

UNIVERSIDAD DE MURCIA
FACULTAD DE MATEMÁTICAS



**MÉTODOS DE REMUESTREO: JACKKNIFE
Y BOOTSTRAP**

Rafael Jordá Muñoz

Grado en Matemáticas
Trabajo Fin de Grado
Curso 2018-2019

Declaración de originalidad

RAFAEL JORDÁ MUÑOZ, autor del TFG "MÉTODOS DE REMUESTREO: JACKKNIFE Y BOOTSTRAP", bajo la tutela de los profesores FÉLIX BELZUNCE TORREGROSA y JOSÉ MARÍA RUIZ GÓMEZ, declara que el trabajo que presenta es original, en el sentido de que ha puesto el mayor empeño en citar debidamente todas las fuentes utilizadas.

Índice general

Resumen	1
Abstract	5
1. Preliminares y resultados básicos	9
1.1. Conceptos generales	9
1.2. Resultados de convergencia	10
1.3. Resultados sobre métricas	13
1.4. Otros resultados	14
2. Método jackknife	17
2.1. Introducción	17
2.2. El método jackknife	18
2.3. Consistencia	23
2.4. Conclusiones	28
3. Método bootstrap	29
3.1. Introducción	29
3.2. El método bootstrap	30
3.3. Estimación bootstrap de la distribución de un estadístico	32
3.3.1. Consistencia	33
3.4. Estimación de la varianza de un estadístico	42
3.4.1. Consistencia	43
3.5. Otras consideraciones	49
3.5.1. Situaciones donde el método bootstrap no se puede aplicar	49
3.5.2. Intervalos de confianza bootstrap	50
3.6. Conclusiones	54
4. Caso práctico	55
4.1. Introducción	55
4.2. Estudio de la media muestral para una distribución exponencial	56
4.2.1. Estimación jackknife y bootstrap de la varianza	57
4.2.2. Estimación bootstrap de la distribución en el muestreo	60
4.3. Estudio de una transformación de la media muestral para una distribución exponencial	63
Bibliografía	67

Resumen

La rama de las matemáticas que trata de obtener conclusiones de una determinada población, a partir de una muestra suficientemente representativa de la misma, se conoce como la Inferencia Estadística. Este mecanismo permite, ante un fenómeno aleatorio, resolver la importante cuestión de cuál es el modelo probabilístico al que se ajusta, o, dado el modelo, cuáles son los parámetros que lo caracterizan.

El primer paso en todo procedimiento de inferencia es la obtención de observaciones de la variable o las variables que estamos estudiando. Cada observación de una variable X da lugar a un dato, y en general realizadas n observaciones se tienen n datos (x_1, \dots, x_n) , que son los que conforman la muestra. En este trabajo se asumirá que las observaciones que provienen de la variable X son independientes. Para modelizar el total de muestras de tamaño fijo n se utiliza una variable aleatoria de dimensión n , (X_1, \dots, X_n) , que cumple que las variables X_i que la conforman son independientes e idénticamente distribuidas, con la misma distribución que X . A lo largo de todo el trabajo se partirá de este tipo de muestras, que definiremos como muestras aleatorias simples.

En cualquier proceso de Inferencia Estadística, para poder extraer conclusiones sobre la población subyacente bajo la muestra de partida, se necesita obtener ciertas características de la muestra, como su media y su varianza, denominadas comúnmente como media muestral y varianza muestral. A las transformaciones de la muestra aleatoria simple se les denomina estadísticos, y son de la forma $h(X_1, \dots, X_n)$, siendo $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$ una función medible Borel. Estos estadísticos son fundamentales en los tres problemas más importantes de la Inferencia Estadística: la estimación puntual, los intervalos de confianza alrededor de un parámetro y la construcción de contrastes de hipótesis sobre variables aleatorias.

Muchos procedimientos estadísticos requieren algún conocimiento de la distribución en el muestreo del estadístico que se esté usando en el análisis y el tipo de conocimiento requerido depende del tipo de análisis. Para construir intervalos de confianza y contrastes de hipótesis, por ejemplo, se necesita el conocimiento de la distribución en el muestreo en sí misma o de los percentiles de la misma. Sin embargo, en un problema de estimación puntual es indispensable tener una medida del error de la estimación dada por el estadístico (en este caso llamado estimador), lo cual está altamente relacionado con la varianza y el sesgo del mismo. Estas medidas son características de la distribución en el muestreo del estimador.

Por tanto, lo ideal en estos problemas es poder conocer la distribución exacta que tienen los estadísticos, pero eso solo ocurre en situaciones muy concretas, como es el caso del Teorema de Fisher, que establece que la media muestral de una muestra aleatoria simple de una variable X cuya distribución es $N(\mu, \sigma^2)$, sigue una distribución $N\left(\mu, \frac{\sigma^2}{n}\right)$, y que el estadístico $\frac{n-1}{n}S^2$ sigue una distribución χ_{n-1}^2 , donde S^2 es la cuasi-varianza muestral que introduciremos en la Sección 1.1 y χ_{n-1}^2 representa una distribución ji-cuadrado de $n - 1$ grados de libertad. Otro ejemplo es el caso de la media muestral en la distribución exponencial, que sigue una distribución gamma.

Salvo en casos particulares, en general no hay resultados para obtener distribuciones exactas de estadísticos. Una alternativa es trabajar con distribuciones asintóticas, es decir, las distribuciones límite de los estadísticos. Uno de los principales resultados sobre distribuciones asintóticas es el Teorema Central del Límite, que establece que, dada una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria X , se tiene que

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow_d N(0, 1),$$

donde $S_n = X_1 + \dots + X_n$, $\mu = E(X)$, $\sigma^2 = Var(X)$ y el símbolo \rightarrow_d representa la convergencia en distribución que definiremos en la Sección 1.2.

Los resultados sobre distribuciones asintóticas representan una buena estrategia cuando no se dispone de la distribución exacta de un estadístico. Sin embargo, no siempre se pueden establecer resultados de convergencia en distribución como el indicado anteriormente, y surge la necesidad de estimar de alguna otra manera la distribución del estadístico, y las características de la misma.

La principal alternativa en la estimación de la distribución en el muestreo de un estadístico y sus características son los métodos de remuestreo. Este Trabajo de Fin de Grado está enfocado al estudio de algunos de ellos. En particular, estudiaremos los métodos jackknife y bootstrap, que están considerados como las técnicas de remuestreo más importantes en el análisis estadístico, y tienen como idea fundamental estimar las características de un estadístico mediante la obtención repetida de muestras de los datos originales, y hacer inferencias sobre esos remuestreos.

El método jackknife obtiene esos remuestreos mediante la eliminación de una observación cada vez de la muestra original, dando lugar a n nuevas muestras de tamaño $n - 1$, a partir de las cuales obtiene la estimación de las características de los estadísticos. Este método fue propuesto inicialmente para reducir y estimar el sesgo de un estimador de un parámetro de la población, pero con el paso del tiempo se ha demostrado que también se puede utilizar como método para estimar la varianza de un estadístico.

Por otro lado, el método bootstrap no elimina observaciones para obtener los remuestreos, sino que crea nuevas observaciones mediante la obtención de un número alto de muestras con reemplazamiento de la muestra original. Este método es especialmente importante por su capacidad de estimar de manera consistente la dis-

tribución en el muestreo de un estadístico, además de estimar también su varianza.

El objetivo principal de este trabajo es explicar en detalle ambas técnicas, junto con sus principales resultados teóricos, y posteriormente llevar esos resultados a la práctica mostrando un ejemplo de aplicación usando el lenguaje de programación *R*. Para ello, hemos recopilado en cuatro capítulos, una parte importante de resultados referentes a ambas técnicas que muestran las posibles aplicaciones de las mismas. Algunas demostraciones han sido modificadas y en ocasiones extendidas o reducidas, con el objetivo de que este trabajo tenga una mayor claridad y coherencia.

El trabajo consta de cuatro capítulos, cuyos contenidos resumimos a continuación:

En el primer capítulo introducimos una serie de definiciones y resultados que serán de utilidad a lo largo del trabajo. Comenzamos definiendo conceptos generales de Inferencia Estadística y a continuación exponemos resultados de convergencia, que serán fundamentales en las demostraciones de la consistencia de los estimadores jackknife y bootstrap, como el Teorema de las Transformaciones Continuas, el Teorema Central del Límite y la Ley Fuerte de los Grandes Números. Introducimos después las métricas K y $\tilde{\rho}_r$, y estudiamos algunas de sus propiedades. Por último, mostramos otros resultados, como el Teorema de Berry-Esseen y las Expansiones de Edgeworth, que serán necesarios en determinados puntos de las demostraciones.

El segundo capítulo está dedicado al método jackknife. Comenzaremos mostrando los conceptos que subyacen bajo este método de remuestreo, entre los que destacan las nociones de muestra de jackknife y pseudovalor de jackknife. Expondremos entonces dos ejemplos de estadísticos concretos en los que calcularemos las estimaciones de jackknife para el sesgo y la varianza. Por último, utilizando el concepto de consistencia, en el caso de estadísticos de la forma $g(\bar{X}_n)$, con \bar{X}_n la media muestral y $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y diferenciable en μ , con g' continua en μ y $g'(\mu) \neq 0$, demostraremos que la estimación dada por el método jackknife para la varianza del estadístico es realmente próxima a la varianza real del estadístico.

El tercer capítulo está dedicado al estudio del método bootstrap, al que dedicaremos una mayor atención, ya que no solo es capaz de estimar características de un estadístico dada una muestra, sino que también permite calcular intervalos de confianza de parámetros desconocidos de la distribución de esa muestra. Iniciaremos el capítulo con un resumen de la idea del método bootstrap, destacando cómo calcular los remuestreos a partir de la muestra original. A continuación obtenemos resultados de consistencia para la estimación de la distribución, centrándonos en estadísticos de la forma $\bar{X}_n - \mu$ y $g(\bar{X}_n) - g(\mu)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y diferenciable en μ , pasando después a la consistencia de la estimación de la varianza, usando los estadísticos $g(\bar{X}_n)$ como en el método jackknife. Por último, mostraremos la teoría referente al cálculo de intervalos de confianza bootstrap respecto a los parámetros desconocidos de la distribución de la muestra de partida. Los tipos de intervalos de confianza bootstrap que se explicarán serán el *bootstrap-t*, el *bootstrap percentil* y el

bootstrap híbrido. Es importante destacar que no son los únicos tipos de intervalos de confianza bootstrap que existen, pero por motivos de espacio y tiempo no se han podido incluir todos.

En el cuarto y último capítulo, mostraremos una aplicación de los resultados teóricos vistos en los anteriores capítulos a un caso en el que estudiaremos la consistencia de los estimadores jackknife y bootstrap. El caso particular se corresponde con tomar una muestra de partida de una distribución exponencial. Los estadísticos con los que se trabajará en este capítulo son, primero, la media muestral \bar{X}_n como estadístico, y segundo, el estadístico $e^{\bar{X}_n}$. El hecho de usar este último estadístico es porque se corresponde con uno del tipo $g(\bar{X}_n)$ como los estudiados en la teoría de los métodos jackknife y bootstrap. El objetivo de usar la distribución exponencial es que sabemos que la media muestral en este caso tiene como distribución exacta una distribución gamma y, por tanto, podremos comparar las estimaciones dadas por los métodos jackknife y bootstrap con los valores reales de las características de la distribución exacta del estadístico.

Abstract

The branch of the mathematics that tries to derive properties of a particular population from a sufficiently representative sample of that population is called Statistical Inference. This mechanism allows to solve the important issue of which probabilistic model adjusts to a random experiment, and, given the model, which are the parameters that characterise it.

The first step in any procedure of statistical inference is the extraction of observations from the variable or variables that are being studied. Each observation of the variable X produces one datum, and given n observations we have n data (x_1, \dots, x_n) , which are the ones that make up the sample. In the present project, we assume that these observations that come from the variable X are independent. To model the complete set of sample of size n we use an n -dimensional random variable (X_1, \dots, X_n) that satisfies that the variables X_i are independent and identically distributed, with the same distribution as X . Throughout this work, we will begin assuming that our samples are of this kind, which will be defined as simple random samples.

In any process of statistical inference, to be able to extract conclusions from the underlying population of the original sample we need to obtain some characteristics of the sample, like its mean or its variance, generally called sample mean and sample variance. This transformations of the simple random sample are known as statistics, and are of the form $h(X_1, \dots, X_n)$, being $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$ a Borel measurable function. These statistics are essential in three most important problems of the statistical inference: point estimation of a parameter of the population, confidence intervals around a parameter and the construction of statistical hypothesis tests.

Many statistical procedures require some knowledge of the sampling distribution of the statistic being used in the analysis. This type of knowledge depends on the type of the analysis. To construct confidence intervals and hypothesis tests it is needed some knowledge in the sampling distribution itself or about the percentiles of it. On the other hand, in an estimation problem it is paramount to have a measure of the error of the estimation given by the statistic (in this case called estimator), which is highly related to its variance and its bias. These measures are characteristics of the sampling distribution of the estimator.

Therefore, the ideal situation in these problems is to know the exact sampling distribution of the statistics, but that only occurs in very particular situations, like

in the case of Fisher's Theorem, which establishes that the sample mean of a simple random variable of a variable X with distribution $N(\mu, \sigma^2)$ follows a normal distribution $N\left(\mu, \frac{\sigma^2}{n}\right)$, and that the statistic $\frac{n-1}{n}S^2$ follows a χ_{n-1}^2 distribution, where S^2 is the cuasi-variance that will be introduced in Section 1.1 and χ_{n-1}^2 represents the chi-squared distribution with $n - 1$ degrees of liberty. Another example is the sample mean of the exponential distribution, which follows a gamma distribution.

Except from particular situations, in general there does not exist results to obtain exact distributions of statistics. An alternative is to work with the asymptotic distribution, that is, the limit distribution of the statistic. One of the main results of asymptotic distributions is the Central Limit Theorem, which, given a simple random sample (X_1, \dots, X_n) of a random variable X , establishes that

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow_d N(0, 1),$$

where $S_n = X_1 + \dots + X_n$, $\mu = E(X)$, $\sigma^2 = Var(X)$ and the symbol \rightarrow_d represents the convergence in distribution that will be defined in Section 1.2.

The results concerning asymptotic distributions represent a good strategy when the exact distribution of a statistic is not known. However, there are cases in which these kind of results of convergence in distribution cannot be applied, and it arises the necessity of estimating the sampling distribution of a statistic and its characteristics.

The principal alternative solving this issue by estimating the sampling distribution of a statistic along with its characteristics are the resampling methods. This Final Degree Project is focused on the study of some of them. In particular, we will study the methods called jackknife and bootstrap, which are considered to be the most important resampling techniques in statistical analysis, and have as their essential idea to estimate the characteristics of a statistic by obtaining samples repeatedly from the original data, and making inferences from those resamples.

The jackknife method produces new resamples by removing an observation of the original sample each time, creating n new samples of length $n - 1$, from which the estimation of the characteristics of the statistics are obtained. This method was proposed initially to reduce and estimate the bias of an estimator of a parameter of the population, but over the years it was proved that it can be also used to obtain an estimation of the variance of a statistic.

On the other hand, the bootstrap method does not remove observations to obtain the resamples, but instead it creates a high number of samples with replacement from the original sample. This method is especially important because of its capacity of estimating consistently the sampling distribution of a statistic, apart from the estimation its variance.

The main goal of this dissertation is to explain in detail both techniques, together with their main theoretical results, and afterwards to get those results into practice showing a practical application using the programming language R. To do this, we have gathered in four chapters, an importante group of results related to both techniques, which show their possible applications. Some of the proofs have been modified and sometimes extended or reduced, for the sake of making this dissertation more clearer and more coherent.

The dissertation consists of four chapters that we summarise down below:

In the first chapter, we introduce a number of definitions and results that will be useful throughout the whole work. We begin by defining general concepts that will be fundamental to prove the consistency of jackknife and bootstrap estimators, like the Continuous Mapping Theorem, the Central Limit Theorem and the Strong Law of the Large Numbers. Next, we introduce the metrics K and $\tilde{\rho}_r$, and we study some of their properties. Finally, we will show other results, such as the Berry-Esseen Theorem and Edgeworth's Expansions, that will be used at some particular points of the proofs.

The second chapter is devoted to the jackknife method. We will begin by explaining the concepts underlying the idea of this method, including the notion of jackknife sample and pseudo-value. We will show then two examples of concrete statistics in which we will obtain the jackknife estimations for their bias and their variance. Lastly, using the concept of consistency, we will prove that the estimation given by the jackknife method is really close to the real variance of the statistic, for statistics of the type $g(\bar{X}_n)$, with \bar{X}_n the sample mean, and $g : \mathbb{R} \rightarrow \mathbb{R}$ continuous and differentiable at μ , with g' continuous at μ and $g'(\mu) \neq 0$.

The third chapter is devoted to the study of the bootstrap method, to which we will devote more attention, because not only it is able to estimate the characteristics of the sampling distribution and the distribution itself of a statistic given a sample, but also it allows to obtain confidence intervals of unknown parameters of the sampling distribution. We will begin the chapter with a summary of the idea of the bootstrap method, focusing on how to obtain the resamples from the original sample. Next, we will obtain consistency results for the estimation of the sampling distribution of a statistic, concentrating on statistics of the type $\bar{X}_n - \mu$ y $g(\bar{X}_n) - g(\mu)$, with $g : \mathbb{R} \rightarrow \mathbb{R}$ continuous and differentiable at μ , passing then to the study of the consistency of the bootstrap variance estimation, using the statistics $g(\bar{X}_n)$ as in the jackknife method. Finally, we will show the theory regarding the interpretation of confidence intervals around unknown parameters of the sampling distribution of the original sample. The types of bootstrap confidence intervals that we will explain are the *bootstrap-t*, the *bootstrap percentile* and the *hybrid bootstrap*. It is important to underline that these are not the only types of bootstrap confidence intervals, but due to our limits in space and time we have not been able to include all of them.

In the fourth and last chapter, we will show the use of the theoretical results presented in the previous chapters with a practical application in which we will study, for a particular case, the consistency of the jackknife and bootstrap estimators. The particular case corresponds to considering a random sample from the exponential distribution. The statistics which we will work with in this chapter are, first, the sample mean \bar{X}_n , and second, the statistic $e^{\bar{X}_n}$. The fact of using this last statistic is because it belongs to one of the kind $g(\bar{X}_n)$ like the ones studied in the theory of the jackknife and the bootstrap methods. The objective of using the exponential distribution is that we know that the sample mean in this case follows a gamma distribution and, therefore, we will be able to compare the estimations given by the jackknife and the bootstrap methods with the real values of the characteristics of the exact sampling distribution of the statistic.

Capítulo 1

Preliminares y resultados básicos

Dedicaremos este primer capítulo a la introducción de una serie de conceptos y resultados, que serán de utilidad en el desarrollo de posteriores capítulos. Comenzaremos viendo conceptos generales de inferencia estadística que es necesario introducir para dar paso al resto de resultados, y que se usan a menudo en el trabajo. Los resultados se han obtenido de [5], [9] y [13].

1.1. Conceptos generales

Uno de los conceptos fundamentales de la inferencia estadística es el de muestra aleatoria simple. Este concepto modeliza todos los conjuntos de observaciones de tamaño fijo que podemos extraer de manera independiente de una variable aleatoria. Definimos primero este concepto.

Definición 1.1. *Sea X una variable aleatoria, se dice que el vector aleatorio de dimensión n , (X_1, \dots, X_n) es una muestra aleatoria simple (m.a.s.) de tamaño n de X si para cada variable X_j esta sigue la misma distribución que X y las variables X_j son independientes entre sí.*

A partir de la muestra aleatoria simple se construyen transformaciones de esta que son utilizadas en la inferencia estadística para la estimación de parámetros, la construcción de intervalos de confianza y el desarrollo de contrastes estadísticos. Estas transformaciones se conocen con el nombre de estadísticos y a continuación recordamos la definición de algunos de los más utilizados.

Definición 1.2. *Sea (X_1, \dots, X_n) una muestra aleatoria simple de una variable X .*

i) Se llama media muestral al estadístico $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Este estadístico tiene como propiedades principales

a) $E(\bar{X}_n) = E(X)$.

b) $Var(\bar{X}_n) = \frac{Var(X)}{n}$.

ii) Se conoce como varianza muestral al estadístico $s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, siendo $E(s^2) = \frac{n-1}{n} Var(X)$.

iii) Se llama *cuasi-varianza muestral* al estadístico $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, siendo $E(S^2) = \text{Var}(X)$.

Un concepto clave en el desarrollo del Capítulo 3 es el de función de distribución empírica asociada a una muestra. Lo definimos a continuación.

Definición 1.3. Sea (X_1, \dots, X_n) una muestra aleatoria simple de una variable aleatoria X , con distribución F . Fijado un valor $x \in \mathbb{R}$, se llama *función de distribución empírica en x al estadístico*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x)}(X_i).$$

Para la sección de intervalos de confianza de bootstrap, necesitamos introducir los conceptos de cuantil y cuantil muestral.

Definición 1.4. Sea F una función de distribución. Se define el *cuantil p -ésimo de F como*

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}.$$

Definición 1.5. Dada una muestra (X_1, \dots, X_n) de observaciones de una variable aleatoria X que sigue una distribución F se define, para $p > 0$, el *cuantil muestral p -ésimo como el cuantil p -ésimo de la función de distribución empírica F_n , es decir, $F_n^{-1}(p)$.*

1.2. Resultados de convergencia

En esta sección se muestran una lista de definiciones y resultados de convergencia que se usarán en los siguientes capítulos. Comenzaremos con las definiciones de convergencia en probabilidad, en distribución, casi segura y en momentos.

Definición 1.6. Sea una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$, se dice que

i) $\{X_n\}_{n \in \mathbb{N}}$ *converge casi seguramente a la variable aleatoria X si*

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1,$$

y denotaremos esta convergencia escribiendo $X_n \rightarrow_{c.s.} X$.

ii) $\{X_n\}_{n \in \mathbb{N}}$ *converge en probabilidad a la variable aleatoria X si*

$$\lim_n P(|X_n - X| < \varepsilon) = 1 \quad \forall \varepsilon > 0,$$

y denotaremos esta convergencia escribiendo $X_n \rightarrow_p X$.

iii) $\{X_n\}_{n \in \mathbb{N}}$ *converge en distribución a la variable aleatoria X si*

$$\lim_{n \rightarrow \infty} G_n(x) = F(x) \quad \forall x \in \text{Cont } F,$$

donde G_n y F son las funciones de distribución de X_n y X , respectivamente, y $\text{Cont } F$ es el conjunto de puntos donde F es continua.

iv) $\{X_n\}_{n \in \mathbb{N}}$ converge en momentos de orden r , con $r > 0$, a la variable aleatoria X si

$$\lim_{n \rightarrow \infty} E|X_n - X|^r = 0,$$

y denotaremos esta convergencia escribiendo $X_n \rightarrow_{r-m} X$.

Una vez vista esta definición, enunciaremos a continuación resultados generales de convergencia que se usarán a la largo del trabajo.

Teorema 1.7. (Teorema de las Transformaciones Continuas). Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias, sea X una variable aleatoria y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua. Entonces se cumple

- i) $X_n \rightarrow_{c.s.} X$ implica que $g(X_n) \rightarrow_{c.s.} g(X)$.
- ii) $X_n \rightarrow_p X$ implica que $g(X_n) \rightarrow_p g(X)$.
- iii) $X_n \rightarrow_d X$ implica que $g(X_n) \rightarrow_d g(X)$.

Teorema 1.8. (Teorema de Slutsky). Sean $\{X_n\}_{n \in \mathbb{N}}$ e $\{Y_n\}_{n \in \mathbb{N}}$ dos sucesiones de variables aleatorias, tal que $X_n \rightarrow_d X$ e $Y_n \rightarrow_p c$, siendo c una constante. Entonces se tiene que

- i) $X_n + Y_n \rightarrow_d X + c$.
- ii) $X_n Y_n \rightarrow_d Xc$.
- iii) $X_n/Y_n \rightarrow_d X/c$ si $c \neq 0$.

Teorema 1.9. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias y X una variable aleatoria tal que $X_n \rightarrow_d X$ y la secuencia $\{X_n^r\}_{n \in \mathbb{N}}$ es uniformemente integrable, donde $r > 0$. Entonces $E|X|^r < \infty$, $\lim_n E\{X_n^r\}_{n \in \mathbb{N}} = E(X^r)$ y $\lim_n E|X_n^r|_{n \in \mathbb{N}} = E|X^r|$.

A continuación definimos la definición de consistencia débil y fuerte de un estimador, que será útil para los resultados de consistencia de los capítulos 2 y 3.

Definición 1.10. Dada una sucesión de estimadores $\{\delta_n\}$ de un parámetro desconocido δ ,

1. Se dice que $\{\delta_n\}$ es débilmente consistente si, cuando $n \rightarrow \infty$,

$$\delta_n \rightarrow_p \delta,$$

para todos los posibles valores de δ .

1. Se dice que $\{\delta_n\}$ es fuertemente consistente si

$$\delta_n \rightarrow_{c.s.} \delta,$$

para todos los posibles valores de δ .

Pasamos ahora a ver una serie de resultados que se corresponden con propiedades de estadísticos particulares sobre los distintos tipos de convergencia vistos en la Definición 1.6.

Teorema 1.11. (*Teorema Central del Límite*). Sea $\{X_i\}_{i=1}^n$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E(X_j) = \mu$ y $\text{Var}(X_j) = \sigma^2$. Llamemos $S_n = X_1 + \cdots + X_n$. Entonces se cumple que

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow_d N(0, 1).$$

Teorema 1.12. (*Ley Fuerte de los Grandes Números*). Sea $\{X_i\}_{i=1}^n$ una sucesión de variables independientes e idénticamente distribuidas con media μ . Llamemos $S_n = X_1 + \cdots + X_n$. Entonces se tiene que

$$\frac{S_n}{n} \rightarrow_{c.s.} \mu.$$

Teorema 1.13. (*Ley Fuerte de los Grandes Números de Marcinkiewicz*). Sea $\{X_i\}_{i=1}^n$ una muestra aleatoria simple de una variable aleatoria X , tal que $E|X_1|^\lambda < \infty$ para $\lambda \in (0, 1)$, entonces

$$\frac{1}{n^{1/\lambda}} \sum_{i=1}^n |X_i| \rightarrow_{c.s.} 0.$$

Proposición 1.14. (*Método Delta*). Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias, sea $\{a_n\}_{n \in \mathbb{N}}$ una sucesión de números y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función. Si

$$a_n(X_n - \mu) \rightarrow_d N(0, 1),$$

para un valor μ , y g es diferenciable en μ , entonces

$$a_n(g(X_n) - g(\mu)) \rightarrow_d N(0, g'(\mu)^2).$$

Teorema 1.15. Sea $0 < p < 1$. Si $F^{-1}(p)$ es la solución única de $F(x-) \leq p \leq F(x)$, entonces $F_n^{-1}(p) \rightarrow_{c.s.} F^{-1}(p)$.

Antes de enunciar los siguientes resultados, hemos de definir previamente los dos conceptos en los que se basan: los momentos centrales y los momentos muestrales.

Definición 1.16. Sea X una variable aleatoria y sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes y con la misma distribución que X , que denotamos por F . Se denomina momento central de orden k de F a

$$\mu_k = E((X - \mu)^k),$$

siempre que dicha esperanza sea finita.

Definición 1.17. Sea X una variable aleatoria y sea $\{X_i\}_{i=1}^n$ una sucesión de variables aleatorias independientes y con la misma distribución que X . Sea F_n la función de distribución empírica asociada a la muestra. Se define el momento muestral de orden k de X como

$$m_k = \int_{-\infty}^{\infty} (x - \bar{X}_n)^k dF_n(x) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k.$$

Teorema 1.18. Sea X una variable aleatoria con distribución F y sea (X_1, \dots, X_n) una muestra aleatoria simple de X . Sean μ_k y m_k el momento central de orden k y el momento muestral de orden k , respectivamente. Entonces se tiene que m_k es fuertemente consistente para μ_k .

Teorema 1.19. Sea X una variable aleatoria y sea $\{X_i\}_{i=1}^n$ una sucesión de variables aleatorias independientes y con la misma distribución que X . Si $\mu_k < \infty$, entonces la variable $\sqrt{n}(m_k - \mu_k)$ converge en distribución a una distribución normal de media 0 y varianza $\sigma_{(k-1)(k-1)}$, donde

$$\sigma_{(k-1)(k-1)} = \mu_{2(k-1)+2} - \mu_k^2 - 2k\mu_{k-1}\mu_{k+1} + k^2\mu_{k-1}^2\mu_2.$$

Para concluir esta sección, vamos a introducir los conceptos de O_p , o_p , O y o , conceptos importantes cuando se quiere determinar el orden de convergencia del método bootstrap como estimación de la distribución de un estadístico.

Definición 1.20. Para dos sucesiones de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ e $\{Y_n\}_{n \in \mathbb{N}}$, se tiene que

$$X_n = O_p(Y_n),$$

si, para cualquier $\varepsilon > 0$, existe una constante $C_\varepsilon > 0$ tal que $P\{|X_n| > C_\varepsilon|Y_n|\} < \varepsilon$, para todo n , y

$$X_n = o_p(Y_n),$$

si $X_n/Y_n \rightarrow_p 0$.

Definición 1.21. Para dos sucesiones de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ e $\{Y_n\}_{n \in \mathbb{N}}$, se tiene que

$$X_n = O(Y_n),$$

si $|X_n| \leq c|Y_n| \forall n$ y una constante c , y

$$X_n = o(Y_n),$$

si $X_n/Y_n \rightarrow 0$ cuando $n \rightarrow \infty$.

1.3. Resultados sobre métricas

En el Capítulo 3, usamos en los resultados del estimador bootstrap de la distribución de un estadístico dos métricas, K y $\tilde{\rho}_r$. Estas métricas están definidas sobre el espacio de distribuciones $\mathcal{F}_{\mathbb{R}^s}$, que definimos a continuación.

Definición 1.22. Sea $s \in \mathbb{N}$. Definimos $\mathcal{F}_{\mathbb{R}^s}$ como el espacio de todas las distribuciones en \mathbb{R}^s . Además, dado $r \in \mathbb{N}$, definimos $\mathcal{F}_{r,s}$ como $\mathcal{F}_{r,s} = \{G \in \mathcal{F}_{\mathbb{R}^s} : \int \|x\|^r dG(x) < \infty\}$

Definición 1.23. Sean F y G dos distribuciones de probabilidad en $\mathcal{F}_{\mathbb{R}}$. Se define la métrica de Kolmogorov como

$$K(F, G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|.$$

Definición 1.24. Sean F y G dos distribuciones de probabilidad en $\mathcal{F}_{\mathbb{R}}$. Se define la distancia de Mallows entre F y G como

$$\tilde{\rho}_r(F, G) = \inf_{\mathcal{T}_{X,Y}} (E|X - Y|^r)^{1/r},$$

donde $\mathcal{T}_{X,Y}$ es la colección de todas las posibles distribuciones conjuntas de los pares (X, Y) cuyas distribuciones marginales son F y G , respectivamente. Además, para dos variables aleatorias U y V con distribuciones F y G respectivamente, se define $\tilde{\rho}_r(U, V) = \tilde{\rho}_r(F, G)$.

Una vez vistas ambas métricas, enunciamos para la métrica $\tilde{\rho}_r$ una serie de propiedades que será necesario usar en la demostración de uno de los teoremas del Capítulo 3.

Proposición 1.25. Sea (X_1, \dots, X_n) una muestra aleatoria simple con función de distribución F en el espacio muestral $\mathcal{F}_{r,s}$. Sea F_n la función de distribución empírica asociada a (X_1, \dots, X_n) . Entonces se tiene

$$\tilde{\rho}_r(F_n, F) \rightarrow_{a.s.} 0.$$

Proposición 1.26. Sean U y V variables aleatorias y sea a una constante. Entonces,

$$\tilde{\rho}_r(aU, aV) = |a| \tilde{\rho}_r(U, V).$$

Proposición 1.27. Sean U y V variables aleatorias. Si $E|U|^2 < \infty$ y $E|V|^2 < \infty$, entonces

$$[\tilde{\rho}_2(U, V)]^2 = [\tilde{\rho}_2(U - E(U), V - E(V))]^2 + |E(U) - E(V)|^2.$$

Proposición 1.28. Sean $\{U_j\}_{j \in \mathbb{N}}$ y $\{V_j\}_{j \in \mathbb{N}}$ dos sucesiones de variables aleatorias cuyas distribuciones están en $\mathcal{F}_{r,s}$ y $E(U_j) = E(V_j) \quad \forall j$. Entonces,

$$\left[\tilde{\rho}_2 \left(\sum_{i=1}^m U_j, \sum_{i=1}^m V_j \right) \right]^2 \leq \sum_{i=1}^m [\tilde{\rho}_2(U_j, V_j)]^2.$$

1.4. Otros resultados

En esta sección se enuncian varios resultados que, aunque no son de convergencia, se necesitan en ciertos puntos de las demostraciones de los capítulos 2 y 3.

Teorema 1.29. (Teorema de Berry-Esseen). Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias con $E(X_i) = 0$, $Var(X_i) = \sigma_i^2 < \infty$, y $V_n^2 = \sum_{i=1}^n \sigma_i^2$, entonces

$$\sup_x \left| P \left\{ \frac{1}{V_n} \sum_{i=1}^n X_i \leq x \right\} - \Phi(x) \right| \leq \frac{c}{V_n^3} \sum_{i=1}^n E|X_i|^3,$$

donde Φ representa la distribución $N(0, 1)$.

Teorema 1.30. (Teorema de Polya). Sea $\{X_n\}_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias con distribución G_n y X es una variable aleatoria con distribución F . Si $G_n \rightarrow F$ puntualmente, se tiene que

$$\lim_{n \rightarrow \infty} \sup_x |G_n(x) - F(x)| = 0.$$

Definición 1.31. (Expansión de Edgeworth). Sea (X_1, \dots, X_n) una muestra aleatoria simple de una variable aleatoria X y sea $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ un estadístico, donde $\mu = E(X)$ y $\sigma^2 = \text{Var}(X)$. Se define la Expansión de Edgeworth de dos términos de T_n como

$$P(T_n \leq x) = \Phi(x) + \frac{c_1(F)p_1(x)}{\sqrt{n}}\varphi(x) + \frac{c_2(F)p_2(x)}{n}\varphi(x) + O(1/n),$$

donde

$$\begin{aligned} p_1(x) &= 1 - x^2, & p_2(x) &= 3x - x^3, \\ c_1(F) &= \frac{\gamma}{6}, & c_2(F) &= \frac{\kappa - 3}{24}. \end{aligned}$$

Capítulo 2

Método jackknife

2.1. Introducción

En el campo de la estimación puntual, cuando se busca un estadístico que aproxime lo máximo posible el valor real de un parámetro asociado a una variable aleatoria, una de las características deseables para ese estadístico es que sea insesgado. Aunque algunas veces es fácil de obtener un estimador de este tipo, por ejemplo por el método de máxima verosimilitud, lo cierto es que en general hay un gran número de estimadores que tienen sesgo.

Maurice Quenouille, en 1949 [7], introdujo una técnica para estimar y reducir el sesgo de un estimador. En particular, dado un estimador T_n que satisface

$$E[T_n] = \theta + \frac{a_1}{n} + O(1/n^2),$$

Quenouille desarrolló un método para mejorar la estimación eliminando los términos de orden $\frac{1}{n}$ de la expresión anterior.

Nueve años más tarde, John Tukey [11] extendió el uso del método, demostrando que se podía usar también para estimar la varianza de un estimador, y fue quien dio nombre a la técnica, llamándola el método jackknife. La traducción literal de jackknife es navaja multiusos. Tukey le dio este nombre al método porque puede ser usado como una herramienta “rápida y fácil” cuando no se dispone de otras técnicas más sofisticadas.

Este capítulo está dedicado al estudio del método jackknife como reductor del sesgo y estimador del sesgo y la varianza. La Sección 2.2 está dedicada a la introducción de los conceptos principales que se usan en la estimación jackknife, dando dos aplicaciones. Por otro lado, en la Sección 2.3, estudiaremos la propiedad teórica básica que se exige a los estimadores jackknife: la consistencia. Esta característica de los estimadores es crucial cuando usamos muestras suficientemente grandes, e indica cómo de bien aproxima el estimador jackknife el valor real del parámetro. Los resultados se han obtenido de [4], [6] y [10].

2.2. El método jackknife

El método de remuestreo jackknife se encuentra dentro del marco de la estimación puntual. Por ello, partiremos de una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria unidimensional X con distribución desconocida F y nuestro objetivo será, dado un parámetro $\theta \in \mathbb{R}$ de la distribución y un estimador T_n del parámetro θ , obtener información de este estimador del parámetro.

En el problema de la estimación puntual, en muchos casos se tienen estimadores que no son insesgados, siendo frecuentes los estimadores que verifican

$$E[T_n] = \theta + \frac{a_1}{n} + O(1/n^2) \quad \forall n \in \mathbb{N}, \quad (2.1)$$

donde el sesgo del estimador es $\frac{a_1}{n} + O(1/n^2)$, y el término $O(1/n^2)$ viene dado en la Definición 1.21. A lo largo de este capítulo supondremos que los estadísticos que estamos tratando cumplen esta propiedad.

El método jackknife da la posibilidad de estimar el sesgo a la vez que elimina el término $\frac{a_1}{n}$ de (2.1). En el Ejemplo 2.8 estudiaremos en detalle esta propiedad para un estadístico particular. Para cumplir su objetivo, el método jackknife se basa en la creación de n muestras, llamadas muestras jackknife, que pasamos a definir.

Definición 2.1. *Dada una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria X y dado $i \in \{1, \dots, n\}$, se define la muestra i -ésima de jackknife como la muestra formada eliminando la observación i -ésima de la muestra original:*

$$X_{[i]} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n).$$

Sea ahora $T_n = T_n(X_1, \dots, X_n)$ un estimador de un parámetro desconocido θ y sean

$$T_{n-1,i} = T_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \quad (2.2)$$

los n estadísticos basados en las muestras jackknife.

Conocidas las muestras $X_{[i]}$ y los valores $T_{n-1,i}$, se definen a continuación los pseudovalores de jackknife.

Definición 2.2. *Sea X una variable aleatoria, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Sea $T_{n-1,i}$ el estadístico basado en la muestra i -ésima de jackknife $\forall i \in \{1, \dots, n\}$. Se define el pseudovalor de jackknife i -ésimo como*

$$\tilde{T}_{n,i} = nT_n - (n-1)T_{n-1,i}, \quad (2.3)$$

que puede interpretarse como una medida de la diferencia entre la estimación de los parámetros obtenida con la muestra total y la estimación obtenida eliminando la observación i -ésima.

La media muestral de estos valores la representaremos por \tilde{T}_n , y su valor viene dado por

$$\tilde{T}_n = \frac{1}{n} \sum_{i=1}^n \tilde{T}_{n,i} = nT_n - (n-1) \frac{1}{n} \sum_{i=1}^n T_{n-1,i}.$$

Este último estadístico tiene la interesante propiedad de que elimina el término a_1/n del sesgo en (2.1). En efecto, si calculamos su esperanza tenemos

$$\begin{aligned} E[\tilde{T}_n] &= nE[T_n] - (n-1) \frac{1}{n} \sum_{i=1}^n E[T_{n-1,i}] \\ &= n \left(\theta + \frac{a_1}{n} + O(1/n^2) \right) - \frac{n-1}{n} \left(n\theta + \frac{na_1}{(n-1)} + O(1/n^2) \right) \\ &= \theta + O(1/n^2), \end{aligned}$$

donde la segunda igualdad se sigue del hecho de que $T_{n-1,i}$ es el estimador basado en una muestra de $n-1$.

Podemos, por tanto, considerar como un estimador del sesgo del parámetro T_n la diferencia entre T_n y \tilde{T}_n , ya que será un estadístico cuya esperanza estará alrededor de a_1/n . La expresión para el estimador del sesgo queda entonces como

$$T_n - \tilde{T}_n = \frac{n-1}{n} \sum_{i=1}^n T_{n-1,i} - (n-1)T_n = (n-1)(\bar{T}_n - T_n),$$

donde $\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_{n-1,i}$.

De un modo más formal, se define a continuación el estimador jackknife del sesgo de T_n .

Definición 2.3. *Sea X una variable aleatoria, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Sea $T_{n-1,i}$ el estadístico basado en la muestra i -ésima de jackknife, dado por (2.2), $\forall i \in \{1, \dots, n\}$. Se define el estimador jackknife del sesgo de T_n como*

$$b_{JACK} = (n-1)(\bar{T}_n - T_n),$$

donde $\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_{n-1,i}$.

El desarrollo anterior que conduce al estimador jackknife del sesgo, también permite obtener el estimador del propio parámetro θ , el cual no es otro que \tilde{T}_n , al ser un estadístico cuya esperanza es muy próxima θ . Este valor lo denotamos a partir de ahora como T_{JACK} .

Definición 2.4. *Sea X una variable aleatoria, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Se define el estimador jackknife de θ como*

$$T_{JACK} = T_n - b_{JACK} = nT_n - (n-1)\bar{T}_n. \quad (2.4)$$

Estudiemos ahora la estimación de la varianza que da el método jackknife. Para ello volveremos a usar los pseudovalores de jackknife definidos en (2.3). Tukey conjeturó que estos pseudovalores son independientes e idénticamente distribuidos, y que cada uno de ellos tiene aproximadamente la misma varianza que $\sqrt{n}T_n$.

A partir de estas suposiciones y como la cuasi-varianza muestral es un estimador insesgado de la varianza, si tenemos los valores $(\tilde{T}_{n,1}, \dots, \tilde{T}_{n,n})$, podemos estimar la varianza de $\sqrt{n}T_n$ usando la cuasi-varianza muestral de estos valores, es decir, podemos aproximar $Var(\sqrt{n}T_n)$ mediante

$$\frac{1}{n-1} \sum_{i=1}^n \left(\tilde{T}_{n,i} - \frac{1}{n} \sum_{j=1}^n \tilde{T}_{n,j} \right)^2.$$

Por tanto, esto lleva a estimar $Var(T_n)$ mediante la cuasi-varianza muestral basada en $\frac{\tilde{T}_{n,i}}{\sqrt{n}}$, es decir, estimamos $Var(T_n)$ mediante

$$\frac{1}{n-1} \sum_{i=1}^n \left(\frac{\tilde{T}_{n,i}}{\sqrt{n}} - \frac{1}{n} \sum_{j=1}^n \frac{\tilde{T}_{n,j}}{\sqrt{n}} \right)^2 = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\tilde{T}_{n,i} - \frac{1}{n} \sum_{j=1}^n \tilde{T}_{n,j} \right)^2. \quad (2.5)$$

Sustituyendo los pseudovalores de la expresión (2.3) en (2.5) se tiene que la expresión anterior se puede escribir en la forma

$$\begin{aligned} & \frac{1}{n(n-1)} \sum_{i=1}^n \left(nT_n - (n-1)T_{n-1,i} - \frac{1}{n} \sum_{j=1}^n (nT_n - (n-1)T_{n-1,j}) \right)^2 \\ &= \frac{n-1}{n} \sum_{i=1}^n \left(T_{n-1,i} - \frac{1}{n} \sum_{j=1}^n T_{n-1,j} \right)^2. \end{aligned}$$

Definimos ahora formalmente este estimador de la varianza de T_n .

Definición 2.5. Sea X una variable aleatoria, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Se define el estimador jackknife de la varianza de T_n como

$$v_{JACK} = \frac{n-1}{n} \sum_{i=1}^n \left(T_{n-1,i} - \frac{1}{n} \sum_{j=1}^n T_{n-1,j} \right)^2. \quad (2.6)$$

Como vemos, el hecho de submuestrear usando el método jackknife da una serie de recursos muy simples para estimar el sesgo y la varianza del estimador de un parámetro, los cuales normalmente son muy difíciles de obtener. El recurso principal son los pseudovalores de jackknife, que se han usado en el desarrollo de ambas estimaciones.

Una vez vista la idea y las definiciones del método jackknife, y antes de continuar, hemos de realizar una observación importante que se usará desde este momento en el resto del trabajo.

Observación 2.6. *En los ejemplos y resultados que siguen se suele usar a menudo igualdades y desigualdades entre variables aleatorias, por lo que son igualdades y desigualdades casi seguras. Para evitar repetirlo continuamente, se omitirá esta indicación en los resultados.*

Pasamos a continuación a dar dos ejemplos de estimación del sesgo y la varianza para dos estadísticos concretos de bastante interés: la media muestral y la media muestral al cuadrado. El primero de ellos representa un caso en el que el sesgo del estimador es 0, mientras que el segundo se corresponde con el tipo de estadísticos que cumplen (2.1), al ser el sesgo del estimador de la forma a_1/n . Veremos en este último caso cómo el estimador T_{JACK} elimina el sesgo del estimador.

Ejemplo 2.7. *Para la media muestral $T_n = \bar{X}_n$, el estadístico basado en las $n - 1$ observaciones $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ se puede expresar como*

$$T_{n-1,i} = \frac{\sum_{j \neq i} X_j}{n-1} = \frac{\sum_{j=1}^n X_j - X_i}{n-1} = \frac{n\bar{X}_n - X_i}{n-1}, \quad (2.7)$$

y podemos obtener fácilmente el valor de \bar{T}_n como

$$\bar{T}_n = \frac{\sum_{i=1}^n T_{n-1,i}}{n} = \frac{\sum_{i=1}^n (n\bar{X}_n - X_i)}{n(n-1)} = \frac{n\bar{X}_n}{n-1} - \frac{\bar{X}_n}{n-1} = \bar{X}_n,$$

por lo que el estimador del sesgo de jackknife viene dado por

$$b_{JACK} = (n-1)(\bar{T}_n - T_n) = (n-1)(\bar{X}_n - \bar{X}_n) = 0.$$

Por otro lado, de (2.3) los pseudovalores de jackknife vienen dados por

$$\tilde{T}_{n,i} = nT_n - (n-1)T_{n-1,i} = n\bar{X}_n - (n\bar{X}_n - X_i) = X_i \quad \forall i \in \{1, \dots, n\},$$

y, al ser los X_i independientes e idénticamente distribuidos, se tiene que los pseudovalores $\tilde{T}_{n,i}$ también lo son, luego la primera conjetura de Tukey se cumple.

Para la segunda conjetura, sabemos, por la Definición 1.2, que se tiene

$$\text{Var}(\sqrt{n}\bar{X}_n) = \text{Var}(X_i),$$

y, como $\tilde{T}_{n,i} = X_i$, se cumple que $\text{Var}(\tilde{T}_{n,i}) = \text{Var}(\sqrt{n}\bar{X}_n)$, luego también se cumple la segunda conjetura de Tukey.

El estimador de la varianza de jackknife dado en (2.6) sería

$$\begin{aligned} v_{JACK} &= \frac{n-1}{n} \sum_{i=1}^n \left(T_{n-1,i} - \frac{1}{n} \sum_{j=1}^n T_{n-1,j} \right)^2 = \frac{n-1}{n} \sum_{i=1}^n \left(\frac{n\bar{X}_n - X_i}{n-1} - \bar{X}_n \right)^2 \\ &= \frac{n-1}{n} \sum_{i=1}^n \left(\frac{\bar{X}_n - X_i}{n-1} \right)^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \end{aligned}$$

donde la segunda igualdad viene por (2.7), y la tercera y la cuarta resultan de transformar la segunda.

Ejemplo 2.8. Tomamos ahora el estadístico $T_n = \overline{X}_n^2$ como estimador de μ^2 . Veamos primero la estimación del sesgo para este estadístico.

Por definición, sabemos que el estimador del sesgo de \overline{X}_n^2 viene dado por

$$b_{JACK} = (n-1) \left(\frac{1}{n} \sum_{i=1}^n \overline{X}_{n-1,i}^2 - \overline{X}_n^2 \right) = \frac{n-1}{n} \sum_{i=1}^n (\overline{X}_{n-1,i}^2 - \overline{X}_n^2). \quad (2.8)$$

Sustituyendo $\overline{X}_{n-1,i}$ en (2.7) y operando se tiene que

$$\begin{aligned} b_{JACK} &= \frac{n-1}{n} \sum_{i=1}^n (\overline{X}_{n-1,i} - \overline{X}_n) (\overline{X}_{n-1,i} + \overline{X}_n) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n (\overline{X}_n - X_i) [2(n-1)\overline{X}_n + (\overline{X}_n - X_i)] \\ &= \frac{1}{n(n-1)} \left[2(n-1) \left(n\overline{X}_n^2 - \sum_{i=1}^n X_i \overline{X}_n \right) + \sum_{i=1}^n (\overline{X}_n - X_i)^2 \right]. \end{aligned} \quad (2.9)$$

Por tanto,

$$b_{JACK} = \frac{\hat{\alpha}_2}{n}, \quad (2.10)$$

donde $\hat{\alpha}_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ y, en general, para $k > 0$, llamaremos $\hat{\alpha}_k$ a

$$\hat{\alpha}_k = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^k. \quad (2.11)$$

Luego el estimador jackknife del parámetro en cuestión será

$$T_{JACK} = \overline{X}_n^2 - \frac{\hat{\alpha}_2}{n}.$$

Puede verse que el sesgo es de la forma a_1/n , donde $a_1 = -\hat{\alpha}_2$. Si calculamos la esperanza de T_{JACK} podemos ver como elimina ese sesgo y es igual al parámetro. En efecto

$$E(T_{JACK}) = E(\overline{X}_n^2) - \frac{\sigma^2}{n} = \text{Var}(\overline{X}_n) + E(\overline{X}_n)^2 - \frac{\sigma^2}{n} = \frac{\sigma^2}{n} + \mu^2 - \frac{\sigma^2}{n} = \mu^2.$$

Vamos a calcular ahora la estimación de la varianza. Para ello, de (2.4), se tiene que

$$T_{JACK} = n\overline{X}_n^2 - \frac{n-1}{n} \sum_{i=1}^n \overline{X}_{n-1,i}^2,$$

de donde

$$\frac{1}{n} \sum_{i=1}^n \overline{X}_{n-1,i}^2 = \frac{n\overline{X}_n^2 - T_{JACK}}{n-1} = \frac{n\overline{X}_n^2 - \overline{X}_n^2 + \frac{\hat{\alpha}_2}{n}}{n-1} = \overline{X}_n^2 + \frac{\hat{\alpha}_2}{n(n-1)},$$

y entonces, de (2.6), para $T_n = \bar{X}_n^2$ se tiene que la varianza de jackknife viene dada por

$$v_{JACK} = \frac{n-1}{n} \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 - \frac{\hat{\alpha}_2}{n(n-1)} \right)^2.$$

Desarrollando el cuadrado y operando obtenemos

$$v_{JACK} = \frac{n-1}{n} \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 \right)^2 + \frac{\hat{\alpha}_2^2}{n^2(n-1)} - \frac{2\hat{\alpha}_2}{n^2} \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 \right). \quad (2.12)$$

Como de (2.8) y (2.10) se tiene

$$\hat{\alpha}_2^2 = (n-1) \left(\sum_{i=1}^n \bar{X}_{n-1,i}^2 - n\bar{X}_n^2 \right) = (n-1) \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 \right),$$

sustituyendo en (2.12) queda

$$v_{JACK} = \frac{n-1}{n} \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 \right)^2 - \frac{\hat{\alpha}_2^2}{n^2(n-1)}.$$

Siguiendo ahora el mismo procedimiento que en (2.9), se tiene que

$$\begin{aligned} v_{JACK} &= \frac{n-1}{n} \sum_{i=1}^n \left(\bar{X}_{n-1,i}^2 - \bar{X}_n^2 \right)^2 - \frac{\hat{\alpha}_2^2}{n^2(n-1)} \\ &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \left(\bar{X}_n - X_i \right)^2 \left[2(n-1)\bar{X}_n + (\bar{X}_n - X_i) \right]^2 - \frac{\hat{\alpha}_2^2}{n^2(n-1)}. \end{aligned}$$

Desarrollando el segundo de los cuadrados de la expresión anterior y operando se obtiene

$$\begin{aligned} v_{JACK} &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \left(\bar{X}_n - X_i \right)^4 + \frac{4\bar{X}_n^2}{n(n-1)} \sum_{i=1}^n \left(\bar{X}_n - X_i \right)^2 \\ &\quad + \frac{4\bar{X}_n}{n(n-1)^2} \sum_{i=1}^n \left(\bar{X}_n - X_i \right)^3 + \frac{\hat{\alpha}_2^2}{n^2(n-1)}. \end{aligned}$$

Por último, usando (2.11), la expresión final para el estimador jackknife de la varianza queda como

$$v_{JACK} = \frac{\hat{\alpha}_4}{n(n-1)^2} + \frac{4\bar{X}_n^2\hat{\alpha}_2}{n} - \frac{4\bar{X}_n\hat{\alpha}_3}{n(n-1)} - \frac{\hat{\alpha}_2^2}{n^2(n-1)}.$$

2.3. Consistencia

Como no tenemos estimadores insesgados de la varianza del estimador desconocido T_n , veremos cuándo nuestro estimador v_{JACK} es fuertemente consistente, según

la Definición 1.10.

Sin embargo, esa definición debe ser modificada cuando estamos estimando la varianza de un estimador T_n , ya que $Var(T_n)$ depende de n y normalmente $Var(T_n) \rightarrow 0$ cuando $n \rightarrow \infty$. La variación de la Definición 1.10 que vamos a usar en esta sección, sobre la consistencia fuerte de un estimador v_n de $Var(T_n)$, es

$$\frac{v_n}{Var(T_n)} \rightarrow_{c.s.} 1.$$

El resultado equivalente para la consistencia débil se corresponde con la convergencia $\frac{v_n}{Var(T_n)} \rightarrow_p 1$.

A continuación, se trata el caso en el que el estadístico es de la forma $T_n = g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y derivable en μ y con g' continua en μ , ya que estos son estadísticos muy usuales y la demostración de la consistencia en este caso se puede extrapolar a casos más generales de estadísticos más complejos.

Sea $\mu = E(X)$ y $\sigma^2 = Var(X)$. Por el Teorema Central del Límite dado en 1.11, sabemos

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow_d N(0, 1).$$

Por otro lado, como g es diferenciable en μ , por la Proposición 1.14 tenemos

$$\frac{g(\bar{X}_n) - g(\mu)}{\sigma/\sqrt{n}} \rightarrow_d N(0, g'(\mu)^2),$$

y por tanto obtenemos que

$$\frac{T_n - g(\mu)}{\sigma_n} \rightarrow_d N(0, 1), \tag{2.13}$$

siendo

$$\sigma_n^2 = \frac{1}{n} \sigma^2 g'(\mu)^2, \tag{2.14}$$

donde σ_n^2 es la varianza asintótica de $Var(T_n)$.

En el siguiente teorema estudiamos la consistencia fuerte del estimador de la varianza de jackknife de $T_n = g(\bar{X}_n)$.

Teorema 2.9. *Sea X una variable aleatoria, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = g(\bar{X}_n)$. Supongamos que σ^2 existe y que la derivada de g está definida en un entorno de μ , con $g'(\mu) \neq 0$, y siendo g' continua en μ . Entonces, el estimador jackknife de la varianza de T_n es fuertemente consistente, es decir*

$$\frac{v_{JACK}}{\sigma_n^2} \rightarrow_{c.s.} 1, \tag{2.15}$$

donde σ_n^2 viene dado por (2.14).

Demostración. Sea $\bar{X}_{n-1,i} = \frac{\sum_{j \neq i} X_j}{n-1}$ la media muestral basada en todas las observaciones de la muestra inicial excepto X_i . Por el Teorema del Valor Medio, tenemos

$$\begin{aligned} T_{n-1,i} - T_n &= g(\bar{X}_{n-1,i}) - g(\bar{X}_n) = g'(\xi_{n,i})(\bar{X}_{n-1,i} - \bar{X}_n) \\ &= g'(\bar{X}_n)(\bar{X}_{n-1,i} - \bar{X}_n) + R_{n,i}, \end{aligned} \quad (2.16)$$

siendo

$$R_{n,i} = (g'(\xi_{n,i}) - g'(\bar{X}_n))(\bar{X}_{n-1,i} - \bar{X}_n), \quad (2.17)$$

y $\xi_{n,i}$ un valor aleatorio intermedio entre $\bar{X}_{n-1,i}$ y \bar{X}_n .

Para estudiar la consistencia de v_{JACK} vamos a realizar en primer lugar unos cálculos previos que ayudarán a simplificar el estudio de la convergencia dada en (2.15). Sabemos, por (2.7), que

$$\bar{X}_{n-1,i} = \frac{n\bar{X}_n - X_i}{n-1},$$

y por tanto,

$$\bar{X}_{n-1,i} - \bar{X}_n = \frac{\bar{X}_n - X_i}{n-1}, \quad (2.18)$$

de lo que se sigue que

$$\sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n) = \sum_{i=1}^n \frac{\bar{X}_n - X_i}{n-1} = \frac{n\bar{X}_n - \sum_{i=1}^n X_i}{n-1} = \frac{n\bar{X}_n - n\bar{X}_n}{n-1} = 0. \quad (2.19)$$

Vamos a calcular ahora a partir de (2.6) el estimador de la varianza de jackknife de T_n . De (2.6) y (2.16) se tiene

$$\begin{aligned} v_{JACK} &= \frac{n-1}{n} \sum_{i=1}^n (T_{n-1,i} - \frac{1}{n} \sum_{j=1}^n T_{n-1,j})^2 = \frac{n-1}{n} \sum_{i=1}^n \left[(g'(\bar{X}_n)(\bar{X}_{n-1,i} - \bar{X}_n) \right. \\ &\quad \left. + R_{n,i} + T_n) - \frac{1}{n} \sum_{j=1}^n (g'(\bar{X}_n)(\bar{X}_{n-1,j} - \bar{X}_n) + R_{n,j} + T_n) \right]^2. \end{aligned}$$

Simplificando la expresión anterior y usando (2.19), obtenemos

$$v_{JACK} = \frac{n-1}{n} \sum_{i=1}^n [g'(\bar{X}_n)(\bar{X}_{n-1,i} - \bar{X}_n) + R_{n,i} - \bar{R}_n]^2,$$

donde $\bar{R}_n = \sum_{i=1}^n R_{n,i}$.

Entonces podemos descomponer v_{JACK} como la suma

$$v_{JACK} = A_n + B_n + C_n,$$

siendo

$$A_n = \frac{n-1}{n} g'(\bar{X}_n)^2 \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2,$$

$$B_n = \frac{n-1}{n} \sum_{i=1}^n (R_{n,i} - \bar{R}_n)^2,$$

$$C_n = \frac{n-1}{n} g'(\bar{X}_n) \sum_{i=1}^n (R_{n,i} - \bar{R}_n)(\bar{X}_{n-1,i} - \bar{X}_n).$$

Para demostrar el teorema, vamos a que probar que

$$i) \quad \frac{A_n}{\sigma_n^2}, \rightarrow_{c.s.} 1 \quad (2.20)$$

y

$$ii) \quad \frac{B_n}{\sigma_n^2}, \rightarrow_{c.s.} 0 \quad (2.21)$$

ya que por la Desigualdad de Cauchy-Schwarz, se tiene que $C_n^2 \leq A_n B_n$, y de lo anterior obtendremos $C_n/\sigma_n^2 \rightarrow_{c.s.} 0$, y por tanto $v_{JACK}/\sigma_n^2 \rightarrow_{c.s.} 1$.

Para probar (2.20), desarrollando A_n y usando (2.18) se tiene

$$A_n = \frac{1}{n(n-1)} g'(\bar{X}_n)^2 \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

pero de la Proposición 1.14, por la continuidad de g' en μ y por $\bar{X}_n \rightarrow_{c.s.} \mu$ se tiene $g'(\bar{X}_n) \rightarrow_{c.s.} g'(\mu)$, y por el Teorema 1.18, sabemos que $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \rightarrow_{c.s.} \sigma^2$, lo que implica que

$$\frac{A_n}{\sigma_n^2} = \frac{\frac{1}{n(n-1)} g'(\bar{X}_n)^2 \sum_{i=1}^n (X_i - \bar{X}_n)^2}{\frac{1}{n} g'(\mu)^2 \sigma^2} \rightarrow_{c.s.} \frac{g'(\mu)^2 \sigma^2}{g'(\mu)^2 \sigma^2} = 1,$$

lo que prueba (2.20).

Veamos ahora (2.21). En primer lugar, tenemos que por (2.18) se verifica que

$$(n-1) \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (2.22)$$

y de esta igualdad se tiene que $(n-1) \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2 = S^2$, que converge casi seguramente a σ^2 , como consecuencia directa del Teorema 1.18 y la Definición 1.2.

Por tanto, de (2.22) se tiene que

$$\sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2 \rightarrow_{c.s.} \frac{\sigma^2}{n-1} \rightarrow_{c.s.} 0,$$

y cada término de la suma anterior, al ser todos positivos, tiende a 0, y en particular el máximo de ellos:

$$\max_{i \leq n} (\bar{X}_{n-1,i} - \bar{X}_n)^2 \rightarrow_{c.s.} 0.$$

Por (2.22), por la continuidad de g' en μ y por tenerse $|\xi_{n,i} - \bar{X}_n| \leq |\bar{X}_{n-1,i} - \bar{X}_n|$ se sigue que

$$u_n = \max_{i \leq n} (g'(\xi_{n,i}) - g'(\bar{X}_n))^2 \rightarrow_{c.s.} 0.$$

Por otro lado, para B_n se cumple la desigualdad

$$\frac{B_n}{\sigma_n^2} \leq \frac{n-1}{\sigma_n^2 n} \sum_{i=1}^n R_{n,i}^2,$$

ya que $\frac{n-1}{n} \sum_{i=1}^n (R_{n,i} - \bar{R}_n)^2 \leq \frac{n-1}{n} \sum_{i=1}^n R_{n,i}^2$, y por (2.17) se tiene

$$\frac{n-1}{\sigma_n^2 n} \sum_{i=1}^n R_{n,i}^2 \leq \frac{u_n}{\sigma_n^2} \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2,$$

luego

$$\frac{B_n}{\sigma_n^2} \leq \frac{u_n}{\sigma_n^2} \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2.$$

Ahora, por (2.14) y (2.22), obtenemos que

$$\frac{\sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2}{\sigma_n^2} \rightarrow_{c.s.} \frac{n\sigma^2}{(n-1)g'(\mu)^2\sigma^2} \rightarrow_{c.s.} \frac{\sigma^2}{g'(\mu)^2\sigma^2} = \frac{1}{g'(\mu)},$$

luego se tiene que

$$\frac{u_n}{\sigma_n^2} \sum_{i=1}^n (\bar{X}_{n-1,i} - \bar{X}_n)^2 \rightarrow_{c.s.} 0,$$

lo que implica que $B_n/\sigma_n^2 \rightarrow_{c.s.} 0$, es decir, hemos probado (2.21), que completa la demostración. \square

En el siguiente ejemplo, vamos a probar de una forma alternativa que para el estimador planteado en el Ejemplo 2.8 se cumple la tesis del teorema que acabamos de probar.

Ejemplo 2.8 (continuación). *La función de la media muestral que se considera en este ejemplo es $g(\bar{X}_n) = \bar{X}_n^2$, continua y derivable, por lo que cumple las condiciones del Teorema 2.9, siempre y cuando $g'(\mu) \neq 0$, y por tanto*

$$\frac{v_{JACK}}{\sigma_n^2} \rightarrow_{c.s.} 1,$$

siendo, en este caso

$$\sigma_n^2 = \frac{4\mu^2\sigma^2}{n},$$

con $\sigma^2 = \text{Var}(X)$.

En este ejemplo, vamos a probar de manera alternativa el teorema para el caso de \overline{X}_n^2 . Primero, dividiendo la expresión final para v_{JACK} en el Ejemplo 2.8 por el valor de σ_n^2 anterior se tiene que

$$\frac{v_{JACK}}{\sigma_n^2} = \frac{\hat{\alpha}_4}{4\mu^2\sigma^2(n-1)^2} + \frac{\overline{X}_n\hat{\alpha}_2}{\mu^2\sigma^2} - \frac{\overline{X}_n\hat{\alpha}_3}{\mu^2\sigma^2(n-1)} - \frac{\hat{\alpha}_2^2}{4\mu^2\sigma^2n(n-1)}. \quad (2.23)$$

Por un lado, sabemos que $\hat{\alpha}_2 = S^2 \rightarrow_{c.s.} \sigma^2$ y, por los teoremas 1.12 y 1.7, se tiene que $\overline{X}_n^2 \rightarrow_{c.s.} \mu^2$. Por tanto, para el segundo sumando de (2.23) se verifica que

$$\frac{\overline{X}_n\hat{\alpha}_2}{\mu^2\sigma^2} \rightarrow_{c.s.} 1.$$

Como, por otro lado, el resto de sumandos converge a 0 casi seguramente, se tiene finalmente (2.15).

2.4. Conclusiones

Hemos estudiado en este capítulo el método de remuestreo jackknife, tanto como técnica para reducir y estimar el sesgo de un estadístico, como su uso para estimar la varianza de un estadístico. Hemos comenzado mostrando los principios y definiciones necesarias que muestran que el método jackknife realmente reduce el sesgo de un estimador, para pasar a dar dos ejemplos de cálculo de las estimaciones b_{JACK} y v_{JACK} , respectivamente. El primero de los ejemplos representaba el caso en el que se tiene un estadístico insesgado, la media muestral, y hemos visto que el método jackknife estimaba correctamente sus propiedades. En el segundo ejemplo, hemos estudiado un estadístico que cumplía (2.1), la media muestral al cuadrado, para ver cómo efectivamente se reducía el sesgo. El resultado fundamental del capítulo es el Teorema 2.9, el cual demuestra que la estimación v_{JACK} para estadísticos de la forma $T_n = g(\overline{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ continua en μ , con $g'(\mu) \neq 0$, cumple

$$\frac{v_{JACK}}{\sigma_n^2} \rightarrow_{c.s.} 1,$$

con σ_n^2 dado en (2.14), siendo este valor un estimador de la varianza de T_n . Por último, se prueba de manera alternativa este resultado para el estadístico del segundo ejemplo, desarrollando los valores de v_{JACK} y σ_n^2 en ese caso.

Capítulo 3

Método bootstrap

3.1. Introducción

Una de las principales características de un estadístico T_n es su distribución en el muestreo. Si tuviéramos muestras replicadas de la población, resultando en una gran serie de valores para el estadístico, entonces podríamos estimar $P(T_n \leq t)$ contando cuántos de los valores evaluados en las muestras son menores que T_n . Pero en un gran número de situaciones esto no es posible y no podemos obtener muestras replicadas, sino un único conjunto de datos de tamaño n .

Sin embargo, sabemos que una muestra suficientemente grande de una población finita debería ser una buena representación de la población, por lo tanto muestras replicadas, con reemplazamiento, de la muestra original, podrían servir como sustitutos de muestras replicadas de la población, para tamaños de muestra suficientemente grandes. Estas muestras replicadas con reemplazamiento, al estar muestreando de la muestra original, serían muestras aleatorias simples de la función de distribución empírica F_n .

Esta es la idea que subyace bajo el método bootstrap, introducido por Efron [3] en 1979. En los últimos años, se ha demostrado que este tipo de remuestreos permite también estimar otras propiedades de un estadístico como su varianza, así como obtener intervalos de confianza sobre un determinado parámetro de la población.

Esta técnica coincide con el método jackknife en el sentido de que reemplaza complejos cálculos teóricos mediante el remuestro repetido de los datos. Sin embargo, la diferencia fundamental entre ambos es que el jackknife recalcula el estadístico para n subconjuntos fijos de la muestra original, mientras que el bootstrap obtiene muestras con reemplazamiento aleatoriamente generadas de la muestra de partida.

En este capítulo se estudia el método bootstrap en sus dos principales vertientes: como estimador de la distribución y la varianza de un estadístico y como generador de intervalos de confianza alrededor de un parámetro de la distribución desconocida. De igual manera que en el método jackknife, estudiaremos la consistencia de los

estimadores bootstrap para cada una de las características anteriores. Los resultados se han obtenido de [2], [4] y [10].

3.2. El método bootstrap

Sea X una variable aleatoria siguiendo una distribución desconocida F , sea $\mathbb{X} = (X_1, \dots, X_n)$ una muestra aleatoria simple de X , y sea $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Nótese que se acaba de introducir la notación \mathbb{X} para las muestras aleatorias simples y se utilizará a lo largo del capítulo. Nuestro objetivo es calcular características del estadístico T_n como pueden ser su varianza, lo que permitirá ver el error en la estimación, o su función de distribución, lo que permite calcular valores críticos o intervalos de confianza basados en el estadístico.

En el caso, por ejemplo, de la varianza de un estadístico, tenemos que esta se obtiene en la forma

$$Var(T_n) = \int_{\mathbb{x} \in Sop(\mathbb{X})} \left(T_n(\mathbb{x}) - \int_{\mathbb{x} \in Sop(\mathbb{X})} T_n(\mathbb{x}) \prod_{i=1}^n dF(x_i) \right)^2 \prod_{i=1}^n dF(x_i).$$

Evidentemente, si F es desconocida, la expresión anterior no se puede evaluar. El método bootstrap plantea como solución, aproximar la expresión anterior sustituyendo la distribución desconocida por una estimación de F . En función de qué estimación tomemos de F , podemos diferenciar entre dos tipos de bootstrap:

- **Bootstrap no paramétrico:** parte de que no conocemos el modelo que sigue la muestra aleatoria simple, y estima F usando la función de distribución empírica F_n .
- **Bootstrap paramétrico:** en este caso sabemos el modelo F_θ que sigue la muestra, pero no sabemos sus parámetros $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^k$. Se basa en obtener una estimación $\hat{\theta}$ de los parámetros y estimar F_θ mediante $F_{\hat{\theta}}$.

En general, es difícil conocer el modelo al cual pertenece el conjunto de datos con el que se está trabajando, por ello, centraremos este trabajo en el bootstrap no paramétrico.

Por tanto, usando la distribución empírica, en el caso de la varianza tendríamos como aproximación el valor

$$v_{BOOT} = \sum_{i=1}^{n^n} \frac{1}{n^n} \left(T_n(\mathbb{x}_i) - \sum_{j=1}^{n^n} \frac{T_n(\mathbb{x}_j)}{n^n} \right)^2,$$

donde \mathbb{x}_j es cualquiera de las n^n posibles variaciones con repetición de las observaciones de la muestra. A esta estimación se la conoce como estimación bootstrap ideal. La razón de llamarlo así es que, si n es grande, este procedimiento no se puede hacer

viable en la práctica. La alternativa es sacar muestras de F_n , calcular el estadístico en esas muestras, y aproximar v_{BOOT} mediante la varianza de esas muestras.

En general, lo que hacemos es aplicar el método de Monte Carlo a F_n . Formalmente, los pasos que hay que seguir son:

1. Tomamos F_n la función de distribución empírica de la muestra (X_1, \dots, X_n) .
2. Tomamos una nueva muestra aleatoria simple de F_n , (X_1^*, \dots, X_n^*) , y calculamos el estadístico $T_n^* = T_n(X_1^*, \dots, X_n^*)$.
3. Repetimos el paso 2 un número B de veces, independientes entre sí, obteniendo las muestras $\{X_{1b}^*, \dots, X_{nb}^*\}, b = 1, \dots, B$, y computamos los valores de los estadísticos $T_{nb}^* = T_n(X_{1b}^*, \dots, X_{nb}^*), b = 1, \dots, B$. La varianza de T_n , usando la distribución empírica, queda como

$$v_{BOOT}^{(B)} = \frac{1}{B} \sum_{b=1}^B \left(T_{nb}^* - \sum_{l=1}^B \frac{T_{nl}^*}{B} \right)^2.$$

Ahora por la ley fuerte de los grandes números vista en el Teorema 1.12, $v_{BOOT}^{(B)} \rightarrow v_{BOOT}$ y sería una buena aproximación de v_{BOOT} . Por tanto, quedaría por probar que el valor de v_{BOOT} es lo suficientemente próximo a $Var(T_n)$.

Veamos otra aplicación del método bootstrap. Si consideramos ahora el caso de la función de distribución de T_n , es decir, de

$$P(T_n(X_1, \dots, X_n) \leq x) = \int_{\substack{T_n(\mathbb{x}) \leq x \\ \mathbb{x} \in Sop(\mathbb{X})}} \prod_{i=1}^n dF(x_i),$$

y usando la estimación de bootstrap basada en las n^n variaciones de la muestra, tendríamos

$$H_{BOOT}(x) = \sum_{\substack{T_n(\mathbb{x}^*) \leq x \\ \mathbb{x}^* \in Sop(\mathbb{X}^*)}} \frac{1}{n^n},$$

donde $\mathbb{X}^* = (X_1^*, \dots, X_n^*)$ es una muestra aleatoria simple de la distribución empírica F_n . La alternativa usando el método de Monte Carlo sería

$$H_{BOOT}^{(B)}(x) = \sum_{b=1}^B \frac{I_{(-\infty, x]}(T_n(X_{1b}^*, \dots, X_{nb}^*))}{B}.$$

El objetivo de las dos siguientes secciones es la utilización del método bootstrap para la estimación de las dos características que acabamos de presentar, que son la varianza y la distribución de un estadístico, probando propiedades de consistencia para estos estimadores.

3.3. Estimación bootstrap de la distribución de un estadístico

Dada una variable aleatoria X que sigue una distribución desconocida F , con $\mathbb{X} = (X_1, \dots, X_n)$ una muestra aleatoria simple de X y dado el estadístico T_n basado en esa muestra, vamos a considerar en este apartado la estimación de la distribución en el muestreo de T_n , es decir, de

$$H_n(x) := P(T_n(X_1, \dots, X_n) \leq x) = \int_{\substack{T_n(\mathbb{x}) \leq x \\ \mathbb{x} \in \text{Sop}(\mathbb{X})}} \prod_{i=1}^n dF(x_i). \quad (3.1)$$

Para ello, comenzaremos definiendo lo que se conoce como *muestra bootstrap*.

Definición 3.1. *Dada una muestra aleatoria simple $\mathbb{X} = (X_1, \dots, X_n)$, se llama muestra bootstrap asociada una función de distribución F a una muestra aleatoria simple $\mathbb{X}^* = (X_1^*, \dots, X_n^*)$ de la distribución empírica F_n asociada a \mathbb{X} .*

Dada una muestra aleatoria simple con función de distribución F desconocida, introducimos la definición de la distribución de bootstrap ideal del estadístico T_n .

Definición 3.2. *Sea X una variable aleatoria que sigue una distribución desconocida F , sea $\mathbb{X} = (X_1, \dots, X_n)$ una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Se define la distribución bootstrap ideal de T_n como*

$$H_{BOOT}(x) = P_*(T_n(X_1^*, \dots, X_n^*) \leq x | \mathbb{X}), \quad (3.2)$$

donde $\mathbb{X}^* = (X_1^*, \dots, X_n^*)$ es una muestra bootstrap y P_* denota la probabilidad condicionada dados (X_1, \dots, X_n) .

Sobre la definición anterior, se observa lo siguiente.

Observación 3.3. *La probabilidad $P_*(\cdot)$ anterior se corresponde con los n^n posibles remuestreos con reemplazamiento de la muestra original (X_1, \dots, X_n) . Como hemos indicado en la Sección 3.2, como recalculamos T_n para los n^n remuestreos es imposible salvo que n sea muy pequeño, se usa un número fijo de B remuestreos y se recalcula T_n solo B veces. Es decir, estimamos $H_{BOOT}(x)$ mediante el método de Monte Carlo.*

En la siguiente definición, en las mismas condiciones que en la Definición 3.2, definimos el estimador bootstrap de la distribución mediante el método de Monte Carlo.

Definición 3.4. *Sea X una variable aleatoria que sigue una distribución desconocida F , sea $\mathbb{X} = (X_1, \dots, X_n)$ una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Sea B un número fijo. Se define la distribución bootstrap de Monte Carlo como*

$$H_{BOOT}^{(B)}(x) = \frac{1}{B} \sum_{i=1}^B I\{T_n(X_{i1}^*, \dots, X_{in}^*) \leq x\},$$

donde $\{X_{i1}^*, \dots, X_{in}^*\}$, $i = 1, \dots, B$ son muestras bootstrap asociadas a \mathbb{X} .

Lo que interesa ahora es qué propiedades tiene como estimador de $H_n(x)$ el valor de $H_{BOOT}(x)$. Para ello, en el siguiente apartado estudiamos la consistencia del estimador bootstrap ideal. Notemos, por lo mencionado en la Sección 3.2, que es suficiente estudiar la consistencia de este tipo de bootstrap al tenerse que $H_{BOOT}^{(B)}$ aproxima H_{BOOT} por la ley fuerte de los grandes números.

3.3.1. Consistencia

Como hemos indicado en el Capítulo 2, una propiedad importante de los estimadores es la consistencia de los mismos. En esta sección obtenemos una serie de resultados de consistencia de la estimación bootstrap de la distribución de un estadístico. Centraremos nuestro estudio, como en el caso del método jackknife, en estadísticos que son funciones de la media muestral al ser estadísticos muy usuales en la práctica.

Si queremos estimar la distribución de un estadístico $T_n = T_n(X_1, \dots, X_n)$, el objetivo es que la distribución H_{BOOT} sea muy próxima a la verdadera distribución $H_n(x)$ de T_n . Para ello consideraremos varias métricas sobre el espacio de las funciones de distribución, $\mathcal{F}_{\mathbb{R}}$. A partir de estas métricas se introducen a continuación los conceptos de consistencia débil y fuerte de la distribución del bootstrap.

Definición 3.5. *Sea X una variable aleatoria que sigue una distribución desconocida F , (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Sea ρ una métrica en el espacio $\mathcal{F}_{\mathbb{R}}$ y sean*

$$H_n(x) = P(T_n(X_1, \dots, X_n) \leq x)$$

$$H_{BOOT}(x) = P_*(T_n(X_1^*, \dots, X_n^*) \leq x | \mathbb{X}),$$

la distribución exacta de T_n y la distribución bootstrap ideal de T_n , respectivamente.

Decimos que el estimador bootstrap es débilmente consistente bajo la métrica ρ para T_n si $\rho(H_n, H_{BOOT}) \rightarrow_p 0$ cuando $n \rightarrow \infty$. Y decimos que el estimador bootstrap es fuertemente consistente bajo la métrica ρ para T_n si $\rho(H_n, H_{BOOT}) \rightarrow_{c.s.} 0$ cuando $n \rightarrow \infty$. Por lo tanto, la consistencia fuerte implica la consistencia débil.

De la definición anterior se puede observar que:

Observación 3.6. *Cuando en la definición anterior se fija $\mathbb{X} = \mathbb{x}$, o lo que es lo mismo, se fija la distribución empírica, $F_n(t)$, tenemos que $H_{BOOT}(x | F_n(t))$ (con $F_n(t)$ un valor fijo), es propiamente una función de distribución y por tanto se puede calcular la métrica que da sobre el espacio de distribuciones respecto de H_n , es decir*

$$h(x, F_n(t)) = \rho(H_{BOOT}(x | F_n(t)), H_n).$$

Evidentemente, cuando después consideramos $F_n(t)$ aleatorio, en este caso $\{F_n(t), t \in \mathbb{R}\}$ es un proceso estocástico, la medida anterior pasa a ser una medida sobre un proceso estocástico y estudiaremos su convergencia.

Observación 3.7. En lo que resta de capítulo, denotaremos a los valores del estadístico T_n de una muestra bootstrap (X_1^*, \dots, X_n^*) como $T_n^* = T_n(X_1^*, \dots, X_n^*)$.

Además, dado T_n^* , cuando usemos la notación $P_*(T_n^*|\mathbb{X} = \mathbb{x})$, $E_*(T_n^*|\mathbb{X} = \mathbb{x})$ y $Var_*(T_n^*|\mathbb{X} = \mathbb{x})$ en futuras demostraciones, estaremos denotando la probabilidad, esperanza y varianza, respectivamente, del estadístico T_n^* , condicionadas a que la muestra aleatoria simple de partida está fija, es decir, $X_1 = x_1, \dots, X_n = x_n$. Cuando eliminemos el condicionamiento anterior, escribiremos $P_*(T_n^*) = P_*(T_n^*|\mathbb{X})$, $E_*(T_n^*|\mathbb{X})$ y $Var_*(T_n^*|\mathbb{X})$.

Notemos también aquí que cuando calculamos la varianza o la esperanza de una variable X^* que sigue la distribución empírica, estas serán

$$i) E_*(X^*) = \bar{X}_n.$$

$$ii) Var_*(X^*) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Por otro lado, cuando necesitemos calcular las características E_* y Var_* , la estrategia que se seguirá consistirá en tomar una realización $(\mathbb{X} = \mathbb{x})$ de la variable aleatoria de partida X , computar la característica correspondiente para esa realización, sabiendo que \mathbb{X} está fija, y finalmente descondicionar (volver a introducir la aleatoriedad) para obtener el valor de la característica.

Por ejemplo, si queremos calcular $Var_*(X_i^* - \bar{X}_n)$, tomamos

$$h(\mathbb{x}) = Var_*(X_i^* - \bar{X}_n|\mathbb{X} = \mathbb{x}) = Var_*(X_i^*|\mathbb{X} = \mathbb{x}),$$

viniendo la última igualdad de que \bar{X}_n es constante al haber tomado una realización de \bar{X} . Por último, descondicionando, es decir, evaluando $h(\mathbb{x})$ en \mathbb{X} , obtenemos que

$$h(\mathbb{X}) = Var_*(X_i^*|\mathbb{X}) = s^2.$$

Observación 3.8. En los siguientes teoremas se usará el estadístico $T_n = \sqrt{n}(\bar{X}_n - \mu)$. Para obtener el estadístico $T_n^* = T_n(X_1^*, \dots, X_n^*)$ correspondiente hemos de reescribir T_n de la forma

$$T_n(X_1, \dots, X_n) = \sqrt{n}(\bar{X}_n - E(X)).$$

De esta manera, al evaluar el estadístico en las muestras bootstrap queda

$$T_n^* = T_n(X_1^*, \dots, X_n^*) = \sqrt{n}(\bar{X}_n^* - E_*(X^*)) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n).$$

Además, en el Teorema 3.11, el estadístico que se tomará es $T_n = \sqrt{n}(g(\bar{X}_n) - g(\mu))$, con $g: \mathbb{R} \rightarrow \mathbb{R}$ continua y diferenciable en μ . El valor de T_n^* en este caso se construye de la misma forma que en el anterior, obteniendo

$$T_n^* = T_n(X_1^*, \dots, X_n^*) = \sqrt{n}(g(\bar{X}_n^*) - g(E_*(X^*))) = \sqrt{n}(g(\bar{X}_n^*) - g(\bar{X}_n)).$$

Para el estudio de la consistencia del método bootstrap trabajaremos principalmente con las dos métricas K y $\tilde{\rho}_2$, dadas en las definiciones 1.23 y 1.24 para $r = 2$, respectivamente.

Para obtener los primeros resultados sobre consistencia del bootstrap consideramos un estadístico importante que hemos mencionado en la observación anterior, es decir, $T_n = \sqrt{n}(\bar{X}_n - \mu)$.

En el siguiente teorema estudiamos dicho estadístico para la métrica $\tilde{\rho}_2$.

Teorema 3.9. *Sea X una variable aleatoria con distribución desconocida $F \in \mathcal{F}_{2,1}$, sea (X_1, \dots, X_n) una muestra aleatoria simple de X y sea el estadístico $T_n = T_n(X_1, \dots, X_n) = \sqrt{n}(\bar{X}_n - \mu)$. Sean H_n y H_{BOOT} dados por (3.1) y (3.2), respectivamente. Entonces, H_{BOOT} es fuertemente consistente en la métrica $\tilde{\rho}_2$ para el estadístico T_n .*

Demostración. Denotemos $Y_n := \tilde{\rho}_2(H_n, H_{BOOT})$. Como se ha comentado en la Observación 3.7, la función de distribución H_{BOOT} es una variable aleatoria, luego el valor de Y_n también es aleatorio.

En este teorema, tenemos que probar, por tanto, que

$$P\left(\lim_{n \rightarrow \infty} Y_n = 0\right) = 1. \quad (3.3)$$

Para ello, primero, por la Definición 1.24 se tiene

$$Y_n = \tilde{\rho}_2(T_n^*, T_n) = \tilde{\rho}_2(\sqrt{n}(\bar{X}_n^* - \bar{X}_n), \sqrt{n}(\bar{X}_n - \mu)),$$

y por la Proposición 1.26 se tiene

$$Y_n = \frac{1}{\sqrt{n}} \tilde{\rho}_2\left(\sum_{i=1}^n (X_i^* - \bar{X}_n), \sum_{i=1}^n (X_i - \mu)\right),$$

y utilizando la Proposición 1.28, dicha métrica puede acotarse superiormente por

$$\begin{aligned} Y_n &\leq \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{\rho}_2(X_i^* - \bar{X}_n, X_i - \mu))^2} r(X_1) = \tilde{\rho}_2(X_1^* - \bar{X}_n, X_1 - \mu) \\ &= \sqrt{(\tilde{\rho}_2(X_1^*, X_1))^2 - (E(X_1) - E_*(X_1^*))^2} = \sqrt{(\tilde{\rho}_2(F_n, F))^2 - (\mu - \bar{X}_n)^2}. \end{aligned}$$

Donde la segunda igualdad se cumple por la intercambiabilidad de las variables, la tercera se debe a la Proposición 1.27 y la última viene dada por la Proposición 1.24 y por definición de las esperanzas de X_1 y X_1^* .

Por otro lado, observamos que, de la Proposición 1.25, $\tilde{\rho}_2(F_n, F) \rightarrow_{c.s.} 0$, y sabemos que $\bar{X}_n - \mu \rightarrow_{c.s.} 0$ por el Teorema Central del Límite. Llamando $B_n = \tilde{\rho}_2(F_n, F)$ y $C_n = \mu - \bar{X}_n$, se tiene, por el Teorema 1.7, que $B_n^2 \rightarrow_{c.s.} 0$ y $C_n^2 \rightarrow_{c.s.} 0$. Además,

$$(B_n^2, C_n^2) \rightarrow_{c.s.} (0, 0),$$

ya que la intersección contable de sucesos casi seguros es un suceso casi seguro. Por tanto, podemos aplicar de nuevo el Teorema 1.7 en varias variables para la función $g(B_n, C_n) = \sqrt{B_n^2 - C_n^2}$, y se obtiene que

$$A_n := \sqrt{(\tilde{\rho}_2(F_n, F))^2 - (\mu - \bar{X}_n)^2} \rightarrow_{c.s.} 0.$$

Por tanto, dado $w \in \Omega$ tal que $\lim_{n \rightarrow \infty} A_n(w) = 0$, como $0 \leq Y_n \leq A_n$, entonces ese w es tal que $\lim_{n \rightarrow \infty} Y_n(w) = 0$. Esto implica que

$$\{w \in \Omega \mid \lim_{n \rightarrow \infty} A_n(w) = 0\} \subseteq \{w \in \Omega \mid \lim_{n \rightarrow \infty} Y_n(w) = 0\},$$

de donde se obtiene que

$$1 \geq P(\{w \in \Omega \mid \lim_{n \rightarrow \infty} Y_n(w) = 0\}) \geq P(\{w \in \Omega \mid \lim_{n \rightarrow \infty} A_n(w) = 0\}) = 1,$$

y por tanto, hemos probado que se cumple (3.3), luego $Y_n \rightarrow_{c.s.} 0$, lo que prueba la consistencia fuerte en $\tilde{\rho}_2$ de H_{BOOT} para $T_n = T_n(X_1, \dots, X_n) = \sqrt{n}(\bar{X}_n - \mu)$. \square

En el siguiente teorema, se prueba la consistencia fuerte en la métrica K para el estadístico $T_n = \sqrt{n}(\bar{X}_n - \mu)$.

Teorema 3.10. *Sea X una variable aleatoria que sigue una distribución desconocida $F \in \mathcal{F}_{2,1}$ y sea (X_1, \dots, X_n) una muestra aleatoria simple de X . Sea $T_n = T_n(X_1, \dots, X_n) = \sqrt{n}(\bar{X}_n - \mu)$, sea $\sigma^2 = \text{Var}(X) < \infty$ y s la desviación estándar muestral asociada a (X_1, \dots, X_n) . Entonces, H_{BOOT} es fuertemente consistente en la métrica K para el estadístico T_n .*

Demostración. Sea $Z_n := K(H_n, H_{BOOT})$. De igual manera que en el Teorema 3.9, tenemos que probar que

$$P\left(\lim_{n \rightarrow \infty} Z_n = 0\right) = 1. \quad (3.4)$$

Usando la Definición 1.23, podemos escribir

$$Z_n = \sup_x |P\{T_n \leq x\} - P_*\{T_n^* \leq x\}|.$$

La expresión anterior puede escribirse como

$$\begin{aligned} Z_n &= \sup_x |P\{T_n \leq x\} - P_*\{T_n^* \leq x\}| \\ &= \sup_x \left| P\left\{\frac{T_n}{\sigma} \leq \frac{x}{\sigma}\right\} - P_*\left\{\frac{T_n^*}{s} \leq \frac{x}{s}\right\} \right| \\ &= \sup_x \left| P\left\{\frac{T_n}{\sigma} \leq \frac{x}{\sigma}\right\} - \Phi\left(\frac{x}{\sigma}\right) + \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(\frac{x}{s}\right) + \Phi\left(\frac{x}{s}\right) - P_*\left\{\frac{T_n^*}{s} \leq \frac{x}{s}\right\} \right|, \end{aligned}$$

donde $\Phi(\cdot)$ representa la función de distribución $N(0, 1)$. Agrupando los términos de dos en dos y considerando el supremo se tiene

$$\begin{aligned} Z_n &\leq \sup_x \left| P\left\{\frac{T_n}{\sigma} \leq \frac{x}{\sigma}\right\} - \Phi\left(\frac{x}{\sigma}\right) \right| + \sup_x \left| \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(\frac{x}{s}\right) \right| + \\ &\quad + \sup_x \left| \Phi\left(\frac{x}{s}\right) - P_*\left\{\frac{T_n^*}{s} \leq \frac{x}{s}\right\} \right| \\ &= A_n + B_n + C_n, \end{aligned}$$

siendo

$$A_n = \sup_x \left| P \left\{ \frac{T_n}{\sigma} \leq \frac{x}{\sigma} \right\} - \Phi \left(\frac{x}{\sigma} \right) \right|,$$

$$B_n = \sup_x \left| \Phi \left(\frac{x}{s} \right) - \Phi \left(\frac{x}{\sigma} \right) \right|,$$

$$C_n = \sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right|,$$

Que $A_n \rightarrow 0$ es una consecuencia directa del Teorema 1.30, ya que $P \left\{ \frac{T_n}{\sigma} \leq \frac{x}{\sigma} \right\}$ converge a $\Phi \left(\frac{x}{\sigma} \right)$ por el Teorema Central del Límite. Esta convergencia es determinista, al no estar tratando con muestras bootstrap.

Por otro lado, veamos la convergencia casi segura de B_n a 0. Como, por el Teorema 1.18, s^2 converge casi seguramente a σ^2 , entonces por el Teorema 1.7 se tiene que s converge casi seguramente a σ . Además, como la función $\Phi(\cdot)$ es una función uniformemente continua, aplicando otra vez el Teorema 1.7 obtenemos que $\Phi \left(\frac{x}{s} \right) \rightarrow_{c.s.} \Phi \left(\frac{x}{\sigma} \right)$, por lo que $B_n \rightarrow_{c.s.} 0$.

Falta comprobar que $C_n \rightarrow_{c.s.} 0$. Para ello, vamos a usar primero el Teorema 1.29 para ver que C_n se puede acotar de la siguiente manera:

$$\sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| \leq \frac{c}{s^3 n^{3/2}} \sum_{i=1}^n |X_i - \bar{X}_n|^3.$$

Primero sustituimos el valor de T_n^* en la expresión anterior y después operamos:

$$\begin{aligned} \sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| &= \sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{s} \leq \frac{x}{s} \right\} \right| \\ &= \sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{1}{\sqrt{ns}} \sum_{i=1}^n (X_i^* - \bar{X}_n) \leq \frac{x}{s} \right\} \right|, \end{aligned}$$

y vamos a aplicar el Teorema 1.29 a $X_i^* - \bar{X}_n$, con $V_n = \sqrt{ns}$. Veamos primero que se cumplen las hipótesis del Teorema 1.7, es decir, que $E_*(X_i^* - \bar{X}_n) = 0$ y que $V_n^2 = \sum_{i=1}^n Var_*(X_i^* - \bar{X}_n)$.

En ambos casos se sigue la estrategia mostrada en la Observación 3.6. Por un lado, en esa observación vimos que $Var_*(X_i^* - \bar{X}_n) = s^2$, de lo que se sigue que

$$\sum_{i=1}^n Var_*(X_i^* - \bar{X}_n) = ns^2 = V_n^2.$$

Por otro lado, siguiendo la misma estrategia que en la Observación 3.6, como

$$E_*(X_i^* - \bar{X}_n | \mathbb{X} = \mathbf{x}) = E_*(X_i^* | \mathbb{X} = \mathbf{x}) - \bar{X}_n,$$

se tiene que

$$E_*(X_i^* - \bar{X}_n | \mathbb{X}) = E_*(X_i^* | \mathbb{X}) - \bar{X}_n = 0.$$

Una vez se cumplen las hipótesis del Teorema 1.29, aplicamos dicho teorema a C_n , es decir

$$\begin{aligned} \sup_x \left| \Phi\left(\frac{x}{s}\right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| &= \sup_x \left| \Phi\left(\frac{x}{s}\right) - P_* \left\{ \frac{1}{\sqrt{ns}} \sum_{i=1}^n (X_i^* - \bar{X}_n) \leq \frac{x}{s} \right\} \right| \\ &\leq \frac{c}{s^3 n^{3/2}} \sum_{i=1}^n E_* |X_i^* - \bar{X}_n|^3 \\ &= \frac{c}{s^3 n^{1/2}} E_* |X_1^* - \bar{X}_n|^3 \\ &= \frac{c}{s^3 n^{3/2}} \sum_{i=1}^n |X_i - \bar{X}_n|^3, \end{aligned}$$

donde la penúltima igualdad se verifica por ser las variables X_i^* idénticamente distribuidas, y la última igualdad de la definición de E_* .

Por tanto, acotando C_n usando el Teorema 1.29 y operando queda

$$C_n \leq \frac{c}{n^{3/2} s^3} \sum_{i=1}^n |X_i - \bar{X}_n|^3 \leq \frac{c}{n^{3/2} s^3} \sum_{i=1}^n |X_i - \mu + \mu - \bar{X}_n|^3,$$

y, usando la propiedad

$$(a + b)^3 \leq (2 \max(a, b))^3 = 2^3 \max(a, b)^3 \leq 2^3 (a^3 + b^3), \quad (3.5)$$

para a y b no negativos, y operando, se tiene que

$$\begin{aligned} C_n &\leq \frac{c}{n^{3/2} s^3} \cdot 2^3 \left[\sum_{i=1}^n |X_i - \mu|^3 + \sum_{i=1}^n |\mu - \bar{X}_n|^3 \right] \\ &= \frac{8}{s^3} \left[\frac{1}{n^{3/2}} \sum_{i=1}^n |X_i - \mu|^3 + \frac{|\mu - \bar{X}_n|^3}{\sqrt{n}} \right]. \end{aligned} \quad (3.6)$$

Para ver que $C_n \xrightarrow{c.s.} 0$, probaremos que los dos últimos sumandos de (3.6) convergen a 0 casi seguramente. Por un lado, aplicando los teoremas 1.18 y 1.7 se tiene que $s \xrightarrow{c.s.} \sigma > 0$ y, añadiendo que $\bar{X}_n \xrightarrow{c.s.} \mu$, se verifica que

$$\frac{|\mu - \bar{X}_n|^3}{\sqrt{ns^3}} \xrightarrow{c.s.} 0.$$

Para el primer término, sea $Y_i = |X_i - \mu|^3$ y $\lambda = 2/3$. Entonces los $\{Y_i\}$ son independientes e idénticamente distribuidos y

$$E|Y_i|^\lambda = E|X_i - \mu|^{3 \cdot 2/3} = \text{Var}(X_i) < \infty.$$

Por el Teorema 1.13, se sigue que

$$\frac{1}{n^{3/2}} \sum_{i=1}^n |X_i - \mu|^3 = n^{-1/\lambda} \sum_{i=1}^n Y_i \xrightarrow{c.s.} 0,$$

y usando que la intersección contable de sucesos casi seguros es un suceso casi seguro y el Teorema 1.7, se tiene, llamando D_n al último término de la desigualdad (3.6), que

$$D_n \rightarrow_{c.s.} 0.$$

Por tanto, dado $w \in \Omega$, se cumple que

$$1 \geq P\left(\left\{w \in \Omega \mid \lim_{n \rightarrow \infty} C_n(w) = 0\right\}\right) \geq P\left(\left\{w \in \Omega \mid \lim_{n \rightarrow \infty} D_n(w) = 0\right\}\right) = 1,$$

y por tanto $C_n \rightarrow_{c.s.} 0$.

Usando otra vez la propiedad de la intersección de sucesos casi seguros y el Teorema 1.7, se tiene que $A_n + B_n + C_n \rightarrow_{c.s.} 0$, y tenemos de nuevo que

$$1 \geq P\left(\left\{\lim_{n \rightarrow \infty} Z_n = 0\right\}\right) \geq P\left(\left\{\lim_{n \rightarrow \infty} A_n + B_n + C_n = 0\right\}\right) = 1,$$

lo que implica (3.4) y prueba que $K(H_n, H_{BOOT}) \rightarrow_{c.s.} 0$. \square

Hemos probado pues que el estimador bootstrap H_{BOOT} es fuertemente consistente para el estadístico $\sqrt{n}(\bar{X}_n - \mu)$ y las métricas K y $\tilde{\rho}_2$.

Pasamos ahora a casos más complejos. Muchos estadísticos pueden escribirse como función de la media muestral, por lo que resulta de interés estudiar la consistencia de los estimadores de la forma $g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$.

En el siguiente teorema probaremos la consistencia fuerte del estimador bootstrap H_{BOOT} con respecto a la métrica K cuando se utiliza el estadístico $T_n = \sqrt{n}(g(\bar{X}_n) - \theta)$, con $\theta = g(\mu)$.

Teorema 3.11. *Sea X una variable aleatoria que sigue una distribución desconocida F y sea (X_1, \dots, X_n) una muestra aleatoria simple de X tal que $E(X_1^2) < \infty$. Sea $T_n = \sqrt{n}(T_n - \theta)$, donde $T_n = g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y diferenciable en $\mu = E(X_1)$, $\theta = g(\mu)$ y $g'(\mu) \neq 0$, y sea H_n la distribución de T_n . Entonces el estimador H_{BOOT} es fuertemente consistente con respecto a K para H_n .*

Demostración. Por la Observación 3.8, sabemos que

- i) $\sqrt{n}(T_n^* - T_n) \sim H_{BOOT}$ y $\sqrt{n}(T_n - g(\mu)) \sim H_n$.
- ii) $\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \sim G_{BOOT}$ y $\sqrt{n}(\bar{X}_n - \mu) \sim G_n$.

Tomando ahora una distribución cualquiera de las anteriores, por ejemplo $H_n(x)$, definimos la función H_n^1 como $H_n(x/g'(\mu))$.

Entonces se tiene que

$$K(H_{BOOT}, H_n) \leq K(H_{BOOT}, G_{BOOT}^1) + K(G_{BOOT}^1, G_n^1) + K(G_n^1, H_n). \quad (3.7)$$

Hemos de probar, por tanto, que cada uno de los sumandos de la desigualdad anterior converge a 0 casi seguramente.

Por un lado, se tiene por el Teorema 3.10 que $K(G_{BOOT}, G_n) \rightarrow_{c.s.} 0$, y por tanto $K(G_{BOOT}^1, G_n^1) \rightarrow_{c.s.} 0$. Para probar ahora que $K(H_{BOOT}, G_{BOOT}^1)$ converge a cero casi seguramente, aplicamos el desarrollo en serie de Taylor a \overline{X}_n^* en \overline{X}_n , y obtenemos que

$$T_n^* - T_n = g'(\overline{X}_n)(\overline{X}_n^* - \overline{X}_n) + \tilde{r}_n.$$

Por otro lado, como \tilde{r}_n es un término que, cuando fijamos $\mathbb{X} = \mathbb{x}$, cumple que

$$P_*(\sqrt{n}|\tilde{r}_n| > \varepsilon | \mathbb{X} = \mathbb{x}) \rightarrow 0,$$

cuando $n \rightarrow \infty$, luego $\forall w \in \Omega$ se tiene que

$$P_*(\sqrt{n}|\tilde{r}_n| > \varepsilon | X_1(w), \dots, X_n(w)) \rightarrow 0,$$

de donde se verifica finalmente que

$$P_*(\sqrt{n}|\tilde{r}_n| > \varepsilon) \rightarrow_{c.s.} 0.$$

Ahora, como g es diferenciable en μ y g' es continua en μ , por el Teorema 1.7 se tiene que $g'(\overline{X}_n) \rightarrow_{c.s.} g'(\mu)$. Por tanto,

$$H_{BOOT}(x) = P_*\{\sqrt{n}(T_n^* - T_n) \leq x\} \rightarrow_{c.s.} P_*\{\sqrt{n}(\overline{X}_n^* - \overline{X}_n) \leq x/g'(\mu)\} = G_{BOOT}^1(x).$$

De manera análoga se obtiene que $K(H_n, G_n^1) \rightarrow_{c.s.} 0$. Por tanto, por la desigualdad (3.7) se tiene que $K(H_n, H_{BOOT}) \rightarrow_{c.s.} 0$, que es lo que queríamos demostrar. \square

Nos preguntamos ahora, si el método bootstrap tiene alguna ventaja sobre el Teorema Central del Límite enunciado en 1.11 cuando este puede usarse. Sea (X_1, \dots, X_n) muestra aleatoria simple de una variable aleatoria X con distribución F , y sea $T_n = T_n(X_1, \dots, X_n) = \sqrt{n}(\overline{X}_n - \mu)$. Si $Var(X) < \infty$, entonces $\sqrt{n}(\overline{X}_n - \mu) \rightarrow_d N(0, \sigma^2)$ por el Teorema 1.11 y, por otro lado, $K(H_n, H_{BOOT}) \rightarrow_{c.s.} 0$ por el Teorema 3.10. Luego dos aproximaciones para $P(T_n(X_1, \dots, X_n) \leq x)$ son $\Phi(x/s)$ y $P_*(\sqrt{n}(\overline{X}_n^* - \overline{X}_n) \leq x)$.

En el siguiente teorema comparamos la convergencia cuando se da el Teorema Central del Límite con la convergencia cuando se puede aplicar el método bootstrap, y obtenemos que si se utiliza el Teorema Central del Límite, la convergencia es del orden $O_p\left(\frac{1}{\sqrt{n}}\right)$, mientras que, por el método bootstrap se obtiene un orden $O_p\left(\frac{1}{n}\right)$.

Teorema 3.12. *Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ un estadístico, donde $\sigma^2 = Var(X) < \infty$. Entonces la aproximación H_{BOOT} converge a H_n con un error $O_p\left(\frac{1}{n}\right)$, mientras que la aproximación dada por el Teorema 1.11 converge con un error $O_p\left(\frac{1}{\sqrt{n}}\right)$.*

Demostración. Dado $T_n^* = T_n(X_1^*, \dots, X_n^*) = \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{s}$, sabemos, para T_n y T_n^* , que la expansión de Edgeworth de dos términos, por la Definición 1.31, viene dada por

$$P(T_n \leq x) = \Phi(x) + \frac{c_1(F)p_1(x)}{\sqrt{n}}\varphi(x) + \frac{c_2(F)p_2(x)}{n}\varphi(x) + O(1/n).$$

$$P_*(T_n^* \leq x) = \Phi(x) + \frac{c_1(F_n)p_1(x)}{\sqrt{n}}\varphi(x) + \frac{c_2(F_n)p_2(x)}{n}\varphi(x) + O(1/n).$$

Por tanto, se tiene que

$$H_n(x) - H_{BOOT}(x) = \left[\frac{(c_1(F) - c_1(F_n))p_1(x)}{\sqrt{n}} + \frac{(c_2(F) - c_2(F_n))p_2(x)}{n} \right] \varphi(x) + O(1/n), \quad (3.8)$$

siendo

$$\begin{aligned} p_1(x) &= 1 - x^2, & p_2(x) &= 3x - x^3, \\ c_1(F) &= \frac{\gamma}{6}, & c_1(F_n) &= \frac{\gamma_*}{6}, \\ c_2(F) &= \frac{\kappa - 3}{24}, & c_2(F_n) &= \frac{\kappa_* - 3}{24}, \end{aligned}$$

donde $\gamma = \frac{E(X_1 - \mu)^3}{\sigma^3}$ y $\kappa = \frac{E(X_1 - \mu)^4}{\sigma^4}$. Denotamos por γ_* a $\frac{E_*(X_1^* - \bar{X}_n)^3}{s^3}$ y por κ_* a $\frac{E_*(X_1^* - \bar{X}_n)^4}{s^4}$, donde, para $k = 3, 4$, se tiene

$$E_*(X_1^* - \bar{X}_n)^k = \sum_{i=1}^n \frac{X_i - \bar{X}_n}{n} = m_k.$$

Queremos ver que (3.8) es de orden $O_p(1/n)$. Para ello, veamos primero que

$$\gamma_* - \gamma = O_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{y} \quad \kappa_* - \kappa = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Para el primer caso, por la Definición 1.20, ha de cumplirse

$$P_*\left(\left|\frac{m_3}{s^3} - \frac{E(X_1 - \mu)^3}{\sigma^3}\right| \geq \frac{1}{\sqrt{n}}C_\varepsilon\right) = P_*\left(\sqrt{n}\left|\frac{m_3}{s^3} - \frac{\mu_3}{\sigma^3}\right| \geq C_\varepsilon\right) < \varepsilon. \quad (3.9)$$

Por un lado, por el Teorema 1.19, sabemos que

$$\sqrt{n}(m_3 - \mu_3) \rightarrow_d N(0, \sigma_{22}),$$

con $\sigma_{22} = \mu_6 - \mu_3^2 - 6\mu_2\mu_4 + 9\mu_2^3$.

Por otro lado, si dividimos por σ^3 , tenemos que

$$\sqrt{n}\left(\frac{m_3}{\sigma^3} - \frac{\mu_3}{\sigma^3}\right) \rightarrow_d N\left(0, \frac{\sigma_{22}}{\sigma^6}\right).$$

y aplicando ahora el Teorema 1.8, como $s^3 \rightarrow_{c.s.} \sigma^3$ y llamando $Z_n = \sqrt{n}(\gamma_* - \gamma)$, se tiene que

$$Z_n = \sqrt{n} \left(\frac{m_3}{s^3} - \frac{\mu_3}{\sigma^3} \right) \rightarrow_d N \left(0, \frac{\sigma_{22}}{\sigma^6} \right).$$

Para ver que se cumple (3.9), hemos de probar que

$$P(Z_n > C_\varepsilon) + P(Z_n < -C_\varepsilon) < \varepsilon.$$

Para ello, tomamos $\bar{F} = 1 - F$, con F la función de distribución de $N(0, \sigma_{22}/\sigma^6)$, y $\bar{G}_n = 1 - G_n$, con G_n la función de distribución de Z_n . Por un lado, por ser F la función de distribución de una normal, se tiene que $\exists C_\varepsilon$ tal que $\bar{F}(C_\varepsilon) \leq \varepsilon/4$.

Por definición de límite de $\bar{G}_n \rightarrow_d \bar{F}$, se tiene que $\exists n_o$, tal que $\forall n \geq n_o$, $\bar{G}_n(C_\varepsilon) \in (F(C_\varepsilon) \pm \varepsilon/4)$. Por tanto, se tiene que

$$\bar{G}_n(C_\varepsilon) \leq \bar{F}(C_\varepsilon) + \frac{\varepsilon}{4} \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2},$$

lo que prueba que $P(Z_n > C_\varepsilon) < \varepsilon/2$. Por simetría, también se tiene que $P(Z_n < -C_\varepsilon) < \varepsilon/2$, luego hemos probado (3.9).

De manera análoga se prueba que $\kappa_* - \kappa = O_p\left(\frac{1}{\sqrt{n}}\right)$. Por tanto, sustituyendo en (3.8), se tiene que

$$H_n(x) - H_{BOOT}(x) = \left[\frac{O_p\left(\frac{1}{\sqrt{n}}\right) p_1(x)}{\sqrt{n}} + \frac{O_p\left(\frac{1}{\sqrt{n}}\right) p_2(x)}{n} \right] \varphi(x) + O\left(\frac{1}{n}\right).$$

Como $\frac{1}{\sqrt{n}} O_p\left(\frac{1}{\sqrt{n}}\right) = O_p\left(\frac{1}{n}\right)$, finalmente queda que

$$H_n(x) - H_{BOOT}(x) = O_p\left(\frac{1}{n}\right),$$

como queríamos probar.

Por otro lado, la aproximación que da el Teorema Central del Límite para el estadístico T_n , siguiendo el mismo razonamiento, es de orden $O_p\left(\frac{1}{\sqrt{n}}\right)$. Por tanto, tomando el estadístico estandarizado $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$, hemos demostrado que el método bootstrap es un orden más preciso que el Teorema Central del Límite. \square

3.4. Estimación de la varianza de un estadístico

El bootstrap puede usarse también para estimar la varianza de un estadístico T_n . Como en el caso de la estimación de la distribución, tenemos dos definiciones para la varianza de bootstrap: la definición ideal que estima la varianza usando el principio de sustitución, y la definición que aplica el método de Monte Carlo cuando

la primera es muy difícil de obtener.

Definimos primero el estimador ideal de la varianza de bootstrap de un estadístico.

Definición 3.13. *Sea X una variable aleatoria que sigue una distribución desconocida F y sea (X_1, \dots, X_n) una muestra aleatoria simple de X . Sea T_n un estadístico. El estimador ideal bootstrap de la varianza de T_n viene dado por*

$$\begin{aligned} \text{Var}(T_n) &= \int_{\mathbb{X} \in \text{Sup}(\mathbb{X})} \left(T_n(\mathbb{X}) - \int_{\mathbb{X} \in \text{Sup}(\mathbb{X})} T_n(\mathbb{X}) \prod_{i=1}^n dF_n(x_i) \right)^2 \prod_{i=1}^n dF_n(x_i) \\ &= \text{Var}_*[T_n(X_1^*, \dots, X_n^*) | \mathbb{X}], \end{aligned} \quad (3.10)$$

donde $\{X_1^*, \dots, X_n^*\}$ es la muestra bootstrap.

Usando el método de Monte Carlo, obtenemos la segunda definición para el estimador de la varianza de bootstrap.

Definición 3.14. *Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = T_n(X_1, \dots, X_n)$ un estadístico. Sea B un entero positivo y sean $\{X_{i1}^*, \dots, X_{in}^*\}$, $i = 1, \dots, B$ muestras bootstrap y T_i^* los valores del estadístico en esas muestras. Se define el estimador bootstrap Monte Carlo de la varianza de T_n como*

$$v_{BOOT}^{(B)} = \frac{1}{B} \sum_{b=1}^B \left(T_b^* - \frac{1}{B} \sum_{l=1}^B T_l^* \right)^2.$$

Como se ha mencionado en la Sección 3.2, al aproximar $v_{BOOT}^{(B)}$ a v_{BOOT} por la ley fuerte de los grandes números, es suficiente probar la consistencia para v_{BOOT} .

3.4.1. Consistencia

Sea una variable aleatoria X que sigue una distribución desconocida F , y sea $(\mathbb{X} = \mathbb{x})$ una realización de una muestra aleatoria simple de X con valores fijos. Dado un estadístico T_n basado en $\mathbb{X} = (X_1, \dots, X_n)$, y una muestra bootstrap (X_1^*, \dots, X_n^*) , la varianza del estadístico $T_n^* = T_n(X_1^*, \dots, X_n^*)$, condicionada a la m.a.s. inicial viene dada por

$$\text{Var}_*(T_n^* | \mathbb{X} = \mathbb{x}) = \text{Var}_*(T_n^* - T_n | \mathbb{X} = \mathbb{x}),$$

donde la última igualdad se sigue del hecho de que T_n es invariante por traslaciones. Descondicionando, de la misma manera que en la Observación 3.6, se tiene que

$$v_{BOOT} = \text{Var}_*(T_n^* - T_n | \mathbb{X}).$$

Por tanto, el estimador de la varianza bootstrap v_{BOOT} definido en (3.10) es realmente la varianza de las variables aleatorias que siguen la distribución H_{BOOT}

en (3.2) pero sustituyendo T_n por $T_n^* - T_n$.

Por tanto, cuando fijamos $\mathbb{X} = \mathfrak{x}$, se tiene, para la métrica K , que

$$K(H_{BOOT}, H_n | \mathbb{X} = \mathfrak{x}) \rightarrow 0,$$

cuando $n \rightarrow \infty$, lo que implica, $\forall w \in \Omega$ que

$$K(H_{BOOT}, H_n | X_1(w), \dots, X_n(w)) \rightarrow 0,$$

de donde se tiene que $K(H_{BOOT}, H_n) \rightarrow_{c.s.} 0$. Esto implica que $K(H_n, H_{BOOT}) \rightarrow_d 0$.

Por otro lado, para la métrica $\tilde{\rho}_2$, se tiene que

$$\begin{aligned} \tilde{\rho}_2(H_{BOOT}(\cdot | \mathbb{X} = \mathfrak{x}), H_n) \rightarrow 0 &\iff \\ H_{BOOT}(\cdot | \mathbb{X} = \mathfrak{x}) \rightarrow_d H_n &\text{ y } H_{BOOT}(\cdot | \mathbb{X} = \mathfrak{x}) \rightarrow_{2-m} H_n, \end{aligned}$$

(ver [1], p. 1197).

Es decir, si la métrica $\tilde{\rho}_2$ es usada, la convergencia de H_{BOOT} implica la convergencia del momento de segundo orden de $T_n^* - T_n$, lo que implica que

$$v_{BOOT}/\sigma_n^2 \rightarrow_{c.s.} 1,$$

en este caso, con σ_n^2 dado en (2.14).

Por tanto, nos centraremos en probar la consistencia para la métrica K , ya que hemos visto que la convergencia en distribución de H_{BOOT} para esta métrica no implica la convergencia en momentos. Una condición suficiente para que esto se tenga, es que la sucesión sea uniformemente integrable, por el Teorema 1.9. Por tanto, mostrar la integrabilidad uniforme de $\{(\sqrt{n}(T_n^* - T_n))^2\}$ es la técnica principal usada para probar la consistencia de v_{BOOT} . El hecho de tomar el estadístico $\sqrt{n}(T_n^* - T_n)$ se debe a que, en la práctica, normalmente tenemos estadísticos de esa forma, para evitar que la varianza del estadístico se vaya a cero conforme n crece.

Una condición suficiente para la integrabilidad uniforme de una sucesión de variables aleatorias es que $\{X_n\}_{n \in \mathbb{N}}$ es que $\sup_n E|X_n|^{1+\delta} < \infty$ para $\delta > 0$ (ver [9], p. 14). Por tanto, si el estimador bootstrap H_{BOOT} es fuertemente consistente, una condición suficiente para la convergencia fuerte de v_{BOOT} es

$$E_*|\sqrt{n}(T_n^* - T_n)|^{2+\delta} = O(1) \quad c.s. \quad (3.11)$$

para un $\delta > 0$.

En el siguiente ejemplo vemos una inconsistencia de v_{BOOT} , que ayudará a ver qué condiciones hemos de exigir para que se verifique su consistencia en el Teorema 3.16.

Ejemplo 3.15. Sea (X_1, \dots, X_n) una muestra aleatoria simple siguiendo una distribución F , la cual satisface que $F(x) = 1 - x^{-1}$ si $x > 10$ y $F(x) = |x|^{-1}$, si $x < -10$. Sea $g(x) = e^{x^2}$. Vamos a probar que la varianza de bootstrap es inconsistente en este caso para $T_n = g(\bar{X}_n)$.

Para ello veremos que se tiene $E_*|\sqrt{n}(T_n^* - T_n)|^2 \rightarrow_{c.s.} \infty$, contradiciendo (3.11). Dado $X_{n:n} = \max(X_1, \dots, X_n)$, como la probabilidad de que en el remuestreo bootstrap se tome en cada variable el máximo es $1/n^n$, se tiene que

$$E_*|(g(\bar{X}_n^*) - g(\bar{X}_n))|^2 \geq \frac{1}{n^n}(g(X_{n:n}) - g(\bar{X}_n)),$$

de lo que se sigue que

$$E_*|\sqrt{n}(g(\bar{X}_n^*) - g(\bar{X}_n))|^2 \geq n^{-n+1}(g(X_{n:n}) - g(\bar{X}_n)),$$

luego bastará con probar que $n^{-(n+1)}[g(X_{n:n})]^2 \rightarrow_{c.s.} \infty$.

Para cualquier $M > 0$, operando y aplicando la definición de F se tiene

$$\begin{aligned} P\{[g(X_{n:n})]^2 < Mn^{n+1}\} &\leq P\{X_{n:n} < [\log(M^{1/2}n^{(n-1)/2})]^{1/2}\} \\ &= \{1 - [\log(M^{1/2}n^{(n-1)/2})]^{-1/2}\}^n. \end{aligned}$$

Usando que la sucesión $(1 - \frac{1}{n})^n$ converge de manera creciente a e^{-1} y acotando, se tiene que

$$\begin{aligned} P\{[g(X_{n:n})]^2 < Mn^{n+1}\} &\leq \exp\{-n[\log(M^{1/2}n^{(n-1)/2})]^{-1/2}\} \\ &\leq \exp\left\{\frac{-n}{(\frac{n-1}{2} \log n)^{1/2}}\right\} \\ &\leq e^{-n^{1/4}}, \end{aligned}$$

donde las dos últimas desigualdades se tienen para n grande. Usando ahora que, por el criterio de la raíz de convergencia de series, $\sum_{n=1}^{\infty} e^{-n^{1/4}} < \infty$, se tiene de la desigualdad anterior que

$$\sum_{n=1}^{\infty} P\{n^{-(n+1)}[g(X_{n:n})]^2 < M\} < \infty.$$

Por el Lema de Borel-Cantelli se tiene entonces que $P\{n^{-(n+1)}[g(X_{n:n})]^2 < M\} = 0$. Luego $n^{-(n+1)}[g(X_{n:n})]^2$ no está acotado y $n^{-(n+1)}[g(X_{n:n})]^2 \rightarrow_{c.s.} \infty$. Por tanto, el estimador bootstrap de la varianza es inconsistente en este caso.

Este ejemplo muestra que el estimador de la varianza de bootstrap puede diverger a infinito mientras que la varianza asintótica de T_n es finita. La inconsistencia de este estimador está causada por el hecho de que $|T_n^* - T_n|$ puede tomar valores excepcionalmente grandes. Por tanto, como veremos en el siguiente teorema, la condición siguiente permite obtener resultados de consistencia para la varianza v_{BOOT} . Consideramos la condición suficiente

$$\max_{i_1, \dots, i_n} |T_n(X_{i_1}, \dots, X_{i_n}) - T_n| / \tau_n \rightarrow_{c.s.} 0, \quad (3.12)$$

donde el máximo está tomado sobre los enteros i_1, \dots, i_n que satisfacen $1 \leq i_1 \leq \dots \leq i_n \leq n$, y $\{\tau_n\}$ es una sucesión de números positivos satisfaciendo $\liminf_n \tau_n > 0$ y $\tau_n = O(e^{n^q})$ con $q \in (0, \frac{1}{4})$.

La condición usada para poder garantizar la consistencia fuerte de v_{BOOT} se ha extraído de [10], p.87. En el siguiente teorema probamos que el estimador v_{BOOT} es fuertemente consistente.

Teorema 3.16. *Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y $T_n = g(\bar{X}_n)$ un estadístico. Supongamos que (3.12) se cumple y que g es continua y diferenciable en un entorno de $\mu = E(X_1)$, con $g'(\mu) \neq 0$. Entonces*

$$v_{BOOT} / \sigma_n^2 \rightarrow_{c.s.} 1,$$

donde $\sigma_n^2 = n^{-1} g'(\mu)^2 \text{Var}(X)$.

Demostración. Para probar que el estimador v_{BOOT} es fuertemente consistente es suficiente probar (3.11). Sea τ_n dado por (3.12) y

$$\Delta_n^* = \begin{cases} \tau_n & \text{si } T_n^* - T_n > \tau_n \\ T_n^* - T_n & \text{si } |T_n^* - T_n| \leq \tau_n \\ -\tau_n & \text{si } T_n^* - T_n < -\tau_n \end{cases} \quad (3.13)$$

De la condición (3.12), para n suficientemente grande se tiene que $|T_n^* - T_n| < \tau_n$, lo que implica que

$$P \{E_* |T_n^* - T_n|^{2+\delta} = E_* |\Delta_n^*|^{2+\delta}\} = 1,$$

con n suficientemente grande. Por tanto, para probar que el estimador de la varianza es fuertemente consistente, es suficiente probar, por (3.12), que para algún $\delta > 0$, se verifica que

$$E_* |\sqrt{n} \Delta_n^*|^{2+\delta} = O(1) \quad c.s. \quad (3.14)$$

Vamos a probar (3.14) para $\delta = 2$. Como g' es continua en un entorno de μ , $\forall \varepsilon > 0$, existen constantes positivas η y M tales que

$$g'(x)^2 \leq (\varepsilon + g'(\mu))^2 = M \quad \text{si } |x - \mu| \leq 2\eta, \quad (3.15)$$

y por el Teorema 1.12, se tiene que

$$\bar{X}_n \rightarrow_{c.s.} \mu. \quad (3.16)$$

Por otra parte se tiene la expresión

$$\begin{aligned} \frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X}_n)^4 &= \frac{1}{n^2} \sum_{i=1}^n (X_i - E(X_i) + E(X_i) - \bar{X}_n)^4 \\ &\leq \frac{2^4}{n^2} \sum_{i=1}^n \left[(X_i - E(X_i))^4 + (E(X_i) - \bar{X}_n)^4 \right] \\ &\rightarrow_{c.s.} \frac{16}{n^2} \sum_{i=1}^n (X_i - E(X_i))^4, \end{aligned}$$

donde la primera desigualdad resulta de (3.5) y la convergencia se debe a (3.16).

Como $(X_i - E(X_i))^2 < \infty$, podemos aplicar entonces el Teorema 1.13 a las variables $(X_i - E(X_i))^4$, para $\lambda = 1/2$, y se tiene por tanto que

$$\frac{16}{n^2} \sum_{i=1}^n (X_i - E(X_i))^4 \rightarrow_{c.s.} 0,$$

lo que implica

$$\frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X}_n)^4 \rightarrow_{c.s.} 0. \quad (3.17)$$

Sea X_1, X_2, \dots una sucesión de variables aleatorias tal que (3.16) y (3.17) se cumplen. Entonces, para n grande, al ser $|\Delta_n^*| \leq \tau_n$ por (3.13), se tiene que

$$\begin{aligned} E_* |\Delta_n^*|^4 &= E_* |\Delta_n^*|^4 I \left\{ |\bar{X}_n^* - \bar{X}_n| \leq \eta \right\} + E_* |\Delta_n^*|^4 I \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\} \\ &\leq E_* |T_n^* - T_n|^4 I \left\{ |\bar{X}_n^* - \bar{X}_n| \leq \eta \right\} + \tau_n^4 E_* I \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\}. \end{aligned} \quad (3.18)$$

La otra parte de la desigualdad se verifica porque

$$\{|T_n^* - T_n| > \tau_n\} \Rightarrow \{|\bar{X}_n^* - \bar{X}_n| > \tau_n\},$$

lo que implica $\{|\bar{X}_n^* - \bar{X}_n| \leq \tau_n\} \Rightarrow \{|T_n^* - T_n| \leq \tau_n\}$. En efecto, sea $w \in \Omega$ tal que $|T_n^*(w) - T_n(w)| > \tau_n$ y supongamos que $|\bar{X}_n^*(w) - \bar{X}_n(w)| \leq \eta$.

Por otro lado, por el Teorema del Valor Medio, se verifica

$$T_n^* - T_n = g'(\xi_n^*) (\bar{X}_n^* - \bar{X}_n),$$

lo que implica que

$$|g'(\xi_n^*) (\bar{X}_n^*(w) - \bar{X}_n(w))| > \tau_n,$$

y por tanto

$$|g'(\xi_n^*)| > \frac{\tau_n}{|\bar{X}_n^*(w) - \bar{X}_n(w)|} > \frac{\tau_n}{\eta},$$

lo que no puede ser posible debido a que $|g'(\xi_n^*)|$ es finito, y tomando n suficientemente grande su valor no estaría acotado, al ser $\tau_n = O(e^{n^q})$, $q \in (0, \frac{1}{4})$.

Por otro lado, para n grande se tiene $|\bar{X}_n - \mu| < \eta$, lo que implica

$$|\xi_n^* - \mu| \leq |\bar{X}_n - \mu| + |\xi_n^* - \bar{X}_n| \leq \eta + |\bar{X}_n^* - \bar{X}_n| \leq 2\eta. \quad (3.19)$$

Por tanto, de (3.15) y (3.19), tenemos que $g'(\xi_n^*)^4 < M^2$ y, utilizando la definición de esperanza de una función indicador, podemos finalmente acotar $E_* |\Delta_n^*|^4$ en (3.18) por

$$E_* |\Delta_n^*|^4 \leq M^2 E_* |\bar{X}_n^* - \bar{X}_n|^4 + \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\}.$$

Por un lado, llamando $Y_i^* = X_i^* - \bar{X}_n$, y siendo \bar{Y}_n^* la media muestral de estas variables, se tiene que

$$\left(\bar{X}_n^* - \bar{X}_n\right)^4 = \left(\frac{\sum_{i=1}^n (X_i - \bar{X}_n)}{n}\right)^4 = \left(\bar{Y}_n^*\right)^4,$$

con $E(Y_i^*) = 0$. Usando entonces la fórmula de los momentos muestrales a la cuarta (ver [8], pág. 315) se tiene que $E_* \left(\bar{X}_n^* - \bar{X}_n\right)^4 = O(n^{-2})$, y queda por probar que

$$n^2 \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\} = O(1) \quad \text{c.s.} \quad (3.20)$$

Notemos que las X_1^*, \dots, X_n^* son independientes e idénticamente distribuidas con media \bar{X}_n y $|X_i^* - \bar{X}_n| \leq Y$, donde $Y = |\max_{i \leq n} X_i| + |\min_{i \leq n} X_i|$. Por tanto, por la desigualdad de Bernstein (ver [9], pág.95), se tiene

$$P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\} \leq c \cdot \exp \left[- \frac{n\eta^2}{2(W + \eta Y)} \right],$$

donde c es una constante y

$$W = \text{Var}_*(X_i^*) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Por tanto, por definición de τ_n se tiene que

$$n^2 \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\} \leq C n^2 \exp \left[n^{4q} - \frac{n\eta^2}{2(W + \eta Y)} \right].$$

Como, fijado $\mathbb{X} = \mathbb{x}$, se tiene que $2(W + \eta Y)$ es constante, entonces, por ser $q < 1/4$, se verifica que

$$n^2 \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta | \mathbb{X} = \mathbb{x} \right\} \rightarrow 0,$$

cuando $n \rightarrow \infty$, al converger la exponencial más rápido que el término n^2 . Por lo tanto, se tiene, para todo w ,

$$n^2 \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta, X_1(w), \dots, X_n(w) \right\} = O(1),$$

de donde finalmente se tiene que

$$n^2 \tau_n^4 P_* \left\{ |\bar{X}_n^* - \bar{X}_n| > \eta \right\} = O(1) \quad \text{c.s.},$$

lo que completa la prueba del teorema. □

3.5. Otras consideraciones

3.5.1. Situaciones donde el método bootstrap no se puede aplicar

A pesar de los teoremas de consistencia estudiados anteriormente, hay ocasiones en las que el método bootstrap basado en muestreo con reemplazamiento de la distribución empírica F_n no funciona. Veamos un par de ejemplos de estas situaciones.

Una de las ocasiones en la que el método bootstrap falla es a la hora de estimar la distribución de T_n cuando $T_n = \sqrt{n}(g(\bar{X}_n) - g(\mu))$ y g no es derivable.

Ejemplo 3.17. Sea (X_1, \dots, X_n) una muestra aleatoria simple de una variable aleatoria X con distribución F y $\sigma^2 = \text{Var}(X)$. Sea $g(x) = |x|$ y $T_n = \sqrt{n}(g(\bar{X}_n) - g(\mu))$. Si el valor de μ es 0, entonces por el Teorema Central del Límite para \bar{X}_n y por el Teorema 1.7, se tiene que $T \rightarrow_d |Z| \sim N(0, \sigma^2)$. Para probar que el bootstrap no funciona en este caso, primero tenemos que hacer dos observaciones:

1. Para casi todas las sucesiones $\{X_1, \dots, X_n\}$, la distribución condicional de $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$, dado \bar{X}_n , converge en distribución a $N(0, \sigma^2)$ por el Teorema Central del Límite de las sucesiones triangulares, obtenido de [12].
2. Se tiene que $(\sqrt{n}(\bar{X}_n - \mu), \sqrt{n}(\bar{X}_n^* - \bar{X}_n)) \rightarrow_d (Z_1, Z_2)$, donde Z_1, Z_2 son independientes e idénticamente distribuidas siguiendo una distribución $N(0, \sigma^2)$.

Por tanto, cuando el valor de μ es 0,

$$\begin{aligned} T_n^* &= \sqrt{n}(|\bar{X}_n^*| - |\bar{X}_n|) = |\sqrt{n}(\bar{X}_n^* - \bar{X}_n) + \sqrt{n}\bar{X}_n| - |\sqrt{n}\bar{X}_n| \\ &\rightarrow_d |Z_1 + Z_2| - |Z_1|, \end{aligned}$$

donde Z_1, Z_2 son independientes e idénticamente distribuidas siguiendo una distribución $N(0, \sigma^2)$. Pero esta suma de variables aleatorias no sigue una distribución $|N(0, \sigma^2)|$. Por tanto, el bootstrap no es consistente en este caso, cuando $\mu = 0$.

Otro caso de inconsistencia del método bootstrap es cuando tenemos una muestra aleatoria simple que sigue una distribución uniforme en el intervalo $(0, \theta)$, es decir $X \sim U(0, \theta)$, y un estadístico $T_n = n(\theta - X_{n:n})$, como vemos en el siguiente ejemplo.

Ejemplo 3.18. Sea (X_1, \dots, X_n) una muestra aleatoria simple de una variable aleatoria $X \sim U(0, \theta)$, y sea $T_n = n(\theta - X_{n:n})$, $T_n^* = n(X_{n:n} - X_{n:n}^*)$. El bootstrap falla en este ejemplo en el sentido de que la distribución condicional de T_n^* dado $X_{n:n}$ no converge casi seguramente a $\text{Exp}(\theta)$. Supongamos $\theta = 1$. Entonces, para $t \geq 0$,

$$\begin{aligned} P_*(T_n^* \leq t) &\geq P_*(T_n^* = 0) = P_*(X_{n:n}^* = X_{n:n}) \\ &= 1 - P_*(X_{n:n}^* < X_{n:n}) = 1 - \left(\frac{n-1}{n}\right)^n \\ &\rightarrow 1 - e^{-1} \quad \text{cuando } n \rightarrow \infty, \end{aligned}$$

Veamos ahora la distribución de T_n . Por un lado, sabemos que la distribución de $X = F(x) = \left(\frac{x}{\theta}\right)$ si $0 < x < \theta$, y la distribución de $Y = X_{n:n}$ es $G(x) = (F(x))^n$, por tanto

$$P(T_n \leq t) = P(n(1 - Y) \leq t) = P\left(Y \leq 1 - \frac{t}{n}\right) = 1 - \left(1 - \frac{t}{n}\right)^n.$$

Tomando $T_n = 0.0001$, se tiene que

$$\begin{aligned}\lim_n P_*(T_n^* \leq t) &\geq 1 - e^{-1}, \\ \lim_n P(T_n \leq t) &= 1 - e^{-0.0001} \approx 0.\end{aligned}$$

Por tanto, $P_*(T_n^* \leq t) \not\rightarrow P(T_n \leq t)$, y el bootstrap no es consistente tampoco en este caso.

3.5.2. Intervalos de confianza bootstrap

Hasta ahora hemos concentrado nuestra atención en estimar la varianza y la distribución de estadísticos, los cuales pueden ser estimadores de parámetros. En muchas situaciones esto es lo que se necesita en la práctica, sin embargo, en muchos casos es importante tener un intervalo de confianza que incluya al parámetro que queremos estimar con una probabilidad alta. El objetivo de esta sección es hacer una introducción, ya que por cuestiones de tiempo y espacio no es posible hacer un desarrollo exhaustivo. Recordamos la definición de intervalo de confianza a continuación.

Definición 3.19. Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y sea θ un parámetro de interés asociado a F . Si $I_n = I_n(X_1, \dots, X_n)$ y $S_n = S_n(X_1, \dots, X_n)$ son dos estadísticos dependientes únicamente de (X_1, \dots, X_n) que verifican

$$P\{I_n \leq \theta \leq S_n\} \geq 1 - \alpha,$$

donde α es una constante que satisface $0 < \alpha < 1$, entonces, (I_n, S_n) se dice que es un intervalo de confianza para θ de nivel de confianza $1 - \alpha$.

En muchos casos, los intervalos de confianza se construyen considerando una función pivote $T_n = T_n(X_1, \dots, X_n, \theta)$ cuya distribución G_n es conocida e independiente de θ .

Definición 3.20. Sea X una variable aleatoria con función de distribución F , de parámetro $\theta \in \Omega \subseteq \mathbb{R}$ y (X_1, \dots, X_n) una muestra aleatoria simple de X . Sea $T : \Psi \times \Omega \rightarrow \mathbb{R}$, se dice que T_n es una función pivote para θ si

1. $\forall x \in \Psi, T_n(x, \theta)$ es una función estrictamente monótona en θ .
2. La distribución de $T_n(x, \theta)$ es independiente de θ .

Por ejemplo, si $T_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$, donde $\hat{\theta}_n$ es un estimador de θ obtenido por ejemplo mediante el método de los momentos y $\hat{\sigma}_n^2$ es el estimador de la varianza de $\hat{\theta}_n$, un intervalo de confianza exacto para θ sería de la forma

$$\left[\hat{\theta}_n - \hat{\sigma}_n G_n^{-1}(1 - \alpha), \hat{\theta}_n - \hat{\sigma}_n G_n^{-1}(\alpha) \right].$$

El problema que se plantea es que no suele ser fácil encontrar una función pivote con una distribución conocida G_n . Si G_n es desconocida, entonces el intervalo anterior no puede usarse como intervalo de confianza y debemos aproximar G_n . El bootstrap se puede aplicar en este caso para obtener un intervalo de confianza simplemente reemplazando G_n con su estimador bootstrap G_{BOOT} .

Este apartado está dedicado a la revisión de algunos de los intervalos de confianza bootstrap más utilizados, y posteriormente estudiaremos la consistencia de cada uno de ellos. Veamos primero cada uno de los tipos de intervalos por separado.

1) Bootstrap-t.

El método *bootstrap-t* tiene como idea fundamental considerar un pivote de la forma $T_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$. Si la distribución G_n de T_n es desconocida, entonces se aproxima usando el estimador bootstrap de la distribución dado por la Definición 3.2, llamándose en este caso G_{BOOT} .

Definición 3.21. Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y sea θ un parámetro de interés asociado a F . Sea X_1^*, \dots, X_n^* una muestra bootstrap y sea también $T_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$, una función pivote para θ con distribución desconocida G_n , donde $\hat{\theta}_n$ es un estimador de θ y $\hat{\sigma}_n^2$ es el estimador de la varianza de $\hat{\theta}_n$. Sea $G_{BOOT} = P_*\{T_n^* \leq x | X_1, \dots, X_n\}$ la distribución bootstrap de T_n , con $T_n^* = (\hat{\theta}_n^* - \hat{\theta})/\hat{\sigma}_n^*$, $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, \dots, X_n^*)$ y $\hat{\sigma}_n^* = \hat{\sigma}_n(X_1^*, \dots, X_n^*)$.

Se define el intervalo de confianza de bootstrap-t como (I_{BT}, S_{BT}) , donde

$$I_{BT} = \hat{\theta}_n - \hat{\sigma}_n G_{BOOT}^{-1}(1 - \alpha), \quad S_{BT} = \hat{\theta}_n - \hat{\sigma}_n G_{BOOT}^{-1}(\alpha). \quad (3.21)$$

2) Bootstrap percentil.

Este método es completamente distinto del anterior. En este caso se toma la distribución

$$K_{BOOT}(x) = P_*\{\hat{\theta}_n^* \leq x\}, \quad (3.22)$$

siendo $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, \dots, X_n^*)$, y $\hat{\theta}_n$ un estimador del parámetro θ , como en el método anterior. Una vez obtenida esa distribución, el método se basa en tomar los percentiles en α y $1 - \alpha$ de esa distribución como intervalo de confianza.

Definición 3.22. Sea X una variable aleatoria que sigue una distribución desconocida F , (X_1, \dots, X_n) una muestra aleatoria simple de X , θ un parámetro de F , X_1^*, \dots, X_n^* una muestra bootstrap y sea $\hat{\theta}_n$ un estimador de θ . Se define el intervalo de confianza bootstrap percentil como (I_{BP}, S_{BP}) , donde

$$I_{BP} = K_{BOOT}^{-1}(\alpha), \quad S_{BP} = K_{BOOT}^{-1}(1 - \alpha).$$

3) Bootstrap híbrido.

Pasamos a considerar también los intervalos de confianza *bootstrap híbrido*. En este caso, tomamos como distribución de bootstrap

$$H_{BOOT}(x) = P_*\{n^l(\hat{\theta}_n^* - \hat{\theta}_n) \leq x\}, \quad (3.23)$$

donde l es una constante fija, normalmente $l = 1/2$. Por los resultados del Capítulo 3, sabemos que H_{BOOT} es una buena aproximación a la distribución de $n^l(\hat{\theta}_n - \theta)$, y podemos escribir, por tanto, la siguiente aproximación:

$$H_{BOOT}(H_{BOOT}^{-1}(1 - \alpha)) \approx 1 - \alpha \approx P\{n^l(\hat{\theta}_n - \theta) \leq H_{BOOT}^{-1}(1 - \alpha)\}.$$

Entonces el intervalo de confianza híbrido de bootstrap se define como:

Definición 3.23. Sea X una variable aleatoria que sigue una distribución desconocida F , sea (X_1, \dots, X_n) una muestra aleatoria simple de X y sea θ un parámetro de interés asociado a F . Sea X_1^*, \dots, X_n^* una muestra bootstrap y sea H_{BOOT} la distribución de bootstrap de $n^l(\hat{\theta}_n - \theta)$ basada en X_1^*, \dots, X_n^* dada por (3.23), donde $\hat{\theta}_n$ es un estimador de θ . Se define el intervalo de confianza híbrido como (I_{HB}, S_{HB}) , donde

$$I_{HB} = \hat{\theta}_n - n^{-l}H_{BOOT}^{-1}(1 - \alpha), \quad S_{HB} = \hat{\theta}_n - n^{-l}H_{BOOT}^{-1}(\alpha). \quad (3.24)$$

Este método trata el percentil de la distribución bootstrap H_{BOOT} como el percentil de H_n , que es la distribución de $n^l(\hat{\theta}_n - \theta)$. Por otro lado, en este caso $n^l(\hat{\theta}_n - \theta)$ hace de función pivote, como en el método *bootstrap-t*. Por ambos motivos a este método se le conoce como el *bootstrap híbrido*.

Pasamos ahora a estudiar la consistencia de los intervalos de confianza que hemos visto, que se basa en ver si la probabilidad de que el intervalo aleatorio cubra al parámetro converge al nivel de confianza cuando $n \rightarrow \infty$, como se observa en la siguiente definición.

Definición 3.24. Un intervalo de confianza $\{I_n, S_n\}$ para un parámetro θ con nivel de confianza $1 - \alpha$ es consistente si, cuando $n \rightarrow \infty$,

$$P\{I_n \leq \theta \leq S_n\} \rightarrow 1 - \alpha.$$

De ahora en adelante, para probar los resultados convenientes, centraremos nuestra atención únicamente en el extremo inferior del intervalo de confianza de bootstrap que estamos tratando, ya que la prueba para el extremo superior será análoga a la de este.

Teorema 3.25. Sea (X_1, \dots, X_n) una muestra aleatoria simple de una distribución desconocida F y sea θ un parámetro de interés asociado a F . Sean H_n y G_n las distribuciones de $n^l(\hat{\theta}_n - \theta)$ y de $(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, respectivamente, ambas continuas y estrictamente crecientes.

Sea X_1^*, \dots, X_n^* una muestra aleatoria simple de F_n y sean H_{BOOT} y G_{BOOT} las distribuciones de bootstrap de H_n y G_n , respectivamente. Entonces se cumple:

- (i) Si G_{BOOT} es fuertemente consistente para la métrica K en el sentido de la Definición 3.5, entonces I_{BT} es consistente.
- (ii) Si H_{BOOT} es fuertemente consistente para la métrica K en el sentido de la Definición 3.5, entonces I_{HB} es consistente.
- (iii) Si H_{BOOT} es fuertemente consistente para la métrica K en el sentido de la Definición 3.5, y además

$$\lim_{n \rightarrow \infty} K(H_n, H) = 0, \quad (3.25)$$

para una función de distribución H continua, estrictamente creciente y simétrica, entonces I_{BP} es consistente.

Demostración. Empezamos probando (i). Por un lado, de (3.21) se tiene

$$P(I_{BT} \leq \theta) = P\left(\hat{\theta}_n - \hat{\sigma}_n G_{BOOT}^{-1}(1 - \alpha) \leq \theta\right) = P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq G_{BOOT}^{-1}(1 - \alpha)\right).$$

Como $G_{BOOT}^{-1} \rightarrow_{c.s.} G_n^{-1}$, por la consistencia de G_{BOOT} , el Teorema 1.15, Teorema 1.7 y la continuidad de G_n tenemos que

$$P(I_{BT} \leq \theta) = G_n(G_{BOOT}^{-1}(1 - \alpha)) \rightarrow_{c.s.} G_n(G_n^{-1}(1 - \alpha)) = 1 - \alpha,$$

Para probar (ii), utilizando (3.24) y por un razonamiento análogo al caso anterior, se tiene que

$$P(I_{HB} \leq \theta) = P\left(\hat{\theta}_n - n^{-l} H_{BOOT}^{-1}(1 - \alpha) \leq \theta\right) = P\left(n^l(\hat{\theta}_n - \theta) \leq H_{BOOT}^{-1}(1 - \alpha)\right),$$

y otra vez usando el Teorema 1.15, el Teorema 1.7 y la continuidad de H_n obtenemos que

$$P(I_{HB} \leq \theta) = H_n(H_{BOOT}^{-1}(1 - \alpha)) \rightarrow_{c.s.} H_n(H_n^{-1}(1 - \alpha)) = 1 - \alpha,$$

Por último, veamos (iii). Por un lado, por (3.22) tenemos

$$P(I_{BP} \leq \theta) = P(K_{BOOT}^{-1}(\alpha) \leq \theta) = P(\alpha \leq K_{BOOT}(\theta)),$$

y como

$$\begin{aligned} K_{BOOT}(\theta) &= P_*\left(\hat{\theta}_n^* \leq \theta\right) = P_*\left(n^l\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \leq n^l\left(\theta - \hat{\theta}_n\right)\right) \\ &= H_{BOOT}\left(n^l\left(\theta - \hat{\theta}_n\right)\right), \end{aligned}$$

se tiene que

$$P(I_{BP} \leq \theta) = P\left(\alpha \leq H_{BOOT}\left(n^l\left(\theta - \hat{\theta}_n\right)\right)\right) = P\left(n^l\left(\hat{\theta}_n - \theta\right) \leq -H_{BOOT}^{-1}(\alpha)\right).$$

Usando ahora la definición de H_n y que, por la consistencia fuerte de H_{BOOT} y por (3.25), se tiene $H_{BOOT} \rightarrow_{c.s.} H$, se tiene por el Teorema 1.7 que

$$P(I_{BP} \leq \theta) = H_n(-H_{BOOT}^{-1}(\alpha)) \rightarrow_{c.s.} H_n(-H^{-1}(\alpha)),$$

y usando de nuevo (3.25) y la simetría y la continuidad de H , obtenemos el resultado

$$P(I_{BP} \leq \theta) \rightarrow 1 - H(H^{-1}(\alpha)) = 1 - \alpha.$$

□

3.6. Conclusiones

Hemos estudiado en este capítulo el método de remuestreo bootstrap como estimador de la varianza y la distribución de un estadístico basado en una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria X , así como varias técnicas para obtener intervalos de confianza alrededor de un parámetro de la distribución asociada a X . Hemos comenzado viendo en qué principios se basa el método bootstrap y el importante rol que juega la función de distribución empírica para conseguir las estimaciones. A continuación se ha descrito formalmente la estimación bootstrap de la distribución de un estadístico, probando una serie de resultados importantes sobre la consistencia del mismo usando las métricas K y $\tilde{\rho}_2$. Tras estudiar la distribución, hemos definido el estimador bootstrap de la varianza de un estadístico y hemos probado un importante resultado de consistencia para estadísticos de la forma $g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y diferenciable en μ . Se han mostrado situaciones en las que el método bootstrap no es consistente y, por último, se han estudiado tres tipos de intervalos de confianza bootstrap, alrededor de un parámetro determinado de la distribución de partida, probando la consistencia de cada uno de ellos.

Capítulo 4

Caso práctico

4.1. Introducción

Una vez vistos los resultados teóricos sobre la estimación de la varianza y la distribución de un estadístico con los métodos jackknife y bootstrap, es interesante analizar de manera práctica estos métodos usando un caso particular.

Una de las herramientas más convenientes y populares que se pueden utilizar para realizar análisis estadísticos es el lenguaje de programación R, que es un entorno de código abierto y gratuito. Contiene una gran cantidad de librerías que ayudan al usuario en tareas tan esenciales como el tratamiento de grandes volúmenes de datos, la generación de gráficas y la escritura de código que se ajuste a sus necesidades.

En este capítulo se va a utilizar el lenguaje R para realizar una serie de pruebas que permitan analizar de manera práctica los resultados vistos en los capítulos 2 y 3. Para ello partiremos de una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria X que sigue una distribución exponencial de parámetro $1/\theta$. La razón de usar una distribución exponencial, como veremos, se debe a las propiedades que tiene la media muestral para este caso.

El capítulo se divide principalmente en dos estudios. El primero corresponde a la comparación de los métodos jackknife y bootstrap al estimar la varianza y la distribución (solo en el caso del bootstrap) para el caso del estadístico \bar{X}_n basado en (X_1, \dots, X_n) . Primero se verá la estimación de la varianza de ambos métodos y después se continuará con la estimación bootstrap de la distribución en el muestreo de \bar{X}_n .

Por otro lado, en el segundo estudio se analiza un estadístico más complejo, del tipo $g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua y derivable en $\mu = E(X)$. La función g elegida ha sido la exponencial. En este estudio se comprobarán en la práctica los resultados 2.9 y 3.16, comparando la estimación jackknife de la varianza con el valor de σ_n^2 dado en (2.14).

4.2. Estudio de la media muestral para una distribución exponencial

En esta sección se parte de una muestra aleatoria simple (X_1, \dots, X_n) de una variable aleatoria $X \sim Exp(1/\theta)$. El estadístico elegido para realizar las pruebas en este caso ha sido la media muestral, \bar{X}_n .

Una variable aleatoria X sigue una distribución exponencial de parámetro $1/\theta$ si su función de densidad es de la forma

$$f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}, \quad x \in (0, \infty).$$

Además, su función característica viene dada por

$$\phi_X(t) = \left(1 - \frac{it}{1/\theta}\right).$$

El hecho de elegir una distribución de este tipo se debe a las interesantes propiedades que tiene la media muestral en el caso de la exponencial. Veamos primero el cálculo de su distribución, la cual conlleva dos pasos:

- i) Por un lado, se calcula la distribución del estadístico $S = \sum_{j=1}^n X_j$ usando la conocida propiedad de la función característica:

$$\phi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t),$$

lo que implica que la función característica del estadístico S se corresponde con

$$\phi_S(t) = \prod_{j=1}^n \left(1 - \frac{it}{1/\theta}\right) = \left(1 - \frac{it}{1/\theta}\right)^n,$$

la cual es la función característica de una función de distribución gamma de parámetros $1/\theta$ y n . Es decir, $S \sim \Gamma(\frac{1}{\theta}, n)$.

- ii) Considerando ahora la variable $\bar{X}_n = \frac{S}{n}$, se obtiene que su distribución es $\bar{X}_n \sim \Gamma\left(\frac{n}{\theta}, n\right)$.

Gracias a que sabemos que la media muestral sigue una distribución gamma, podemos obtener también su esperanza y su varianza:

- i) $E(\bar{X}_n) = \theta$.
ii) $Var(\bar{X}_n) = \frac{\theta^2}{n}$.

Estas propiedades serán de utilidad a lo largo del capítulo. Al ser conocidos todos estos valores, podremos comparar nuestros resultados aproximados del método jackknife y el bootstrap con los valores exactos de la distribución en el muestreo y la varianza de la media muestral.

En la siguiente sección mostraremos una comparación de las estimaciones de la varianza jackknife y bootstrap con el valor real de la varianza de la media muestral.

4.2.1. Estimación jackknife y bootstrap de la varianza

Para llevar a cabo el estudio de las estimaciones jackknife y bootstrap de la varianza frente a su valor exacto, generamos utilizando el software R cuatro distribuciones exponenciales de parámetros $\theta = 2, 1, 1/2$ y $1/4$. Dadas estas distribuciones, se ha variado el tamaño de muestra n para obtener una comparación de las varianzas en cada par de valores (n, θ) , siendo los tamaños elegidos $n = 100, 200, 500$ y 750 .

Además, a la hora de obtener la varianza de bootstrap para cada posible par $(n, 1/\theta)$, se han realizado cuatro replicaciones bootstrap de tamaños $B = 1000, 2500, 5000$ y 10000 , para poder contrastar la diferencia entre la varianza de bootstrap y la varianza real en función del valor de B .

El objetivo ha sido realizar una gráfica para cada posible par (n, θ) , siendo el eje x los posibles valores de B , que permita configurar las estimaciones de ambos métodos. En ella, se muestra en negro el valor de la varianza bootstrap para cada número de replicaciones, acompañado de dos líneas horizontales que muestran la varianza exacta de la media muestral (azul) y la estimación jackknife (rojo).

El hecho de generar una gráfica para cada n y cada θ hace que obtengamos 16 gráficas en total. Por ello hemos considerado agruparlas de cuatro en cuatro y que cada figura represente la comparación de la varianza para cada valor de θ .

Como se puede observar en las figuras 4.1, 4.2, 4.3 y 4.4, concluimos que la estimación bootstrap ofrece en general, mejores aproximaciones que el método jackknife, siendo este más efectivo cuanto mayor es el tamaño muestral. Se observa también que el método bootstrap, conforme aumenta el tamaño de muestra inicial, necesita más replicaciones para acercarse lo suficiente al valor exacto de la varianza.

En el siguiente apartado se realiza la comparación entre la estimación bootstrap de la distribución de \bar{X}_n frente al valor real de esta distribución, que sabemos que se corresponde con una distribución $\Gamma\left(\frac{n}{\theta}, n\right)$.

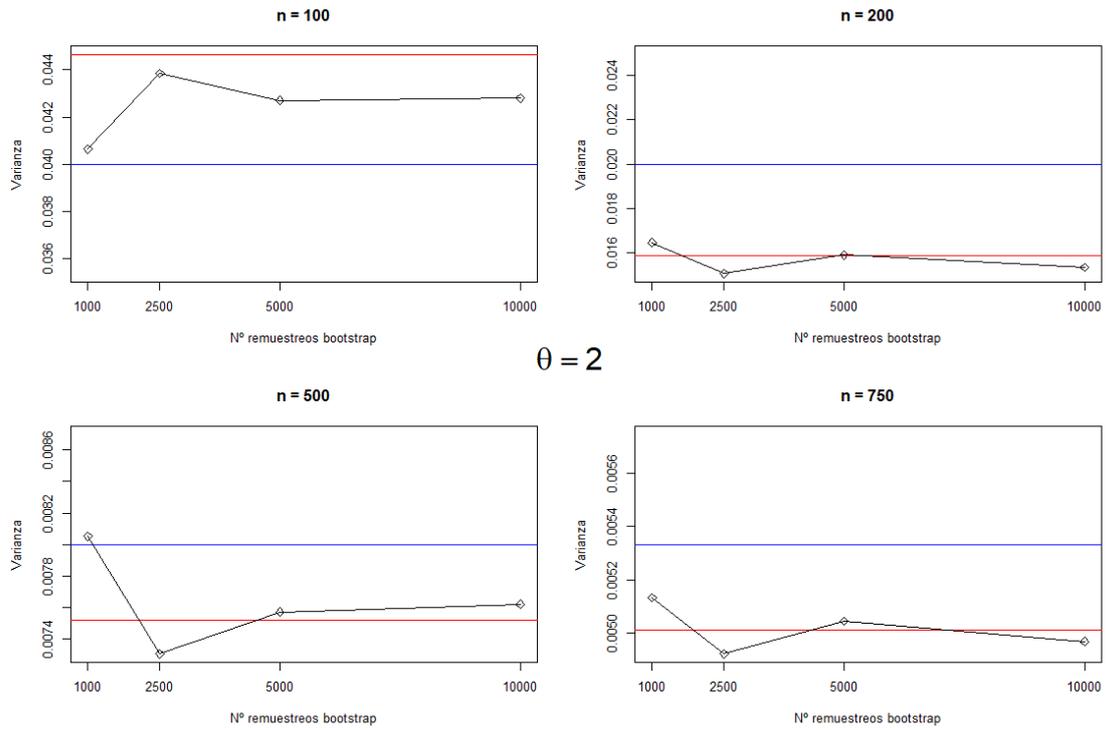


Figura 4.1: Comparación de la estimación de la varianza jackknife (azul) y bootstrap (negro) de la media muestral con respecto a la varianza real (rojo) para el caso $\theta = 2$.

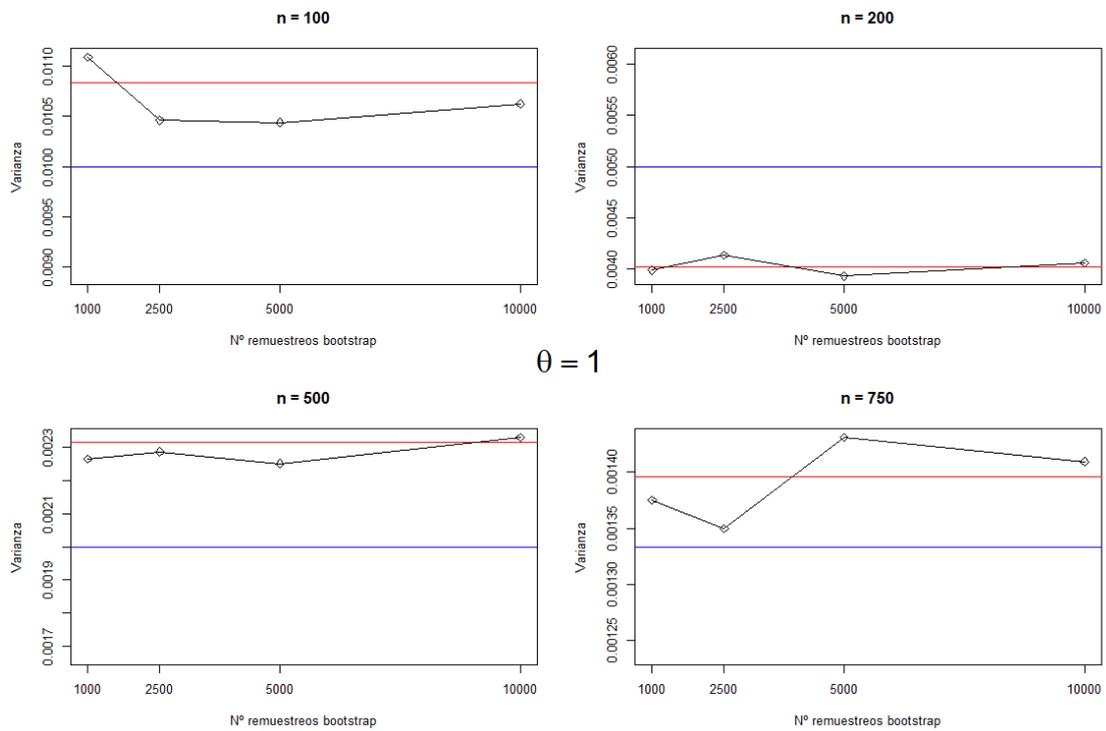


Figura 4.2: Comparación de la estimación de la varianza jackknife (azul) y bootstrap (negro) de la media muestral con respecto a la varianza real (rojo) para el caso $\theta = 1$.

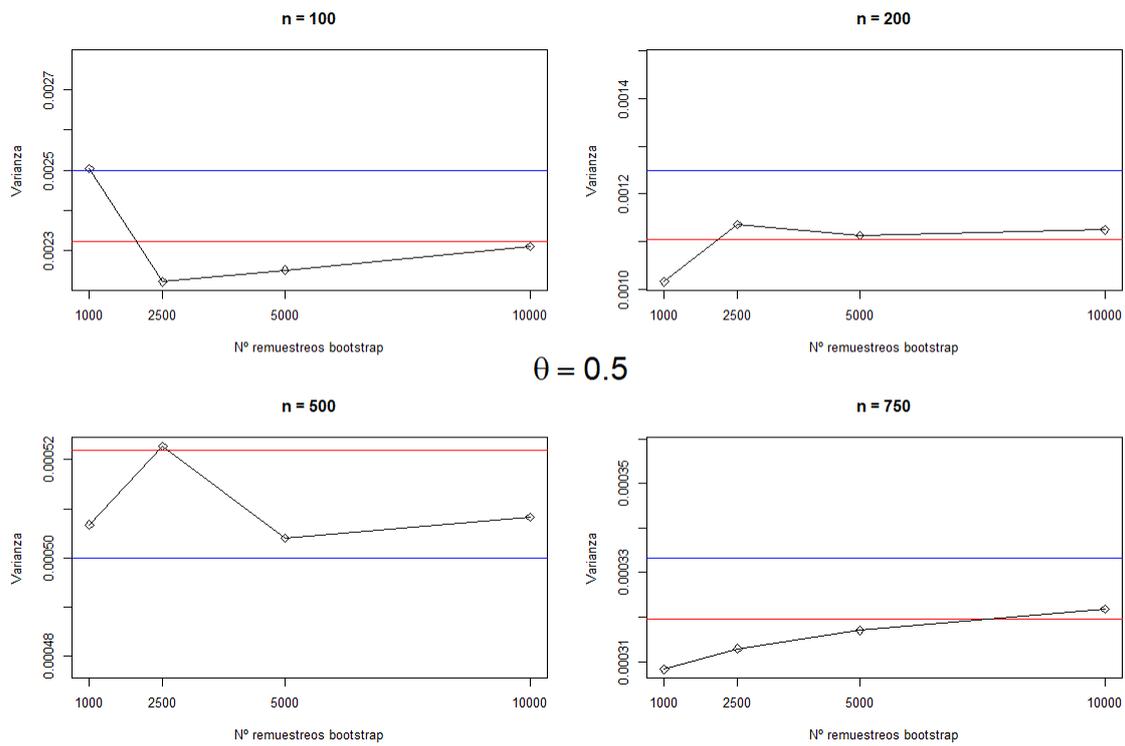


Figura 4.3: Comparación de la estimación de la varianza jackknife (azul) y bootstrap (negro) de la media muestral con respecto a la varianza real (rojo) para $\theta = 1/2$.

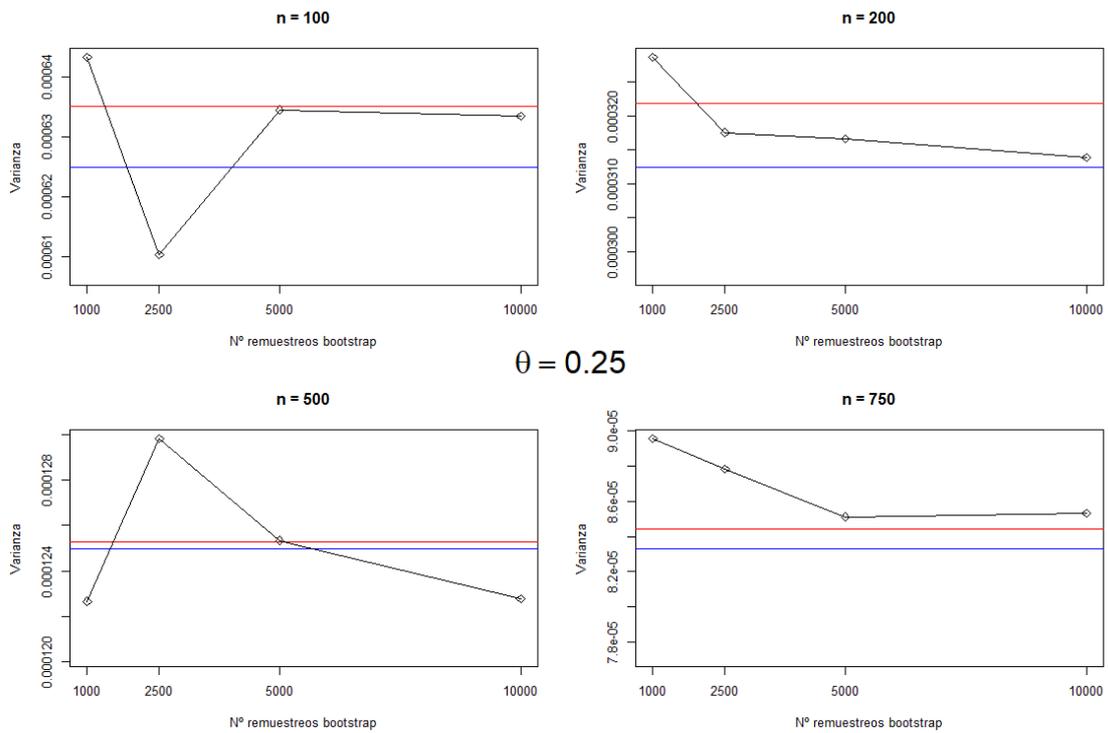


Figura 4.4: Comparación de la estimación de la varianza jackknife (azul) y bootstrap (negro) de la media muestral con respecto a la varianza real (rojo) para $\theta = 1/4$.

4.2.2. Estimación bootstrap de la distribución en el muestreo

El objetivo de este apartado es mostrar gráficamente la comparación entre el valor real de la distribución de la media muestral y la estimación H_{BOOT} que ofrece el método bootstrap. De igual manera que en el apartado anterior, se irán variando los valores de θ, n y B para observar las comparaciones en diferentes casos.

A través de \mathbb{R} podemos representar la distribución gamma asociada a los valores de los estadísticos $T_b^* = T_n(X_{1b}^*, \dots, X_{nb}^*), b = 1, \dots, B$, siendo T_n la media muestral. Sin embargo, por cada posible valor de B la distribución gamma asociada al vector $\{T_b^*\}$ es distinta, luego no podemos mostrar la distribución H_{BOOT} correspondiente a cada valor de B en una misma gráfica.

Por esta razón, si hiciéramos como en el apartado anterior y tomáramos cuatro valores de θ , cuatro valores de n y cuatro valores de B se tendrían 64 gráficas, un número excesivamente alto. Por ello, se ha decidido tomar únicamente dos valores para n y dos para B , manteniendo los cuatro valores de θ , lo que genera 16 gráficas. Volviendo a agruparlas de 4 en 4 obtenemos las figuras 4.5, 4.6, 4.7 y 4.8. Los valores elegidos han sido $\theta = 5, 10, 15$ y 20 , $n = 100$ y 500 , y $B = 1000$ y 5000 .

Además, en cada una de las gráficas se ha añadido también el valor del error cuadrático medio (MSE) que tiene H_{BOOT} al aproximar la distribución gamma correspondiente al vector de las medias muestrales $\{T_b^*\}$. La línea verde muestra la distribución real y la azul la estimación bootstrap.

Es importante resaltar que por cada valor de n y θ la muestra exponencial generada es distinta y por tanto, no se pueden comparar las cuatro gráficas de cada figura, sino únicamente las que están en la misma horizontal, en las que solo cambia el valor de B .

En cada una de las figuras se puede apreciar como, excepto en algún caso particular, el error cuadrático medio disminuye cuando aumenta el valor de B , siendo más notable la diferencia conforme disminuye θ . Podemos observar que las estimaciones son muy próximas al valor real del parámetro, teniendo la peor de ellas un MSE de 0.1327 para $\theta = 20$, $n = 500$ y $B = 5000$.

En la siguiente sección se realiza un estudio similar al realizado en el apartado 4.2.1, considerando un estadístico de la forma $g(\bar{X}_n)$, como los utilizados en los teoremas 3.16 y 2.9, siendo la función g en nuestro caso la exponencial.

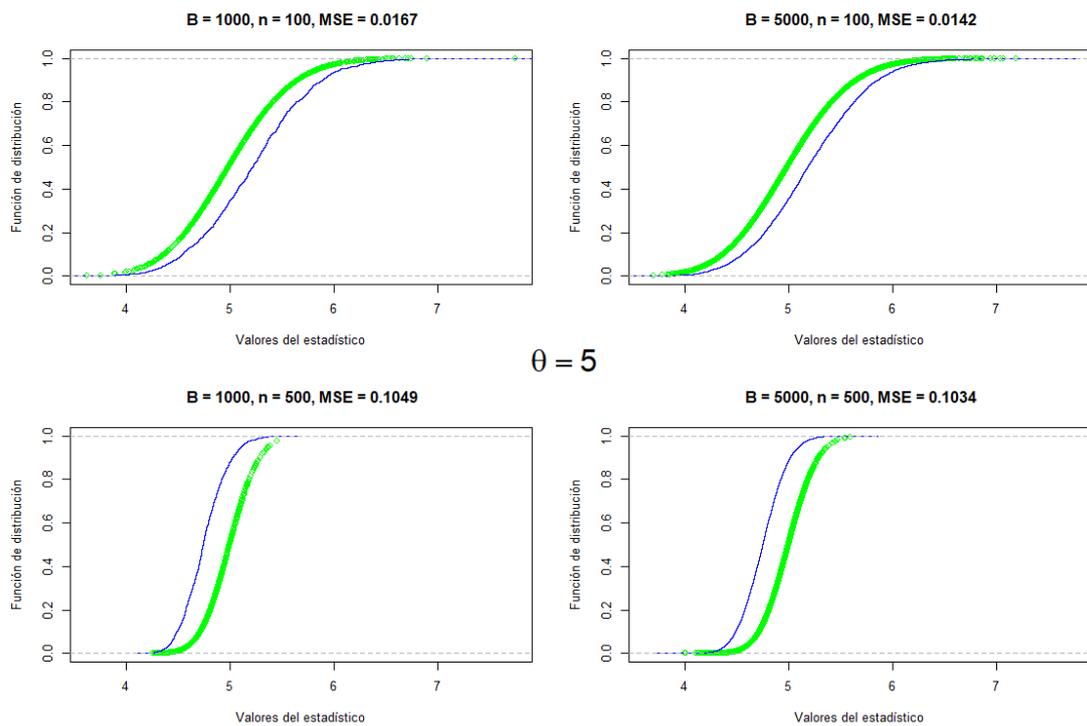


Figura 4.5: Comparación del estimador bootstrap de la distribución de \bar{X}_n (azul) frente a su distribución gamma exacta (verde), para $\theta = 5$.

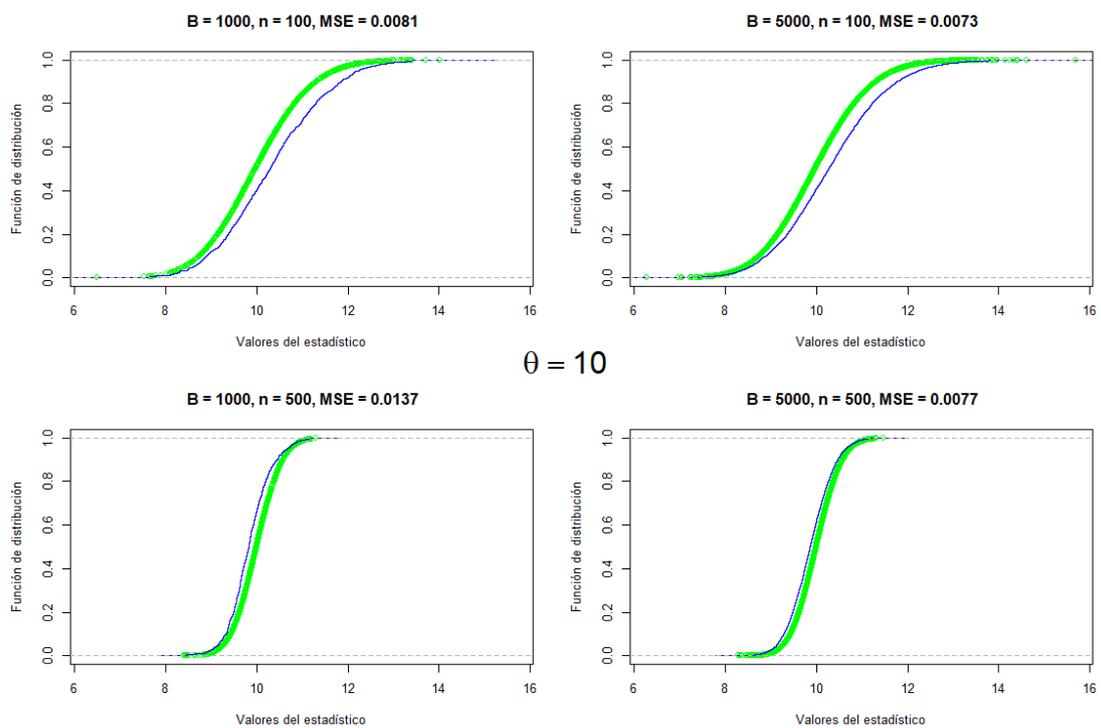


Figura 4.6: Comparación del estimador bootstrap de la distribución de \bar{X}_n (azul) frente a su distribución gamma exacta (verde), para $\theta = 10$.

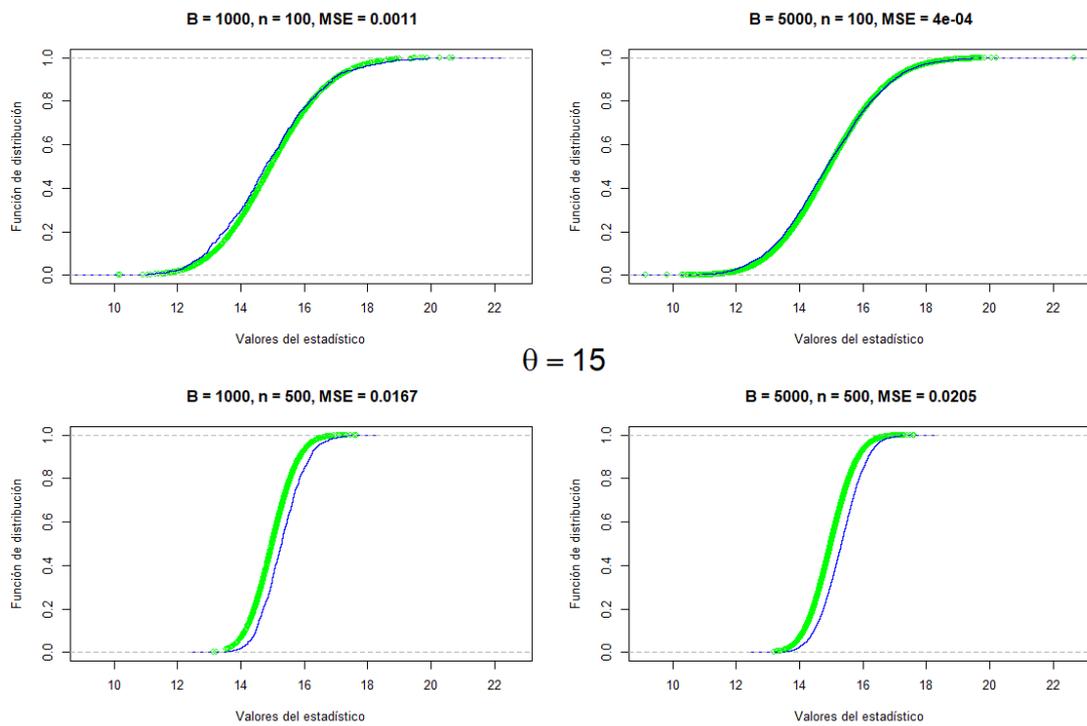


Figura 4.7: Comparación del estimador bootstrap de la distribución de \bar{X}_n (azul) frente a su distribución gamma exacta (verde), para $\theta = 15$.

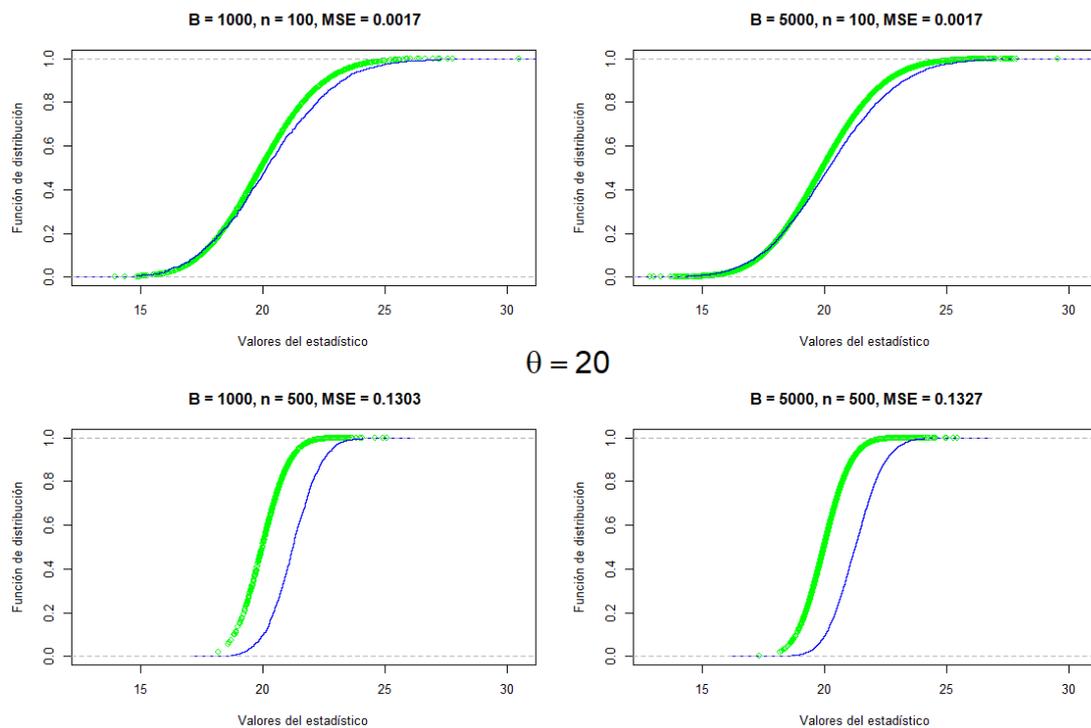


Figura 4.8: Comparación del estimador bootstrap de la distribución de \bar{X}_n (azul) frente a su distribución gamma exacta (verde), para $\theta = 20$.

4.3. Estudio de una transformación de la media muestral para una distribución exponencial

A lo largo del trabajo hemos visto cómo los estadísticos de la forma $g(\bar{X}_n)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ tenían la interesante propiedad de que las varianzas de jackknife y bootstrap para ellos cumplían

$$\frac{v_{JACK}}{\sigma_n^2} \rightarrow_{c.s.} 1,$$

$$\frac{v_{BOOT}}{\sigma_n^2} \rightarrow_{c.s.} 1,$$

donde σ_n^2 viene dado por (2.14).

Por ello, es interesante estudiar el caso de una transformación de la media muestral en la práctica. La transformación elegida ha sido

$$g(\bar{X}_n) = e^{\bar{X}_n}.$$

Como conocemos la media μ y la varianza σ^2 de la media muestral, y existe la derivada de g en μ podemos calcular el valor de σ_n^2 :

$$\sigma_n^2 = \frac{1}{n} \sigma^2 g'(\mu)^2 = \frac{(e^\theta)^2 \theta^2}{n}.$$

Para este apartado se ha seguido el mismo procedimiento que en la Sección 4.2.1. Se han utilizado los mismos cuatro valores para θ , n y B , lo que ha generado 16 gráficas agrupadas de cuatro en cuatro, que pueden observarse en las figuras 4.9, 4.10, 4.11 y 4.12.

Lo primero que observamos es que el método bootstrap aproxima mucho mejor el valor de σ_n^2 que el método jackknife, salvo en un único caso particular, cuando $n = 100$ y $\theta = 2$, como se puede ver en la Figura 4.9.

Además, si nos fijamos en los ejes y de cada una de las gráficas, ambos estimadores dan una mejor aproximación conforme θ disminuye, llegando a tomar valores del orden de $1 \cdot 10^{-4}$ cuando $\theta = 0.25$.

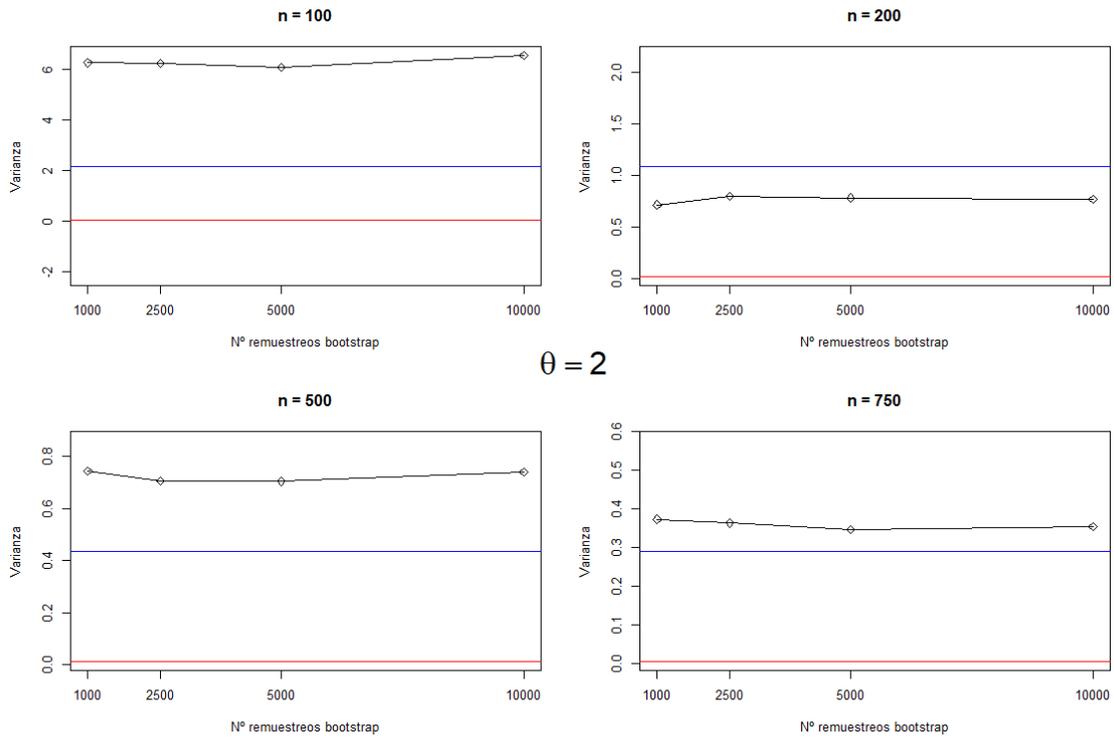


Figura 4.9: Comparación de la estimación de la varianza jackknife y bootstrap de el estadístico $e^{\bar{X}_n}$ para el caso $\theta = 2$.

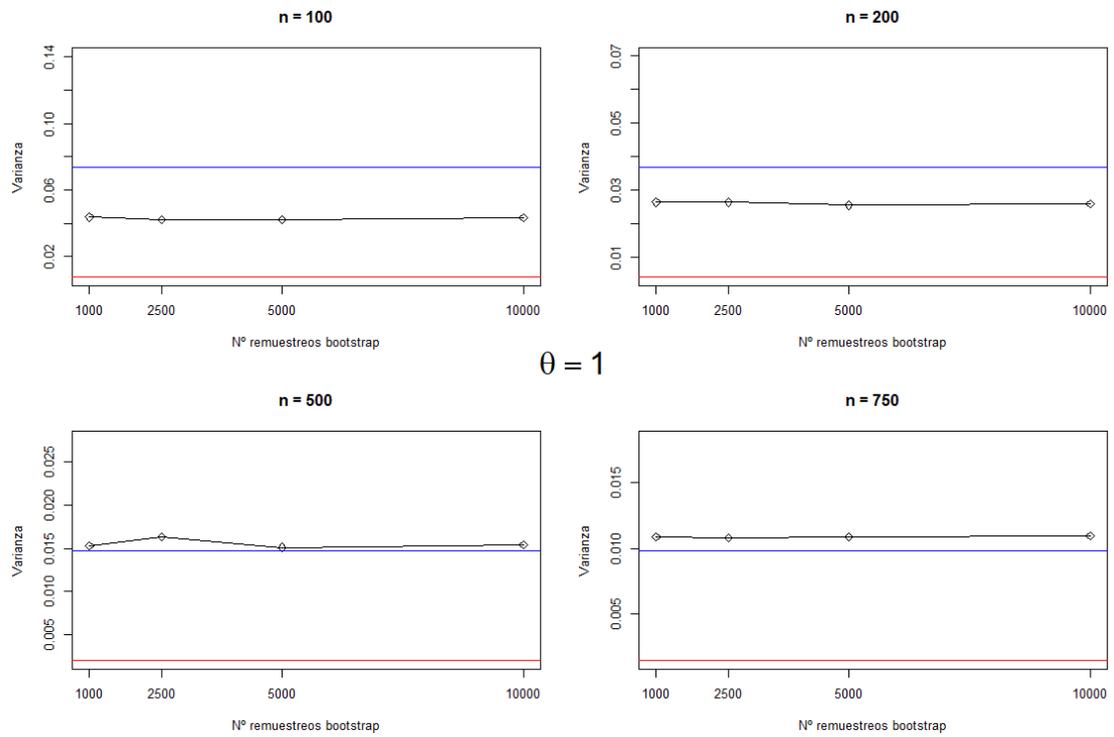


Figura 4.10: Comparación de la estimación de la varianza jackknife y bootstrap de el estadístico $e^{\bar{X}_n}$ para el caso $\theta = 1$.

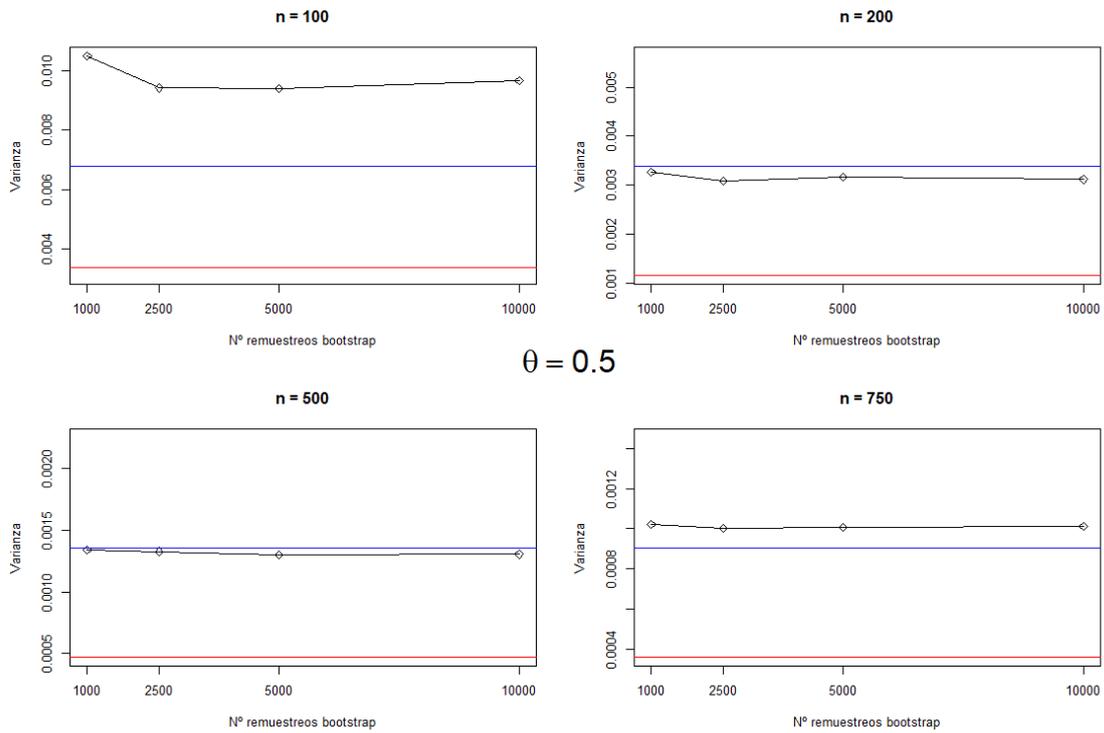


Figura 4.11: Comparación de la estimación de la varianza jackknife y bootstrap de el estadístico $e^{\bar{X}_n}$ para el caso $\theta = 1/2$.

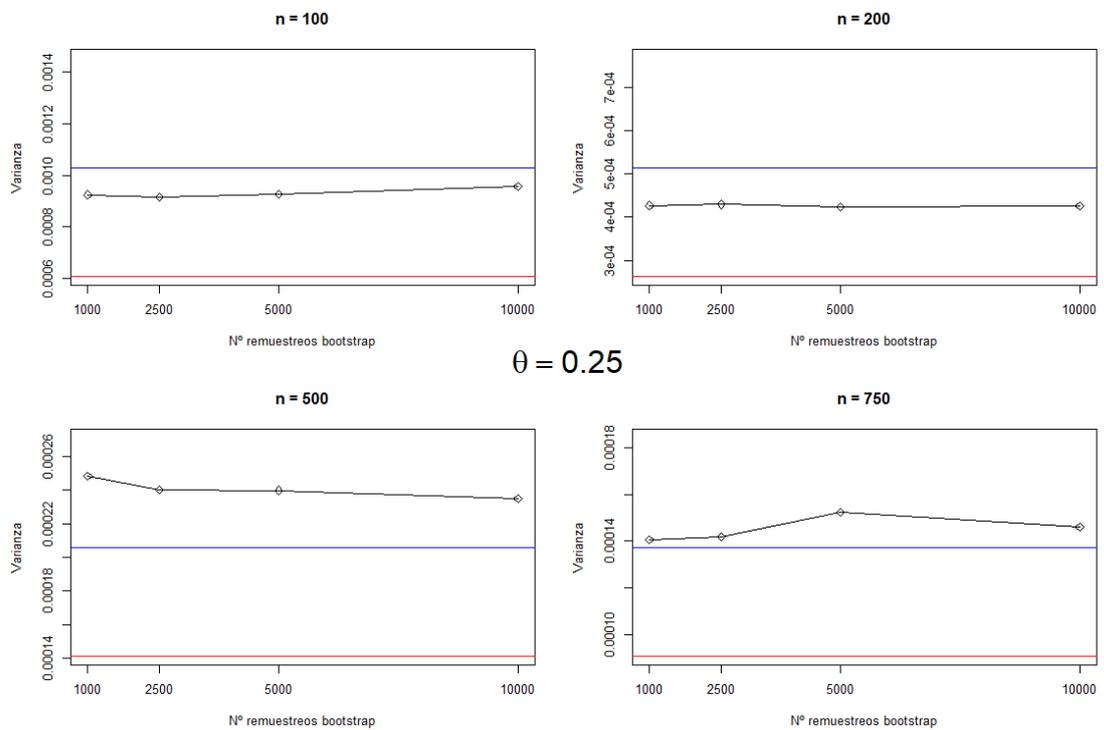


Figura 4.12: Comparación de la estimación de la varianza jackknife y bootstrap de el estadístico $e^{\bar{X}_n}$ para el caso $\theta = 1/4$.

Bibliografía

- [1] Bickel, P. J. y Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *The Annals of Statistics*, **9**, 1196-1217.
- [2] Dasgupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer, Nueva York.
- [3] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**, 1-26.
- [4] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia.
- [5] Hall, P. (1992). *The Bootstrap and Edgeworth expansion*, Springer, Nueva York.
- [6] Miller, R. G. (1974). The jackknife—a review, *Biometrika*, **61**, 1-15.
- [7] Quenouille, M. (1949). Approximate tests of correlation in time series, *Journal of the Royal Statistical Society B*, **11**, 18-84.
- [8] Rohatgi, V.K. (2001). *An Introduction to Probability and Statistics*, Wiley, New York.
- [9] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, Nueva York.
- [10] Shao, J. y Tu., D. (1995). *The Jackknife and the Bootstrap*, Springer, Nueva York.
- [11] Tukey, J. (1958). Bias and confidence in not quite large samples, *Annals of Mathematical Statistics*, **29**, 614.
- [12] Van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press. Cambridge.
- [13] Zoroa, P. y Zoroa, N. (2008). *Elementos de Probabilidades*, Diego Marín Librero Editor, Murcia.