UNA REVISIÓN DE LOS MÉTODOS DE REMUESTREO Y SU ROBUSTEZ

Carlos N. Bouza[1]
Universidad de La Habana

RESUMEN:
En este trabajo se trata de establecer como se engarzan las teorías que sustentan los modelos teóricos de la . estadística robusta, con los métodos intensivos de computación. Para ello se discute como el enfoque . funcional permite discernir sobre la convergencia de estimadores y predictores al parámetro, expresado como un funcional de la función de distribución desconocida. Esto permite establecer las condiciones generales que nos permiten usar Jacknife y Boostrap para estimar intervalos confidenciales, hacer pruebas de hipótesis etc, utilizando solo la información brindada por la muestra observada. Se hace hincapié en algunos métodos derivados del principio Boostrap.

ABSTRACT:
This paper looks for establishing how robust statistics theoretical models and intensive computation methods interact. The functional approach is discussed for establishing the convergence of estimators and predictors to the parameter, expressed as a functional of the unknown distribution function. It allows the establishment of general conditions that sustain the use of Jacknife and Boostrap for estimating confidence intervals, hypothesis testing etc., using only sampling information. Some Boostrap derived methods are rekamarked

.

Key Words: Functional, differentiability, almost sure convergence, probability convergence, Edgerworth Series.

MSC 62G35

1 INTRODUCCIÓN

"Las leyes científicas no avanzan mediante un principio dictatorial o pueden justificarse mediante fe o filosofía medieval. La Estadística es el único tribunal hacia el nuevo conocimiento." (P.C. Mahalanobis).

Las ciencias tratan de conocer los fenómenos naturales y usar el nuevo conocimiento en mejorar la vida de la humanidad. Para ello se establece un proceso de abstracción que trata de establecer leyes y teorías que permitan establecer el comportamiento de los fenómenos con cierta exactitud. Actualmente las hipótesis se apoyan en hechos observables los que permiten determinar leyes. Estos hechos pueden ser caracterizados por conjuntos de datos los que reflejan patrones, relaciones o interdependencias que permiten construir modelos que expliquen las fluctuaciones aleatorias y errores observan en la información obtenida.

Los programas estadísticos actualmente son muy interactivos y permiten analizar los datos, usando un *software.* Es muy sencillo su uso por lo que cualquier profesional puede explorar sus datos sí mismo, pero para el uso adecuado de la información se requiere de conocimientos de la teoría Estadística. La estadística es un instrumento de uso casi ineludible para tratar de estudiar y cuantificar la incertidumbre lo que hace que su uso sea inevitable. En nuestra era las actividades humanas están basadas en predicciones, y es se predice bajo incertidumbre. La incertidumbre está presente en nuestra realidad generada por diversas causas como el desconocimiento de parte de la información necesaria, los errores producidos por mediciones defectuosas etc. Es importante reflexionar sobre cómo cuantificar la vaguedad formulando estrategias para reducir, controlar y modelar la incertidumbre. En esta búsqueda podemos considerar que la Estadística puede permitir

[1] bouza@matcom.uh.cu

tomar decisiones adecuadas. Sus métodos dependen de cual es la distribución de las variables y el error asociado al método que usamos. Errores en la distribución asumida, el uso de medidas ineficientes del error son comunes en el quehacer estadístico.

Es usual que cursos y teorías se basen en la hipótesis de normalidad, por ejemplo. Sin

embargo esta distribución casi nunca esta presente en la vida real. Se acude al Teorema Central del Límite para argüir que la distribución asintótica es la normal. Sin embargo, aun para grandes tamaños de muestra es posible que esta aproximación no sea válida, sobre todo en fenómenos económicos.

Las nuevas metodologías: técnicas de remuestreo como el "Jacknife" o como el "Bootstrap", la idea de robustez y el estudio de procedimientos de validación y diagnóstico abren nuevas áreas a la metodología estadística y con ello al papel del estadístico en problemas complejos que no son abordables solo con un conocimiento básico.

En este trabajo revisamos el uso de métodos modernos que han permitido la creación de métodos que se apoyan en una sólida teoría matemática. Esta permite establecer criterios generales de convergencia a partir de la estructura y diferenciabilidad del funcional que representa al parámetro de interés. La convergencia puede ser descrita con elegancia mediante representaciones en Series de Edgeworth. La convergencia de la distribución de los funcionales utilizando criterios de suavidad descritos por la diferenciabilidad del tipo Fréchet, Gateaux o Hadamar permite el tratamiento unificado de la estimación u docimasia.

La próxima sección presenta el problema del remuestreo como es tratado comúnmente: son procedimientos que se fundan en la posibilidad de utilizar las capacidades de cómputo existentes actualmente. En este se describen los dos procedimientos mas estudiados: Jacknife y Boostrap. La siguiente se dedica presentar el modelar de estimadores como funcionales de la distribución empírica y destacar sus propiedades generales. Algunos casos son utilizados para ilustrar el comportamiento de la teoría en la implementación práctica. En la cuarta sección se estudian las características como funcionales de los métodos de Jacknife y Booststrap dentro de este enfoque, destacándose su convergencia bajo condiciones de diferenciabilidad. Finalmente se presentan algunas derivaciones de métodos de remuestreo basadas en el procedimiento Bootstrap que buscan utilizar más eficientemente los medios de cómputo garantizando también la convergencia de los estimadores.

Podemos dar como objetivo general del trabajo el dar una visión unificada de las conexiones entre el enfoque funcional de la estadística, los métodos de remuestreo y como estos justifican el uso de Jacknife o Boostrap.

2. LOS MÉTODOS AUTOSUFICIENTES

"La vida es el arte de extraer conclusiones suficientes de insuficientes premisas." (Samuel Butler).

Los dos problemas estadísticos de la inferencia más populares son la estimación de parámetros y el de pruebas (dócimas o contrastes) de hipótesis.. En general los modelos suponen que la distribución de probabilidad para los datos es normal . En la realidad casi nunca conocemos ni la distribución que genera las observaciones ni la desviación típica de la población. La forma más popular de objetivizar la precisión de un parámetro es construir un intervalo de confianza para un determinado nivel de probabilidad, usualmente mayor que 0,95. Si se acepta la normalidad todo se puede resolver con elegancia. El ejemplo clásico es el de usar una muestra de tamaño n, estimar la media μ mediante la media

muestral m y el error estándar de la media $\frac{s}{\sqrt{n}}$ y el intervalo de confianza para la media es $m \pm t_{\alpha n} \cdot \frac{s}{\sqrt{n}}$

En algunos los libros de inferencia se estudia el papel de las llamadas Pruebas Aleatorizadas. Estas consisten en calcular las diferencias de cada valor con respecto a la media y generar muestras de esas diferencias. Cada valor se identifica con un signo positivo o negativo. La distribución obtenida permite calcular un intervalo de confianza al fijar los quantiles de orden 0,025 y 0,0975, si se fija α=0,05. Esta línea considera que, en ausencia de cualquier otra información respecto a la población que no sea la contenida en la propia muestra, la distribución de los valores encontrados en una muestra aleatoria constituyen la mejor orientación en cuanto a la distribución de esa población. Los métodos auto-suficientes se basan solo en la información muestral. La idea que subyace en métodos autosuficientes es generar a partir de los valores observados en la muestra aleatoria un modelo específico que se está contrastando. Por ejemplo supongamos que no disponemos de un método confiable para calcular intervalos de confianza para la media *m* y aceptemos que la distribución es normal con media μ desconocida. Si generamos de acuerdo a ese modelo un gran número de muestras aleatorias de tamaño n, podremos analizar la distribución de esas muestras para calcular los percentiles requeridos y determinar un intervalo de confianza. Si lo hacemos comprobaremos que los valores obtenidos concuerdan con los que se obtendrían utilizando la fórmula teórica basada en las propiedades de la distribución normal.

Este método de simulación se conoce como método de Monte Carlo. Las pruebas aleatorizadas se consideran casos particulares.

La investigación de lo que constituye una prueba ha influido en el concepto actual de algoritmo. En los últimos tiempos se ha hecho evidente que el desarrollo computacional (capacidad de cómputo, accesibilidad a la programación de modelos teóricos, estabilidad de los métodos numéricos a los que puede acceder, etc.) ha influenciado en la estadística cambiando drásticamente algunos puntos de vista y casi todo el quehacer de la estadística. El comportamiento de un procedimiento estadístico puede ser analizado y estudiado, bajo distintos escenarios, para obtener una visión de lo que subyace en el fenómeno que se estudia. Por ello el comportamiento de los modelos estadísticos depende cada vez más de estudios de simulación que de desarrollar una teoría asintótica. La experimentación numérica permite fijar que es un valor de n `grande` en forma efectiva. Usando generadores de números pseudo aleatorios en forma repetida suplanta o suplementa el modelo probabilístico, generalmente desconocido, y permite determinar velocidades de convergencia. Esto ha hecho muy borrosa la frontera existente entre usar los datos existentes, establecer la validez de un modelo probabilístico generador de variables aleatorias y usar los datos para generar un modelo probabilístico aproximado.

Esto ha llevado al desarrollo de principios eminentemente numéricos para estudiar problemas de la estadística, los que son denominado autosuficientes por solo utilizar los datos y métodos numéricos. Procedimientos como el Jacknife y el Bootstrap proveen de algoritmos que satisfacen intuitivamente principios sobre los que se basa el razonamiento estadístico: muestrear repetidamente, promediar etc. Su uso nos lleva a resolver de forma general no solo la estimación de errores de los estimadores, sino también construir regiones de confianza y hacer pruebas de hipótesis. Estos funcionan bien en casos muy complejos en los que los modelos teóricos, a lo más proveen de reglas sobre la convergencia.

Cuando se dispone de un modelo teórico de un problema no tiene sentido utilizar la simulación de Monte Carlo. Sin embargo en ocasiones la solución analítica del problema puede ser muy compleja y un único camino más inteligente es utilizar técnicas de simulación. Esto es algo que se viene empleando desde hace mucho y con éxito en problemas de ingeniería y también en estudios de biología, ciencias naturales, economía etc.

Debido a que se trata de métodos que sólo son posibles gracias a la moderna potencia de cálculo de las computadoras estos métodos son llamados "métodos de uso intensivo" del ordenador (computer intensive methods). Si tenemos una fórmula teórica para un caso muy concreto, como lo es que los datos sean extraídos de una población normal, es mejor no usarles. Pero el método de simulación permite obtener estimaciones para cualquier parámetro y cualquier modelo. Por ello ante el desconocimiento del modelo probabilístico una solución acertada es usar un método de intensivo.

Como todos los parámetros poblacionales pueden ser estimados a partir de técnicas de estimación. La mayoría de los estimadores se basan en la distribución de los estadísticos en el muestreo y toman como base algunas propiedades deseables del teorema Central del Límite . Este fija unas propiedades de convergencia deseables y es la base de gran parte la estadística analítica. Este establece fundamentalmente dos cosas:

1.- Las muestras individualmente son diferentes de las poblaciones pero en su conjunto son muy parecidas

2. Las muestras no son solo gobernadas por el azar, sino que en su conjunto siguen, no importa de lo que estemos tratando, las leyes universales fijadas por funciones teóricas de probabilidad.

3. La función de probabilidad normal rige como modelo aceptable para funciones del tipo media aritmética , en la mayoría de las ocasiones cuando las muestras 'grandes'

Para la estimación de los estadísticos de posición se puede tomar como base el cálculo combinatorio y permutacional. Estas técnicas no fueron suficientemente desarrolladas y utilizadas aunque históricamente fueron usadas por matemitso ilustres como Edgeworth. En sus inicios ellos empleaban sumatorias y las medias pero no otras medidas de posición como medianas, quantiles, percentiles o cuartiles.

El método Jacknife fue propuesto por Tukey en 1958. Este le dio este nombre al aludir jocosamente a las navajas multiuso, que lo mismo sirven de sacacorchos, que de abrelatas, que permiten destornillar etc. Estas son la panacea del excursionista y del operario que tiene lejos la herramienta adecuada. Estas navajas son adecuadas para toda tarea si no se tiene algo a mano la herramienta óptima.

El método de Bootstrap se ocupa de los mismos asuntos que la estadística paramétrica pero bajo otro enfoque, vea Efron y Tibshirani (1993). Las ideas básicas de la estadística no han cambiado, lo que ha cambiado es la posibilidad de su implementación y hacer los cómputos con rapidez. El desarrollo de la potencia de las computadoras es lo que ha aportado la rapidez y flexibilidad necesaria para hacer posible la aplicación de viejas ideas inaplicables, permitiendo superar la dependencia de las soluciones analíticas.

La idea básica del Bootstrap es tratar la muestra como si fuera la población, y aplicar el muestreo Monte Carlo para generar una estimación empírica de la distribución muestral del estadístico. La del Jacknife la de hacer un censo de las sub muestras de un tamaño reducido.

La verdadera estimación Monte Carlo requiere un conocimiento total de la población, pero por supuesto este no es generalmente disponible en la investigación aplicada. Sólo tenemos la información proveída por una muestra extraída de esa población, debido a lo cual necesitamos, inferir sobre θ a partir de la estimación $\hat{\theta}$.

En el Bootstrap la muestra es tratada como si fuera la población y realizamos un procedimiento del estilo Monte Carlo seleccionando de ella muestras independientes de tamaño *n. C*ada remuestra tiene el mismo número de elementos que la original. Mediante el remuestreo con reposición cada remuestra podría tener algunos de los datos originales representados en ella más de una vez, y algunos que no aparecerán. Por lo tanto, cada una de estas remuestras probablemente será leve y aleatoriamente diferente de la muestra original. Como los elementos en estas remuestras varían levemente , un estadístico $\hat{\theta}^*$, calculado a partir de una de ellas tomará un valor ligeramente diferente de los otros $\hat{\theta}^*$ y del $\hat{\theta}$ original . La afirmación fundamental del Bootstrap es que una distribución de frecuencias relativas de las estimaciones $\hat{\theta}^*$, calculada a partir de las remuestras, es una estimación de la distribución muestral de $\hat{\theta}$.

Más formalmente, los pasos básicos en la estimación Bootstrap son los siguientes Efron, (1979; Efron y Tibshirani, (1993):

1.- Construir una distribución de probabilidad empírica, $\hat{F}(x)$, a partir de la muestra asignando una probabilidad de 1/n a cada punto, $x_1, x_2, ..., x_n$. Esta es la función de distribución empírica (FDE) de x, la cual es el estimador no-paramétrico de máxima verosimilitud de la función de distribución de la población, F(X).

2.- A partir de la FDE, $\hat{F}(x)$, se extrae una muestra aleatoria simple de tamaño n con reposición. Esta es una "remuestra", x^*_b.

3.- Se calcula el estadístico de interés, $\hat{\theta}$, a partir de esa remuestra, dando $\hat{\theta}^*_b$.

4.- Se repiten los pasos 2 y 3 B veces, donde B es un número grande.

5.- Construir una distribución de probabilidad a partir de los B $\hat{\theta}^*_b$ asignando una probabilidad de 1/B a cada punto, $\hat{\theta}^*_1, \hat{\theta}^*_2, ..., \hat{\theta}^*_B$. Esta distribución es la estimación Bootstrap de la distribución muestral de $\hat{\theta}$, $\hat{F}^*(\hat{\theta}^*)$. Esta distribución puede usarse para hacer inferencias sobre θ.

La magnitud de B en la práctica depende de las pruebas que se van a aplicar a los datos. En general, B debería ser de entre 50 a 200 para estimar el error típico de $\hat{\theta}$, y al menos de 1000 para estimar intervalos de confianza alrededor de $\hat{\theta}$, vea Efron y Tibshirani, (1986) y (1993).

El estimador Bootstrap del parámetro θ se define como:

$$\hat{\theta}^*_{(.)} = \frac{\sum_{b=1}^{B} \hat{\theta}^*_b}{B}$$

Este no es sino la media de los valores del estadístico calculados en las B remuestras Bootstrap.

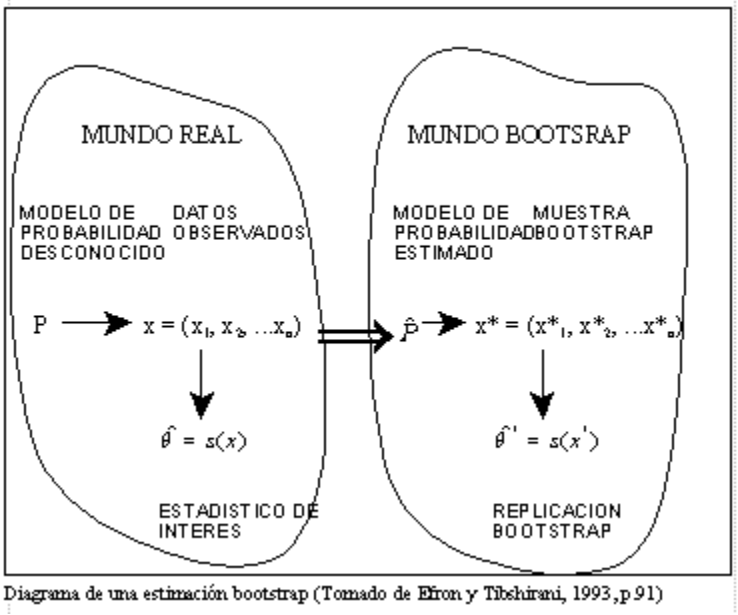
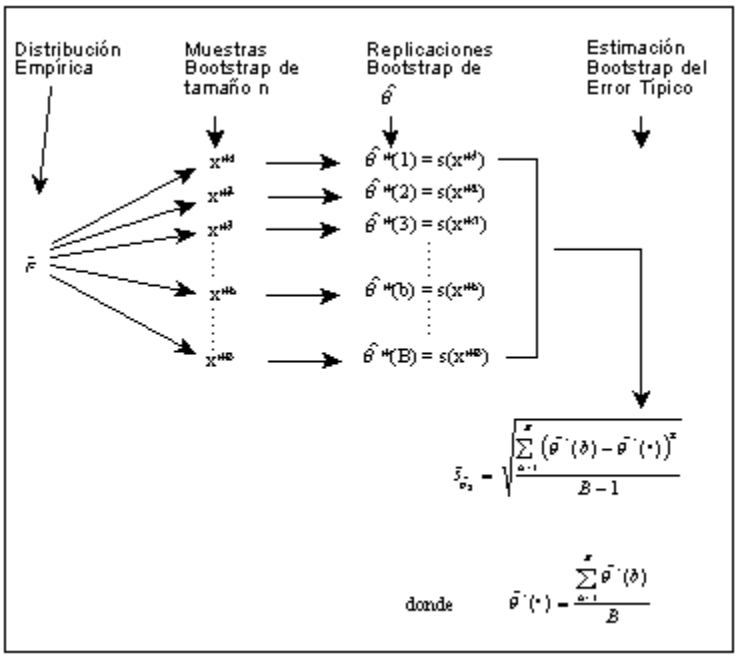


Figura 1. Relación entre los universos de discurso

En algunas ocasiones se tiene algún conocimiento más que el estrictamente aportado por la muestra, por ejemplo se conoce la función de distribución de la variable objeto de estudio pero se desconocen los parámetros, que deben ser estimados. En estos casos se estimarían los parámetros a partir de la muestra y el remuestreo se realizaría a partir de la función teórica conocida, con los parámetros estimados, en lugar de a partir de la FDE construida a partir de la muestra. En este caso hablamos de *Bootstrap Paramétrico*. Para algunos casos como el de la media aritmética existen fórmulas sencillas que nos permiten estimar el valor del error típico. Pero no siempre es este el caso, como ocurre con la mediana o con las medias recortadas, por citar dos estadísticos relativamente simples. La estimación Bootstrap del error típico es posible para todos los estimadores y no requiere cálculos teóricos.

El algoritmo Bootstrap trabaja extrayendo muchas muestras Bootstrap independientes, calculando en cada una de ellas el correspondiente estadístico, y estimando el error típico de $\hat{\theta}$ mediante la desviación típica de las replicaciones. Es decir, se aplica directamente la definición de error típico (desviación típica de la distribución muestral de un estadístico) sobre la distribución muestral Bootstrap.



Algoritmo de estimación bootstrap del Error Típico de un estadístico.

Figura 2. Esquema del desarrollo de un algoritmo típico de Bootstap

Se puede usar el criterio paramétrico (Bootstrap paramétrico) en este se considera F como el modelo verdadero que genera los datos, el funcional T(F), el estimador de Bootstrap $T(F^*_n)$, y la clase donde pertenece F esta indexada por el parámetro θ. Esto es un modelo paramétrico y $T(F^*_n)$ es llamado Bootstrap Paramètrico. Se dice que es el Bootstrap no paramètrico si la familia a la que pertenece F no está indexada paramétricamente. Por ejemplo el problema de la regresión es considerado semiparamètrico.

3. CONSIDERACIONES GENERALES SOBRE LOS ESTADÍSTICOS FUNCIONALES

"El azar es acaso un pseudónimo de Dios cuando no quiere estampar su firma." (Anatole France).

3.1. El uso de criterio de la Estadística Robusta

Desde los 70´s se ha venido desarrollando la Estadística Robusta. Su impacto en las concepciones y en la teoría estadística es muy importante. Afianza esto el hecho de que el estudio de la robustez no niega entra en contradicción con métodos y modelos sino que establece bajo que condiciones si se puede confiar en ellos. Además la posible robustificación de estos puede ser realizada con alguna sencillez en muchos casos. Lamentablemente gran parte de los estimadores usualmente usados no son robustos. Tomemos el caso de las funciones de la media. La media es afectada por la existencia de outliers dando una visión deformada del fenómeno estudiado. Algo similar ocurre incluso con los mínimos cuadráticos.

El razonamiento de usual parte de considerar que se trabaja con una familia de funciones de distribución (FD), Υ. En la estadística paramétrica se considera que $\Upsilon=\{F_\theta \mid \theta\in\Theta\}$ y solo desconocemos el valor del parámetro θ_0 , pero en general el espacio paramétrico Θ es conocido. Las observaciones hechas del fenómeno de interés es llevada a cabo al ser generada una variable aleatoria $X\in\Re^p$ por una FD particular $F^*_\theta\in\Upsilon$. La información brindada por la muestra nos permite indagar, usando diversos criterios, sobre las

propiedades de la FD que la generó. La validez del método es evaluado a través de criterios de eficiencia.

El espacio paramétrico es particionado y representarlo como $\Theta=\cup_{j=1}^{J}\Theta_j$ y el interés es determinar, cual es el j adecuado. Algo similar se puede hacer con Υ, ($\Upsilon=\cup_{j=1}^{J}\Upsilon_j$). Esto último plantea la necesidad de determinar cual es la famílla de FD´s a la que pertenece la que generó nuestras observaciones, que es un problema de la estadística no paramétrica. El modelo normal asume que Υ={F tales que $dF(x)/dx= (2\pi\sigma^2)^{-1/2}\ \exp\{-(x-\mu)^2/2\sigma^2$. | $(\mu,\sigma)\in\Re\times\Re^+$} y el Binomial que

Υ={F tales que $Prob(X=x)= C^n_x P^x (1-P)^{n-x}$ | $P\in(0,1)$}.

En ambos casos el problema estadístico es el desconocer cual es el valor del parámetro, $\theta=(\mu,\sigma)$ o $\theta=P$, y se dice que el problema es caracterizado por una FD de familia conocida.

Para estudiar cada problema real es necesario hacer observaciones. El desconocimiento de la familia a la que pertenece la FD puede llevar a hacer pruebas de la Bondad del Ajuste a los datos. Es claro que si usamos una FD, denotada por F, y esta no es adecuada las inferencias no lo serán tampoco. La estadística robusta trata de desarrollar criterios para establecer como se comportan los procedimientos cundo no usamos la FD real. Desde el punto de vista general se considera que, si asumimos que F es la distribución que genera los datos debemos considerar que quizás estemos equivocados y sea realmente otra FD de Υ la que los describe. Si se considera que para una cierta métrica D, definida en el espacio de las funciones, cuando D(F,G) es pequeña es natural considerar que el método será robusto si la distancia entre sus inferencias también lo es. O sea un buen método será aquel que no se ve seriamente afectado por utilizar una distribución incorrecta bajo ciertas condiciones métricas.

Sea T: $\Upsilon\rightarrow\Re^m$, tal que T(F) es el funcional de interés sobre el cual queremos inferir. En el ejemplo más sencillo es el de estimar la media de una distribución, en cuyo caso T(F)=E(X|F). La FD empírica es:

$$F_n(x)=\sum_{i=1}^{n} I(X_i<x)/n,$$

donde $I(X_i<x)$ =1 {0} si $X_i<x$ {si no}, nos estima F(x). Si usamos el funcional T(F) para describir el parámetro de interés θ, dada la convergencia de la distribución empírica a la FD que genera las observaciones, se espera que para una muestra suficientemente grande y si D es una métrica adecuada , metriza a la Topología Débil, consideraremos que es válido que $D(F,F_n)<\delta\Rightarrow D(T(F),T(F_n))<\varepsilon$ si $n\rightarrow\infty$. $T(F_n)$ es un predictor de T(F) consistente. Note que en el caso de la distribución normal si $T(F)=E(X|F)=\mu$ como $dF_n(x)/dx=1/n$ el estimador es la media muestral:

$$\mu_n=T(F_n)=\sum_{i=1}^{n} X_i/n.$$

Es bien conocido que $\lim_{n\rightarrow\infty}\mu_n=\mu$.

Es práctico definir clases de estimadores. Estos van a determinarse a partir de estructuras que soporten el estudio de la convergencia de los funcionales de la FD empírica. La invarianza es una de las propiedades convenientes.

Si $T(F_n)$ es evaluado en $X+\varepsilon$ este es equivariante cuando se cumple para $\varepsilon=a^Tb$, $\forall b\in\Re^p$, que $T(F_n(X+e))=T(F_n)+ a_n^T b_n$.

Una clase de estimadores deducidos de a partir de un criterio de robustez es la de los M-estimadores. Ella está asociada a una función $\rho(x,\theta)$: $\Re^h \rightarrow \Re^q$. la que está conectada con un criterio de distancia con el parámetro de interés.

Consideraremos $E(\rho(x,t)|F_\theta)$ con mínimo en $t=\theta$. Entonces un M-estimador de θ es la solución del problema de optimización:

$$M_n = \text{argmin } \{\sum_{i=1}^{n} \rho(x_i ,t)|t\in\Re^q\}$$

Por ejemplo en la regresión lineal se tiene que $x=A+BZ+\varepsilon$ y tomamos $\rho(x_i ,t)=(x_i-(A+Bz_i))^2$

Es claro que el obtenido M_n será el estimador mínimo cuadrático usual. Sin embargo si acometemos el estudio usando $T(F_n)$ no asumimos ninguna hipótesis sobre F excepto que pertenezca a Υ. Nos despreocupamos de la invalidez de la hipótesis sobre la forma específica de F_θ pues esperamos que se trabajará con una $G_\theta\in\Upsilon$ suficientemente cerca de ella, $D|F_\theta , G_\theta|<d$, lo que el Teorema de Glivenko-Cantelli garantiza que, al usar la FD empírica:

$$\text{Lim}_{n\rightarrow\infty} \{||F_n - G_\theta ||=\text{Sup}_x |F_n(x)-G_\theta (x)|\}=0$$

Los M-estimadores son invariantes para la posición pero no para la escala. El problema asociado al uso práctico de los M-estimadores es si hay una solución única o no del problema de optimización. En particular es más simple aceptar que se garantiza la existencia de una sucesión de soluciones consistentes. Para ello es necesario que $d\rho(x,t)/dt=\psi(x,t)$ sea monótona en t. Esto garantiza que los M-estimadores sean Asintóticamente Óptimos Normales (AON). La existencia de un M-estimador $M(F_n)$ puede ser garantizada utilizando condiciones de regularidad muy suaves. Esta suavidad es una condición de suficiencia para el caso en que la monotonía no sea aceptable. En todo caso la representación de "un-paso" de un M-estimador es:

$$(M(F_n)-M(F))n^{1/2} =(\int\psi (x,M(F)df(x))^{-1} \sum_{i=1}^{n} \psi(x_i ,M(F))/n^{1/2} +o_P(1)$$

Esta expresión permite estudiar si $M(F_n)$ es AON pues

$$(M(F_n)-M(F))n^{1/2} \approx\Rightarrow N\{0, (\int\psi (x,M(F)df(x))^{-2} \text{Var}(\psi(x ,M(F))\}$$

Otra clase de estimadores deducidos de los criterios de robustez es la de los L-estimadores. Estos son expresados como una combinación lineal de estadígrafos de orden, EO. En la Estadística no Paramétrica estos juegan un papel fundamental en el desarrollo de estadígrafos de prueba. La completitud de los estadísticos lineales de rango hace que los L-estimadores también lo sean. Estos están definido como

$$L(F_n)=\sum_{i=1}^{n} c_{ni} X_{n(i)}$$

Las constantes c_{ni} , i=1,…,n, son conocidas y $X_{n(i)}$ es el EO "i" de la muestra. Los coeficientes se buscan tratando que tengan ciertas propiedades. Muy popularmente se utilizan los quantiles:

$$Q(t)=\inf \{x| F_\theta(x)\geq t\}, t\in(0,1).$$

Para la FD empírica tendremos como quantil a

$$Q_n(t)=\inf \{x| F_n(x)\geq t\}, t\in(0,1).$$

Considerando una sucesión $\{J_n(u)|u\in(0,1)\}$, con $J_n(u)$ constante en ((i-1)/n, i/n) y como F_n tiene saltos de magnitud 1/n en los n EO´s tomamos como estructura general de las constantes a

$$c_{ni} = J_n(i(n+1)^{-1})/n$$

Entonces

$$L(F_n)=\sum_{i=1}^{n} Q_n\,(i(n+1)^{-1}J_n(i(n+1)^{-1}/n=\int Q_n\,(F_n(x)J_n(F_n(x)dF_n(x)$$

es el estimador de un parámetro, el que es expresable como una función de los EO´s dada por :

$$L(F)=\int Q\,(t)J(t)dF(t)$$

Donde t denota el EO t-ésimo de F. Si F admite una función de densidad

$$f(x) \text{ y } f(x)=\sigma^2 f((x-\mu)/\sigma),$$

con $\theta=(\mu,\sigma)$.

Consideremos que podemos escribir $X_{n(i)} =\mu+\sigma Z_{n(i)}$, i=1,…,n. $Z_{n(i)}$ es un EO cuya esperanza es $E(Z_{n(i)})=a_{n(i)}$, $Cov\ (Z_{n(i)} , Z_{n(j)})=b_{n(i,j)}$, donde $a_n=(\ a_{n(1)},..., a_{n(n)})^T$ y $B_n=(\ b_{n(i,j)})^T{}_{nxn}$ son independientes de θ si existe la inversa de B_n. La Estadística Paramétrica buscaría un estimador óptimo a partir de conocer $f(x|\theta)$. Note que $J_n(u)\rightarrow J(u)$ para $u\in(0,1)$ fijo, pero el óptimo J* depende de f (x) que es independiente de θ.

Una amplia clase de L-estadísticos es aquella definida como :

$$L^*=\{L(F\ )=\sum c_{n(i)}=1, \text{ para todo i } c_{n(i)}= c_{n(n-i+1)}\}$$

Tomando

$$L^{**}=\{T_{nk} =\sum_{i=k+1}^{n} Ck^{i-1}\ C_k^{n-i}\ X_{n(i)}\ /C_{2k+1}^{n}, \text{ para todo i,k=0,...,n+1/2}\}$$

Note que $L^{**} \subset L^*$ y que si k=0 entonces T_{nk} es la media muestral y si k=(n+1)/2 será la mediana.

Otra clase interesante de L-estimadores es la de las medias censuradas.

$$L(F_n,\alpha)=\{\sum^{\ n-\lfloor m\rfloor}{}_{i=\lfloor n\rfloor+1}\ X_{n(i)}\ /n-\lfloor m\rfloor \mid m=n\ \alpha,\ \alpha\in(0,1/2\ )\}$$

En general muchos parámetros son combinaciones convexas de los EO como es el caso del kernel del índice de Fager que es $\sum^{K}_{i=1} R_i\ \pi_i$ donde R es el rango asignado a la especie i. Usando los EO maestrales un pequeño cambio de notación nos lleva a que $\sum^{K}_{i=1} R_i\ n_{(i)}/n$. Este es un L-estimador.

Cuando $J(Fn)\rightarrow J(F(t)\ )$ se garantiza la consistencia de los L-estimadores. Bajo ciertas condiciones suaves de regularidad.

Los rangos pueden ser utilizados en otra forma para derivar estimadores robustos de parámetros de posición. Esto es un punto de referencia dado que la eficiencia asociada a los estadígrafos basados en rangos es común que sean globalmente robustos.

Consideremos que la FD es simétrica respecto al parámetro de interés θ, $F(x)=F(x-\theta)$, nuestra hipótesis a probar es H0: $\theta=\theta$. Buscando una prueba no paramétrica el estadístico de prueba nos llevara a un buen estimador de los parámetros usando rangos. Un Estadígrafo de los rangos por ejemplo es:

$$S_n = \sum_{i=1}^{n} \text{signo}\,(x_i - \theta_0)\, a_n\,(R_{n(i)}^{+}\,(\theta_0))$$

Donde $R_{n(i)}^{+}\,(\theta_0)=\text{rank}\,|x_i-\theta_0|$, $a_n(h)$ es un score asignado a h . Un caso importante es aquel en que $a_n(i)=i/(n+1)$ que es el estadígrafo de Wilcoxon.

Tomando los estadísticos $R^{-}_n=\text{Sup}\,\{t|S_n(t)>0\}$ y $R^{+}_n=\text{Inf}\,\{t|S_n(t)<0\}$ un R-estimador de θ es

$$R_n = \tfrac{1}{2}[R^{-}_n + R^{-}_n].$$

Note que si $a_n(h)=1$ entonces S_n es el estadístico de la Prueba de los signos y R_n es la mediana.

El problema de la estimación basada en el Criterio de Minimización de una Distancia es muy intuitivo planteándose el minimizar una distancia entre el estimado y el parámetro de interés. Si sabemos que la distribución de las variables aleatorias pertenece a la familia proyectiva del espacio paramétrico dada por $\tilde{A}=\{F_\theta\ |\theta\in\Theta\}$, y definimos una distancia $D(F,F')$ entre un par de FD´s de $\tilde{A}$ buscaremos un estimador tal que sea igual a :

$$\theta_n = \text{Arg inf}\,\{d(F_n, F_t)|t\in\Theta\}.$$

Este estimador es robusto. Cuando se utiliza ciertas distancias como las de Crámer-von Mises y la de Kolmogorov-Smirnov se gana en robustez al compararse con el uso de la Euclidiana, ver Donoho-Liu (1988 a y 1988 b). Por otra parte el mejor resultado es asociado al uso de la distancia de Hellinger, ver Vajda (1984 a-e). El problema con estos estimadores es la solución del problema de optimización, lo que en el caso de más de un parámetro puede ser muy laborioso. Por ejemplo cuando nos planteamos que la distribución es simétrica respecto a θ , $F_\theta(x)=F_\theta(x-\theta)$, $\theta\in\Re$, y usamos la distancia Euclidiana $D(F, F_t)=\int(x-\int y dF_t(y))^2 dF(x)$ diferenciándola e igualando a cero obtenemos que $\int x dF(x)=\int y dF_t(y)$. Como obtenemos el mínimo a partir de esta solución el estimador de distancia mínima està determinado. Una solución estadística es hacer $F=F_n$ y $\int x dF_n(x)=\sum^{n}_{i=1}X_i/n$

Podemos hacer una visita a la clase de estimadores basada en el criterio de Pitman. Este propuso un método general para obtener estimadores óptimos. Este método esta ligado con el criterio de Máxima Verosimilitud desarrollado en detalle después de su trabajo. Estos estimadores son utilizados para establecer la eficiencia de estimadores particulares como los que aparecen en la estimación no paramétrica. La hipótesis base es la simetría de las distribuciones

$$F_\theta(x)=F_\theta(x-\theta),\ \theta\in\Re.$$

Para esta FD debe existir una función de densidad con momentos de primer orden finitos. Al tomar una muestra aleatoria independiente el estimador del parámetro es

$$T_P(F_n)=[\int t\Pi^{n}_{i=1}f(x_i-t)dt][\int\Pi^{n}_{i=1}f(x_i-t)dt]^{-1}$$

Como depende del conocimiento de f(x) su robustez no es válida pues una equivocación en su determinación determinará la no eficiencia del estimador. Sin embargo este puede ser robustificado.

Al no estar seguros de cual es la FD utilizamos el criterio de los M-estimación determinando:

$$\rho(x)=-\text{Log}\ \eta(x)$$

donde $\eta(x)=dG(x)/dx$, $G\in\tilde{A}$. Se espera que para un distancia preseleccionada $D(G,F)\leq\varepsilon$. η se considera como una densidad apropiada para aproximarse a f. El estimador robustificado de Pitman es:

$$T^*_P(F_n)=[\int t\Pi^n_{i=1}\eta(x_i-t)dt][\int\Pi^n_{i=1}\eta(x_i-t)dt]^{-1}$$

Note que este toma la forma de un función del tipo von Mises.

3.2 La estimación funcional

Trabajaremos con algunos conceptos bàsicos del análisis funcional para poder establecer las propiedades de los estimadores. Esto debe ser explicitado para lograr una transparencia en el procedimiento que se seguirá en lo adelante.

Partimos de un funcional lineal que depende de una FD como $T(F)=\int_{\Re}\varphi(x)dF(x)$. En el caso en que tenemos un vector $\mathbf{x}=(x_1,\ldots,x_p)^T$

$$T(F)=\int_{\Re^p}\varphi(x_1,\ldots,x_p)dF(x_1)\ldots dF(x_p)$$

Podemos trazar su propuesta en los trabajos de von Mises publicados en 1936 y 1947, vea Jurečková-Sen (1996). Este proponía aproximar estos funcionales usando la función distribución empírica . Entonces en general un estimador sería expresado por:

$$T(F_n)=\int_{\Re^p}\varphi(x_1,\ldots,x_p)dF_n(x_1)\ldots dF_n(x_p).$$

En el caso univariado se tiene la insesgadez de este tipo de estimadores . En los trabajos de von Mises se consideró la expansión en Series de Taylor de $T(F_n)$ alrededor de $T(F)$ utilizando la hipótesis sobre la diferenciabilidad del funcional obteniéndose $T(F_n)=T(F)+T'_F(F_n-F)+R_n(F_n,F)$, donde $T'_F(F_n-F)$ es la derivada del funcional en F y $R_n(F_n,F)$ es un resto. El estudio de tales funcionales requiere de herramientas del Análisis Funcional pues es muy elegante y además fácil considerar que se trabaja con el conjunto de transformaciones lineales continuas de A en B, L(A,B), sobre dos espacios topológicos. Note que tanto F_n como F pertenecen a $\tilde{A}$ y T es una función de un abierto A* de A en B , $T:A^*\longrightarrow B$, con F y H pertenecientes a A*. A partir de la topología del espacio se determina la diferenciabilidad del funcional. Tres tipos de diferenciabilidad son utilizables en este contexto y dependen, básicamente, de los conjuntos de hipótesis siguientes

C1.
- -C es una clase de todos los subconjuntos acotados de A.

C2
- C es una clase de subconjuntos compactos de A.
- $\exists$ la diferencial $T'_F\in L(A,B)$ tal que para $\forall C^*\subset C$ y $H\in C^*$, C* tal que $\lim_{t\to 0}\{[T(F+tH)-T(F)-T'_F(tH)]/t\}=0$, uniformmente.

- C3
  - C es una clase de puntos sencillos de A.

Los tipos de diferenciabilidad están encadenados de acuerdo al esquema siguiente

C1$\Rightarrow$la Fréchet diferenciabilidad

C2⇒la Hadamard diferenciabilidad
C3⇒la Gateux

El orden dado anteriormente es el de implicación existente entre ellas. En la práctica es la Hadamard diferenciabilidad la más usada en la estadística.

Regresando a $T(F_n)=T(F)+T'_F(F_n-F)+R_n(F_n,F)$ tenemos, bajo la diferenciabilidad que $R_n(F_n,F)=o(||F_n-F||)$ sobre el conjunto en que la distancia de Kolmogorov satisface que:

$$||F_n-F||=Sup\ \{||F_n(x)-F(x)||;\ x\in\Re\}\rightarrow 0.$$

Al usar el estadístico de prueba de Kolmogorov Smirnov se tenía que $||F_n-F||n^{1/2}=O_P(1)$, por lo que para $n^{-1}\rightarrow 0$

$$||F_n-F||=o_P(n^{1/2}).$$

Se usa esta distancia por razones de claridad del razonamiento pero esto es válido para otras métricas igualmente.

Reanalizando el funcional T(F) usando estos resultados tenemos que $T'_F(F_n-F)\cong T(F_n)-T(F)$. Por lo que $\int T'(F,x)\,d(F_n(x)-F(x))\cong T'_F(F_n-F)$. Si aceptamos que $\int T'(F,x)\,dF(x)=0$, hipótesis usual en los desarrollos en Series de Taylor,

$$T'_F(F_n-F)\cong\Sigma\ T'(F,x)/n=T(F_n)-T(F)$$

Entonces como trabajamos con una media, si $\sigma^2_F=V(T'(F,x))$ es finita el Teorema Central del Límite garantiza la convergencia a una N(0,1) de $(T(F_n)-T(F))\ \sigma^{-1}_F\ n^{1/2}$. Si el funcional es Hadamard diferenciable al equivocarnos y asumir que la FD es $G\in\tilde{A}$ se espera que $||T(G)-T(F)||<\varepsilon(n)$ si $||G-F||<\delta$, donde ambas cotas tienden a cero cuando $n\rightarrow\infty$. De ahí que si trabajamos en un espacio metrizable por ||a|| y utilizamos un estimador funcional Hadamard-diferenciable no es importante el conocer la distribución pues este convergerá al verdadero funcional, parámetro. Esto nos permite utilizar estimaciones de F basadas en estimadores de la FD como es el caso al usar directamente estimadores de esta como es el caso al estimar la función de densidad usando criterios no paramétricos en los que la convergencia L1 es aceptada. Bajo la hipótesis de continuidad asumida se cumple que :

$$Lim_{\ n\rightarrow\infty}\ V(\Sigma\ T'(F,x)/n)=\sigma^2_G$$

Podemos entonces utilizar un subconjunto de las variables medidas $X_1,\ldots,X_n$, que son iid. Denotemos por s(m) el subconjunto de vectores formados con $m\geq 1$ de estas variables. La función de kernel evaluada en un vector $x\in s(m)$, $\varphi(x)$ es llamado pseudo valor.

Sea la pseudo variable $Z_{i(1),\ldots,\ i(m)}=\varphi(X_{\ i(1)}\ ,\ldots,X_{\ i(m)})$, $i(j)\in\{1,\ldots,n\}$, $j=01,\ldots,m$. Tendremos $N=C^n_m$ posibles pseudo variables. Sea H la FD de Z y H_N la FD empírica. Por l Ley de los grande números si $n\rightarrow\infty$ tenemos que $n^{1/2}\ (H_N-H)$ tiende a tener una distribución normal de media cero.

Este marco nos permite re-escribir algunos estimadores usando las pseudo variables a través del kernel del funcional para considerar su convergencia usando otras herramientas.

Para los L-estimadores se deriva con facilidad la expresión $\int T'(H,x)\,d(H_N(x)-H(x))$, por lo que sus propiedades asintóticas son deducibles nuevamente al considerar su estructura a partir de pseudo variables

Por su parte si tenemos un M-estimador la condición de suficiencia para ser Hadamard diferenciable, utilizando la estructura basada en las pseudo variables, en el estudio de la robustez es un tanto más complicada. Deben cumplirse:

- $\varphi(t)$ es continua y diferenciable a trozos...
- Ψ' es acotada y $\Psi'(t)=0$, $\forall t$ en un intervalo compacto que contiene al 0 .

Así al estudiar un L-estimador lo que se debe cumplir es:

- $J(t)=0$ para un $t \leq \alpha$ y $t \geq \alpha$, con $\alpha \in (0,1/2)$.
- $J(t)$ continua y diferenciable a trozos.
- $J'(t)$ es acotada en $t \in (\alpha, 1-\alpha)$
- $F(t)$ es absolutamente continua
- $(F'(t))^2$ es integrable en $(\alpha, 1-\alpha)$.

Para los R-estimadores del tipo Wilcoxon descritos como

$\phi^*:[0,1] \rightarrow \Re^+$, siendo $\phi^*(u)= \phi^+(u)=$signo (u)

tendremos como estadístico funcional a:

$$S_n = \sum^n_{i=1} \text{signo}\,(X_i - \theta_0) a_n(R_n^+, \theta_0) = n\int \phi^*(F_n(x)-F_n(2\theta_0 - x)dF_n(x), \theta_0 \in \Re.$$

Condiciones suficientes para la Hadamard diferenciabilidad están dadas porque se cumplan solamente

- $\phi^*(u)$, $u \in (-1.1)$, es continua y diferenciable a trozos.
  $\phi^{*\prime}(u)=$ es continua a trozos y acotada

Estas condiciones de suficiencia permiten determinar con facilidad la diferenciabilidad de muchos estadísticos para dar por sentada su robustez sin embargo son muy restrictivas.

En la práctica si nuestro estimador no es diferenciable indagaremos. Usando otras vías, su convergencia. Estadígrafos importantes como los basados en la función de verosimilitud y en los scores normales no satisfacen las condiciones de diferenciabilidad y sin embargo su robustez es garantizada bajo ciertas condiciones particulares.

4. MÉTODOS DE RE-MUESTREO

"Adivinar es barato, adivinar erróneamente es caro." (Antiguo proverbio chino).

4.1. La consistencia de los estadísticos funcionales.

Al estimar un parámetro usando $T(F_n)$ nos interesa conocer su error. Un medida de particular importancia es la varianza $V(T(F_n))$, especialmente si es insesgado. El caso de la media es el más estudiado. Este fija las ideas básicas de la convergencia de un estimador funcional a través de los Teoremas de convergencia que han sido demostrados para funciones de la medias. Es bien conocido que si estimamos $T(F)=\theta=\mu$, la media aritmética, usamos la muestra $X_1, \ldots, X_n$ para evaluar el predictor $T(F_n)=\mu_n$. Si la FD posee momento de segundo orden finito podemos calcular la varianza de la población $T(F)=\sigma^2 = \int(x-\theta)dF(x)$. Para tener una idea de la dispersión de la variable aleatorio y es natural usar como estimación a

$$T(F_n)=\sigma^2_n = \sum^n_{i=1} (x_i - \mu_n)^2 / n-1.$$

Se espera que $\sigma^2_n \rightarrow \sigma^2$.

La precisión es medida a partir del uso de la FD al estimar los límites de confianza. Estos son función de los predictores

$$L(X_1,\ldots,X_n)=L_n=\text{Sup}\{t|\ F_\theta(x)\geq t\},\ U(X_1,\ldots,X_n)=U_n=\text{Inf}\{t|\ F_\theta(x)\leq t\}$$

El intervalo de confianza. IC, (L_n, U_n) posee cotas aleatorias y el coeficiente de confianza, probabilidad de cubrimiento, del parámetro de interés esπ. Esto es
$\text{Prob}_\theta\{T(F)=\theta\in(L_n,U_n)\}\geq\pi$, para todo $\theta\in\Theta$. Usualmente hacemos $\pi=1-\alpha$. Si hacemos un cambio en la notación y tomamos como funcional a $\theta(F)$ con $F\in\Upsilon$ lo correcto será denotar el intervalo de confianza como

$$\text{Prob}_F\{\theta(F)\in(L_n,U_n)=I(\theta)\mid F\in\Upsilon\}\geq\pi$$

Esto denota que se depende de conocer la FD para poder calcular el intervalo de confianza. Si $\theta\in\Re^k$ el mismo razonamiento nos lleva a determinar un subespacio del espacio euclidiano de k-dimensiones tal que $\text{Prob}_F\{\theta(F)\in I(\theta)\subset\Re^k\mid F\in\Upsilon\}\geq\pi$

La estadística aborda el problema de obtener el $I(\theta)$ de volumen mínimo para un π fijo para una cierta FD. Si k=1 aunque $I(\theta)$ depende de F lo usual es acudir a Teoría Asintótica y a partir de un Teorema Central del Limite buscar condiciones para que

$$I_{asint}(\theta)=(\theta(F_n)-z_{1-\alpha/2}\,\sigma_n(\theta_n),\ (\theta(F_n)+z_{1-\alpha/2}\,\sigma_n(\theta(F_n))$$

Donde $\theta(F_n)$ es el predictor , $z_{1-\alpha/2}$ es el percentil de la distribución Normal Standard y $\sigma_n(\theta(F_n))$ es el estimador de la varianza del predictor de $\sigma(\theta(F))$. La robustez de $\theta(F_n)$ y $\sigma_n(\theta(F_n))$ juega un rol importante en la validez de π como probabilidad de cubrimiento.

En general usando un predictor y tomando $Z_n=\theta(F_n)-\theta(F)$ tal que $G_n(t,F)=\text{Prob}_F(Z_n\leq t)$ podemos fijar $\alpha=\alpha_1+\alpha_2$, $\alpha_r\in[0,\alpha]$, r=1,2. Las cotas del intervalo confidencial son

$$L(\alpha_1, X_1,\ldots,X_n)=L_n(\alpha_1)=\text{Sup}\{t|\ G_n(t,F)\leq\alpha_1\}$$

$$U(\alpha_2, X_1,\ldots,X_n)=U_n(\alpha_2)=\text{Inf}\{t|\ G_n(t,F)\geq 1-\alpha_2\}$$

Sustituyendo convenientemente tenemos que $\text{Prob}_F\{\theta(F)\in(\theta(F_n)-L_n,\ \theta(F_n)+U_n)=I(\theta)\}\geq\pi$. Este método general solo funciona si $\sigma(\theta(F)$ es conocida. De no serlo utilizaremos $\sigma_n(\theta(F_n))$. Suponiendo que la variable $T_{no}=\theta(F_n)-\theta(F)$ tiene una distribución simétrica respecto al origen vale usar $\alpha/2=\alpha_1=\alpha_2$

Cuando $\sigma_n(\theta(F_n))$ es un estimador consistente de $\sigma(\theta(F_n))$, bajo ciertas condiciones de regularidad, $\sigma_n^2(\theta(F_n))/\sigma^2(\theta(F_n))\longrightarrow_P 1$, se acepta que la distribución limite de $n^{1/2}(\theta(F_n)-\theta(F))\ \sigma_n^{-1}(\theta(F_n))$ sea la normal $N\{0,\ \sigma(\theta(F_n))\}$. Entonces $L(\alpha_1, X_1,\ldots,X_n)=L_n(\alpha_1)=-z_{1-\alpha/2}\ \sigma_n(\theta(F_n))$ y

$$U(\alpha_2, X_1,\ldots,X_n)=U_n(\alpha_2)=z_{1-\alpha/2}\,\sigma_n(\theta(F_n))\ .$$

Por lo que $\text{Lim}_{n\longrightarrow\infty}\ \text{Prob}_F\{\theta(F)\in(\theta(F_n)-L_n,\ \theta(F_n)+U_n)=I(\theta)\}\geq\pi$. Pues el límite de la amplitud de este intervalo es $2z_{1-\alpha/2}\sigma(\theta(F_n))$.

El enfoque paramétrico se dedica a buscar un intervalo de amplitud mínima , o lo que es lo mismo que un estimador con mínima varianza bajo la normalidad asintótica, Best Asimptotically Normal. En el enfoque robusto se busca un preeditor que sea robusto, en el sentido local o global, que satisfaga la convergencia del estimador de la varianza. Por ello debemos prestar particular atención a la consistencia de $\sigma_n(\theta(F_n))$ buscando una rápida

convergencia a la normal. Las propiedades de los estimadores del tipo M, L y R garantizan la convergencia y con ello que podamos trabajar con intervalos confidenciales asintóticamente normales. Este hecho es el que justifica el uso de los métodos de remuestreo .

4.2 El método de Jacknife

Al considerar aceptable que la ley de $(T(F_n)-T(F))n^{1/2}$ converge a la $N(0, \sigma(T(F))$ podemos desarrollar la esperanza de $T(F_n)=T_n$ en Series de Taylor como:

$$E_F(T_n)=T(F)+an^{-1} -bn^{-2} +\ldots.$$

Donde a , b.. son en principio constantes desconocidas que dependen de F. Consideremos el uso del kernel

$$\phi(\mathbf{x}\therefore x_i)= \phi(x_1 ,\ldots x_{i-1} ,x_{i+1} ,\ldots,x_n)= T_{n-i}$$

Cuya esperanza es $E_F(\phi(\mathbf{x}\therefore x_i))=0$, $E_F(\phi(\mathbf{x}\therefore x_i)^2)=\sigma^2(T)\in(0,\infty)$. Como $V_n^J =V_n^* +V_{nR} +2C_n$, entonces $T_{ni}-\theta= \phi(x_{i,}\theta)+R_{ni}$. Donde

$$V_n^*=(n-1)^{-1}\,\Phi_{i=1}^{\,n}\,(\phi(x_{i,}\theta)-\phi^*)^2,\ V_{nR}^*=(n-1)^{-1}\,\Phi_{i=1}^{\,n}\,(R_{ni}-R^*)^2,\ C_n=(n-1)^{-1}\,\Phi_{i=1}^{\,n}\,(\phi(x_{i,}\theta)-\phi^*)(R_{ni}-R)$$

Siendo $\phi^*=n^{-1}\,\Phi_{\,i=1}^{\,n}\,\phi(x_{i,}\theta)$, $R_{ni}=nR_n -(n-1)R_{(n-1)I}$ y $R^*=n^{-1}\,\Phi_{i=1}^{\,n}\,R_{ni}$. Como cuando $n\to\infty$ se tiene que $V_{nR}^*\to 0$ y $C_n\to 0$ se acepta la consistencia del estimador de la varianza, $V_n^*\to\sigma^2(T)$. O sea que el resto de la serie no tiene peso en el error esperado. Cuando el estimador de Jacknife admite esta representación podemos aplicar el Teorema de Slutsky y probar la convergencia a la normal.

En algunos casos extremos el estimador de Jacknife no puede ser representable así por no poseer una FD suficientemente suave. Tal es el caso con los estadísticos de orden. Un paliativo es utilizar el llamado Jacknife de las k eliminaciones (k-deleted Jacknife). En este se eliminan $k\in\{2,.., n-1\}$. El caso k=1 es el ya discutido. Para hacer su definición consideremos :

$$\boldsymbol{\tau}=(t_1 ,\ldots,t_k)\in I(k)=\{\tau \mid 1\le t_1 \le\ldots t_k \le n\},\ \eta=\{1,\ldots,n\},\ s(\tau|n-k)=\eta-\tau.$$

Note que $|I(k)|=C^n_k$.

El ejemplo clásico que apunta la no diferenciabilidad de un funcional, y con ello ilustrar las advertencias hechas respecto al usar Jacknife, es el de la mediana. Veamos la discusión hecha por Ghosh et.al. (1984).

$$T(F_n)= \text{inf. } \{x|F_n (x)\ge 0{,}5\}$$

Consideremos que F tiene densidad f continua en una vecindad de la mediana poblacional

$$T(F)= \text{inf } \{x| F (x)\ge 0{,}5\}$$

Entonces $Z(n)=(T(F_n)-T(F))n^{1/2}$ se distribuye asintóticamente $N(0, (4f^2(T(F))^{-1})$ si $n\to\infty$. Esto nos puede permitir hacer un tratamiento paramétrico de las inferencias sobre la mediana de F pero esta es desconocida y con ello su densidad f. La estimación de la varianza σ_T^2 usando Jacknife tiende a la variable aleatoria $\xi=(4f^2(T(F))^{-1}(\chi^2 /2)^{2,}$ donde χ^2 se distribuye $\chi^2(2)$. Entonces la FD del k-deleted Jacknife (k-Jacknife) es:

$$F_{n-k,\tau}(x)=(n-k)^{-1}\,\Phi_{\,j\in s(\tau|n-k)}\,I(X\le x),\ x\in\Re.$$

Consecuentemente tendremos a $T_{n-k,\tau}=T(F_{n-k,\tau})$, $\tau \in I(k)$ y $T_{n,k,t}= k^{-1}\ (nT(F_n)-(n-k)T(F_{n-k,\tau}))$, $\tau \in I(k)$ es el τ-ésimo pseudo valor por lo que el estimador de k-Jacknife de T(F) està dado por

$$T_{(n,k)J} = (C_k^n)^{-1}\ \Phi_{\tau\ \in I(k)}\ T_{n,k(\tau)}$$

Naturalmente la varianza ingenua es $V_{(n,k)} =(C_k^n)^{-1}\ \Phi_{\tau\ \in I(k)}\ (T_{n,k(\tau)}-T_{(n,k)J})^2$. Pero la de Jacknife a utilizar, por razones de velocidad de convergencia, es $V_{(n,k)J}=k(n-1)(n-k)^{-1}\ V_{(n,k)-}$

Es importante fijar que si T(F) es suficientemente suave, lo que se asocia a la diferenciabilidad del funcional $(n-k)(T(F_n)-T_{(n,k)J})\rightarrow B(F)$, donde B(F) es el llamado ´sesgo asintótico de T_n´. El uso de Jacknife elimina el sesgo de primer orden, como vimos anteriormente, pero solo si T no es un estadístico funcional robusto el uso del método de Jacknife tampoco lo es. Más aún, este es más sensible a la no diferenciabilidad dada que la descripción hecha usando ϕ. Las condiciones fijadas para que la representación sea consistente indica claramente que los pseudo valores son afectados si no se cumplen fuertes condiciones de regularidad . Esto también afecta a los estimadores de σ^2_T en su consistencia.

4.3 El método Bootstrap

El uso de F_n como aproximación de la FD desconocida es la solución usualmente aceptada por los estadísticos. El método de Bootstrap parte de considerar que la información a la mano es la brindada por F_n. Si seleccionamos una muestra de la FD conocida F_n podremos hacer estimaciones de $T(F_n)$ y con ello de T(F). Una muestra de Bootstrap es una replicación de la muestra observada $(X_1 ,..,X_n)$ al seleccionar una muestra simple aleatoria con reemplazo de ella $(X^*_1 ,..,X^*_n)$. Esto nos garantiza que trabajamos condicionalmente con variables aleatorias con igual distribución e independientes entre si. La FD de Bootstrap la denotamos por F^*_n. Entonces el funcional:

$$T^*_n=T(F_n^*)=T(X^*_1 ,...,X^*_n)$$

Estima $T(F_n)$. De la definición dada está clara la independencia condicional de las X*, ya que

$$Prob(X_i^* =X_t\ |F_n\ )=1/n\ ,\ \forall t=1,...,n,\ i\geq 1.$$

El espacio muestral de las muestras generadas por el método de Bootstrap, denotémosla S(BS), tiene como cardinal n^n. Cada muestra de Bootstrap $(X^*_{b1},..,X^*_{bn})$ tiene la probabilidad de ser observada $1/n^n$. De ahí que

$$E(T^*(F^*_n\ )|F_n)= n^{-n}\ \Phi_{s(b)\in S(BS)}\ T(X^*_{1,...,}X^*_n)_b= n^{-n}\ \Phi_{s(b)\in S(BS)}\ T_{nb}$$

Tiene como error condicional a $E(T^*(F^*_n\ )-T_n\ |F_n)^2= n^{-n}\ \Phi_{s(b)\in S(BS)}\ (T_{nb}-T_n)^2$. Este converge a σ_T^2 si $n\rightarrow\infty$.

Computar todas las muestras de Bootstrap tiene sentido formal pero hacer un censo seria muy costoso. En la práctica lo que hacemos es tomar una muestra de B muestras aleatoria e independientemente de S(BS) y calcular T_{nb}, b=1,…,B. Entonces un estimador de la varianza es

$$V^*_{nB}=(nB)^{-1}\Phi_{b=1}^{B}\ (T_{nb}-T_n)^2$$

Usando la FD de Bootstrap podemos calcular sus quantiles

$$F^*_n(t) = B^{-1} \Phi_{b=1}{}^{B} I((T_{nb}-T_n)n^{1/2} \leq t), \quad t \in \Re$$

convergen, bajo condiciones de regularidad muy débiles, V^*_{nB} a $\sigma_T{}^2$ y los quantiles de F^*_n a los de la FD generadora de los datos υ si para $n \to \infty$, $P_F\{(T(F_n)-T(F)n^{1/2} \leq t\} \longrightarrow \upsilon(t)$

Otra representación de primer orden que podemos hacer es la basada en series de Edgeworth en la que se valida la relacion

$$H_n(x)=P(n^{1/2}\,(T(F_n)-T(F)) \leq x)=\Phi(x|\sigma(F))+n^{-1/2}\,p(x,F)\phi(x|\sigma(F))+o(n^{-1/2})$$

Siendo $\Phi(x|\sigma(F))$ y $\phi(x|\sigma(F))$ la FD y la de densidad asociadas a $N(0,\sigma(F))$, $p(x,F)$ un polinomio en x que depone de algunos momentos y $o(n^{-1/2})$ el orden de convergencia de H_n. Al utilizar un estimador $\sigma^2(F_n)$ consistente para $\sigma^2(F)$ la representación para la FD de Boostrap es

$H_{nB}(x)=P^*(n^{1/2}\,(T(F^*_n)-T(F_n)) \leq x)=\Phi(x|\sigma(F_n))+n^{-1/2}\,p(x,F_n)\phi(x|\sigma(F_n))+o(n^{-1/2})$ cs (as).

Cuando $P(|p(x,F)-p(x,F_n)|) \to 1$ se tiene que se cumple casi seguramente (cs)

$$H_n(x)-H_{nB}(x)=P(n^{1/2}(T(F_n)-T(F)) \leq x)-P^*(n^{1/2}(T(F^*_n)-T(F_n)) \leq x)=\phi(x|\sigma(F))-\phi(x|\sigma(F_n))+o(n^{-1/2})$$

Entonces podemos usar H (x) para obtener un quantil en forma aproximada pues la convergencia de $\sigma(F_n)$ garantiza la de estos a los obtenibles usando F_n.
En muchas ocasiones los quantiles de Bootstrap son más consistentes que los de F_n para estimar los de la FD desconocida. Entonces podemos calcular IC´s usando directamente los quantiles de F^*_n al determinar

$$.IC(T(F))=(Sup\ \{t|F^*_n(t) \leq \alpha'\},\ Inf\{t|F^*_n(t) \geq 1-\alpha''\})=(t_{BS}\,(\alpha'),t_{BS}\,(1-\alpha''))$$

El usual convenio es utilizar $\alpha'=\alpha''=\alpha/2$. Estos IC en general no tienen por centro a $T(F^*_n)$.

Cuando la representación de primer orden

$$T(F_n)-T(F)=n^{-1}\Phi^{n}_{i=1}\,\phi(X_i\,,T(F))+R_n$$

es válida $\upsilon(t)$ es la $N(0,\ \sigma_T)$ y $F^*_n(t)$ converge a ella. Entonces podremos determinar IC´s usando los quantiles de la N(0,1) de acuerdo a las fórmulas asociadas a un Teorema Central del Limite o directamente de la FD Bootstrap. Estos en general van a ser diferentes.

Veamos como funcionan en la práctica los resultados discutidos.

Consideremos que T(F) es la media de la población. Al usar la aproximación normal basada en un Teorema Central del Límite el IC será:

$$IC\{\mu=T(F)\}=(T(F_n)-z_{1-\alpha/2}\,\sigma(F_n)n^{-1/2},\ T(F_n)+z_{1-\alpha/2}\,\sigma(F_n)n^{-1/2})$$

Es sabido que la probabilidad de cubrimiento es:

$$P\{T(F) \in IC(T(F)\}=1-\alpha-\{\mu_3\,(2\,z^2_{1-\alpha/2}+1)(3\sigma^2(F)n^{1/2})^{-1}\phi(-z_{1-\alpha/2}\,)\}+o\,(n^{-1/2})$$

El término entre llaves representa en cuanto disminuye el valor supuesto de la probabilidad de cubrimiento $1-\alpha$ asumida.

Por su parte si usamos el IC de Bootstrap

$$IC_B(T(F))=(T(F_n)-n^{-1/2}H^{-1}_{nB}(\alpha/2),\ T(F_n)+n^{-1/2}H^{-1}_{nB}(1-\alpha/2))$$

Y

$$P\{T(F)\in IC_B(T(F)\}=1-\alpha-\{\mu_3(2z^2_{1-\alpha/2})(\sigma^3(F)n^{1/2})^{-1}\}+o(n^{-1/2})$$

Si comparamos las dos probabilidades de cubrimiento tenemos que :

$$|P\{T(F)\in IC(T(F)\}-P\{T(F)\in IC_B(T(F)\}|=\phi(z_{1-\alpha/2})|\mu_3|(z^2_{1-\alpha/2}-1)(3\sigma^2(F)n^{1/2})^{-1}+o(n^{-1/2})$$

Está claro que si $n\rightarrow\infty$ esta expresión tiende a cero. Por tanto las probabilidades de cubrimiento son aproximadamente iguales. Estudios para problemas más complejos, (regresión, uso de estadísticos robustos, etc.) arrojan resultados similares para otras distribuciones y problemas. Esto nos da una cierta confianza en que al desconocer la FD el aplicar el método de Bootstrap para calcular los IC's es una solución cuyos resultados serán similares a lo obtenible al realizar un estudio paramétrico y usar una la aproximación a la FD desconocida.

Las probabilidades asociadas a una prueba de hipótesis pueden ser estimadas. Usando p como el p-value de la prueba podemos estimarle usando la proporción de veces en que se tomó la decisión correcta Sea π la probabilidad de aceptar H0 siendo falsa y π_n su estimador. Este es sesgado cuando usamos un estimador particular pues su esperanza es, tomando I(•) como la función indicadora, pues

$$E(\pi_n)=P(p\leq\alpha|H1)=EE(I(p\leq\alpha)\pi)\neq E(I(E(p\leq\alpha))=E(I(p\leq\alpha))=\pi.$$

Para estimar el p-value usaremos p_n . Una prueba modificada será aceptar H0 si $p_n\leq\alpha$.

El uso de Bootstrap nos lleva a estimar p utilizando para H0: $T\leq T_0$ utilizando a:

$$.p_B=\sum_{b=1}^{B} I(T^*_b\geq T_0)/B$$

El Bootstrap se adapta mejor que el de Jacknife a muchos problemas donde la diferenciabilidad del funcional no es aceptable. Sin embargo su robustez depende también de la de T.

Revisitando el problema de la mediana tenemos que si usamos $T(F^*_n)=$ inf. $\{x|F^*_n(x)\geq 0,5\}$ el procedimiento de Bootstrap es:

Procedimiento bootstrap para la mediana

1. Generar la muestra de Bootstrap b.
2. Computar $Z^*(n)_b=(T(F^*_n)-T(Fn))n^{1/2}$.
3. Mientras b<B b=b+1.
4. Ir a 2.
5. Computar $\sigma_b^2=V(Z^*(n)_b)$.

Este procedimiento funciona bien para funciones de la media pues

$$V(Z^*(n)_b)\rightarrow(4f^2(T(F))^{-1}\text{ si :}$$

i) F tiene una sola mediana

ii) .f es positiva continua en una vecindad de la mediana.
iii) Existe h>0 tal que $E|X^h|$ es finito

O sea que la estimación de la varianza usando Boostrap es un funcional consistente lo que no es satisfecho al usar Jacknife. Esto ejemplifica el comportamiento robusto del bootstrap en el estimar quantiles .

5. VARIANTES DE ALGORITMOS QUE USAN EL PRINCIPIO DEL BOOTSTRAP.

"Solo la cuchara sabe lo que pasa en la olla"(Provebio judío)

Veremos algunas variaciones basadas en el Bootstrap que son usadas para problemas de tipo específico.

5. 1. El método de los bloques móviles en el Bootstrap (Moving Blocks Bootstrap, MBB).

El procedimiento asociado los bloques móviles utiliza el principio de Bootstrap para tomar en cuenta las características de un proceso estocástico tal como los α-mixing o los presentes comúnmente en las Series de Tiempo. El procedimiento básico se puede describir como sigue

Procedimiento de MBB.

1. Fijar m, b, T
2. Dada la muestra $(X_1,...,X_n)$ calcular $\theta(F_n)=T(F_n)$
3. Determinar X_j y con este los bloques $b_j=(X_j,...,X_{j+b+1})$, j=1,...,q.
4. Tomar mediante muestreo simple aleatorio con reemplazo m de los q bloques formados.
5. La muestra de Bootstrap es $\cup_{j\in s} b_j=\{X^*_1,...,X^*_{mb}\}$
6. Calcular $T_t(F^*_B)$ donde F^*_B es la distribución empírica de los vectores X^*_h en la muestra.
7. Si t<T entonces t=t+1 y regresar a 4.
8. $T_B=\sum_{t=1}^{T} T_t(F^*_B)/T$ y $S^2_B=\sum_{t=1}^{T}(T_t(F^*_B)-T_B)^2/T-1$

En este procedimiento las X^*_h se distribuyen de acuerdo a la distribución de Bootstrap F^*_B, por lo que son independientes al condicionar a la muestra inicial $\{X_1,...,X_n\}$. Este y los procedimientos que se derivan de él capturan la dependencia de los datos y es robusta frente a la falta de homocedasticidad. Los trabajos pioneros son los de Künsch (1989), Liu-Singh (1992), Polites-Romano (1992 y 1995) y Fitzenberger (1997.

Uno de los problemas presentes en el uso del MBB es la determinación del b adecuado.

5.2. Uso de probabilidades desiguales (Importance Sampling)

El uso de probabilidades desiguales permite seleccionar muestras independientes con la misma distribución. Solo se afecta la hipótesis de equiprobabilidad. Tomando como probabilidad de seleccionar la variable X_i de la muestra $\{X_1,...,X_n\}$ a $\pi_i>0$ con

$$1=\sum_{i=1}^{n}\pi_i$$

el procedimiento de Bootstrap, ver Efron (1990), Johns (1988), Hinkley-Shi (1989) y Do-Hall (1991) es:

Procedimiento de Bootstrap con probabilidades desiguales

1. Seleccionar una muestra con reemplazo de tamaño n usando $\{\pi_1,...,\pi_n\}$.
2. Calcular $T(F^*_{n\pi})$ en la selección b-ésima-

3. Calcular $m_{(bi)}$= Numero de veces que X_i fue observado en la b-ésima muestra de Bootstrap.
4. Calcular
$$T(F^*_{nB\pi})=\sum_{b=1}^{B} T(F^*_{n\pi})_b \prod_{i=1}^{n}(n\pi_i)^{-m(bi)}$$

Como $E(T(F^*_{nB\pi}|F_n)=T(F_n)$, tomando

$$.m(i)^*=\sum_{b=1}^{B} m(bi)$$

$$V(T(F^*_{nB\pi})=\sum_{b=1}^{B} (T(F^*_{n\pi})_b^2 /\prod_{i=1}^{n}(n\pi_i)^{-m(i)^*})-(T(F_n))^2$$

Este procedimiento ha sido de particular utilidad en el estudio de problemas multivariados.

Note que cuando π_i=1/n para todo i tenemos el procedimiento clásico.

5.3. Bootstrap suavizado.

Usualmente tomamos la muestra de Bootstrap para estimar T (F). En ocasiones F no es una distribución naturalmente suave y la diferenciabilidad del funcional puede ser dudable. Para robustificar el Bootstrap respecto a esta inconveniencia se propone su suavizamiento . Para ello estimamos F a través de un estimador de Kernel, ver Silverman (1986) de su función de densidad. Este se denota:

$$F_K(x)=\int_{-\infty}^{x} f_K(x)dx$$

Donde

$$f_K(x)=(nh)^{-1}\sum_{I=1}^{n} K((x-X_i)/h)$$

La función de ancho de banda h es elegida para suavizar convenientemente la estimación. Por su parte K, la función de kernel (núcleo), es acotada y simétrica que se anula fuera de un cierto intervalo (-c, c). Además K´(x) existe y es una función continua, absolutamente integrable de variación acotada y

$$\int |x^r| K(x)dx<\infty.$$

El error del estimador de Bootstrap de la varianza de F es de orden $n^{-r/(2r+1).}$ El error del estimador de Bootstrap $\sigma_B^2=\sigma^2(F^*_n)$ es de orden $n^{-1/4}$ por lo que es mejor usar el estimador suavizado . Como f puede tomar valores negativos en algunos casos, puede ser necesario establecer una cierta corrección y trabajar en esos casos con $f^*_K=\delta|f^*_K|$ en los que la constante garantiza que $\int f^*_K(x)dx=1$. Una discusión amplia sobre este problema puede obtenerse en Silverman-Young (1987)y Lee-Young (1992).

5.4 Bootstrap Ponderado

En ocasiones no todas las observaciones tienen la misma importancia. La lógica subyacente en la selección con probabilidades desiguales es aplicable a partir de un procedimiento de Monte Carlo llamado “Importance Sampling”. La propuesta puede trazarse ya en le trabajo de Efron (1979). Mason- Newton (1990) comenzaron estudiando esta posibilidad en el contexto de los estadísticos lineales de rango. Podemos describirle en general de la siguiente forma:

Procedimiento básico de Bootstrap Ponderado.

1. Fijar B
2. Seleccionar una muestra aleatoria independiente $X_n=(X_1,...,X_n)$

de la FD desconocida F.

3. Generar un vector $W_n = (W_{11}, ..., W_{nn})$ de un distribución multinomial con parámetros $(n, \pi_1, ..., \pi_n)$.
4. $T(F_{nb} | W_n) = \sum_{i=1}^{n} W_{ii} T(X_i)/n = T_b$.
5. Mientras b<B entonces b=b+1
6. Ir a 3.
7. Computar

$$T^*_{nWB} = \sum_{b=1}^{B} T(F_{nb} | W_n)/B$$

$$\sigma_{WB}^2 = \sum_{b=1}^{B} (T(F_{nb} | W_n) - T^*_{nWB})^2 / B.$$

$$F^*_{nWB}(x) = \sum_{1=1}^{n} I(X_i \leq x) W_{ni}/n.$$

Este se basa en que la distribución asociada a $H(F_n) = T(F_n) - T(F)$ debe ser usada para hacer las inferencias pero no es conocida F. Por ello se utiliza la distribución subrogada $H^*(F_n) = T(F_{nB}) - T(F_n)$, que puede ser aproximada usando métodos de Monte Carlo. Por ejemplo si $T(F) = \mu$ utilizaremos a $(T(F_{nWB}) - T(F_n)) n^{1/2} / \sigma_{WB} = n^{1/2} \sum_{i=1}^{n} W_{ni} (X_i - \mu_n)^2 / \sigma_{WB}$.

Si F no es un làtice y el momento de tercer orden es finito el desarrollo en Series de Edgeworth de un paso fija que al condicionar a la muestra observada, ver Praestgaard (1990), Swanepoel (1986) por ejemplo:

$$Sup_{-\infty<\kappa<\infty} \{P\{(T(F_{nWB}) - T(F_n)) n^{1/2} / \sigma_{WB} < \kappa\} - (\Phi(\kappa) - \phi(\kappa)(\kappa^3 - 1) B_n / 6 n^{1/2} \sigma^3_{WB}\} = o(1/n^{1/2})$$

donde $B_n = \sum_{i=1}^{n} (X_i - \mu_n)^3 / n$. Por tanto podemos evaluar la convergencia a la N(0,1) y optar por utilizar la distribución de Bootstrap para hacer las inferencias o basarnos en la aproximación normal al analizar el valor de n.

5.5. El Bootstrap en el muestreo de poblaciones finitas

Sean **X** el vector de los datos en la población y **x** el de las observaciones. La muestra se selecciona para estimar θ**(X)** y θ**(x)** su estimador. Para evaluar su precisión se calcula su error cuadrático medio $E(\theta(\mathbf{x}) - \theta(\mathbf{X}))^2$

El enfoque usual lleva a derivar una fórmula explícita de este error para hacer después su estimación utilizando un estimador adecuado. Para ello se considera la distribución como la generada por el diseño de muestreo. El uso del Bootstrap funciona de acuerdo a la misma filosofía que para una distribución del caso infinito, estimando el modelo que produjo los datos observados. Sin embargo, debemos hacer algunas consideraciones especiales.

Generando B conjuntos de datos de Bootstrap, $\mathbf{x_1}, \ldots, \mathbf{x_B}$ del modelo estimado podemos valorar la variabilidad del mismo. Así el bootstrap usa las B muestras para estimar el error de muestreo reemplazando la desviación teórica derivada analíticamente por la de Bootstrap.

Ha sido reportado, ver Shao (2003) que este método produce estimaciones con un comportamiento muy similar al obtenido al estimar la expresión analítica del error.
Esto motiva que cuando la expresión analítica lleva a una complicada expresión de su estimador lo mejor sea usar un método de remuestreo , como el Bootstrap o el Jacknife.

El Jacknife y el uso de muestras balanceadas replicadas (balanced repeated replication, BRR), tienen una larga historia en el muestreo de poblaciones finitas. En el BRR, otro tipo de remuestreo se aplica al formarse B subconjuntos propios que satisfagan una cierta

condición de balanceo. Esto le hace similar al Jacknife pues no se aplica un método de aleatorización.

Ambos son muy útiles para aproximar los primeros dos momentos por lo que pueden utilizarse en estimar el sesgo y la varianza. Los problemas asociados a la no-suavidad de algunos funcionales conlleva a que nuevamente estos métodos tengan peor comportamiento que el Boostrap.

Si el muestreo es sin reemplazo debe tomarse en cuenta la fracción de muestreo f= *n/N.*
Cuando N es grande podemos considerar que la población finita estudiada es una de una sucesión de poblaciones finitas indexadas por q = 1, 2, ... Entonces la muestra observada es una $\mathbf{x} = \mathbf{x_q}$ *de tamaño* n_q seleccionada de la q-ésima población de tamaño N_q. Se considera que $n \to \infty$ y también $N \to \infty$ cuando $q \to \infty$.

Métodos de remuestreo son también aplicados en el tratamiento de las no respuestas. Saigo, Shao y Sitter (2001) propusieron el uso de las medias muestras y tomar muestras de bootstrap del mismo tamaño. Este método es asintotícamente válido en caracterizar la variabilidad, vea las experiencias de Saigo, Shao y Sitter (2001).

Ante los datos faltantes es usual hacer imputaciones. También es usado el método de Bootsrap en el análisis de datos con imputaciones por no-respuesta. Sea **x"** el conjunto de los datos imputados y $\theta'(\mathbf{x})$ un estimador sesgado utilizado usando la información obtenida. El método de imputación típico desarrolla una estimador $\theta''(\mathbf{x})$ que debe tener una esperanza muy similar en valores al basado en los respondientes. El error de muestreo de este debe ser mayor debido al error generado por imputar. Tomando los datos imputados, y dándole el mismo tratamiento, podemos aplicar un procedimiento de Bootstrap para generar conjuntos del tipo imputado $\mathbf{x}'_b$. Como es sabido la variabilidad del estimador de Bootstrap será menor, en este caso por ignorar el proceso de imputación. Vea una amplia discusión en Efron (1994) y Shao y Sitter (1996) quienes propusieron re-imputar los datos de Bootstrap tal y como se hizo con los datos originales al identificar que datos fueron de respondientes y cuales fueron no respondientes e imputar. Tal y como se hizo con los datos originales. La distribución condicional generada va a ser válida en el sentido que lo sea el método de imputación en hacer substituciones adecuadas. Por su parte Efron (1994) estableció que el método clásico de Bootstrap genera una distribución asintótica mente valida para los datos re-imputado al ´bootstrapear´.

5.6 . La no universalidad de la convergencia

Diversas modificaciones son implementadas sobre la base de este procedimiento. Una de los más populares generar un número diferente de muestras y no solo n. Sin embargo no es una panacea el método de Bootstrap. Por ejemplo al trabajar con estimadores robustos de la regresión aparecen problemas denominados de `inestabilidad numérica`. La distribución de Bootstrap puede ser un estimador muy malo de la los residuos en la regresión debido a la influencia de los outliers pues, en algunas muestras de Bootstrap, esta va ser mucho mayor que en los datos originales. Por ello F_N^* poseerá colas mas pesadas que F si hay outliers dando origen a una no convergencia rapida. Por otra parte como se plantea un problema de optimización no convexa, al usar un método robusto, puede ser muy caro de resolver la regresión en cada muestra de Bootstrap, incluso cuando B es pequeño, para poco más de 10 variables controladas.

Algo similar ocurre con la dificultad de estimar el parámetro de escala. Este debe ser calculado para cada muestra y este es uno de los problemas mas complejos en muchos problemas. Si esto no se hiciese no se garantizaría la convergencia a F y con ello fallaría la asumida robustez.

Reconocimientos: Esta monografía fue desarrollada durante una visita del autor al CIDET de la Facultad de Matemática y Computación de la UAG

Received June 2005
Revised April 2006

REFERENCIAS

BABU, C.J. y A. BOSE (1989): Accuracy of the bootstrap approximation. **Statist. and Probability Letters**. 7, 151-160

BABU, C.J. y K. SINGH (1983): Non parametric inferences on means using bootstrap. **Ann. Statist. 11,** 999-1003

BABU, C.J. y K. SINGH (198a4): Asymptotic representations related to jackniffing and bootstrapping L-statistics. **Sankhya A**. 46, 195-204.

BABU, C.J. y K. SINGH (1984b): On one term Edgeworth correction by Efron´s bootstrap. Sankhya A. 46, 219-232.

BAI, Z.D., X. R. CHEN, Y. WU y L.C. ZHAO (1990): Asymptotic normality of minimum L1-norm estimates in linear models. **Chinese Sciences. A,** 33, 440-463.

BASSET, G. y KOWENKER R. (1978): Asymptotic theory of least absolute error regression. J. **Amer. Stat. Ass**. 38, 439-443.

FITZENBERGER, B. (1997): The moving blocks bootstrap and robust inference for linear least squares and quantile regression. **J. of Econometrics**. 82, 235-287

HINKLEY D. V. y S. SHI (1989): Importance sampling and the nested bootstrap. **Biometrika** 76, 435-446.

JOHNs, M.V. (1988): Importance resampling for bootstrap confidence intervals. **J. Amer. Stat. Ass**. 83, 709-714

LAHIRI, S.N. (1992): Second order optimality of stationary bootstrap. En "**Exploring the Limits of Bootstrap",** Lepage, R y L. Billard (Eds.). Wiley, N. York. 183-214.

POLITIS D., J. P. ROMANO y M. WOLF (1992): Sub sampling non stationary time series. **Technical Report, Dept. Of Statistics, Stanford University**

POLITIS D., y J. P. ROMANO (1992): Large sample confidence regions based on sub samples under minimal assumptions. **Ann. Stat**. 22, 390-426.

PRAESTGAARD, J. (1990): Bootstrap with general weights and multiplier central limit theorems. **Technical Rep. 195. Department of Stat. University of Washington**

SAIGO, H., J., SHAO Y R. SITTER. (2001). A repeated halfsample bootstrap and balanced repeated replications for randomly imputed data. **Survey Methodology** 27 189–196

SHAO, J. Y H. WANG. (2002). Sample correlation coefficients based on survey data under regression imputation. **J. Amer.Statist. Assoc***.* 97 544–552

SHAO, J. y R. R. SITTER (1996). Bootstrap for imputed survey data. **J. Amer. Statist. Assoc***.* 91 1278–1288

SHAO, J.(2003)Impact of the Bootstrap on Sample Surveys. **Statistical Science***.* 182, 191–198

SILVERMAN, B. W. (1986): **Density estimation for Statistics and Data Analysis**. Chapman and Hall. Londres

SILVERMAN, B. W. (1987): The Bootstrap: to smooth or not to smooth. **Biometrika**, 74, 469-479.

SINGH, K. (1981): On the asymptotic accuracy of Efron´s bootstrap. **Ann. Statitst**. 9, 1187, 1195

SWNEPOEL, J.W.H. (1986): A note on proving that the modified bootstrap works. **Commun. Stat. Theory and Methods.** 15, 3193-3203

WU, C.F.J. (1986): Jacknife, bootstrap and other resampling methods in regression analysis. **The Annals of Stat**. 14, 1261-1295

YUNG, W y J.N.K. RAO (1986): Jacknife linearization of variance estimators under stratified multistage sampling. **Survey Methodology**, 22, 23-39.
YUNG, W Y J.N.K. RAO (2000): Jacknife variance estimation under imputation for estimators using poststratified information. **J. Amer. Stat. Ass.** 95, 903-919