar a su forma sesgada por longitud que la $Beta(9,1)$. Esto se debe a que en la $Beta(9,1)$ los valores de x tienen probabilidad 0 hasta que x es mayor que 0,6; mientras que en el otro caso x siempre toma probabilidades positivas, incluso en valores próximos a cero.

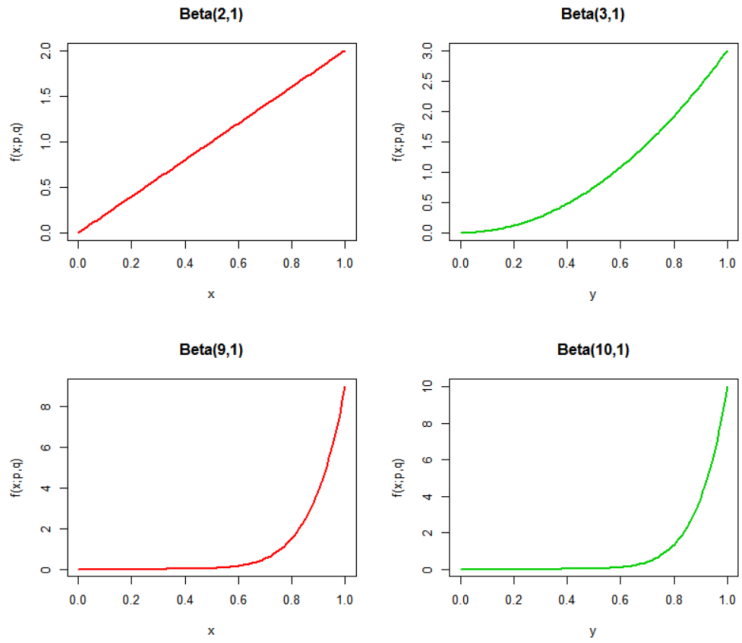


Figura 2.3: Distribuciones Beta no sesgadas y sesgadas

2.4. Distribuciones discretas

En esta sección vamos a simular muestras sesgadas por longitud de una Poisson. Como hicimos en el caso de las distribuciones continuas, aproximaremos propiedades de la media armónica, y haremos una comparación de cómo varían los resultados obtenidos en función del tamaño muestral.

Veremos que la media armónica también es consistente en distribuciones discretas, excepto si el cero tiene probabilidad positiva (no nula). Esto es debido a que en estos casos el cero dejaría de ser observable y resultaría irrecuperable. En el caso de que se observe con mayor o menor probabilidad, la media armónica seguiría siendo el mejor estimador para el verdadero valor de la media.

Por lo tanto si el cero tiene probabilidad positiva, el sesgo aproximado no tenderá a cero a medida que aumenta el tamaño muestral. Como consecuencia, la varianza y el ECM difieren más en este caso que en el caso de trabajar con distribuciones continuas, en las que el valor aproximado del sesgo siempre se iba a cero a medida que aumentaba n.

Poisson(λ)

Esta distribución es una de las más importantes de variable discreta. El parámetro de la distribución, λ, representa el número promedio de eventos esperados por unidad de tiempo o de espacio, por lo que se suele hablar de λ como "la tasa de ocurrencia" del fenómeno que se observa. En este caso vamos a considerar una $Poisson(\lambda = 2)$, luego su forma sesgada será $1 + Poisson(2)$, tal y como vimos en el Cuadro 2.1.

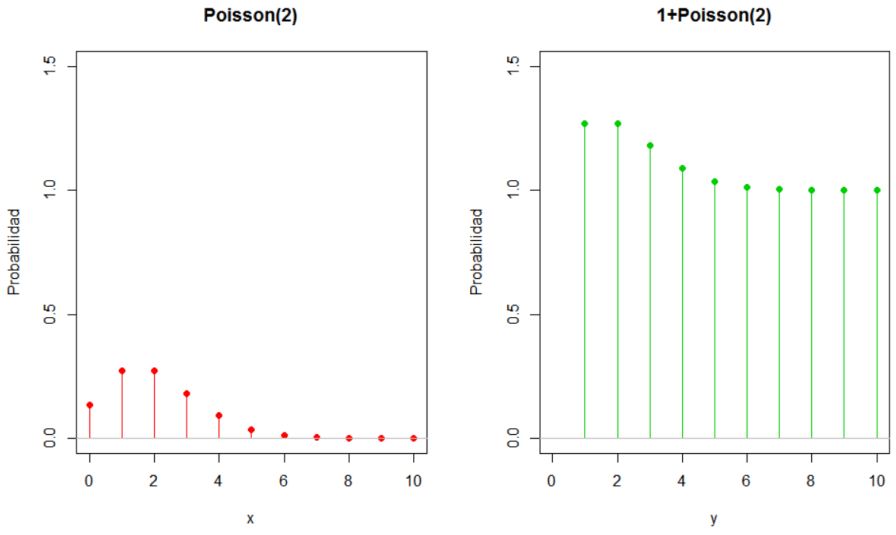


Figura 2.4: Distribución Poisson no sesgada y sesgada

En la Figura 2.4 tenemos una representación de la función de probabilidad de la $Poisson(2)$ y su forma sesgada. Como vemos en la $Poisson(2)$, el cero tiene probabilidad positiva no nula, luego al pasar a su forma sesgada por longitud deja de ser observable. Por lo tanto la media armónica no es consistente, tal y como se refleja en el Cuadro 2.8 (el sesgo no se va a cero a medida que aumenta n). Y en consecuencia la media armónica no converge al verdadero valor de la media.

En estos casos, en lugar de considerar la variable X que toma $k+1$ valores, $x_0 = 0, x_1, ..., x_k$, con probabilidades $p_0, p_1, ..., p_k$, consideraremos una variable aleatoria a la que llamaremos X^t, obtenida al suprimirle a X el valor $x_0 = 0$. Es decir, X^t toma k valores

posibles $x_1, ..., x_k$ con las probabilidades resultantes de eliminar la asociada a $x_0 = 0$, es decir, suprimir p_0, y repartirla al resto de probabilidades $p_1, ..., p_k$ de forma proporcional.

Es fácil ver que $E(X) = 2$, mientras que $E(X^t) = \dfrac{E(X)}{1 - P(X = 0)} = \dfrac{2}{1 - 0{,}1353} = 2{,}3130$. De ahí que la aproximación del sesgo de la media armónica considerando como tamaño muestral $n = 500$ sea 0,3130.

n	Sesgo	Varianza	ECM	Varianza asintótica	σ^2/n
20	35.58	9.05	21.72	111.5	10
50	32.61	3.81	14.4	44.63	4
100	31.45	1.82	11.76	22.31	2
500	31.30	0.36	10.16	4.46	0.4

Cuadro 2.8: Valores de sesgo, varianza, ECM y varianza asintótica de la media armónica en función del tamaño muestral, n, para una Poisson(2). (Los resultados vienen en centésimas). Se añade el valor σ^2/n, también en centésimas.

Como vemos, la varianza asintótica no se parece mucho a la varianza de la media aritmética simple. Esto es debido a que el sesgo no tiende a cero. También lo vemos reflejado en los valores aproximados de la varianza de la media armónica y del ECM que defieren más que cuando hablabamos de distribuciones continuas.

Capítulo 3

Regresión con datos sesgados por longitud

Como hemos visto en la asignatura de *Modelos de Regresión y Análisis Multivariante*, para representar la dependencia de una variable Y (variable dependiente, variable respuesta) con respecto a otra variable X (variable independiente, variable explicativa) se utilizan los modelos de regresión. La relación de causalidad entre estas variables es unidireccional, las variables explicativas pueden influir en la respuesta, pero no a la inversa. Nótese que a partir de ahora habrá un cambio de notación, pues las variables X e Y van a representar respectivamente a la variable explicativa y respuesta de un modelo de regresión, en lugar de ser Y la versión sesgada por longitud de X como hacíamos en los capítulos anteriores.

La regresión se suele formalizar como la media condicionada de la variable respuesta en función del valor que tome la variable explicativa. Es decir,

$$m(x) = \mathbb{E}\left(Y/X = x\right) \qquad para\ \ cada\ \ x.$$

Luego podemos descomponer la variable respuesta en función del resultado de X, más un error de media cero, es decir $Y = m(X) + \varepsilon$. Para construir un modelo de regresión específico en cada caso, tenemos que tener en cuenta si hay una o varias variables explicativas, o variables respuesta, si éstas son discretas o continuas, la forma de la función de regresión (lineal, polinómica, u otras), el tipo de distribución del error, la forma de obtener los datos muestrales, y otros aspectos.

En este tema vamos a tratar modelos de regresión considerando que la variable respuesta Y está sesgada por longitud. El sesgo en la variable respuesta no solo causa una distorsión en la determinación de la función de regresión verdadera, sino que también afecta

a la estimación de la varianza, ver Vardi (1982). A pesar de que el problema de la regresión lineal bajo supuestos de observación de datos habituales se ha estudiado ampliamente en la literatura, el desarrollo de métodos para hacer frente a los datos sesgados se ha llevado a cabo principalmente en el último medio siglo, y gran parte de la teoría se dedica a censurar o truncar datos. Aunque el sesgo por longitud no es una situación extrema, debemos tener en cuenta que, como ocurre con el truncamiento y la censura, los métodos de estimación paramétrica estándar no son adecuados. En este sentido, los métodos que proponemos para estimar la función de regresión se basan en compensar el efecto que produce el sesgo de longitud en las observaciones.

En la primera sección de este tema vamos a hablar del modelo lineal simple con datos no sesgados y veremos como a partir de él obtenemos el modelo de regresión lineal múltiple y el modelo lineal general. Posteriormente desarrollaremos los resultados de inferencia para el modelo lineal general, y veremos su aplicación e interpretación en el modelo de regresión lineal múltiple. En la segunda sección, abordaremos los mismos temas, pero con la diferencia de que trataremos con datos sesgados por longitud. Y en la última, empleando el programa R generaremos muestras sesgadas de un modelo de regresión y estimaremos los coeficientes del modelo con y sin ponderación, para hacer una comparación de los sesgos y desviaciones típicas aproximadas.

3.1. Modelo lineal general con datos no sesgados

En el modelo de regresión lineal simple tanto la variable respuesta Y, como la variable explicativa X, se suponen univariantes. Cada una de ellas refleja el valor de una sola característica. Se suele escribir como:

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

donde $Var(\varepsilon/X = x) = \sigma^2 \quad \forall x$, $\varepsilon \in N(0, \sigma^2)$, y $\varepsilon_1, ..., \varepsilon_n$ son independientes. Es decir, el modelo cumple las hipótesis de linealidad (la función de regresión es una línea recta), homocedasticidad (la varianza del error es la misma cualquiera que sea el valor de la variable explicativa), normalidad ($\varepsilon \in N(0, \sigma^2)$) e independencia (las variables aleatorias que representan los errores $\varepsilon_1, ..., \varepsilon_n$ son mutuamente independiente).

Para poder hacer una estimación de los parámetros del modelo (β_0, β_1), necesitamos datos experimentales. Distinguiremos dos tipos de diseño experimental:

• Diseño fijo. Los valores de la variable explicativa están fijados por el experimentador, de acuerdo a un diseño conveniente. En este caso los valores de la variable explicativa no son aleatorios, sólo es aleatorio el error y, en consecuencia, la variable respuesta. La muestra resultante de un diseño fijo sería del tipo:

$$(x_1, Y_1), ..., (x_n, Y_n)$$

• Diseño aleatorio. Tanto la variable explicativa como la variable respuesta son aleatorias. La muestra resultante de un diseño aleatorio sería del tipo:

$$(X_1, Y_1), ..., (X_n, Y_n)$$

En resumen, un modelo de regresión lineal simple, homocedástico, con errores normales e independientes, del que extraemos una muestra bajo diseño fijo nos proporciona datos del tipo $(x_1, Y_1), ..., (x_n, Y_n)$ donde $x_1, ..., x_n$ son valores fijados por el experimentador, mientras que $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ para $i \in \{1, ..., n\}$, siendo $\varepsilon_1, ..., \varepsilon_n$ independientes.

Si consideramos el modelo de regresión lineal simple y lo extendemos a situaciones más complejas, en las que hay más de una variable explicativa, obtenemos el modelo de regresión lineal múltiple. Sea entonces una variable respuesta Y y una colección de variables explicativas $X_1, X_2, ..., X_{p-1}$. Para obtener un modelo de regresión lineal múltiple, basta con considerar una combinación lineal de las variables explicativas, de la siguiente manera:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_{p-1} X_{p-1} + \varepsilon,$$

donde $\beta_0, \beta_1, ..., \beta_{p-1}$ son los parámetros que acompañan a las variables (β_0 es el intercepto, $\beta_1, ..., \beta_{p-1}$ acompañan a las variables explicativas) y ε el error.

Una muestra bajo diseño fijo de este modelo se puede expresar como

$$Y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_{p-1} x_{i,p-1} + \varepsilon_i,$$

siendo Y_i la variable respuesta del i-ésimo individuo, $x_{i,1}, ..., x_{i,p-1}$ las variables explicativas del mismo y ε_i el error asociado a dicho individuo. Expresado matricialmente:

$$\begin{pmatrix} Y_1 \\ ... \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & . & . & . & x_{1,p-1} \\ ... & ... & . & . & . & ... \\ 1 & x_{n,1} & . & . & . & x_{n,p-1} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ ... \\ \varepsilon_n \end{pmatrix}.$$

Sustituyendo cada vector o matriz por un símbolo, llegaríamos a la siguiente expresión abreviada:

$$Y = X\beta + \varepsilon,$$

donde Y es el vector de respuestas, la matriz X es una matriz $n \times p$, (cada fila representa a un individuo y cada columna a cierta característica) β es el vector de parámetros y ε es un vector que contiene los errores y verifica $\varepsilon \in N_n(0, \sigma^2)$.

La representación del modelo de regresión múltiple permite considerar un contexto más general, denominado modelo lineal general, en el cual caben modelos de regresión con variable explicativa discreta:

$$\begin{pmatrix} Y_1 \\ ... \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & . & . & . & x_{1,p} \\ ... & . & . & . & ... \\ x_{n,1} & . & . & . & x_{n,p} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ ... \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ ... \\ \varepsilon_n \end{pmatrix}.$$

Bajo el modelo lineal general, X es una matriz no aleatoria, β es un vector de parámetros que hay que estimar y $\varepsilon \in N_n(0, \sigma^2 I_n)$, siendo σ^2 la varianza del error, que también hay que estimar, e I_n la matriz identidad de orden n. Esta última expresión aglutina las suposiciones de homocedasticidad, normalidad e independencia de los errores.

3.1.1. Estimación de los parámetros del modelo : β y σ^2.

Nuestro problema ahora se va a centrar en la estimación del vector de parámetros β y de la varianza del error σ^2. Comenzaremos con la estimación de β y plantearemos este procedimiento desde el método de mínimos cuadrados.

Escogeremos como estimador aquel $\hat{\beta}$ donde se alcance

$$\min_{\beta} \sum_{i=1}^{n} (Y_i - x_i\beta)^2,$$

siendo x_i la fila i-esima de la matriz del diseño X.

En notación matricial, el problema de minimización se puede expresar de manera equivalente así:

$$\min_{\beta}(Y - X\beta)'(Y - X\beta) = \min_{\beta} \phi(\beta),$$

donde ϕ es la función objetivo. Si derivamos la función objetivo respecto de β, e igualamos a cero, se obtiene lo que se conoce como las ecuaciones normales de regresión,

$$X'X\beta = X'Y,$$

cuya solución es el estimador de β por mínimos cuadrados, dado por:

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

Una vez que se han obtenido los estimadores de los parámetros, $\hat{\beta}$, se pueden calcular los ajustes o predicciones para los individuos de la muestra, de la siguiente manera:

$$\hat{Y}_i = x_i\hat{\beta}, i \in \{1, ..., n\},$$

o lo que es lo mismo

$$\hat{Y} = X\hat{\beta}.$$

Antes de seguir con la estimación de la varianza, vamos a dar una interpretación geométrica a las predicciones obtenidas con $\hat{\beta}$. Para ello partimos de la expresión matricial

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY,$$

donde $H = X(X'X)^{-1}X'Y$ se conoce como matriz hat. Así, las predicciones $\hat{Y}$ se obtienen aplicando la matriz hat a las observaciones Y. Además $\hat{Y}$ es una combinación lineal de las columnas de X, y dentro de todas las posibles combinaciones, es la que tiene menor distancia a Y. Por ello, podemos decir que la predicción $\hat{Y}$ es la proyección de Y sobre el espacio formado por todas las combinaciones lineales de las columnas de X. Además, también podemos decir que H es una matriz de proyección sobre dicho espacio. Como corresponde a una matriz de proyección, es una matriz $n \times n$ simétrica, idempotente y de rango p.

Vamos cerrar esta sección con la estimación de la varianza.

Es conocido que los residuos se definen como la diferencia entre las observaciones y las predicciones, esto es,

$$\hat{\varepsilon_i} = Y_i - \hat{Y_i} = Y_i - x_i\hat{\beta}, i \in \{1, ..., n\}.$$

Por lo tanto, se puede formar un vector de residuos:

$$Y - \hat{Y} = (I_n - H)Y = MY,$$

donde $M = I_n - H$ se conoce como la matriz generadora de residuos. Es fácil ver que M es una matriz simétrica, idempotente, de rango $(n - p)$ y ortogonal a la matriz hat, es decir, $MH = 0$.

Como los errores no se observan, para estimar su varianza σ^2, emplearemos los residuos que acabamos de definir. Así, el estimador de la varianza del error sería

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\hat{\varepsilon_i}^2 = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - x_i\hat{\beta})^2 = \frac{RSS}{n-p}.$$

Hemos extendido la notación RSS para la suma residual de cuadrados, y hemos empleado como denominador $(n - p)$ para conseguir un estimador insesgado de la varianza del error.

3.1.2. Propiedades de los estimadores

En este apartado vamos a deducir las propiedades de los estimadores de los parámetros asociados al modelo lineal genera, es decir de $\hat{\beta}$ y de $\hat{\sigma}^2$. Estas propiedades serán su comportamiento en media (veremos que son insesgados), su varianza y su distribución e independencia.

En primer lugar vamos a ver que $\hat{\beta}$ es un estimador insesgado de β. Para ello es suficiente emplear que $\mathbb{E}(\varepsilon) = 0$.

$$\mathbb{E}(\hat{\beta}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\beta = \beta$$

Vamos a obtener la matriz de covarianzas del vector de aleatorio $\hat{\beta}$. Basta tener en cuenta la hipótesis de homocedasticidad, es decir , $Var(Y_i) = \sigma^2$. Por lo que la matriz de covarianzas del vector de respuestas se puede escribir como $Cov(Y, Y) = \sigma^2 I_n$.

$$Cov(\hat{\beta}, \hat{\beta}) = Cov\left((X'X)^{-1}X'Y, (X'X)^{-1}X'Y\right) =$$

$$= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Para la estimación de cada coeficiente del modelo, que es una componente del vector β, se toma la componente correspondiente de $\hat{\beta}$, que es un estimador insesgado. Y su varianza se obtiene multiplicando σ^2 por el elemento correspondiente de la diagonal de $(X'X)^{-1}$.

Vamos a mencionar un par de propiedades que utilizaremos en la demostración del siguiente teorema.

Propiedad 1. Si $X \in N_m(\mu, \Sigma)$ y C es una matriz $p \times m$ de rango p, entonces

$$CX \in N_p(C\mu, C\Sigma C').$$

Propiedad 2. Sea $X_1, ..., X_m \in N(0, \sigma^2)$ una muestra aleatoria simple formada por m observaciones independientes de una misma distribución normal de media cero y varianza σ^2, y $X = (X_1, ..., X_m)' \in N_m(0, \sigma^2 I_m)$ el vector aleatorio construido con las observaciones.

(i) Si A es una matriz simétrica de orden $m \times m$, idempotente $(A^2 = A)$ y de rango $r \leq m$, entonces

$$X'AX \in \sigma^2 \chi^2_r.$$

(ii) Si A es una matriz en las condiciones anteriores y b es un vector de dimensión m, tal que $Ab = 0$, entonces

$$X'AX \quad y \quad b'X \quad son\ independientes.$$

(iii) Si A y B son dos matrices en las condiciones anteriores y $AB = 0$, entonces

$$X'AX \quad y \quad X'BX \quad son\ independientes.$$

Teorema 3.1. *Supongamos que los errores $\varepsilon_1, ..., \varepsilon_n$ son independientes y tienen distribución común $N(0, \sigma^2)$, y X es una matriz $n \times p$ de rango p. Entonces*

(i) $\hat{\beta} \in N_p(\beta, \sigma^2(X'X)^{-1})$

(*ii*) $\dfrac{RSS}{\sigma^2} = \dfrac{(n-p)\hat{\sigma}^2}{\sigma^2} \in \chi^2_{n-p}$

(*iii*) *$\hat{\beta}$ y RSS (ó $\hat{\sigma}^2$) son independientes.*

Demostración. Para obtener (*i*) simplemente tenemos que aplicar la Propiedad 1 (sobre transformaciones lineales de vectores aleatorios normales), teniendo en cuenta que $\hat{\beta} = (X'X)^{-1}X'Y$, e $Y \in N_n(X\beta, \sigma^2 I)$.

Por lo tanto $\hat{\beta}$ tendrá distribución normal, cuyo vector de medias y matriz de covarianzas han sido calculados anteriormente.

Para obtener (*ii*) tenemos que emplear el apartado (*i*) de la Propiedad 2 (sobre transformaciones cuadráticas aplicadas a una muestra de variables normales), pues según hemos visto $RSS = \varepsilon'(I_n - H)\varepsilon$, y como $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ está en las condiciones de la Propiedad 2, y la matriz $(I_n - H)$ es simétrica, idempotente y de rango $(n-p)$, entonces $RSS \in \sigma^2\chi^2_{n-p}$.

Finalmente el apartado (*iii*) se obtiene también de la Propiedad 2, en este caso de su apartado (*ii*), para lo cual basta con observar que

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon = \beta + (X'X)^{-1}X'\varepsilon,$$

mientras que $RSS = \varepsilon'(I_n - H)\varepsilon$. Como $(X'X)^{-1}X'(I_n - H) = 0$, entonces $\varepsilon'(I_n - H)\varepsilon$ y $(X'X)^{-1}X'\varepsilon$ son independientes, y por tanto, también lo son $\hat{\beta}$ y RSS. □

3.2. Modelo lineal general con datos sesgados por longitud

En esta sección vamos a seguir tratando con el modelo lineal general, con la diferencia de que ahora consideraremos que la variable respuesta Y está sesgada por longitud. Daremos una estimación de los parámetros del modelo, que van a ser diferentes a los vistos en la sección anterior debido a la presencia del sesgo, y mencionaremos algunas de sus propiedades.

Hemos visto que un modelo de regresión lineal, homocedástico, con errores normales e independientes, del que extraemos una muestra bajo diseño aleatorio nos proporciona datos del tipo $(X_1, Y_1), ..., (X_n, Y_n)$. Seguimos suponiendo el modelo lineal para las variables originales, pero ahora suponemos que el proceso de observación está sometido a cierto sesgo. De este modo, la muestra resultante será del tipo $(X_1^w, Y_1^w), ..., (X_n^w, Y_n^w)$, siendo w una función que indica la probabilidad de observar cada dato bajo el mecanismo de sesgo.

Por ejemplo, en el caso de sesgo por longitud, w es proporcional al valor de Y. Por lo tanto, bajo sesgo por longitud, nuestra muestra $(X_1^w, Y_1^w), ..., (X_n^w, Y_n^w)$, es una muestra i.i.d. de una variable aleatoria con distribución F^w y cuya densidad viene dada por:

$$dF^w(x,y) = \frac{y}{\mu_Y} dF(x,y).$$

siendo $F(x,y)$ la distribución bivariada de (X,Y) y $\mu_Y = E(Y)$.

3.2.1. Estimación de los parámetros del modelo

Uno de los procedimientos más utilizados en la literatura para obtener una estimación de β, es el método de mínimos cuadrados ponderados, que consiste en:

$$\min_{\beta} \sum_{i=1}^{n} \frac{1}{w_i} (Y_i^w - x_i^w \beta)^2, \tag{3.1}$$

siendo $w_i = \omega(x_i^\omega, Y_i^\omega)$. La introducción de los recíprocos de ω_i, como pesos en los mínimos cuadrados ponderados, es la manera de corregir el sesgo de los datos para obtener un estimador consistente de los coeficientes del modelo lineal en la distribución original de (X,Y). En concreto, en el caso de sesgo por longitud, basta con tomar el recíproco de las respuestas (ver Cristóbal y Alcalá (2000)), obteniendo así el siguiente problema de optimización:

$$\min_{\beta} \sum_{i=1}^{n} \frac{1}{Y_i^w} \left(Y_i^w - x_i^w \beta\right)^2.$$

La solución al problema anterior es

$$\hat{\beta} = \left((X^w)' W X^w\right)^{-1} \left((X^w)' W Y^w\right),$$

donde Y^w es un vector columna con las observaciones Y_i^w, X^w es una matriz $n \times p$ dada por

$$X^w = \begin{pmatrix} x_1^w \\ \vdots \\ x_n^w \end{pmatrix} = \begin{pmatrix} x_{1,1}^w & \cdots & x_{1,p}^w \\ \vdots & \cdots & \vdots \\ x_{n,1}^w & \cdots & x_{n,p}^w \end{pmatrix}$$

y W es la siguiente matriz diagonal

$$W = \begin{pmatrix} (Y_1^w)^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (Y_n^w)^{-1} \end{pmatrix}.$$

3.2.2. Propiedades de los estimadores

Aunque el estimador por mínimos cuadrados ponderados, $\hat{\beta}$, es fácil de calcular, sus propiedades de sesgo y varianza no son sencillas de obtener, pues involucra unos pesos aleatorios, $(Y_i^w)^{-1}$ en el caso de sesgo por longitud en Y. Aquí vamos a proporcionar las propiedades asintóticas que obtuvieron Ojeda Cabrera y Van Keilegom (2009) para el estimador $\hat{\beta}$ con función de sesgo arbitraria $\omega(x, y)$ y en un contexto de regresión paramétrica no necesariamente lineal.

Así, se considera un modelo de regresión del tipo

$$Y = m(X, \beta) + \sigma(X)\varepsilon,$$

donde Y es la variable respuesta, X la variable explicativa, ε el error, que se supone independiente de X (cumple que $\mathbb{E}(\varepsilon) = 0$,$Var(\varepsilon) = 1$), $m(x, \beta)$ es la función de regresión , siendo β un parámetro desconocido que debemos estimar, y $\sigma^2(x) = Var(Y/X = x)$ es la función de varianza condicional. Claramente, el modelo de regresión puede no ser lineal. Cabría, por ejemplo, un modelo periódico como éste $m(X, \beta) = \beta_0 + \beta_1 \cdot \text{sen}\left(\frac{2\pi X}{\beta_2}\right)$, donde la función de regresión no es lineal respecto del vector de parámetros $\beta = (\beta_0, \beta_1, \beta_2)$.

A continuación enunciamos el resultado sobre la distribución límite del estimador β, que obtuvieron Ojeda Cabrera y Van Keilegom (2009), página 2839. En su Lemma 3.1 se considera una función de peso $\omega(x, y)$ arbitraria.

Lema 3.2. *Consideremos el modelo de regresión*

$$Y = m(X, \beta) + \sigma(X)\varepsilon,$$

del cual se observa una muestra bajo sesgo de selección. Entonces, bajo ciertas condiciones de regularidad en m y σ, el estimador por mínimos cuadrados ponderados, β, presenta la siguiente distribución límite cuando n tiende a infinito:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma),$$

siendo

$$\Omega = \mathbb{E}\left[\frac{\partial m_\beta(X)}{\partial \beta}\frac{\partial m_\beta(X)^T}{\partial \beta}\right]$$

y

$$\Sigma = \mu_w \Omega^{-1} \mathbb{E}\left[\frac{(Y - \partial m_\beta(X))^2}{w(X, Y)}\frac{\partial m_\beta(X)}{\partial \beta}\frac{\partial m_\beta(X)^T}{\partial \beta}\right]\Omega^{-1}.$$

En el Lema 3.2 se denota $\mu_w = \mathbb{E}(\omega(X,Y))$. Observamos que la matriz Ω juega el papel de la matriz $X'X$ de los modelos lineales. De hecho, en la matriz de covarianzas aparece Ω junto con una matriz ponderada en función del cuadrado del erro y la función de ponderación por sesgo de selección, ω.

3.3. Simulaciones

En esta sección se presenta un estudio de simulación en el que se han estudiado las propiedades de sesgo y varianza del estimador por mínimos cuadrados ponderados $\hat{\beta}$, y se han comparado con el estimador por mínimos cuadrados ordinarios (sin ponderación).

Hemos considerado el modelo de regresión lineal simple

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

donde $\beta_0 = \beta_1 = 3$, X tiene distribución uniforme en el intervalo $[0,1]$ y $\varepsilon \in \chi^2_2 - 2$, esto es, el error tiene distribución ji-cuadrado con dos grados de libertad, a la cual le hemos restado 2 para que tenga media cero.

Generar muestras de este modelo, añadiendo sesgo por longitud en Y, no es tan sencillo como en el tema anterior, en el cual no había regresión, y había muchos casos de distribuciones cuya versión sesgada era conocida o fácilmente deducible. En el contexto de regresión que nos ocupa, la versión sesgada del vector (X,Y) no tiene distribución conocida. En estas circunstancias, hemos optado por generar una cantidad ingente de datos del modelo original (un millón en este caso) y tomar muestras de esta "pseudo-población" con probabilidades proporcionales a los valores de Y.

De este modo, resulta una muestra $(X_1^w, Y_1^w), \ldots, (X_n^w, Y_n^w)$ obtenida bajo sesgo por longitud de Y. Realmente no hemos tomado una muestra sino mil, para obtener así mil réplicas del estimador por mínimos cuadrados ponderados, $\hat{\beta}$, cada una calculada sobre una muestra. A partir de estas mil réplicas podremos aproximar el sesgo y la varianza del estimador.

En el Cuadro 3.1 se muestran el sesgo y la varianza aproximados en mil simulaciones del modelo, para los estimadores de los coeficientes de regresión, por mínimos cuadrados ponderados (bajo el título "Con ponderación" en el cuadro) y por mínimos cuadrados ordinarios (bajo el título "Sin ponderación" en el cuadro). Se han considerado dos tamaños de muestra, $n = 50$ y $n = 100$.

Respecto del sesgo, observamos que el estimador sin ponderación presenta un sesgo considerable (tanto para la ordenada en el origen como para la pendiente), bastante mayor que el sesgo del estimador con ponderación, y lo que es más grave, ese sesgo no se reduce al aumentar el tamaño muestral. Esto es coherente con la previsible falta de consistencia del estimador sin ponderación, pues no corrige el sesgo de observación, y por tanto, no converge a los verdaderos valores de los coeficientes.

Por el contrario, el estimador con ponderación reduce su sesgo con el tamaño muestral, y como era previsible, converge a los valores verdaderos de los coeficientes. Sin embargo, debemos destacar que es un estimador sesgado, particularmente para tamaños de muestra pequeños o moderados, pues la corrección del sesgo de observación no es perfecta.

En relación con la varianza, tanto el estimador con ponderación como el estimador ordinario presentan varianzas decrecientes con el tamaño muestral. Aun así, el estimador con ponderación también muestra varianzas menores que el estimador ordinario.

En consecuencia, hemos obtenido las propiedades que cabía esperar de los estimadores, con ponderación por sesgo y sin ella. Podemos concluir que la ponderación por sesgo es necesaria para obtener estimaciones correctas de los coeficientes de regresión, como ya ocurría con la media armónica respecto de la media usual en capítulos anteriores.

Método	Coeficientes	Sesgo		Varianza	
		n=50	n=100	n=50	n=100
Sin ponderación	Ordenada en el origen	1.2135	1.2131	0.6697	0.3462
	Pendiente	-0.5839	-0.5909	1.7435	0.8760
Con ponderación	Ordenada en el origen	0.0779	0.02024	0.3309	0.1713
	Pendiente	-0.0767	-0.0102	0.8806	0.4463

Cuadro 3.1: Sesgo y varianza de los estimadores de los coeficientes de regresión, con ponderación y sin ponderación.

Apéndice A

Código R para las simulaciones

A.1. Distribuciones continuas

Modelo: F es uniforme en [a,b]

```
#- - - Generación de los datos
set.seed(123456)
a=9
b=10
media=(a+b)/2
varianza=(b-a)^2/12
n=20 # Tamaño muestral
ns=1000 # Mil muestras simuladas
vm=c() # Vector que recogerá las mil medias aritmética
vma=c() # Vector que recogerá las mil medias armónica

#- - - Generamos muestras
for (is in 1:ns){
    u=runif(n,0,1)
    x=sqrt(a^2+u*(b^2-a^2)); x # Muestra sesgada
    # Media aritmética simple
    vm[is]=mean(x)
    # Media armónica
    vma[is]=1/mean(1/x)
    }
```

```
vm[is]
vma[is]
mean(vm)
mean(vma)

#- - - Propiedades de la media armónica
sesgo=mean(vma)-media # Sesgo
et_hat=sd(vma) # Error típico aproximado por simulación
var=(et_hat)^2 # Varianza aproximada por simulación
ECM=var+sesgo^2 # Error cuadrático medio
sigma^2/n # Var(X̄)
new=log(10)-log(9)
varasin=((media2̂)*(media*new-1))/n # Varianza asintótica

#- - - Representación gráfica
par(mfrow=c(3,2))

#- - Uniforme (0,1) no sesgada
a=0
b=1
curve(dunif(x, min = 0, max = 1),xlab = "x", ylab ="Probabilidad",xlim=c(0,1),ylim=c(0,1),
0,1,col = 2, lwd = 2, xaxt="n", main = "Uniforme(0,1)")
axis(1, at = c(0:1), cex.axis=1)
abline(v=0, col="gray")
abline(v=1, col="gray")

#- - Uniforme (0,1) sesgada
curve(x*dunif(x, min = 0, max = 1)/(1/2),xlab = "y", ylab = "Probabilidad",xlim=c(0,1),
ylim=c(0,2),0,1,col = 3, lwd = 2, xaxt="n", main = "Uniforme(0,1) sesgada")
axis(1, at = c(0:1), cex.axis=1)
abline(v=0, col="gray")
abline(v=1, col="gray")

#- - Uniforme (9,10) no sesgada
a=9
b=10
```

```
curve(dunif(x, min = 9, max = 10),xlab = "x", ylab = "Probabilidad",xlim=c(9,10),ylim=c(0,1),
9,10,col = 2, lwd = 2, xaxt="n", main = "Uniforme(9,10)")
axis(1, at = c(9:10), cex.axis=1)
abline(v=9, col="gray")
abline(v=10, col="gray")

#- - Uniforme (9,10) sesgada
curve(x*dunif(x, min = 9, max = 10)/(19/2),xlab = "y", ylab ="Probabilidad",xlim=c(9,10),
ylim=c(0,1.5), 9,10,col = 3, lwd = 2,xaxt="n", main = "Uniforme(9,10) sesgada")
axis(1, at = c(9:10), cex.axis=1)
abline(v=9, col="gray")
abline(v=10, col="gray")

#- - Uniforme (100,101) no sesgada
a=100
b=101
curve(dunif(x, min = 100, max = 101),xlab = "x", ylab = "Probabilidad",xlim=c(100,101),
ylim=c(0,1), 100,101,col = 2, lwd = 2, xaxt="n", main = "Uniforme(100,101)")
axis(1, at = c(100:101), cex.axis=1)
abline(v=100, col="gray")
abline(v=101, col="gray")

#- - Uniforme (100,101) sesgada
curve(x*dunif(x, min = 100, max = 101)/(201/2),xlab = "y", ylab = "Probabilidad",
xlim=c(100,101), ylim=c(0,1.5), 100,101,col = 3, lwd = 2,xaxt="n", main =
"Uniforme(100,101) sesgada")
axis(1, at = c(100:101), cex.axis=1)
abline(v=100, col="gray")
abline(v=101, col="gray")
```

Modelo: F es Gamma(p,a)

```
#- - - Generación de los datos
set.seed(123456)
p=8
```

```
a=1
media=p/a
sigma=sqrt(p)/a
sigma^2
n=20
et=sigma/sqrt(n)
et

#- - - Generamos muestras
ns=1000 # Mil muestras simuladas
vm=c() # Vector que recogerá las mil medias aritmética
vma=c() # Vector que recogerá las mil medias armónica

for (is in 1:ns){
    # x=rgamma(n,shape=p,scale=1/a) # Muestra no sesgada
    x=rgamma(n,shape=p+1,scale=1/a) # Muestra sesgada
    # Media aritmética simple
    vm[is]=mean(x)
    # Media armónica
    vma[is]=1/mean(1/x)
    }

#- - - Propiedades de la media aritmética simple
sesgo=mean(vm)-media
et_hat=sd(vm) # Error típico aproximado por simulación
et_hat^2

#- - - Propiedades de la media armónica
sesgo=mean(vma)-media
et_hat=sd(vma) # Error típico aproximado por simulación
var=(et_hat)^2 # Varianza aproximada por simulación
ECM=var+sesgo^2
sigma^2/n # Var(X̄)
et^2
new=a^2/(p-1)
```

```
((media^2)*(media*new-1))/n # Varianza asintótica

#- - - Representación gráfica
par(mfrow=c(2,2))

#- - Gamma(2,1) no sesgada
p=2
curve(dgamma(x, shape=p, scale=1/a), xlab = "x", ylab = "f(x;p,a)", 0, 10, col = 2,
lwd = 2, main = "Gamma(2,1)")

#- - Gamma(2,1) sesgada
p=3
curve(dgamma(x, shape=p+1, scale=1/a), xlab = z", ylab = "f(x;p,a)", 0, 10, col = 3,
lwd = 2, main = "Gamma(3,1)")

#- - Gamma(8,1) no sesgada
p=8
curve(dgamma(x,shape=p, scale=1/a), xlab = "x", ylab = "f(x;p,a)", 0,10,col = 2,
lwd = 2, main = "Gamma(8,1)")

#- - Gamma(8,1) sesgada
p=9
curve(dgamma(x, shape=p+1, scale=1/a), xlab = z", ylab = "f(x;p,a)", 0,10,col = 3,
lwd = 2, main = "Gamma(9,1)")
```

Modelo: F es Beta(a,b)

```
#- - - Generación de los datos
set.seed(123456)
a=1
b=1
media=a/(a+b)
sigma=(a*b)/((a+b)^2*(a+b+1))
n=20
et=sigma/sqrt(n)
```

```
#- - - Generamos muestras
ns=1000 # Mil muestras simuladas
vm=c() # Vector que recogerá las mil medias aritmética
vma=c() # Vector que recogerá las mil medias armónica

for (is in 1:ns){
    # x=rbeta(n,shape1=a, shape2=b, ncp = 0) # Muestra no sesgada
    x=rbeta(n,shape1=a+1, shape2=b, ncp = 0) # Muestra sesgada
    # Media aritmética simple
    vm[is]=mean(x)
    # Media armónica
    vma[is]=1/mean(1/x)
    }

#- - - Propiedades de la media aritmética simple
sesgo=mean(vm)-media
et_hat=sd(vm); et_hat # Error típico aproximado por simulación

#- - - Propiedades de la media armónica
sesgo=mean(vma)-media
et_hat=sd(vma); et_hat # Error típico aproximado por simulación
var=(et_hat)^2
ECM=var+sesgo^2
sigma^2/n
new=(a+b-1)/(a-1)
((media^2)*(media*new-1))/n # Varianza asintótica

#- - - Representación gráfica
par(mfrow=c(2,2))

#- - Beta(2,1) no sesgada
a=2
b=1
curve(dbeta(x, shape1=a, shape2=b,ncp=0,log = FALSE), xlab = "x", ylab = "f(x;p,q)",
0,1,col = 2, lwd = 2, main = "Beta(2,1)")
```

```
#- - Beta(2,1) sesgada
curve(dbeta(x, shape1=a+1, shape2=b,ncp=0,log = FALSE), xlab = ", ylab ="f(x;p,q)",
0,1,col = 3, lwd = 2, main = "Beta(3,1)")

#- - Beta(9,1) no sesgada
a=9
b=1
curve(dbeta(x, shape1=a, shape2=b,ncp=0,log = FALSE), xlab = "x", ylab = "f(x;p,q)",
0,1,col = 2, lwd = 2, main = "Beta(9,1)")

#- - Beta(9,1) sesgada
curve(dbeta(x, shape1=a+1, shape2=b,ncp=0,log = FALSE), xlab = ", ylab ="f(x;p,q)",
0,1,col = 3, lwd = 2, main = "Beta(10,1)")
```

A.2. Distribuciones discretas

Modelo: F es Poisson(λ)

```
set.seed(123456)
lambda=2
sigma=sqrt(lambda)
n=20
et=sigma/sqrt(n)

#- - - Generamos muestras
ns=1000 # Mil muestras simuladas
vm=c() # Vector que recogerá las mil medias aritmética
vma=c() # Vector que recogerá las mil medias armónica

for (is in 1:ns){
    # x=rpois(n,lambda=lambda) # Muestra no sesgada
    x=1+rpois(n,lambda=lambda) # Muestra sesgada
    # Media aritmética simple
    vm[is]=mean(x)
    # Media armónica
    vma[is]=1/mean(1/x)
```

```
    }

#- - - Propiedades de la media aritmética simple
sesgo=mean(vm)-lambda
et_hat=sd(vm); et_hat # Error típico aproximado por simulación

#- - - Propiedades de la media armónica
sesgo=mean(vma)-lambda
et_hat=sd(vma); et_hat # Error típico aproximado por simulación
var=(et_hat)^2
ECM=var+sesgo^2
sigma^2/n
new=(pi2̂/6)*lambda
((media^2)*(media*new-1))/n # Varianza asintótica

#- - - Representación gráfica
par(mfrow=c(1,2))

#- - Poisson(2) no sesgada
lambda=2
.x <- 0:30
plot(.x, dpois(.x, lambda=lambda,log = FALSE),col=2,xlim=c(0,10),ylim=c(0,1.5),
xlab="x", ylab="Probabilidad", main="Poisson(2)", type="h")
points(.x, dpois(.x, lambda=lambda), col=2,pch=16)
abline(h=0, col="gray")

#- - Poisson(2) sesgada
plot(.x, 1+dpois(.x, lambda=lambda,log = FALSE),col=3,xlim=c(0,10),ylim=c(0,1.5),
xlab=ƨ", ylab="Probabilidad", main="1+Poisson(2)", type="h")
points(.x, 1+dpois(.x, lambda=lambda), col=3,pch=16)
abline(h=0, col="gray")
```

A.3. Regresión con datos sesgados por longitud

```
set.seed(123456)

nn=1000000 # Tamaño poblacional de referencia
beta=c(3,3) # Coeficientes de regresión

xg=runif(nn)
eps=rchisq(nn,df=2)-2
yg=beta[1]+beta[2]*xg+eps
#plot(xg,yg,ylim=c(0,max(y)))
#abline(a=2,b=3)

p=yg/sum(yg)

n=50

# Generamos muestras
ns=1000 # Mil muestras simuladas
m_sin=matrix(0,nrow=ns,ncol=2) # Matriz que recogerá los mil pares de coeficientes (sin ponderación)
m_con=matrix(0,nrow=ns,ncol=2) # Matriz que recogerá los mil pares de coeficientes (con ponderación)
for (is in 1:ns){
    #- - - Generación de la muestra sesgada
    ind=sample(nn,n,prob=p)
    x=xg[ind]
    y=yg[ind]
    # plot(x,y,ylim=c(0,max(y)))

    #- - - Ajuste lineal sin ponderación
    beta_sin=coef(lm(yx̃))
    m_sin[is,]=beta_sin

    #- - - Ajuste lineal con ponderación
    beta_con=coef(lm(yx̃,weights=1/y))
```

```
    m_con[is,]=beta_con

    if (is==floor(is/5)*5){
    cat("Sample",is,"Coef_sin",beta_sin,"Coef_con",beta_con,"\n")}
}

# Propiedades de los coeficientes sin ponderación
sesgo_sin=colMeans(m_sin)-beta; sesgo_sin
et_sin=c(sd(m_sin[,1]),sd(m_sin[,2])); et_sin # Error típico aproximado por
simulación

# Propiedades de los coeficientes con ponderación
sesgo_con=colMeans(m_con)-beta; sesgo_con
et_con=c(sd(m_con[,1]),sd(m_con[,2])); et_con # Error típico aproximado por
simulación
```

Bibliografía

[1] Feller,W. (1971), *An Introduction to Probability Theory and Its Applications. Vol II,* John Wiley & Sons Inc, New York.

[2] Patil, G. P. (1984). Studies in statistical ecology involving weighted distributions. *Statistics: Applications and New Directions*, Indian Stat. Inst., Calcutta, 478-503.

[3] Cox, D.R.. (1969). Some Sampling Problems in Technology. In: *New Developments in Survey Sampling*, John Wiley, New York, 506-527.

[4] Vardi, Y. (1982). Nonparametric estimation of a regression function from recurrence times. *Ann. Stat.*, **10**, 616-620.

[5] Patil, G. P. (1978). Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families. *Biometrics*, **34**, 179-189.

[6] Vélez, R. y A. García (1993) *Principios de Inferencia Estadística.*, UNED.

[7] Jorge L. Ojeda Cabrera and Ingrid Van Keilegom (2009).Goodness-of-fit tests for parametric regression with selection biased data *Journal of Statistical Planning and Inference Vol 139,*, **8**, 2836 - 2850.

[8] Cristóbal, J. A., Ojeda, J. L., Alcalá, J. T (2004).. Confidence bands in nonparametric regression with length biased data. *Annals of the Institute of Statistical Mathematics*, **56 (3)**, 175 - 196.