Evaluación del Bootstrap para la estimación de percentiles extremos. Aplicación en intervalos de referencia

Pablo Andrés Guzmán González

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2011

Evaluación del Bootstrap para la estimación de percentiles extremos. Aplicación en intervalos de referencia

Pablo Andrés Guzmán González

Tesis presentada como requisito parcial para optar al título de:
Magister en Estadística

Director(a):
Juan Carlos Salazar, PhD.

Línea de Investigación:
Bioestadística

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2011

(Dedicatoria o un lema)

A mis padres

Agradecimientos

Un especial agradecimiento a la Escuela de Estadística de la Universidad Nacional de Colombia, Sede Medellín, que me permitió hacer parte de su Escuela y brindarme nuevo conocimiento. La programación del enfoque bayesiano en `R` fue posible gracias a las notas de clase de Juan Carlos Correa, y gracias a Juan Carlos Salazar por su comprensión y por apoyar desde un principio la idea. Finalmente, a mis padres y a mi esposa, por su apoyo y ánimo en todo momento, infinitas gracias !!!

Resumen

En medicina clínica (o de laboratorio), un *intervalo de referencia* (IR), para alguna variable biológica (e.g., la concentración de hemoglobina), es un conjunto de valores cuyos límites corresponden a dos percentiles simétricos con respecto a la mediana (usualmente, los percentiles 2.5 y 97.5) y es la herramienta de decisión más ampliamente usada en medicina para catalogar un sujeto como "saludable" o no. Debido a su importancia, diferentes técnicas estadísticas paramétricas y no paramétricas se han propuesto para su estimación. Este trabajo comparó, por simulación, el desempeño del bootstrap con el de otros métodos disponibles para estimar los límites del IR. En la comparación se propone aplicar una aproximación bayesiana donde el bootstrap se usa para estimar la verosimilitud del percentil, que al ser combinada con una distribución a priori adecuada, produce una densidad a posteriori estimada del percentil de interés. Esta densidad a posteriori puede ser usada para hacer las estimaciones respectivas. El bootstrap se encontró con un menor error cuadrático medio que los otros métodos disponibles en poblaciones con menor sesgo. De otro lado, el enfoque bayesiano mostró igual o menor error cuadrático medio que todos los otros métodos evaluados, incluyendo el bootstrap. En la simulación se usaron aprioris medianamente informativas. Los resultados sugieren que el método bayesiano propuesto es una alternativa viable cuando se tienen tamaños de muestra pequeños (30 ó 40) y una información apriori insesgada y con baja dispersión.

Palabras clave: Percentil extremo, Bootstrap, Estimación, Intervalo de referencia, Bayes, Estimación Robusta.

Abstract

In clinical medicine, a *reference interval* (IR), for some biological variable (e.g., hemoglobine concentration), is a set of values whose boundaries correspond to two symmetric percentiles from the median (usually, the percentiles 2.5 and 97.5); then, IR's are decision tool the most widely used in medicine to classify a subject as "healthy" or not. Because of its importance, different techniques parametric and nonparametric statistics have been proposed for estimation. This study compared, by simulation, the performance of the bootstrap with the other available methods to estimate the reference limits. It also plans to implement a Bayesian approach where the bootstrap is used to estimate the likelihood of the percentile, which when combined with a proper prior distribution, produces a posterior density estimate of the percentile of interest. This posterior density can be used to make the respective estimates. The bootstrap was found with a lower MSE than other methods available in less biased populations. On the other hand, the Bayesian approach showed equal or lower RMSE than

all other methods evaluated, including the bootstrap. The simulation used aprioris fairly informative. The results suggest that the proposed Bayesian method is a viable alternative when you have small sample sizes (30 or 40) and a priori information unbiased and low dispersion.

Keywords: Extreme Percentil, Bootstrap, Estimation, Reference Interval, Bayes, Robust Estimation

Contenido

Agradecimientos vii

Resumen ix

Lista de Figuras xiii

Lista de Tablas xiv

1. Introducción 1

2. Marco Teórico y Estado del Arte 3
2.1. Estado del Arte y Conceptos Básicos 3
2.2. Bootstrap 7
2.3. Enfoque bayesiano 10

3. Objetivos 11
3.1. Objetivo general 11
3.2. Objetivos específicos 11

4. Metodología 12

5. Resultados 14
5.1. Desempeño de la estimación puntual 14
5.2. Cobertura de los intervalos de confianza 15
5.3. Ejemplo de aplicación 17

6. Discusión 21

7. Conclusiones 25

A. Anexo: Programas computacionales para los métodos evaluados 26

Bibliografía 35

Lista de Figuras

5-1. Desempeño (medido con la raíz del error cuadrático medio, $\sqrt{\mathsf{ECM}}$) de cuatro métodos para la estimación puntual del percentil 2.5 % ($\psi_{0.025}$) desde el muestreo de (A) una población $\mathsf{normal}(0, 1)$ y (B) una población t_5. Los métodos son: ($\circ$) Bayesiano con una apriori $\mathsf{normal}(\psi_{0.025}, 0.38\psi_{0.025})$; ($\triangle$) Bayesiano con una apriori $\mathsf{normal}(\psi_{0.025}, 0.28\psi_{0.025})$; ($+$) Bootstrap del percentil empírico tipo 5 (media de las replicas bootstrap); ($\times$) percentil empírico tipo 5; ($\diamond$) Harrell & Davis. 15

5-2. Desempeño (medido con la raíz del error cuadrático medio, $\sqrt{\mathsf{ECM}}$) de cinco métodos para la estimación puntual del percentil 2.5 % ($\psi_{0.025}$) (arriba: A y C) y 97.5 % ($\psi_{0.975}$) (abajo: B y D) desde el muestreo de (A y B) una población χ^2_6 y (C y D) una población χ^2_{12}. Los métodos son: ($\circ$) Bayesiano con una apriori $\mathsf{normal}(\psi_{2.5}, 0.38\psi_{0.025})$ (A y C) y con una apriori $\mathsf{normal}(\psi_{0.975}, 0.20\psi_{0.975})$ (B y D); ($\triangle$) Bayesiano con una apriori $\mathsf{normal}(\psi_{0.025}, 0.28\psi_{0.025})$ (A y C) y con una apriori $\mathsf{normal}(\psi_{0.975}, 0.10\psi_{0.975})$ (B y D) ; ($+$) Bootstrap del percentil empírico tipo 5 (media de las replicas bootstrap); ($\times$) percentil empírico tipo 5; ($\diamond$) Harrell & Davis; ($\square$) Robusto de Horn (sólo para el percentil 97.5 %, B y D) 16

5-3. Cobertura (medida como la proporción de intervalos que incluyen el verdadero valor del percentil en las 1000 simulaciones Monte Carlo) para intervalos tipo bootstrap y por el método bayesiano, del percentil 2.5 % ($\psi_{0.025}$) (izquierda: A, C, E, G) y 97.5 % ($\psi_{0.975}$) (derecha: B, D, F, H) desde el muestreo de (A y B) una población $\mathsf{normal}(0, 1)$, (C y D) t_5, (E y F) χ^2_6 y (G y H) una población χ^2_{12}. Los métodos son: (+) Bayesiano con una apriori $\mathsf{normal}(\psi_p,\ 0.38\psi_p)$ para las poblaciones simétricas en ambos percentiles y para el $\psi_{0.025}$ en las poblaciones χ^2s. Para el $\psi_{0.975}$ en las χ^2s se usó la apriori $\mathsf{normal}(\psi_{0.975},\ 0.20\psi_{0.975})$; (×) Bayesiano con una apriori más informativa, $\mathsf{normal}(\psi_p,\ 0.28\psi_p)$ en el caso de las poblaciones simétricas y el $\psi_{0.025}$ de las χ^2s y con una apriori $\mathsf{normal}(\psi_{0.975},\ 0.10\psi_{0.975})$ para el $\psi_{0.975}$ de las χ^2s; (∘) Bootstrap del percentil empírico tipo 5 con el método del percentil; (△) Bootstrap del percentil empírico tipo 5 con el método BC_a. La línea discontinua (- - -) se ubica en 0.9, el nivel de confianza nominal con el cual se pidieron los intervalos en cada simulación. 18

5-4. Distribución, con histograma (arriba) y boxplot (abajo), de valores de glucosa en plasma de una muestra de 46 hombres (edad > 80 años), reportada por Horn et al. (1998). Los rectángulos sombreados ubicados en cada cola de la distribución señalan, en cada caso, un posible rango de valores para los límites inferior (3.890 a 4.110 mmol/L) y superior (5.551 a 6.400 mmol/L) del intervalo de referencia para la glucosa en ayunas (ver texto). 19

6-1. Distribución (mediante histogramas) de las réplicas bootstrap, $\hat{\psi}_p^{*(b)}$, (con $B = 1000$ réplicas) para muestras de tamaño n extraídas de una población $\mathsf{normal}(0, 1)$. La línea vertical discontinua identifica en cada caso el valor del estadístico obtenido con la muestra original. La banda negra horizontal en cada cola, indica el intervalo de confianza tipo BC_a obtenido. 23

6-2. Distribución (mediante histogramas) de las réplicas bootstrap, $\hat{\psi}_p^{*(b)}$, (con $B = 1000$ réplicas) para muestras de tamaño n extraídas de una población χ^2_6. Los cuantiles 0.025 y 0.975 verdaderos son 1.24 y 14.45 respectivamente. La línea vertical discontinua identifica en cada caso el valor del estadístico obtenido con la muestra original. La banda negra horizontal en cada cola, indica el intervalo de confianza tipo BC_a obtenido. 24

Lista de Tablas

5-1. Cobertura estimada de los intervalos de confianza obtenidos bajo el bootstrap y el enfoque bayesiano para cada límite de un intervalo de referencia del 95 %. La cobertura nominal fue 0.900 (ver Figura **5-3**). 17

5-2. Estimación de un intervalo de referencia del 95 % para la glucosa basado en los diferentes métodos evaluados y usando la muestra de 46 hombres (edad $>$ 80 años) reportada por Horn et al. (1998) (ver Figura **5-4**). 20

1. Introducción

Los cuantiles (o percentiles) tienen una amplia aplicación en las ciencias. Un ejemplo de ello se encuentra en la determinación de *Intervalos de referencia* (IR) en química clínica (o medicina de laboratorio). Un IR, para alguna variable biológica (e.g., la concentración de hemoglobina), es un conjunto de valores cuyos límites corresponden a dos percentiles de la distribución de probabilidad de la población, y que son simétricos con respecto a la mediana (Wright & Royston, 1999). Los límites de mayor uso en química clínica son el percentil 2.5 y el 97.5 que definen un IR con una cobertura del 95 % (Wright & Royston, 1999). Los IR sirven como patrón de comparación para identificar si una medición (realizada sobre una persona en particular de la cual, por ejemplo, se sospecha que padece alguna enfermedad) es demasiado alta, demasiado baja o se encuentra en un rango aceptable (Wright & Royston, 1999; Koduah et al., 2004; Solberg, 2004). En este sentido, son una de las herramientas más ampliamente usadas en medicina de laboratorio (Horn & Pesce, 2003). Como ejemplo considere el estudio de Balconi et al. (2006) en el cual se determinan IR's para tres variables asociadas con el riesgo a sufrir infarto agudo del miocardio en un grupo poblacional argentino; tales variables fueron *Troponina I*, *Creatina Quinasa Fracción MB Masa* y la *Mioglobina*. Estos autores estimaron, en mujeres, un IR del 95 % para la Mioglobina de 8.3 a 40.3 ng/mL, mientras que para los hombres de 11.5 a 56.5 ng/mL. Así, por ejemplo, cualquier medición de la mioglobina en una mujer por debajo de 8.3 o por encima de 40.3 ng/mL debe considerarse indicador de que podría haber algún problema de salud. En la literatura sobre química clínica se pueden encontrar numerosos estudios que reportan IR's para diferentes variables bioquímicas ("analitos", en la jerga de química clínica). Algunos ejemplos se pueden encontrar en Tanyalcin et al. (2000); Jiménez Díaz & Schosinsky Nevermann (2000); Heider et al. (2004); Fernandez et al. (2006); Yip et al. (2006); en medicina veterinaria, los IR también son necesarios, ver por ejemplo Paul Murphy et al. (1994); Nordøy & Thoresen (2002).

La estimación estadística de los límites de un IR (como se definió arriba) ha tenido un alto desarrollo teórico (ver sección 2.1) y se han propuesto diferentes técnicas, paramétricas y no paramétricas, (Wright & Royston, 1999; Horn & Pesce, 2003; Solberg, 2004) para tal fin. Entre las no paramétricas se encuentra el bootstrap, método de remuestreo que ha sido bien recomendado (Solberg, 2004). De hecho, el bootstrap se ha visto en los últimos tiempos como la solución estadística a diversos problemas de estimación, no obstante, debe tratarse

con cuidado; Haukoos & Lewis (2005) no recomiendan el bootstrap para hacer estimaciones puntuales, sin embargo este se ha utilizado con ambos propósitos, estimar puntualmente los límites del IR y generar intervalos de confianza para cada límite (Solberg, 2004). Linnet (2000) encontró deficiencias en el bootstrap respecto a la estimación de la incertidumbre en los límites estimados del IR sobre todo cuando se tienen tamaños de muestra pequeños. Si bien existen diferentes métodos para obtener intervalos de confianza vía bootstrap (con los cuales se estima la incertidumbre alrededor de cada límite estimado), tales como el "percentil", "BC_a" , etc. (Efron & Tibshirani, 1993; Carpenter & Bithell, 2000), hacen falta estudios que evalúen el desempeño de tales métodos boostrap en la estimación de los percentiles del IR.

Con frecuencia la estimación de los límites de un intervalo de referencia para una variable en una población particular, supone algún conocimiento a priori de los límites de referencia debido a mediciones realizadas en otras poblaciones o debido a que el propio kit de laboratorio para medir la variable trae consigo unos límites de referencia (e.g., Balconi et al., 2006; Horn & Pesce, 2003; Barreto Penie et al., 2003, entre otros). La inferencia bayesiana, un marco conceptual distinto de la inferencia clásica frecuentista, permite incluir de forma explícita tal información previa. Bajo este enfoque, se construiría una distribución a posteriori del percentil 2.5 (ó del 97.5) dado los datos, proporcional al producto de una distribución a priori del percentil 2.5 (que se construye con base en la información previa) y una función de verosimilitud del percentil, la cuál depende del conocimiento de la forma paramétrica para la distribución que genera los datos. Sin embargo, Boos & Monahan (1986) proponen utilizar una función de verosimilitud estimada vía bootstrap de los propios datos. A partir de la distribuición a posteriori de cada percentil, se obtiene una estimación puntual y un intervalo de confianza para el mismo. El enfoque bayesiano puede resultar particularmente útil cuando se tienen tamaños de muestra pequeños.

El presente trabajo evalúa el desempeño del bootstrap en la estimación de los percentiles 2.5 % y el 97.5 % que definen los bien conocidos IR en medicina de laboratorio, así como también, dos métodos para obtener intervalos de confianza vía bootstrap: el método del percentil y el BC_a, y se implementa una estimación bayesiana semi-paramétrica usando un estimador de la verosimilitud construido con base en el bootstrap.

2. Marco Teórico y Estado del Arte

2.1. Estado del Arte y Conceptos Básicos

Desde muy temprano existe preocupación por llevar registros de valores para variables biológicas a nivel bioquímico para la población humana (e.g., Jellinek & Looney, 1939). Inicialmente y más o menos hasta los 70s se utilizaba el término "valores normales" o "rango normal" en lugar de intervalo de referencia (o valores de referencia) con la idea de que los sujetos con mediciones por fuera del rango eran "anormales". Sin embargo un defecto de esta terminología es que un intervalo del 95 % deja, por definición, al 5 % de los individuos "normales" fuera del rango. Además, nada impide el determinar intervalos de referencia para poblaciones de enfermos con una patología concreta, donde por tanto el término "normal" constituye en cierta medida un contrasentido. Por otro lado, la palabra "normal" induce rápidamente a pensar en una distribución de probabilidad normal o gaussiana, cuando es habitual encontrar variables bioquímicas cuyos datos no se ajusten a este modelo probabilístico (Wright & Royston, 1999; Molinero, 2007). El trabajo de Gräsbeck & Saris (1969) llama la atención sobre esto y promueve el nuevo término "intervalo de referencia". Desde aquí, el uso del término "rango normal" ha ido decayendo, no obstante, a veces se utiliza.
A partir de la década de los 80, diferentes comités nacionales e internacionales han promovido el desarrollo y estandarización del tema de valores de referencia y en ellos se hacen recomendaciones sobre el tratamiento estadístico para determinar los intervalos. Por ejemplo, en 1987, la Federación Internacional de Química Clínica (IFCC) publicó una serie de recomendaciones (divididas en seis partes) relacionadas con intervalos de referencia, de las cuales la quinta parte es sobre el tratamiento estadístico (Solberg, 1987), mientras que en 1992, el subcomité sobre intervalos de referencia del Comité Nacional para Estándares de Laboratorio Clínico (NCCLS, de su nombre en ingles) publicó el documento NCCLS (1992) el cual ha servido de guía en el tema. En el trabajo de Solberg (1987) se hacen las siguientes sugerencias para calcular un intervalo de referencia: (1) Tener presente métodos paramétricos (distribución normal) y no paramétricos. (2) En caso de aplicar métodos paramétricos, tener presente transformaciones (e.g., Box–Cox) para alcanzar normalidad; y (3) que los límites de referencia estimados deben acompañarse por una medida de su incertidumbre (e.g., un intervalo de confianza). Estas recomendaciones han sido bien aceptadas al tiempo que han motivado un mayor desarrollo sobre el tema (Solberg, 2004).

El tratamiento estadístico de intervalos de referencia se puede considerar desde dos enfoques: primero, determinación de intervalos de referencia sin considerar su cambio en la distribución de las mediciones con la edad (intervalos independientes del tiempo) y segundo, intervalos de referencia relacionados con la edad (Wright & Royston, 1997, 1999; Koduah et al., 2004; Molinero, 2007). El segundo enfoque no será abordado en este proyecto; para una mayor profundización ver por ejemplo, Wright & Royston (1997); Virtanen et al. (1998); Royston & Wright (1998); Griffiths et al. (2004). Respecto al primer aspecto, el problema consiste en estimar dos percentiles que definan el intervalo que incluya un porcentaje deseado de la población de referencia.

Definición de los parámetros a estimar Sea $f_Y(y \mid \boldsymbol{\theta})$ la función de densidad de probabilidad, con el vector de parámetros $\boldsymbol{\theta}$, para la variable aleatoria continua Y de la población de *referencia* y de la cual se desea conocer su verdadero *intervalo de referencia* (IR). Y es una variable bioquímica tal como la concentración de hemoglobina. Sea ψ_p (con $0 \leq p \leq 1$) el $p\,100\,\%$ percentil verdadero de la población f_Y; se busca estimar los percentiles $\psi_{(1-q)/2}$ y $\psi_{(1+q)/2}$, (con $0 \leq q \leq 1$), los cuales definen un IR del $q\,100\,\%$. Aquí nos interesa la cobertura habitual para un IR, que es de $q = 0.95$ (Geffre et al., 2009), en este caso el objetivo es estimar dos límites, $\psi_{0.025}$ y $\psi_{0.975}$, es decir, los percentiles (o cuantiles) $2.5\,\%$ y $97.5\,\%$.
Existen diferentes métodos para estimar $\psi_{0.025}$ y $\psi_{0.975}$ usando una muestra aleatoria tomada de la población de referencia. Protocolos para seleccionar la muestra y en general para conducir estudios sobre determinación de IR's están bien establecidos (e.g., Horn & Pesce, 2003; Geffre et al., 2009). En este estudio nos centramos en los métodos inferenciales para estimar $\psi_{0.025}$ y $\psi_{0.975}$. En seguida se describen algunos de estos métodos siguiendo las revisiones de Horn et al. (1998), Wright & Royston (1999), Horn & Pesce (2003), Correa (2004) y Solberg (2004), haciendo énfasis en las técnicas evaluadas en este estudio.

Métodos paramétricos Tales métodos requieren conocer (o asumir) una forma para $f_Y(y \mid \boldsymbol{\theta})$ (por ejemplo, normal, gamma, etc.). Una vez que se estima $\boldsymbol{\theta}$ por métodos ya bien establecidos (e.g., máxima verosimilitud), ψ_p se estima resolviendo para y

$$F_Y(\hat{\psi}_p|\hat{\boldsymbol{\theta}}) = \int_{-\infty}^{\hat{\psi}_p} f_Y(y \mid \hat{\boldsymbol{\theta}})dy = p \tag{2-1}$$

donde $F_Y(y)$ es la función de distribución acumulada de f_Y y $\hat{\boldsymbol{\theta}}$ es un estimador de $\boldsymbol{\theta}$ basado en una muestra aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$. La propuesta inicial para estimar ψ_p fue suponer un modelo normal para Y. No obstante es frecuente encontrar variables bioquímicas con una distribución no normal (Elveback et al., 1970; Gräsbeck, 2004), por lo cual, se ha recomendado el uso de transformaciones que permitan alcanzar la normalidad, ya sea con una sola transformación (e.g., con logaritmo) o con una transformación en dos etapas donde

se remueva primero el sesgo (e.g., con la función de Box & Cox (1964)) y luego cualquier curtosis remanente (e.g., con la función módulo (John & Draper, 1980)) (Boyd & Lacher, 1982; Linnet, 1987; Solberg, 1987, 2004).
La alternativa de buscar el ajuste a algún otro modelo probabilístico, tal como Gamma, Weibull, etc. y luego aplicar la ecuación (2-1) se ha explorado muy poco (para un ejemplo, ver Thompson & Theron (1990)). Esto puede ser debido precisamente al desconocimiento del modelo real de los datos. Recientemente, Trujillano et al. (2007) proponen usar la familia de distribuciones GS (Muiño et al., 2006) la cual está basada en una sóla ecuación diferencial con cinco parámetros:

$$\frac{\mathrm{d}F(y)}{\mathrm{d}y} = \alpha F(y)^g[1 - F(y)^k]^\gamma, \qquad F(x_0) = 0.5$$

donde $F(y)$ es la función de distribución acumulada, con x_0, α, g, k y γ como los parámetros del modelo. Esta familia de distribuciones puede representar una buena multitud de patrones, incluyendo las formas paramétricas comunes (normal, T-student, Ji-cuadrado, exponencial, gamma, etc.) y de aquí, es una buena alternativa para modelar datos observados en la práctica. Usando esta familia, se puede ajustar una distribución GS a datos unimodales sin hacer asunciones sobre la distribución real subyacente (Trujillano et al., 2007). Esta distribución aún no ha tenido una aplicación importante en el contexto de los IR's, posiblemente debido a que la implementación computacional no es fácil.
Otros métodos paramétricos para estimar cuantiles, que no requieren el uso directo de la ecuación (2-1), tampoco se han aplicado en el contexto de los IR's. Correa (2004) revisa algunos de ellos. Por ejemplo, se puede usar la distribución del estadístico de orden correspondiente al cuantil deseado y usar la misma para obtener su promedio e intervalo de confianza. No obstante, la construcción de tal distribución requiere conocer f_Y.

Métodos no paramétricos Los métodos no paramétricos no requieren conocer (o asumir) una forma para f_Y. La aproximación más sencilla para estimar ψ_p es usar los percentiles empíricos, los cuales se obtienen a partir de los estadísticos de orden. Suponga la muestra aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$; los estadísticos de orden estan definidos como $\{Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}\}$, donde $Y_{(j)}$, con $j = 1, \ldots, n$, indica el dato que esta en la posición j después de ordenar la muestra de forma ascendente. De acuerdo a Hyndman & Fan (1996), una fórmula general para estimar el percentil (o cuantil) ψ_p, a partir de los estadísticos de orden, es:

$$\hat{\psi}_p = Q_t(p) = (1 - \gamma)Y_{(j)} + \gamma Y_{(j+1)}, \qquad \text{donde } \frac{j - m}{n} \leq p < \frac{j - m + 1}{n} \tag{2-2}$$

para algún $m \in \Re$ y $0 \leq \gamma \leq 1$. El valor de γ es una función de $j = \lfloor pn + m \rfloor$ y de $g = pn + m - j$. Aquí, $\lfloor u \rfloor$ denota el entero más grande no mayor que u, y t indica el tipo de función cuantil a utilizar. Dependiendo del valor de m, Hyndman & Fan (1996) reunen

en la ecuación (2-2) hasta nueve métodos distintos para el cálculo de percentiles (es decir, $t = 1, 2, \ldots, 9$), esto en un intento por estandarizar la notación alrededor de los diferentes métodos existentes en los programas de estadística tales como SAS, SPSS, S-Plus, BMDP, GLIM y Minitab, y también, describir las propiedades de cada método. Por ejemplo, para los métodos 5, 6 y 8, el valor de γ esta dado por:

$$\gamma = \begin{cases} 1 & \text{si } pn + m < 1 \\ g & \text{si } 1 \leq pn + m \leq n \\ 0 & \text{si } pn + m > n \end{cases}$$

y el valor de m es

Tipo (t)	m
5	0.5
6	p
8	$\frac{1}{3}(p+1)$

La IFCC (Solberg, 1987, 2004) recomienda usar el método o tipo 6 para el cual, $m = p$; con esto, $j = \lfloor p(n+1) \rfloor$, $g = p(n+1) - j$. Este método corresponde al siguiente procedimiento de cálculo: (i) ordenar los datos en forma ascendente, (ii) Asignar las posiciones $j = 1, \ldots, n$ a cada dato en la muestra ordenada, (iii) Obtener la posición de los datos correspondientes al percentil deseado como $j = p(n+1)$, y (iv) si j es entero, calcular $\hat{\psi}_p = Y_{(j)}$, si j no es entero, obtener $\hat{\psi}_p$ por interpolación entre los valores $Y_{(\lfloor j \rfloor)}$ y $Y_{(\lfloor j \rfloor + 1)}$. Por otra parte, los métodos 5 y 8 son los mejores de acuerdo con Hyndman & Fan (1996). Linnet (2000) comparó los métodos 5 y 6, para estimar $\psi_{2.5}$ y $\psi_{97.5}$ en un experimento de simulación encontrando el método 5 con un mejor desempeño para tamaños muestrales de 40 y 120; en tamaños muestrales superiores mostraron resultados similares.

La aproximación basada en percentiles empíricos puede ser robusta en muestras grandes pero, en muestras pequeñas estos percentiles son ineficientes y de varianza grande (Wright & Royston, 1999). Estimaciones más precisas se pueden obtener calculando el promedio ponderado de los estadísticos de orden, donde aquellos más cercanos al percentil de interés tendrán las ponderaciones más altas. Tales ponderaciones pueden ser obtenidas por estimadores kernel, por ejemplo, el de Parzen (1979)

$$\hat{\psi}_p = \sum_{i=1}^{n} \left[\int_{(i-1)/n}^{i/n} K\left(\frac{t - \tilde{p}}{h} \right) \mathrm{d}t \right] Y_{(i)}$$

donde K es la función de densidad kernel, simétrica respecto al 0, $h \to 0$ cuando $n \to \infty$ y $\tilde{p} = p/100$; note que el tamaño de los pesos se controla a través del parámetro de suavizado,

h. Las ponderaciones también se pueden obtener mediante el estimador de Harrell & Davis (1982)

$$\hat{\psi}_p = \sum_{i=1}^{n} \left\{ \int_{(i-1)/n}^{i/n} \text{Beta}\,[\,(n+1)p, \quad (n+1)(1-p)\,]\,\mathrm{d}y \right\} Y_{(i)} \tag{2-3}$$

donde Beta[$\cdot$] es la función de densidad Beta(α, β); en este caso, se ponderan los estadísticos de orden usando dos funciones beta incompletas. Más recientemente Horn et al. (1998) proponen un método robusto basado en estimadores biponderados ("biweight") de localización y escala. El método parte de la ecuación para un intervalo de predicción para una observación futura Y_{n+1}

$$\bar{y} \pm t_{n-1}(1-\alpha/2)\sqrt{s^2 + s^2/n}$$

donde $t_{n-1}(1-\alpha/2)$ es el cuantil $1-\alpha/2$ de una variable aleatoria t con $n-1$ grados de libertad. El método consiste en reemplazar $\bar{y}$, s^2 y s^2/n por estimadores robustos de localización y escala. Específicamente, un intervalo de predicción biponderado de $(1-\alpha)100\,\%$ para poblaciones simétricas esta dado por (Horn et al., 1998):

$$T_{bi}(c_1) \pm t_{n-1}(1-\alpha/2)\sqrt{S_T^2(c_1) + s_{bi}^2(c_2)} \tag{2-4}$$

donde $T_{bi}(c_1)$ es un estimador biponderado de localización con constante de afinamiento c_1, $S_T^2(c_1)$ es un estimador biponderado de la variabilidad de $T_{bi}(c_1)$ y $s_{bi}^2(c_2)$ es un estimador biponderado de la variablidad de Y_{n+1}. Para detalles del cálculo de estas cantidades ver Horn (1990) y Horn et al. (1998). Esta aproximación ofrece límites de referencia más exactos que usando la transformación a la normalidad y que el estimador de Harrell & Davis (1982) para tamaños de muestra pequeños en poblaciones con sesgo positivo (Horn et al., 1998).
Finalmente se encuentran los intervalos obtenidos mediante bootstrap, método que de acuerdo a Solberg (2004), es quizá el mejor método disponible para ajustar intervalos de referencia.

2.2. Bootstrap

La técnica del Bootstrap se puede encontrar bien explicada en referencias clásicas sobre el tema como Efron (1979), Diciccio & Romano (1988), Efron & Tibshirani (1993), Manly (1997), Davison & Hinkley (1997), adicionalmente, una excelente revisión del tema de los intervalos de confianza vía bootstrap se puede ver en Carpenter & Bithell (2000). El algoritmo, adaptado a la estimación de un cuantil ψ_p, es como sigue:

1. Dada una muestra aleatoria $\mathbf{x} = (x_1, \ldots, x_n)$, calcule $\hat{\psi}_p$ (con e.g., $p = 0.025$ ó $p = 0.975$ y usando alguno de los tipos de percentil empírico dados por la ecuación (2-2)).

2. Tome una muestra aleatoria con reemplazo de la muestra original $\mathbf{x}$ para obtener $\mathbf{x}^{*(b)} = (x_1^{*(b)}, \ldots, x_n^{*(b)})$

3. Calcule el estadístico $\hat{\psi}_p$ usando $\mathbf{x}^{*(b)}$ para obtener una réplica bootstrap $\hat{\psi}_p^{*(b)}$.

4. Repita los pasos b y c, B veces (con B grande, $B \geq 500$ es razonable (Solberg, 2004)).

5. Suponga que la secuencia de estimadores bootstrap para el percentil ψ_p es $\hat{\psi}_p^{*(1)}, \hat{\psi}_p^{*(2)}, \ldots, \hat{\psi}_p^{*(B)}$. Esta secuencia se puede usar para estudiar la distribución del estadístico y hacer todo tipo de inferencias sobre ψ_p.

En particular, la media de las réplicas bootstrap, $\hat{\psi}_p^{*(1)}, \hat{\psi}_p^{*(2)}, \ldots, \hat{\psi}_p^{*(B)}$, produce una estimación puntual de ψ_p, es decir,

$$\hat{\psi}_p = \frac{1}{B} \sum_{b=1}^{B} \hat{\psi}_p^{*(b)} \tag{2-5}$$

La técnica bootstrap, si bien permite encontrar una estimación puntual, su mayor utilidad se centra en generar intervalos de confianza (Haukoos & Lewis, 2005) como medida de incertidumbre de la estimación realizada. En el contexto de los IR's se ha usado con ambos objetivos (Solberg, 2004). Existen varios métodos de intervalos de confianza bootstrap. Aquí se describen dos: el método percentil y el método BC_a. El primero se utiliza con mucha frecuencia y en particular es el empleado para determinar intervalos de confianza para los límites del IR cuando no hay otro método disponible. El segundo es bien recomendado por Efron & Tibshirani (1993), más no es frecuente en el contexto de los IR's.

Método percentil. El método percentil es simple; luego de tener las réplicas bootstrap, $\hat{\psi}_p^{*(1)}, \hat{\psi}_p^{*(2)}, \ldots, \hat{\psi}_p^{*(B)}$, un intervalo de confianza tipo percentil para ψ_p del $(1-\alpha)100\,\%$ está dado por los cuantiles $(q_t(p_1),\ q_t(p_2))$ de las B réplicas bootstrap, con

$$p_1 = \alpha/2 \ \text{ y } \ p_2 = 1 - \alpha/2 \tag{2-6}$$

Los cuantiles $q_t(p_1)$ y $q_t(p_2)$ se obtienen por alguno de los métodos $(t = 1, 2, \ldots, 9)$ descritos por Hyndman & Fan (1996) a través de la ecuación (2-2).

Método BC_a. El método BC_a corrige las probabilidades $\alpha/2$ y $1-\alpha/2$ del método percentil, teniendo en cuenta el sesgo y la variabilidad de las réplicas bootstrap con dos coeficientes llamados de "sesgo" (z_0) y de "acelaración" (a) respectivamente. Las nuevas probabilidades p_1 y p_2 para los límites del intervalo de confianza de $(1-\alpha)100\,\%$ para ψ_p de acuerdo al método BC_a son (Efron & Tibshirani, 1993):

$$p_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right); \qquad p_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right) \tag{2-7}$$

donde $\Phi(\cdot)$ denota la función de distribución normal estándar acumulada y z_τ un cuantil τ de la normal estándar (con $0 < \tau < 1$), es decir $\mathsf{P}(Z < z_\tau) = \Phi(z_\tau) = \tau$. La corrección por el sesgo la realiza el coeficiente z_0 el cual se estima como:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\psi}_p^{*(b)} < \hat{\psi}_p)}{B}\right) \tag{2-8}$$

donde $\Phi^{-1}(\cdot)$ es la inversa de la función de distribución acumulada normal estándar y $\#(\hat{\psi}_p^{*(b)} < \hat{\psi}_p)$ es el número de réplicas boostrap que son mayores que el estadístico original (o calculado de la muestra original). El coeficiente de aceleración (a), el cual da cuenta por la variabilidad en las réplicas bootstrap, se aproxima (Davison & Hinkley, 1997) usando el procedimiento Jacknife, como:

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\overline{\hat{\psi}^{(J)}} - \hat{\psi}^{(-i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{\hat{\psi}^{(J)}} - \hat{\psi}^{(-i)}\right)^2\right]^{3/2}} \tag{2-9}$$

donde $\hat{\psi}^{(-i)}$ es el valor del estadístico (es decir, del cuantil p, $\hat{\psi}_p$) calculado usando la muestra original con el i–ésimo dato removido (esto es, la i–ésima réplica jacknife del estadístico, con $i = 1, 2, \ldots, n$) y

$$\overline{\hat{\psi}^{(J)}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}^{(-i)}$$

es la media las n réplicas Jacknife. De acuerdo a Efron & Tibshirani (1993), el coeficiente de sesgo, $\hat{z}_0$ dado en la ecuación (2-8), es una medida de la diferencia entre la mediana de las réplicas bootstrap y el estadístico calculado de la muestra original ($\hat{\psi}_p$) en unidades normales. Si la mitad de las réplicas bootstrap son menores o iguales que $\hat{\psi}_p$ entonces no hay sesgo por la mediana y $\hat{z}_0 = 0$. Por otro lado, el coeficiente de aceleración, $\hat{a}$ dado en la ecuación (2-9), mide la tasa de aceleración del error estándar de $\hat{\psi}_p$ respecto al valor verdadero ψ_p sobre una escala normalizada. Para mayores detalles y justificación teórica de estas correcciones revisar Efron & Tibshirani (1993) o Davison & Hinkley (1997). Considere la situación particular en la cual $\#(\hat{\psi}_p^{*(b)} < \hat{\psi}_p)/B = 0.5$ con lo cual el coeficiente de sesgo, $\hat{z}_0 = \Phi^{-1}(0.5) = 0$, y además, que el coeficiente de aceleración sea $\hat{a} = 0$. En este caso las probabilidades p_1 y p_2 quedan como:

$$p_1 = \Phi\left(0 + \frac{0 + z_{\alpha/2}}{1 - 0(0 + z_{\alpha/2})}\right) = \alpha/2; \qquad p_2 = \Phi\left(0 + \frac{0 + z_{1-\alpha/2}}{1 - 0(0 + z_{1-\alpha/2})}\right) = 1 - \alpha/2$$

Es decir, cuando el sesgo y la aceleración son cero, el intervalo BC_a queda igual al intervalo percentil dado por (2-6).

2.3. Enfoque bayesiano

Supongamos que la información a priori sobre algún ψ_p este contenida en la densidad a priori $\zeta(\psi_p)$. Si $\mathbf{y} = (y_1, \ldots, y_n)$ es una muestra aleatoria de f_Y, entonces la densidad posteriori está dada por:

$$\begin{array}{rcl} \zeta(\psi_p \,|\, \mathbf{y}) & \propto & \zeta(\psi_p) \times f(\mathbf{y} \,|\, \psi_p) \\ \text{posteriori} & \propto & \text{a priori} \times \text{verosimilitud} \end{array} \tag{2-10}$$

Así, cualquier inferencia sobre ψ_p se puede obtener de $\zeta(\psi_p \,|\, \mathbf{y})$. Por ejemplo, la media a posteriori de $\zeta(\psi_p \,|\, \mathbf{y})$ será un estimador puntual bayesiano de ψ_p. Boos & Monahan (1986) proponen reemplazar (2-10) por

$$\hat{\zeta}(\psi_p \,|\, \mathbf{y}) \propto \zeta(\psi_p) \times \hat{f}_{NB}(\mathbf{y} \,|\, \psi_p) \tag{2-11}$$

donde $\hat{f}_{NB}(\mathbf{y} \,|\, \psi_p)$ es la estimación basada en bootstrap de la densidad de la cantidad pivotal $\hat{\psi}_p - \psi_p$ utilizando un estimador kernel. Correa (2004) da ejemplos utilizando esta aproximación para el percentil 85. El primer paso para aplicar este enfoque bayesiano es obtener la verosimilitud, $\hat{f}_{NB}(\mathbf{y} \,|\, \psi_p)$, de la ecuación (2-11). Para esto Boos & Monahan (1986) proponen lo siguiente:

1. Estimar la función de distribución poblacional mediante la función de distribución empírica, $\hat{F}_n$, desde los Y_i's.
2. Generar B muestas aleatorias bootstrap de tamaño n desde $\hat{F}_n$ y calcular $\hat{\psi}_p^{*(b)}$ desde la muestra b (con $b = 1, 2, \ldots, B$).
3. Desde las B estimaciones bootstrap $(\hat{\psi}_p^{*(1)}, \hat{\psi}_p^{*(2)}, \ldots, \hat{\psi}_p^{*(B)})$ calcular el estimador de densidad kernel
$$\hat{g}_{\mathrm{NB}}(u) = \frac{1}{Bh_B} \sum_{b=1}^{B} K\left[\frac{u - (\hat{\psi}_p^{*(b)} - \hat{\psi}_p)}{h_B}\right] \tag{2-12}$$

donde $K[\,\cdot\,]$ es la función kernel. Si se reemplaza $u = x - \psi_p$ en (2-12), $\hat{g}_{\mathrm{NB}}(x - \psi_p)$ es un estimado de la densidad de muestreo de $\hat{\psi}_p$ dado ψ_p, y evaluando esto en $x = \hat{\psi}_p$ produce una función de ψ_p para usarse como verosimilitud,

$$\hat{f}_{NB}(\mathbf{y} \,|\, \psi_p) = \frac{1}{Bh_B} \sum_{b=1}^{B} K\left[\frac{2\hat{\psi}_p - \psi_p - \hat{\psi}_p^{*(b)}}{h_B}\right] \tag{2-13}$$

La cantidad (2-13) es reemplazada en la ecuación general para la densidad posteriori (2-11) y la constante de normalización (para quitar la proporcionalidad) se halla por integración numérica. La revisión bibliográfica realizada para este estudio no encontró aplicación de esta aproximación bayesiana en el contexto de los IR's, de manera que el presente trabajo constituye un aporte al tema de la estimación de los IR's utilizando este método.

3. Objetivos

3.1. Objetivo general

Estudiar el uso del bootstrap en la estimación clásica frencuentista y bajo un enfoque bayesiano de los percentiles 2.5 y 97.5, los cuales tienen un uso directo en la determinación de intervalos de referencia para variables biológicas.

3.2. Objetivos específicos

- Implementar una estimación bayesiana de los percentiles 2.5 % y 97.5 % usando verosimilitudes estimadas con bootstrap.
- Evaluar el desempeño de la rutina bayesiana (sección 2.3) frente al bootstrap (sección 2.2) y frente a otros métodos disponibles para estimar cuantiles poblacionales, como son, el percentil empírico (ecuación 2-2), el estimador de Harrell & Davis (1982) (ecuación 2-3) y el estimador robusto de Horn et al. (1998) (ecuación 2-4).

4. Metodología

Para evaluar el bootstrap y la propuesta bayesiana en la estimación de percentiles extremos, se comparó su desempeño con el de otros métodos disponibles para el mismo propósito. En detalle, se evaluaron los siguientes métodos no paramétricos:

- Percentil empírico obtenido con la ecuación (2-2) y el tipo 5.
- Estimador de Harrell & Davis (1982) de la ecuación (2-3).
- Estimador robusto de Horn et al. (1998) para poblaciones sesgadas. Este consiste en aplicar la ecuación (2-4) sobre una pseudo-muestra simétrica con respecto a la mediana (ver Horn et al. (1998) para detalles).
- Bootstrap sobre el percentil empírico tipo 5 con la ecuación (2-5). Para los intervalos de confianza, el método percentil y el método BC_a.
- Enfoque Bayesiano usando una versosimilitud estimada por Bootstrap (Boos & Monahan, 1986).

La programación fue realizada por completo en el software `R` (R Development Core Team, 2006). El comando `quantile(...)` con la opción `type = 5` genera el percentil empírico (2-2) tipo 5. El estimador de Harrell & Davis (1982) fue obtenido de la librería `Hmisc` (Harrell, 2010) con el comando `hdquantile(...)`. La programación del estimador robusto de Horn et al. (1998) fue tomada del código para SAS proporcionado en el mismo trabajo. El código fue trasladado a `R`. Para el bootstrap se utilizó el comando `boot(...)` de la librería del mismo nombre (Canty & Ripley, 2010; Davison & Hinkley, 1997). En todos los casos se usó $B = 1000$ replicas bootstrap. Los métodos percentil y BC_a fueron aplicados con el comando `boot.ci(...)` de la misma librería. Un intervalo de confianza del 90 % (esto es, la probabilidad nominal) fue pedido en todos los casos.
La verosimilitud estimada desde las replicas bootstrap mediante una densidad kernel, para el enfoque bayesiano, fue obtenida usando un kernel $\mathsf{normal}(0, 1)$, y se utilizó un algoritmo de validación cruzada sesgado para buscar el ancho de banda (h_B en la ecuación 2-13). Esto se hizó con el comando `density(..., bw = "bcv", kernel = "gaussian")` disponible desde la librería `stats`. El método `bw = "bcv"` generó en todos los casos un densidad más suave que las otras opciones disponibles en `bw`. La constante de normalización fue obtenida con el

comando `integrate(...)`, aunque el resumen posteriori obtenido sin la constante y con la constante fue exactamente igual, de modo que el cálculo de la constante se obvio para las simulaciones que se describen en seguida. Para mayor simplicidad, se establecieron aprioris normales en todos los casos. El resumen posteriori desde la densidad posteriori estimada se obtuvo tomando una muestra de 10000 datos sobre la escala del percentil evaluado usando como probabilidades de muestreo la densidad posteriori estimada, esto se hizó con los comandos `sample(...)` para obtener intervalos de confianza[1] del 90 % y `weighted.mean(...)` para obtener la media posteriori. El código en `R` para realizar los cálculos de cada método descrito aquí se detalla en el apéndice de este documento.
Usando los comandos `rnorm`, `rt` y `rchisq` del `R` (librería `stats`), se simularon 1000 muestras de una población $\mathsf{normal}(0,1)$, de una t_5, de una χ^2_6 y de una χ^2_{12} para tamaños de muestra 30, 40, 80, 120 y 240. Sobre cada muestra se aplicaron los métodos descritos arriba. La precisión y exactitud de la estimación puntual fue comparada usando la raíz del error cuadrático medio, $\sqrt{\mathsf{ECM}}$, donde,

$$\mathsf{ECM}(\hat{\psi}_p) = \mathsf{E}[(\hat{\psi}_p - \psi_p)^2]$$

es el error cuadrático medio poblacional del estimador $\hat{\psi}_p$. Esta cantidad se estima desde las replicas simuladas a partir de

$$\widehat{\mathsf{ECM}}(\hat{\psi}_p) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\psi}_p^{(r)} - \psi_p)^2$$

donde $r = 1, 2, \ldots, R$ son las muestras simuladas. En este caso, $R = 1000$. Por otra parte, se comparó la cobertura de los intervalos de confianza percentil, BC_a y de aquellos producidos por el enfoque bayesiano mediante la proporción de intervalos que incluyen el verdadero valor del percentil evaluado.

[1] Bajo un marco conceptual bayesiano estricto, los intervalos obtenidos a partir de la densidad posteriori son conocidos como *intervalos de credibilidad*; sin embargo, al no estar aquí bajo un tema netamente bayesiano, seguiremos usando el término *intervalos de confianza*

5. Resultados

5.1. Desempeño de la estimación puntual

La Figura **5-1** compara el resultado de los métodos evaluados para las muestras provenientes de las dos poblaciones simétricas: normal y T-student, y esto sólo para el percentil 2.5. No se muestran los resultados del percentil 97.5 debido a que los patrones de cambio fueron exactamente iguales. Para el enfoque bayesiano se utilizaron dos modelos apriori normales. Ambos centrados en el verdadero valor del percentil pero con diferente desviación estándar. Una apriori "poco informativa" se estableció con un coeficiente de variación del 38 % y otra apriori "más informativa" se fijó con un coeficiente de variación del 28 %. Para muestras de tamaño 30, 40, en el caso normal (Figura **5-1**A), el método bayesiano presentó el mejor desempeño seguido por el bootstrap y luego por los percentiles empírico y de Harrell & Davis. Estos dos últimos se comportaron muy similar. Para tamaños de muestra mayores, el bootstrap se comparta mejor que los otros métodos y en general, en tamaños de muestra grandes (240) parecen comportarse todos muy similar. Los resultados desde el modelo t_5 (Figura **5-1**B) muestran mayor diferencia entre los métodos con un posicionamiento más o menos similar que el observado para el caso normal. El enfoque bayesiano tuvo un mejor desempeño (esta vez, durante todo el espectro de tamaños muestrales), seguido por el bootstrap, luego el percentil empírico y finalmente el estimador de Harrell & Davis.

Los resultados para los dos modelos con sesgo positivo, χ^2_6 y χ^2_{12}, se muestran en la Figura **5-2**. En este caso, de nuevo, se usaron aprioris normales centradas en el valor verdadero del percentil respectivo, pero, dado la asimetría en las magnitudes de los dos cuantiles del mismo modelo, se usaron aprioris con diferente coeficiente de variación para las colas izquierda y derecha, intentando en ambos casos, representar modelos "poco informativos" y "más informativos". Entonces, para la cola izquierda en ambos modelos (χ^2_6 y χ^2_{12}), se usó una apriori normal "poco informativa" con un coeficiente de variación del 38 % y "más informativa" con un coeficiente de variación del 28 %, y para la cola derecha, en ambos modelos, un apriori "poco informativa" con un coeficiente de variación del 20 % y "más informativa" con un coeficiente de variación del 10 %.

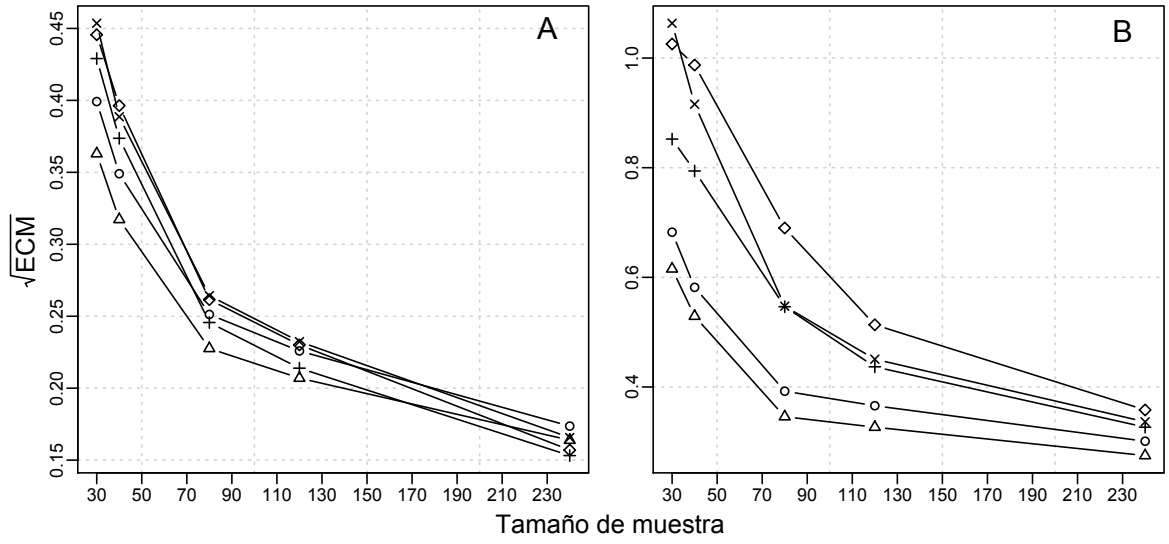

Figura 5-1.: Desempeño (medido con la raíz del error cuadrático medio, $\sqrt{\text{ECM}}$) de cuatro métodos para la estimación puntual del percentil 2.5 % ($\psi_{0.025}$) desde el muestreo de (A) una población $\text{normal}(0, 1)$ y (B) una población t_5. Los métodos son: (∘) Bayesiano con una apriori $\text{normal}(\psi_{0.025}, 0.38\psi_{0.025})$; (△) Bayesiano con una apriori $\text{normal}(\psi_{0.025}, 0.28\psi_{0.025})$; (+) Bootstrap del percentil empírico tipo 5 (media de las replicas bootstrap); (×) percentil empírico tipo 5; (◇) Harrell & Davis.

Otra vez, el enfoque bayesiano con ambos tipos de apriori sobresalen con un error cuadrático medio menor en tamaños de muestra pequeños. Los otros métodos tienen un comportamiento variable en su desempeño dependiendo de la cola del modelo y del modelo mismo. Por ejemplo el estimador de Harrell & Davis tuvo un mejor desempeño en ambas colas de la χ^2_6. El bootstrap por su parte, se comporta mejor en la χ^2_{12}. Finalmente, el estimador robusto de Horn sobresale en la estimación del percentil en la cola derecha (Figura **5-2**B y **5-2**D) con un desempeño igual o mejor que el enfoque bayesiano con la apriori "poco informativa".

5.2. Cobertura de los intervalos de confianza

La Figura **5-3** y la Tabla **5-1** muestran la cobertura real de los intervalos de confianza obtenida bajo los métodos bootstrap percentil, bootstrap BC_a y el enfoque bayesiano. En este último caso, tales intervalos se obtuvieron desde la densidad posteriori estimada usando las aprioris "poco informativas" y "más informativas" descritas en la sección anterior. La cobertura buscada o nominal fue 0.9 (ó 90 %).

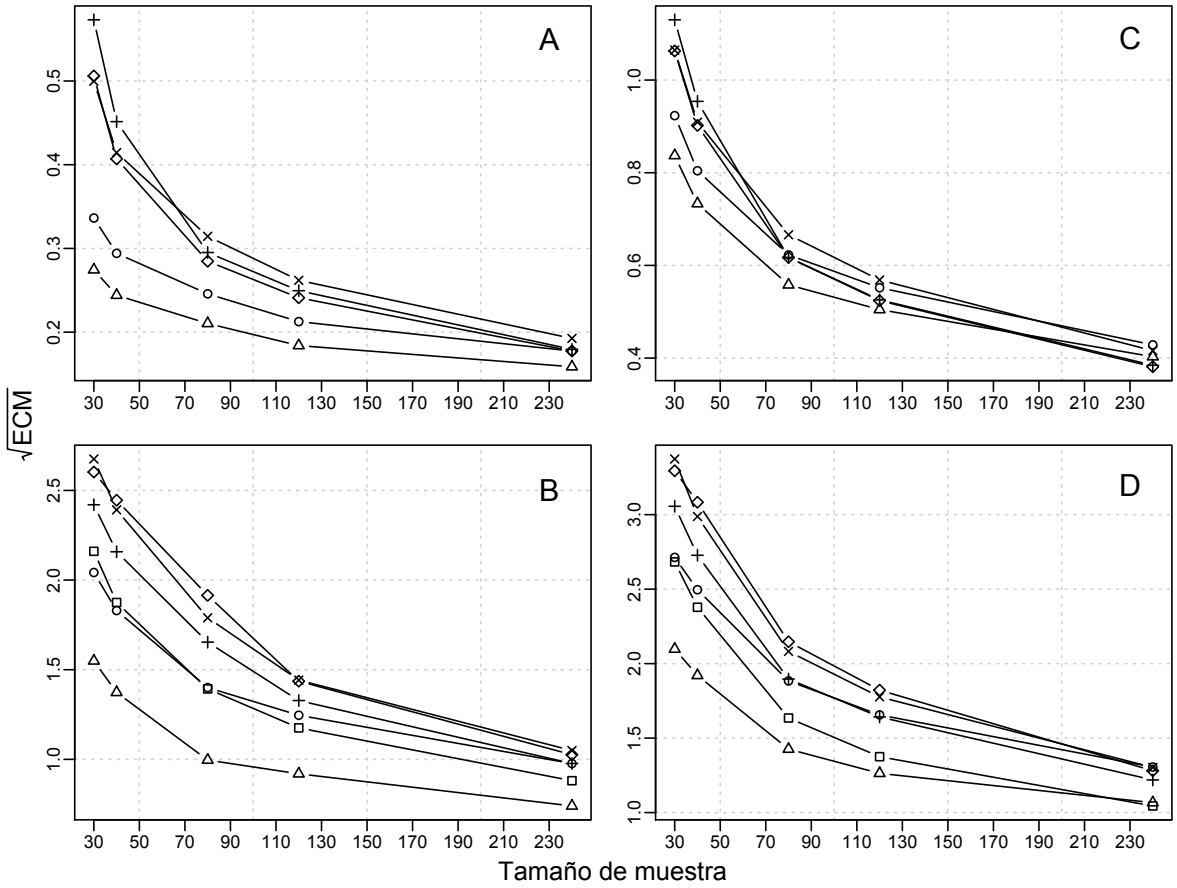

Figura 5-2.: Desempeño (medido con la raíz del error cuadrático medio, $\sqrt{\mathsf{ECM}}$) de cinco métodos para la estimación puntual del percentil 2.5 % ($\psi_{0.025}$) (arriba: A y C) y 97.5 % ($\psi_{0.975}$) (abajo: B y D) desde el muestreo de (A y B) una población χ^2_6 y (C y D) una población χ^2_{12}. Los métodos son: (∘) Bayesiano con una apriori $\mathsf{normal}(\psi_{2.5}, 0.38\psi_{0.025})$ (A y C) y con una apriori $\mathsf{normal}(\psi_{0.975}, 0.20\psi_{0.975})$ (B y D); (△) Bayesiano con una apriori $\mathsf{normal}(\psi_{0.025}, 0.28\psi_{0.025})$ (A y C) y con una apriori $\mathsf{normal}(\psi_{0.975}, 0.10\psi_{0.975})$ (B y D) ; (+) Bootstrap del percentil empírico tipo 5 (media de las replicas bootstrap); (×) percentil empírico tipo 5; (◇) Harrell & Davis; (□) Robusto de Horn (sólo para el percentil 97.5 %, B y D)

En general, las coberturas en tamaños de muestra pequeños resultaron bajas para todos los métodos. Los intervalos generados por el enfoque bayesiano fueron los únicos con coberturas que tan solo alcanzaron el 80 % para el tamaño de muestra más grande en todos los modelos. El intervalo BC_a produjo una cobertura muy baja para tamaños de muestra pequeños en las

colas derechas, mientras que el intervalo percentil fue el que más rápido alcanzó la cobertura nominal del 90 % en las colas izquierdas. Ambos, el percentil y el BC_a fueron similares en las colas derechas.

Tabla 5-1.: Cobertura estimada de los intervalos de confianza obtenidos bajo el bootstrap y el enfoque bayesiano para cada límite de un intervalo de referencia del 95 %. La cobertura nominal fue 0.900 (ver Figura **5-3**).

		Para $\psi_{0.025}$				Para $\psi_{0.975}$			
		Bootstrap		Bayesiano		Bootstrap		Bayesiano	
Población	n	Percentil	BC_a	Bayes1†	Bayes2‡	Percentil	BC_a	Bayes1†	Bayes2‡
normal$(0,1)$	30	0.532	0.176	0.597	0.620	0.519	0.509	0.631	0.644
normal$(0,1)$	40	0.612	0.304	0.664	0.673	0.618	0.607	0.684	0.692
normal$(0,1)$	80	0.854	0.746	0.798	0.817	0.847	0.837	0.791	0.800
normal$(0,1)$	120	0.900	0.855	0.824	0.828	0.876	0.870	0.807	0.816
normal$(0,1)$	240	0.882	0.895	0.807	0.811	0.896	0.895	0.826	0.830
T_5	30	0.529	0.140	0.598	0.618	0.521	0.512	0.610	0.617
T_5	40	0.629	0.281	0.668	0.680	0.600	0.580	0.646	0.656
T_5	80	0.856	0.746	0.824	0.831	0.841	0.830	0.817	0.824
T_5	120	0.900	0.852	0.838	0.836	0.903	0.895	0.807	0.820
T_5	240	0.880	0.880	0.839	0.840	0.891	0.889	0.818	0.824
χ^2_6	30	0.543	0.231	0.726	0.757	0.523	0.507	0.625	0.673
χ^2_6	40	0.648	0.385	0.766	0.783	0.643	0.627	0.675	0.698
χ^2_6	80	0.864	0.751	0.788	0.805	0.833	0.820	0.791	0.828
χ^2_6	120	0.899	0.860	0.837	0.848	0.879	0.875	0.799	0.826
χ^2_6	240	0.891	0.877	0.811	0.822	0.897	0.900	0.819	0.857
χ^2_{12}	30	0.529	0.208	0.657	0.678	0.500	0.489	0.645	0.685
χ^2_{12}	40	0.603	0.344	0.693	0.708	0.657	0.639	0.681	0.698
χ^2_{12}	80	0.862	0.751	0.794	0.807	0.842	0.831	0.772	0.795
χ^2_{12}	120	0.897	0.851	0.787	0.797	0.883	0.880	0.796	0.823
χ^2_{12}	240	0.883	0.880	0.812	0.817	0.873	0.879	0.799	0.823

† Con apriori "poco" informativa.

‡ Con apriori "más" informativa.

5.3. Ejemplo de aplicación

Para mostrar la aplicación de las técnicas evaluados usaremos una muestra de datos de glucosa de 46 hombres (con edades mayores a 80 años) reportados por Horn et al. (1998). Los datos, en milimoles por litro (mmol/L) se han graficado en la Figura **5-4**.

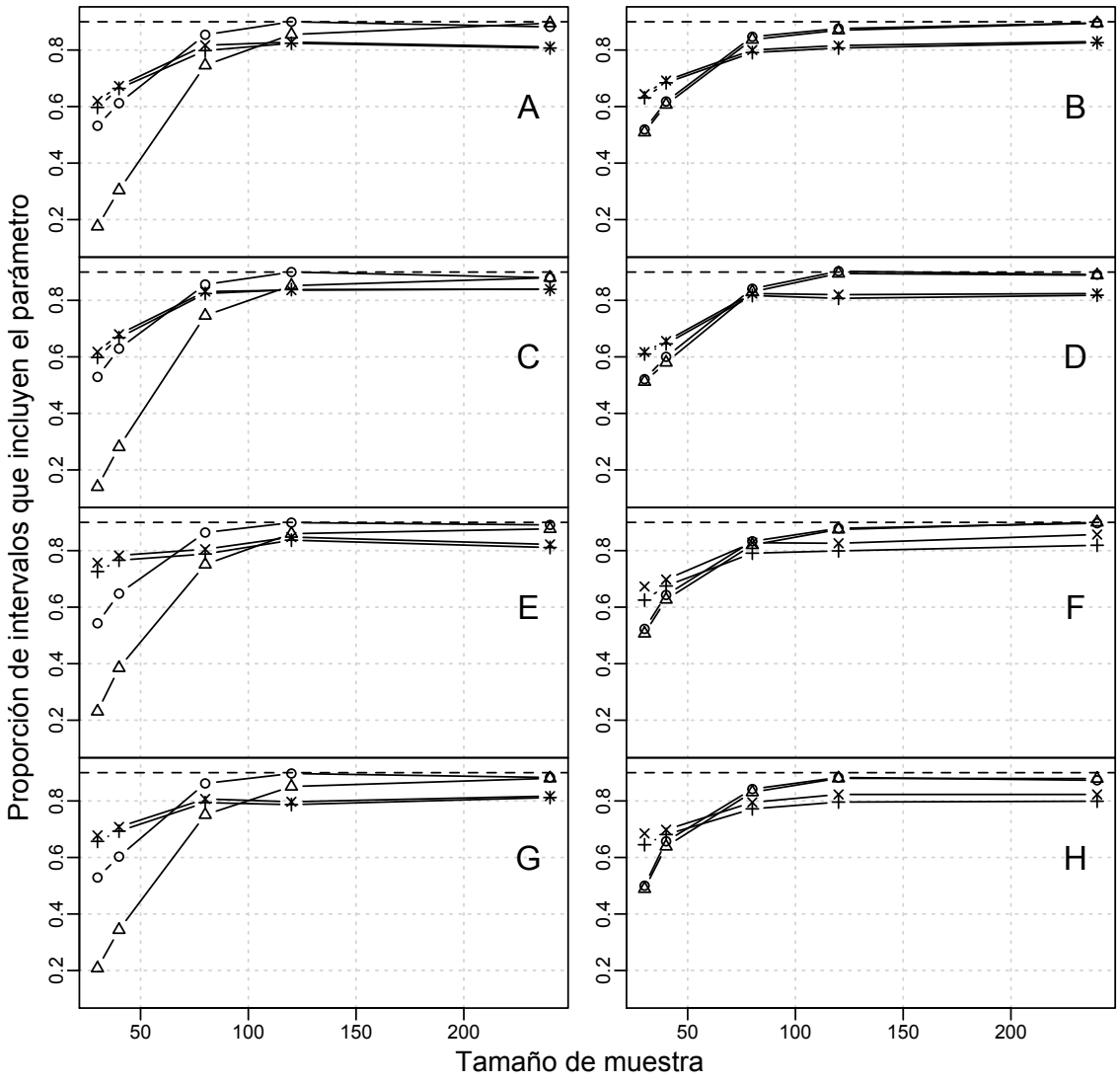

Figura 5-3.: Cobertura (medida como la proporción de intervalos que incluyen el verdadero valor del percentil en las 1000 simulaciones Monte Carlo) para intervalos tipo bootstrap y por el método bayesiano, del percentil 2.5 % ($\psi_{0.025}$) (izquierda: A, C, E, G) y 97.5 % ($\psi_{0.975}$) (derecha: B, D, F, H) desde el muestreo de (A y B) una población $\mathsf{normal}(0, 1)$, (C y D) t_5, (E y F) χ^2_6 y (G y H) una población χ^2_{12}. Los métodos son: (+) Bayesiano con una apriori $\mathsf{normal}(\psi_p,\, 0.38\psi_p)$ para las poblaciones simétricas en ambos percentiles y para el $\psi_{0.025}$ en las poblaciones χ^2s. Para el $\psi_{0.975}$ en las χ^2s se usó la apriori $\mathsf{normal}(\psi_{0.975},\, 0.20\psi_{0.975})$; (×) Bayesiano con una apriori más informativa, $\mathsf{normal}(\psi_p,\, 0.28\psi_p)$ en el caso de las poblaciones simétricas y el $\psi_{0.025}$ de las χ^2s y con una apriori $\mathsf{normal}(\psi_{0.975},\, 0.10\psi_{0.975})$ para el $\psi_{0.975}$ de las χ^2s; (∘) Bootstrap del percentil empírico tipo 5 con el método del percentil; (△) Bootstrap del percentil empírico tipo 5 con el método BC_a. La línea discontinua (- - -) se ubica en 0.9, el nivel de confianza nominal con el cual se pidieron los intervalos en cada simulación.

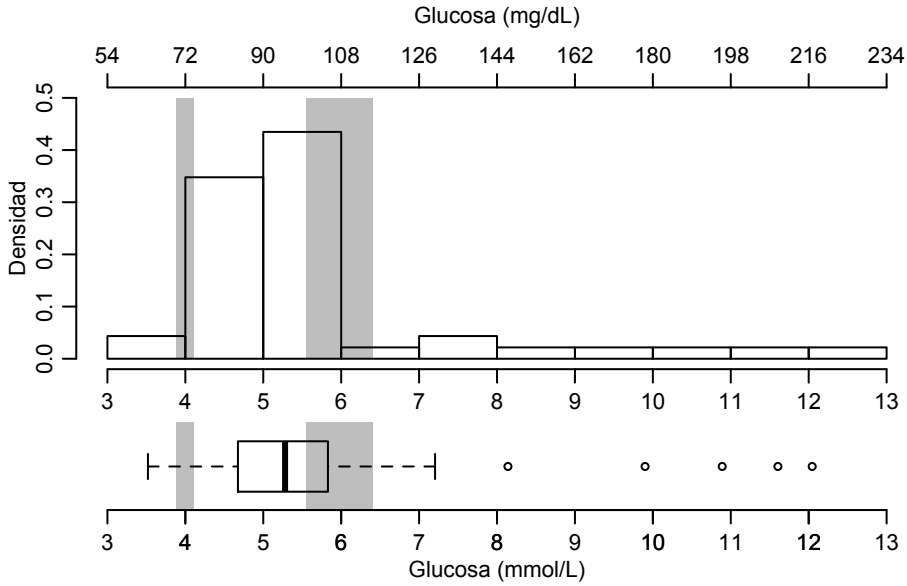

Figura 5-4.: Distribución, con histograma (arriba) y boxplot (abajo), de valores de glucosa en plasma de una muestra de 46 hombres (edad > 80 años), reportada por Horn et al. (1998). Los rectángulos sombreados ubicados en cada cola de la distribución señalan, en cada caso, un posible rango de valores para los límites inferior (3.890 a 4.110 mmol/L) y superior (5.551 a 6.400 mmol/L) del intervalo de referencia para la glucosa en ayunas (ver texto).

En esta muestra, la glucosa evidencia una distribución con una fuerte asimetría positiva (coeficiente de asimétrica, $g_1 = 2.18$). Sus valores se encuentran en un rango de 3.520 a 12.040 mmol/L con un promedio de 5.738 y una mediana de 5.280 mmol/L. El rango intercuartílico va de 4.716 a 5.816 mmol/L. La distribución y rango de esta muestra concuerda con lo reportado en internet y diferentes trabajos (e.g., Harwood & Cole, 1978; Tirosh et al., 2005). Para recopilar información sobre cada límite de referencia, se realizó una revisión en internet sobre los valores de referencia adoptados para la glucosa en ayunas. Esta revisión muestra que el límite inferior puede estar alrededor de 3.890 y 4.110 mmol/L, mientras que el límite superior puede moverse entre 5.551 y 6.400 mmol/L. La Figura **5-4** señala estos límites con rectángulos sombreados. Teniendo en cuenta esta información, se usó una apriori $\mathsf{normal}(\mu = 3.9725, \sigma = 0.2229)$ para la estimación bajo el enfoque bayesiano del límite inferior y una apriori $\mathsf{normal}(\mu = 5.9542, \sigma = 0.3342)$ en la estimación del límite superior bajo el mismo método. La desviación estándar escogida fue seleccionada desde la variabilidad de los valores para los límites hallados en la revisión.

Los datos fueron leídos en el `R` y, mediante las funciones descritas en el Apéndice, se aplicaron cada uno de los métodos evaluados en este estudio. La Tabla **5-2** muestra los resultados

Tabla 5-2.: Estimación de un intervalo de referencia del 95 % para la glucosa basado en los diferentes métodos evaluados y usando la muestra de 46 hombres (edad > 80 años) reportada por Horn et al. (1998) (ver Figura **5-4**).

	Límites de referencia (mmol/L)	
Método	$\hat{\psi}_{0.025}$	$\hat{\psi}_{0.975}$
Percentil empírico tipo 5	3.770	11.759
Harrell & Davis	3.772	11.622
Robusto de Horn	3.772	10.596
Bootstrap con IC tipo Percentil †	3.818 (3.520 – 4.213)	11.261 (8.909 – 12.045)
Bootstrap con IC tipo BC_a †	3.818 (3.520 – 3.905)	11.261 (9.900 – 12.045)
Bayesiano †	3.881 (3.583 – 4.123)*	10.917 (10.891 – 10.982)**

† Entre paréntesis se indica un intervalo de confianza del 90 %

$*$ Se usó una apriori $\mathsf{normal}(\mu = 3.9725, \sigma = 0.2229)$

$**$ Se usó una apriori $\mathsf{normal}(\mu = 5.9542, \sigma = 0.3342)$

obtenidos. Note como el método robusto de Horn et al. (1998) presenta el menor valor para el límite superior, seguido por la estimación obtenida desde el enfoque bayesiano. Esto es debido a que los otros métodos tienen una influencia total de los datos originales, mientras la estimación robusta de Horn et al. (1998) y el enfoque bayesiano no.

6. Discusión

Con frecuencia existe información previa sobre la distribución de variables bioquímicas susceptibles de incluirse en estudios sobre determinación de IR's y también existe la necesidad de realizar este tipo de estudios con bajos tamaños de muestra (20 a 60) (Horn et al., 1998). El enfoque bayesiano evaluado en este trabajo presentó para los cuatro modelos poblacionales evaluados, y con bajos tamaños muestrales, un desempeño igual o superior a otros métodos no paramétricos disponibles para estimar los límites del IR, esto, si se cuenta con una información apriori valiosa en cuanto a posición del límite y grado de dispersión alrededor del mismo. Los resultados de la simulación mostraron que el desempeño mejora si el grado de dispersión se hace más pequeño, pero con frecuencia, la calidad de la información encontrada en la literatura al respecto es mucho mejor (menor dispersión) que la supuesta para las dos apriori trabajadas en este estudio, de tal manera que el enfoque bayesiano se presenta aquí como una alternativa adicional para tener en cuenta a la hora de analizar los datos sobre IR's.

La cobertura del intervalo de confianza obtenido con el enfoque bayesiano resultó entre 22 y 33 % más baja que la cobertura nominal dependiendo del tamaño de la muestra. Si bien se observa que la cobertura estimada mejora con un incremento en n, no se alcanza por completo dicho valor nominal. Pero, en general la cobertura obtenida por ambos métodos, el bayesiano y el bootstrap, fue baja y no se alcanza de forma contundente el valor nominal. Esto puede ser debido a que, si bien los cuantiles pueden comportarse asintóticamente normales, cuantiles ubicados en las colas de la distribución (comparados, por ejemplo, con la mediana) requieren de tamaños de muestra mucho más grandes para lograrlo[1]. Las bajas coberturas obtenidas en tamaños de muestra pequeños están de acuerdo con los resultados de Linnet (2000) quien evaluó el bootstrap usando los cuantiles tipo 5 y 6 con IC tipo percentil. Simulaciones adicionales con $n = 600$ y $n = 1600$ (no descritas en la sección anterior) muestran que, en el caso bayesiano, las coberturas alcanzan valores de 0.890 para $n = 1600$, mientras que para los intervalos bootstrap tipo percentil y BC_a, las coberturas alcanzan valores de 0.900 para $n = 600$. Tentativamente, para mejorar la cobertura del intervalo de confianza generado por el enfoque bayesiano se puede pedir una probabilidad (nominal) del 95 %. Estudios posteriores pueden evaluar el desempeño de aprioris informativas sesgadas

[1] Comunicación personal con Juan Carlos Correa, PhD. Profesor Asociado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.

dado que los cuantiles en las colas parecen seguir una distribución con algún sesgo (positivo o negativo, según la cola de la distribución), y entonces es factible pensar en este tipo de patrones a priori para los límites del IR.

Por otra parte, el desempeño de la estimación puntual del bootstrap tuvo un mejor desempeño, comparado a los otros métodos (con excepción del enfoque bayesiano) en poblaciones simétricas (normal y T). Sin embargo, si el tamaño de muestra es pequeño (e.g., 30 o 40) y se tienen indicios de que la población tiene un sesgo positivo (como las χ^2's con bajos grados de libertad), no es recomendable usar el bootstrap (ver Figura **5-2**). Sumado a esto, la cobertura observada de los intervalos de confianza bootstrap tipo percentil y tipo BC_a es muy baja en tamaños de muestra pequeños, y es más baja aún para el límite inferior del IR cuando se utiliza el método BC_a. Por ejemplo, note como para $n = 30$, la cobertura real del IC tipo BC_a fue de 0.231 para $\psi_{0.025}$ en la población χ^2_6 y de 0.507 para $\psi_{0.975}$ en la misma población (ver Tabla **5-1**). Esta asimetría en cobertura de los IC's tipo BC_a entre los dos límites del intervalo de referencia tiene su origen principalmente en el coeficiente de sesgo (2-8) (más que el coeficiente de aceleración),

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\psi}_p^{*(b)} < \hat{\psi}_p)}{B}\right)$$

y en la distribución de las réplicas bootstrap de cada cuantil. Para explicar esto, en la Figura **6-1** se ha graficado la distribución de las $B = 1000$ réplicas bootstrap de cada cuantil obtenidas a partir de muestras de una $\mathsf{normal}(0, 1)$ desde una de las 1000 simulaciones realizadas. La Figura **6-2** gráfica lo mismo para la χ^2_6. Note que la distribución de las réplicas bootstrap del cuantil izquierdo, $\hat{\psi}^{*(b)}_{0.025}$, presenta asimetría positiva mientras que la distribución de las réplicas bootstrap del cuantil derecho, $\hat{\psi}^{*(b)}_{0.025}$, tiene asimetría negativa. Tales distribuciones también resultan asimétricas respecto al valor del estadístico calculado de la muestra original, $\hat{\psi}_p$. Como consecuencia de esto, en la ecuación (2-8), la mayoría de las réplicas bootstrap resultan con valores mayores que $\hat{\psi}_{0.025}$, y por tanto, la proporción $\#(\hat{\psi}^{*(b)}_{0.025} < \hat{\psi}_{0.025})/B$ tiende a ser muy baja. Por otro lado, en la cola izquierda ocurre lo contrario, la mayoría de réplicas bootstrap tienen valores menores que $\hat{\psi}_{0.975}$, y por ende, la proporción $\#(\hat{\psi}^{*(b)}_{0.975} < \hat{\psi}_{0.975})/B$ tiende a ser más alta que su análoga de la cola derecha.

Se recomienda entonces, usar el intervalo tipo percentil, el cual, si bien tiene coberturas bajas en tamaños de muestra pequeño, son mejores que las mostradas por el intervalo BC_a. Finalmente, el bootstrap ha sido bien recomendado en estudios de IR's (Solberg, 2004), no obstante, los resultados de este estudio advierten que para muestras pequeñas y poblaciones muy sesgadas no es recomendable. En estos casos, es mejor usar el método de Harrell & Davis en combinación con el estimador robusto de Horn.

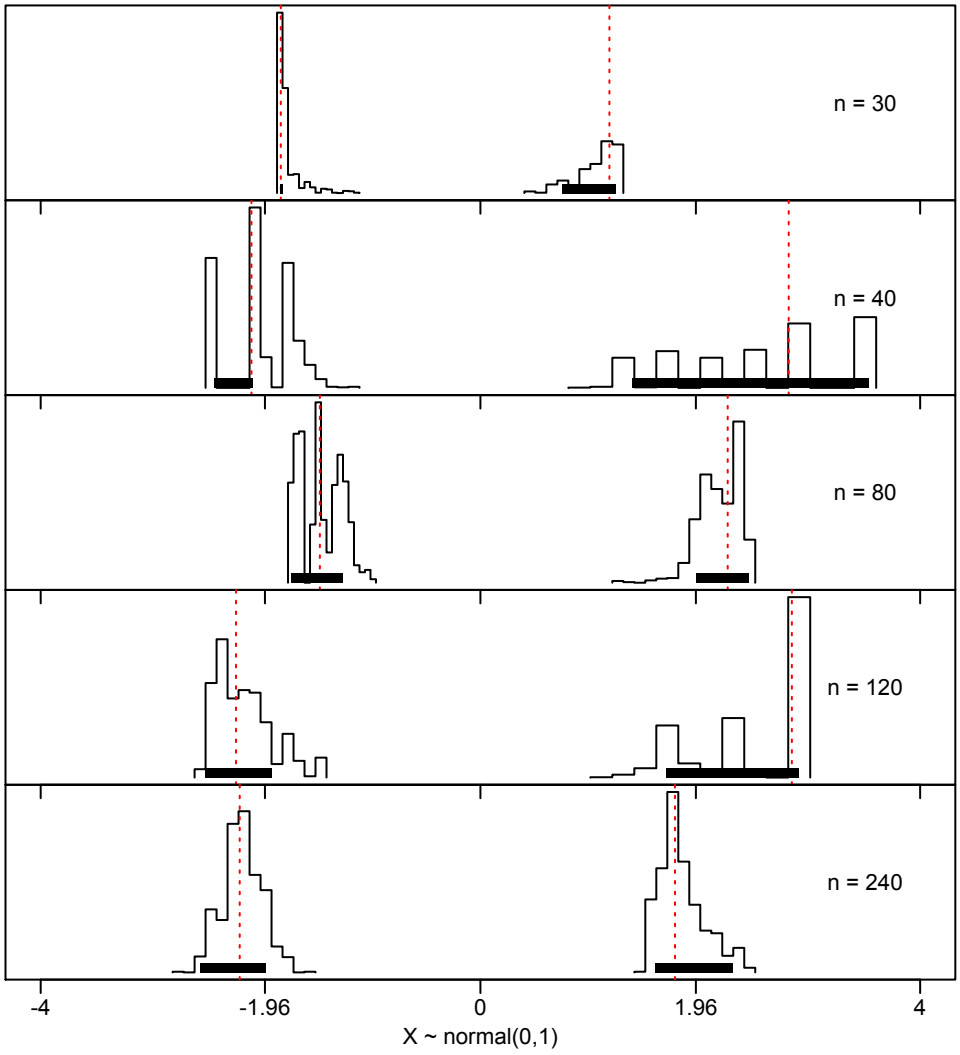

Figura 6-1.: Distribución (mediante histogramas) de las réplicas bootstrap, $\hat{\psi}_p^{*(b)}$, (con $B = 1000$ réplicas) para muestras de tamaño n extraídas de una población $\mathsf{normal}(0, 1)$. La línea vertical discontinua identifica en cada caso el valor del estadístico obtenido con la muestra original. La banda negra horizontal en cada cola, indica el intervalo de confianza tipo BC_a obtenido.

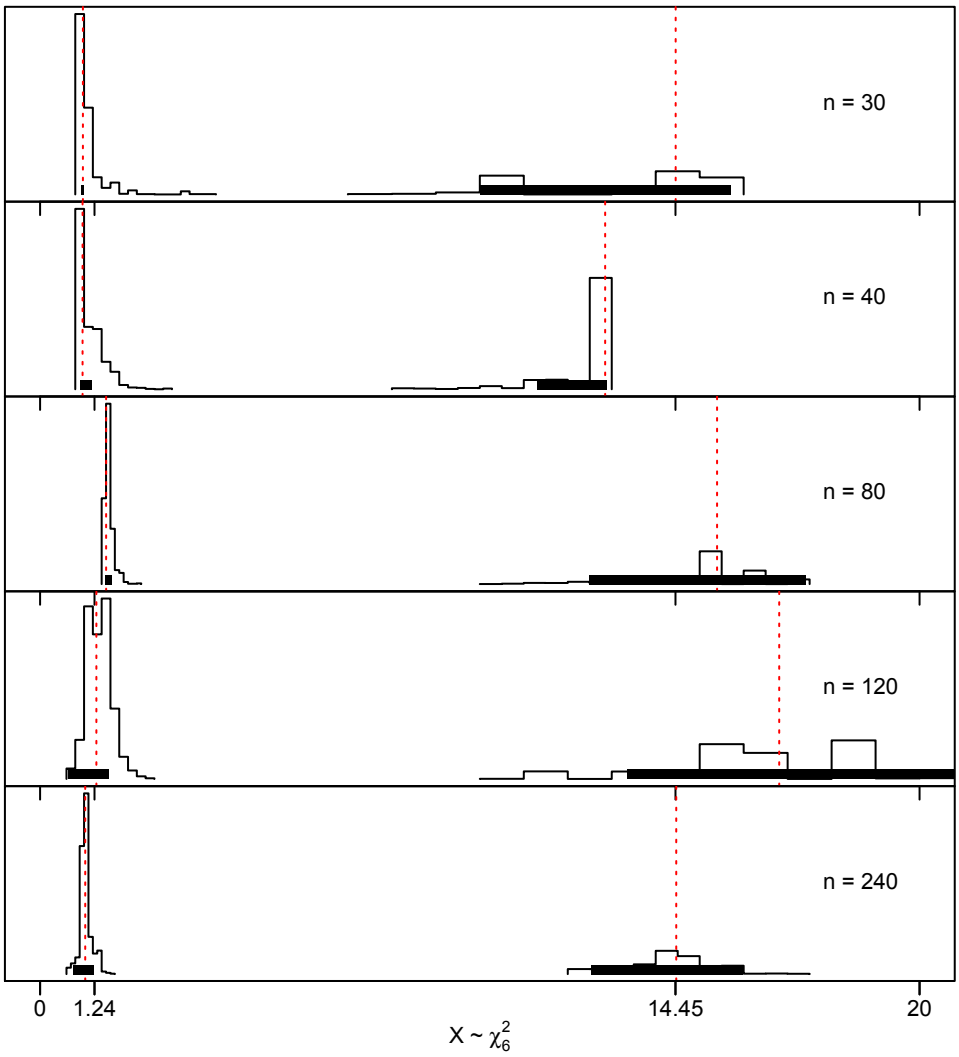

Figura 6-2.: Distribución (mediante histogramas) de las réplicas bootstrap, $\hat{\psi}_p^{*(b)}$, (con $B = 1000$ réplicas) para muestras de tamaño n extraídas de una población χ^2_6. Los cuantiles 0.025 y 0.975 verdaderos son 1.24 y 14.45 respectivamente. La línea vertical discontinua identifica en cada caso el valor del estadístico obtenido con la muestra original. La banda negra horizontal en cada cola, indica el intervalo de confianza tipo BC_a obtenido.

7. Conclusiones

- Si se cuenta con información apriori valiosa y bajos tamaños muestrales (30 ó 40), el enfoque bayesiano puede ser útil, e incluso mejor que métodos alternativos, para estimar los límites inferior y superior de un intervalo de referencia. Sin embargo, la cobertura real con la cual queda el intervalo de confianza para cada límite puede quedar hasta 33 % más baja que la nominal. Esto podría mejorarse si se incrementa la cobertura nominal a un 95 %.
- El bootstrap presenta un buen desempeño en poblaciones con poco sesgo aún con bajos tamaños de muestra. En poblaciones asimétricas y con bajos tamaños muestrales otros métodos presentan un mejor desempeño como el estimador de Horn et al. (1998) en combinación con el estimador de Harrell & Davis (1982).

A. Anexo: Programas computacionales para los métodos evaluados

Aquí se describe la parte computacional de los cinco métodos evaluados en este estudio. Todo el código presentado es para el software `R` (R Development Core Team, 2006). Además se utilizaron las siguientes librerías: `stats` (base), `boot` (base), `Hmisc` (no base) y `hdrcde` (no base), donde las "no base" no hacen parte de la distribución base del `R` y por ende deben descargarse de forma adicional.

Percentil empírico obtenido con la ecuación (2-2) y el método $t = 5$. El comando `quantile` con la opción `type = 5` fue utilizada para calcular el cuantil empírico (2-2) bajo el método o tipo $t = 5$ (Hyndman & Fan, 1996). No fue requerido código adicional para aplicar esta función. Por ejemplo, si `x` tiene el vector con la muestra $x_1, x_2, \ldots, x_n$, de acuerdo a la notación de la ecuación (2-2), para pedir $Q_5(0.025)$ y $Q_5(0.975)$ se utiliza el siguiente código:

```
quantile(x, probs = c(0.025, 0.975), type = 5 )
```

Estimador de Harrell & Davis (1982) de la ecuación (2-3) El estimador de Harrell & Davis (1982) fue obtenido de la librería `Hmisc` (Harrell, 2010) con el comando `hdquantile`. No fue requerido código adicional para aplicar esta función. Por ejemplo, si `x` tiene el vector con la muestra $x_1, x_2, \ldots, x_n$, las estimaciones de los cuantiles 0.025 y 0.975 de acuerdo a este estimador se obtienen con el siguiente código:

```
library(Hmisc)
hdquantile(x, probs = c(0.025, 0.975) )
```

Estimador Robusto de Horn et al. (1998) Este método se basa en la ecuación de un intervalo de predicción para una observación futura Y_{n+1}

$$\bar{y} \pm t_{n-1}(1 - \alpha/2)\sqrt{s^2 + s^2/n}$$

donde $t_{n-1}(1 - \alpha/2)$ es el cuantil $1 - \alpha/2$ de una variable aleatoria t con $n - 1$ grados de libertad. El método consiste en reemplazar $\bar{y}$, s^2 y s^2/n por estimadores robustos de

localización y escala. Específicamente, un intervalo de predicción biponderado de $(1-\alpha)100\,\%$ para poblaciones simétricas esta dado por:

$$T_{bi}(c_1) \pm t_{n-1}(1-\alpha/2)\sqrt{S_T^2(c_1) + s_{bi}^2(c_2)} \qquad \text{(A-1)}$$

donde $T_{bi}(c_1)$ es un estimador biponderado de localización con constante de afinamiento c_1, $S_T^2(c_1)$ es un estimador biponderado de la variabilidad de $T_{bi}(c_1)$ y $s_{bi}^2(c_2)$ es un estimador biponderado de la variablidad de Y_{n+1}. El cálculo de estas cantidades se realiza de acuerdo a un proceso iterativo descrito en detalle por Horn (1988, 1990) y Horn et al. (1998). El programa para `SAS` dado en el apéndice de Horn et al. (1998) ha sido traducido al lenguaje `R` para lo cual se construyó la función `biw.pred.int` que recibe un vector de datos y la probabilidad de cobertura deseada para el IR, y con esto, entrega los dos límites para el IR dado por la ecuación (A-1). Esta función para `R` se muestra en seguida:

```
biw.pred.int <- function(dat, conf = 0.9){
# Se genera un intervalo de predicción biponderado para una observación futura y[n+1]
# de acuerdo a Horn(1988) y Horn et al. (1998). El código es traducido del programa en
# SAS dado por Horn et al. (1998).
#
# dat = vector con los datos de la muestra
# conf = nivel de confianza deseado. Por defecto en 0.90.
#
# Ordenación de los datos y tamaño de la muestra:
y <- sort(dat); n <- length(dat)
#
# Posicion en la muestra ordenada necesaria como punto de comienzo
# en las iteraciones. Depende de si n es par o impar, pues esta posición
# es la de la mediana:
m <- n/2
m1 <- ifelse(n %% 2 == 0, m+1, m+0.5)
#
# La mediana
ymed <- median(dat)
#
# Constantes de Afinamiento (Tunning)
c1 <- 3.7
c2 <- 1/(0.58173-0.607227*(conf))
#
# ---------
if(m %% 2 == 0) mad <- 0.5*(y[m+m/2]+y[m+1+m/2]) - ymed else mad <- y[m+(m+1)/2] - ymed
mad <- mad/0.6745
# ---------
s <- mad; s1 <- 0; s2 <- 0
for(j in m1:n){
```

```
u2 <- ((y[j]-ymed)/(c1*s))^2
if(u2 < 1) {
s1 <- s1+2*u2*(1-u2)^4
s2 <- s2+2*(1-u2)*(1-5*u2)
}
}
# ----------
s3 <- ifelse(s2-1 < 1, 1, s2-1)
if(s2 > 1) s <- c1*s*sqrt((n*s1)/(s2*s3))
# ----------
sbi <- s; s1 <- 0; s2 <- 0
for(j in m1:n){
u2 <- ((y[j]-ymed)/(c1*sbi))^2
if(u2 < 1){
s1 <- s1+2*u2*(1-u2)^4
s2 <- s2+2*(1-u2)*(1-5*u2)
}
}
# ----------
s3 <- ifelse(s2-1<1, 1, s2-1)
if(s2>1) s <- c1*s*sqrt((s1)/(s2*s3))  # note que aqui no va una "n"
# ----------
st2 <- s^2; s <- mad; s1 <- 0; s2 <- 0
for(j in m1:n){
u2 = ((y[j]-ymed)/(c2*s))^2  # note que aqui se usa c2 (no c1)
if(u2 < 1) {
s1 <- s1+2*u2*(1-u2)^4
s2 <- s2+2*(1-u2)*(1-5*u2)
}
}
# ----------
s3 <- ifelse(s2-1<1, 1, s2-1)
if(s2>1) s <- c2*s*sqrt((n*s1)/(s2*s3))  # note que aqui se usa c2 (no c1)
sbi2 <- s^2
# ----------
# Se calculan los límites del intervalo de predicción:
L1 <- ymed - qt(p=1-(1-conf)/2,df=n-1)*sqrt(st2+sbi2)
L2 <- ymed + qt(p=1-(1-conf)/2,df=n-1)*sqrt(st2+sbi2)
c(L1, L2)
} # Fin de la función !!!
# ----------
```

El método robusto de Horn et al. (1998) funciona sobre datos simétricos de modo que estos autores aplican su procedimiento sobre datos con previa transformación de Box-Cox para asegurar la simetría de los mismos.

Además de esto, Horn et al. (1998) presentan otra alternativa que usa el mismo estimador cuantil robusto descrito arriba por (A-1). La idea es examinar sólo datos mayores que la mediana muestral. Para esto, crean una "pseudo-muestra" simétrica que incluye todos los datos mayores que la mediana y sus "pseudo-valores" que son equidistantes y menores que la mediana. Por ejemplo, si $n = 20$ y los datos ordenados son $x_{(1)} < \cdots < x_{(20)}$, entonces la mediana muestral es $M = (x_{(10)} + x_{(11)})/2$ y la pseudo-muestra simétrica es:

$$2M - x_{(20)} < 2M - x_{(19)} < \cdots < \underbrace{2M - x_{(11)}}_{(= x_{(10)})} < x_{(11)} < x_{(12)} < \cdots < x_{(20)}$$

Desde esta "pseudo-muestra", se calcula un intervalo de predicción simétrico usando el procedimiento `biw.pred.int`, pero sólo se usa el límite superior para la construcción del intervalo de referencia. El límite inferior no se usa debido a que en la mayoría de los casos la población estudiada tiene asimetría positiva, y entonces, la mediana será mayor que la moda. La pseudo-muestra en estos casos, aunque simétrica, sera indicativa de una población bimodal. Así, el límite inferior para el intervalo de referencia se estima usando el estimador no-paramétrico de Harrell & Davis (1982).
En este estudio sólo se incluyó el método robusto de Horn con la pseudo-muestra y el límite inferior no-paramétrico. El siguiente código para `R` muestra la función `pseudo.muestra` que permite obtener la pseudo-muestra como la describen Horn et al. (1998):

```
# Función para obtener la pseudo-muestra:
pseudo.muestra <- function(x){  # x es un vector con la muestra original
xmed <- median(x); n <- length(x)
x <- sort(x)  # se ordenan los datos
if(n %% 2 == 0) {  # Si n es par
pos <- n/2+1
res <- c(2*xmed - x[n:pos], x[pos:n])
}
else {  # si n es impar
pos <- n/2 + 0.5
res <- c(2*xmed - x[n:pos], x[(pos+1):n])
}
res
}
```

Bootstrap para un percentil empírico e Intervalos de confianza tipo percentil y BC$_a$
El bootstrap como se describe en la sección 2.2 se programó usando el comando `boot(...)` de la librería del mismo nombre (Davison & Hinkley, 1997). Para usar este comando se construyó la función `boot.cuantil` que recibe los datos de la muestra y permite realizar bootstrap y calcular intervalos de confianza tipo percentil y tipo BCa para cada límite del intervalo de referencia al mismo tiempo. El código para esta función se muestra enseguida:

```
# Función para calcular el intervalo de confianza bootstrap tipo BCa
# y tipo percentil para dos cuantiles al tiempo sobre un unico vector
# de datos (la muestra).
boot.cuantil <- function(x, B=1000, alfa=0.1, p=c(0.025,0.975), tipo=5){
#
# Argumentos:
# x = vector de datos
# B = nro. de réplicas bootstrap
# alfa = probabilidad complementaria a la confianza deseada para el IC
# bootstrap. Esto es, si se desea un IC de 0.9, entonces, alfa = 0.1.
# p = probabilidades para los limites del Intervalo de referencia
# tipo = método de calculo del percentil empírico (estadístico)
# para el bootstrap
#
# 1. Se realiza bootstrap sobre cada cuantil
library(boot)
fun.quant <- function(x,i,p) quantile(x[i], probs=p, type = tipo, names = F)
boot.res1  <- boot(data = x, statistic = fun.quant, R = B, p=p[1])
boot.res2  <- boot(data = x, statistic = fun.quant, R = B, p=p[2])
#
# 2. Función que calcula la muestra jacknife del percentil:
jacknife <- function(x, p, tipo){
matrix(1:length(x), ncol=1) -> indices
q.jack <- function(i, x) quantile(x[-i], probs=p, type=tipo, names=F)
apply(indices, 1, q.jack, x=x)
} # fin de funcion: jacknife
#
# 3. Función que calcula el coeficiente de acelaración:
coef.ace <- function(x) sum((mean(x)-x)^3)/(6*sum((mean(x)-x)^2)^(3/2))
#
# 4. Función que calcula el coeficiente de sesgo:
coef.sesgo <- function(x, estad) {
dem <- length(x)+1
p <- sum(ifelse(x < estad, 1, 0))/dem
q <- qnorm(p=ifelse(p==0, 1/dem, p) )
list(p=p,q=q)
}
#
# 5. Valores de alfa1 y alfa2 para obtener los límites de los IC en el BCa
alfas <- function(coe, alfa){
m1 <- coe[1] + qnorm(alfa/2); m2 <- coe[1] + qnorm(1-alfa/2)
Z1 <- coe[1] + m1/(1-coe[2]*m1); Z2 <- coe[1] + m2/(1-coe[2]*m2)
c(alfa1=pnorm(Z1), alfa2 = pnorm(Z2))
} # Fin de la función alfas
#
# 6. Calculos varios:
```

```
# estimadores puntuales bootstrap
media.boot <- colMeans(boot.res <- cbind(boot.res1$t,boot.res2$t))
est.org <- c(boot.res1$t0, boot.res2$t0)
sesgo.boot <- media.boot - est.org
de.boot <- apply(boot.res, 2, sd)
ICperc  <- apply(boot.res, 2, quantile, probs=c(alfa/2, 1-alfa/2), type=6, names=T)
#
# Calculo de IC BCa:
res.z01 <- coef.sesgo(boot.res1$t[,1], boot.res1$t0)
z01 <- res.z01$q
ace1 <- coef.ace(jacknife(x, p=p[1], tipo=tipo))
alfas1 <- alfas(c(z01,ace1), alfa=alfa)
IC.BCa1 <- quantile(boot.res1$t[,1], probs=alfas1, type=tipo, names=F)
#
res.z02 <- coef.sesgo(boot.res2$t[,1], boot.res2$t0)
z02 <- res.z02$q
ace2 <- coef.ace(jacknife(x, p=p[2], tipo=tipo))
alfas2 <- alfas(c(z02,ace2), alfa=alfa)
IC.BCa2 <- quantile(boot.res2$t[,1], probs=alfas2, type=tipo, names=F)
#
# Calculo de los intervalos desde boot.ci:
error1 <- try( icboot1 <- boot.ci(boot.res1, conf = 1-alfa), silent = T )
if(is.character(error1) ) { icboot1 <- boot.ci(boot.res1, conf = 1-alfa,
type = c("basic","stud","perc")) }
#
error2 <- try( icboot2 <- boot.ci(boot.res2, conf = 1-alfa), silent = T )
if(is.character(error2) ) { icboot2 <- boot.ci(boot.res2, conf = 1-alfa,
type = c("basic","stud","perc")) }
#
ic1.basic   <- icboot1$basic[1,-1]
if(is.null(ic1.basic)) ic1.basic <- rep(NA,4)
ic1.student <- icboot1$student[1,-1]
if(is.null(ic1.student)) ic1.student <- rep(NA,4)
ic1.perc    <- icboot1$percent[1,-1]
if(is.null(ic1.perc)) ic1.perc <- rep(NA,4)
ic1.bca     <- icboot1$bca[1,-1]
if(is.null(ic1.bca)) ic1.bca <- rep(NA,4)
#
ic2.basic   <- icboot2$basic[1,-1]
if(is.null(ic2.basic)) ic2.basic <- rep(NA,4)
ic2.student <- icboot2$student[1,-1]
if(is.null(ic2.student)) ic2.student <- rep(NA,4)
ic2.perc    <- icboot2$percent[1,-1]
if(is.null(ic2.perc)) ic2.perc <- rep(NA,4)
ic2.bca     <- icboot2$bca[1,-1]
if(is.null(ic2.bca)) ic2.bca <- rep(NA,4)
```

```
#
# Se entregan resultados:
res <- c(est.org, media.boot, de.boot, ICperc[,1], ICperc[,2], IC.BCa1, IC.BCa2,
res.z01$p, res.z02$p, ace1, ace2, alfas1, alfas2, B*alfas1, B*alfas2,
ic1.basic, ic1.student, ic1.perc, ic1.bca,
ic2.basic, ic2.student, ic2.perc, ic2.bca)
res <- round(res,5)
names(res) <- c("est.org1", "est.org2", "media1", "media2", "de1", "de2",
"L1perc1", "L2perc1", "L1perc2", "L2perc2",
 "L1.BCa1","L2.BCa1","L1.BCa2","L2.BCa2","p.z01","p.z02","ace1","ace2",
 "alfa11","alfa12", "alfa21", "alfa22",
 "pos1.bca1","pos2.bca1","pos1.bca2","pos2.bca2",
 "ic1.basic.pos1", "ic1.basic.pos2", "ic1.basic.L1", "ic1.basic.L2",
 "ic1.student.pos1", "ic1.student.pos2", "ic1.student.L1", "ic1.student.L2",
  "ic1.perc.pos1", "ic1.perc.pos2", "ic1.perc.L1", "ic1.perc.L2",
  "ic1.bca.pos1", "ic1.bca.pos2", "ic1.bca.L1", "ic1.bca.L2",
  "ic2.basic.pos1", "ic2.basic.pos2", "ic2.basic.L1", "ic2.basic.L2",
 "ic2.student.pos1", "ic2.student.pos2", "ic2.student.L1", "ic2.student.L2",
  "ic2.perc.pos1", "ic2.perc.pos2", "ic2.perc.L1", "ic2.perc.L2",
  "ic2.bca.pos1", "ic2.bca.pos2", "ic2.bca.L1", "ic2.bca.L2" )
#
res
}
```

Enfoque Bayesiano usando una verosimilitud estimada por Bootstrap (Boos & Monahan, 1986) Este procedimiento se describe en detalle en la sección 2.3 y se programó en `R` mediante la función `bbayes.org.simul` cuyo código se muestra enseguida:

```
bbayes.org.simul <- function(datos, pq = c(0.025, 0.975),  tipoq = 5, rb = 1000,
                        transf = TRUE, metodo.h = "bcv", cambiar.h = TRUE, ph = 80,
                        conf = 0.9, amedia, ade ){
#
# ARGUMENTOS:
# datos = datos
# pq = prob para los cuantiles que se desean estimar; máximo dos
# rb = nro. de replicas bootstrap
# transf = lógico. ¿Se desea que las replicas bootstrap se transformen de acuerdo
#    con Boos & Monaham: 2*tetha - tetha[i]?
# metodo.h = metodo para encontrar h. Ver ayuda de density(.).
# También puede ser un valor deseado para la amplitud de banda.
# cambiar.h = lógico. Se desea que el valor de la amplitud de banda se cambia
# hasta hallar una densidad para la verosimilitud bootstrap con una sola
# región de alta densidad
# ph = prob (entre 0 y 100) para la región de alta densidad en caso de que
# cambiar.h = T
```

```
# amedia y ade = vectores de longitud 1 ó 2. Con las medias y desv. estándar
# de la densidad apriori normal para cada cuantil en pq.
# conf = confianza (entre 0 y 1) para el IC de conf*100%
# tipq = argumento "type" para la función "quantile" en el bootstrap
#
#
# CÓDIGO:
# 1. Bootstrap no paramétrico para los dos cuantiles:
library(boot)
fun.quant <- function(x,i) quantile(x[i], probs = pq, type = tipoq, names = F)
boot.res  <- boot(data = datos, statistic = fun.quant, R = rb)
#
# 2. Verosimilitud bootstrap de acuerdo a Boos & Monaham para cada cuantil:
#
# Se sacan las replicas bootstrap de cada cuantil
if(transf) {
bq1 <- 2*boot.res$t0[1] - boot.res$t[,1]
bq2 <- 2*boot.res$t0[2] - boot.res$t[,2]
} else {
bq1 <- boot.res$t[,1]
bq2 <- boot.res$t[,2]
}
#
# Densidades no parametricas desde las replicas bootstrap
# (que serán las verosimilitudes):
#
den1 <- density(bq1, bw = metodo.h)
h01 <- den1$bw
den2 <- density(bq2, bw = metodo.h)
h02 <- den2$bw
#
library(hdrcde)
if(cambiar.h){
# En seguida se verifica si den1 y den2 tienen un sólo "high density region" de la
# probabilidad escogida en la entrada de la función con el argumento ph. Si no,
# se escoge un ancho de banda que si lo haga. Note que este procedimiento permite
# una densidad final con más de una moda local
#
# Funcion que cuenta el nro. de "high density region" para una prob = ph
nro.hdr <- function(dens, p = ph){
res.hdr <- hdr(prob = p, den = list(x = dens$x, y =dens$y) )
length(res.hdr$hdr)/2
}
#
# La siguiente función selecciona una ancho de banda que genera una densidad
# con un solo hdr de probabilidad ph
```

```
selh <- function(datos, dens, p = ph){
h <- dens$bw
nhdr <- nro.hdr(dens)
#
while(nhdr != 1){
h <- h + 0.01
nhdr <- nro.hdr(density(datos,bw=h))
}
h  # entrega h
}
#
if(nro.hdr(dens = den1) != 1) {
h1 <- selh(bq1, dens = den1)
den1 <- density(bq1, bw = h1)
}
if(nro.hdr(dens = den2) != 1) {
h2 <- selh(bq2, dens = den2)
den2 <- density(bq2, bw = h2)
}
}
#
h11 <- den1$bw
h12 <- den2$bw
#
# 3. Construcción de densidad posteriori
#
veros1 <- den1$y/sum(den1$y)
apriori1 <- dnorm(den1$x, mean = amedia[1], sd = ade[1])
apriori1 <- apriori1/sum(apriori1)
apost1 <- veros1*apriori1
apost1 <- apost1/sum(apost1)
#
veros2 <- den2$y/sum(den2$y)
apriori2 <- dnorm(den2$x, mean = amedia[2], sd = ade[2])
apriori2 <- apriori2/sum(apriori2)
apost2 <- veros2*apriori2
apost2 <- apost2/sum(apost2)
#
# 3. Se resume la densidad posteriori:
#
# media posteriori:
med.post1 <- weighted.mean(den1$x, apost1)
med.post2 <- weighted.mean(den2$x, apost2)
#
# Resumen con sample:
p1 <- (1-conf)/2; p2 <- p1+conf  # se calculan probs para la confianza de cada cuantil
```

```
res.post1 <- quantile(sample(den1$x,size=10000,prob=apost1,replace=T), probs = c(p1,0.5,p2))
res.post2 <- quantile(sample(den2$x,size=10000,prob=apost2,replace=T), probs = c(p1,0.5,p2))
#
   #
# Entrega de resultados:
res1 <- c(med.post1,res.post1, h01, h11,
as.numeric(c(cambiar.h, transf)), tipoq, amedia[1], ade[1])
names(res1) <- c("q1.media",paste("q1.s",c(p1,0.5,p2)*100,sep=""),
   "h01","h11","cambiar.h1","transf1","tipoq1","amedia1","ade1")
#
res2 <- c(med.post2,res.post2, h02, h12,
        as.numeric(c(cambiar.h, transf)), tipoq, amedia[2], ade[2])
names(res2) <- c("q2.media",paste("q2.s",c(p1,0.5,p2)*100,sep=""),
    "h02","h12","cambiar.h2", "transf2","tipoq2","amedia2","ade2") #
res <- c(res1,res2)
res
#
}  # Fin de la función
```

Bibliografía

Balconi, S. M., Posse, G. B., & Hentschel, S. (2006). Intervalos de referencia para Troponina I, Creatina quinasa, Fracción MB masa y Mioglobina. *Acta Bioquímica Clínica Latinoamericana*, *40*(2), 197–204. Disponible en la URL: http://www.scielo.org.ar/pdf/abcl/v40n2/v40n2a06.pdf.

Barreto Penie, J., Santana Porben, S., & Consuegra Silverio, D. (2003). Intervalos de referencia locales para la excreción urinaria de creatinina en una población adulta. *Nutrición Hospitalaria*, *18*(2), 65–75. Disponible en la URL: http://wwwscielo.isciii.es/pdf/nh/v18n2/original1.pdf.

Boos, D. D. & Monahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika*, *73*(1), 77–83.

Box, G. E. P. & Cox, D. R. (1964). The analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211–252.

Boyd, J. C. & Lacher, D. A. (1982). A multi-stage gaussiantransformation algorithm for clinical laboratory data. *Clinical Chemistry*, *28*(8), 1735–1741.

Canty, A. & Ripley, B. (2010). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.2-43.

Carpenter, J. & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine*, *19*, 1141–1164.

Correa, J. C. (2004). Método bayesiano bootstrap y una aplicación en la estimación del percentil 85 en ingeniería de tránsito. *Revista Colombiana de Estadística.*, *27*(2), 99–107.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap methods and their applications.* Cambrige: Cambrige University Press.

Diciccio, T. J. & Romano, J. P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society. Series B (Methodological)*, *50*(3), 338–354.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, *7*, 1–26.

Efron, B. & Tibshirani, R. (1993). *An introducction to the bootstrap.* New York: Chapman & Hall.

Elveback, L. R., Guillier, C. L., & Keating, F. R. J. (1970). Health, normality, and the ghost of gauss. *Journal of the American Medical Association*, *211*, 69?–75.

Fernandez, L. E., Bustamante, Y., & Garcia, G. (2006). Valores de referencia obtenidos con el autoanalizador coulter gen-s. *RFM*, *29*(1), 38–43.

Geffre, A., Friedrichs, K., Harr, K., Concordet, D., Trumel, C., & Braun, J.-P. (2009). Reference values: a review. *Veterinary Clinical Pathology*, *38*(3), 288—-298.

Griffiths, J. K., Iles, T. C., Koduah, M., & Nix, A. B. (2004). Centile charts II: Alternative nonparametric approach for establishing time-specific reference centiles and assessment of the sample size required. *Clinical Chemistry*, *50*(5), 907?–914.

Gräsbeck, R. (2004). The evolution of the reference value concept. *Clin Chem Lab Med*, *42*(7), 692–697.

Gräsbeck, R. & Saris, N. E. (1969). Establishment and use of normal values. *Scandinavian Journal of Clinical Laboratory Investigations*, *26*(Suppl 110), 62–63.

Harrell, F. E. (2010). *Hmisc: Harrell Miscellaneous.* R package version 3.8-3.

Harrell, F. E. & Davis, C. E. (1982). A new distribution free quantile estimator. *Biometrika*, *69*(3), 635–640.

Harwood, S. J. & Cole, G. W. (1978). Reference values based on hospital admission laboratory data. *JAMA*, *240*(3), 270–274.

Haukoos, J. S. & Lewis, R. J. (2005). Advanced statistics: Bootstrapping confidence intervals for statistics with "difficult" distributions. *Acad. Emerg. Med.*, *12*(4), 360–365.

Heider, E. C., Davis, B. G., & Frank, E. L. (2004). Nonparametric determination of reference intervals for plasma Metanephrine and Normetanephrine. *Clinical Chemistry*, *50*(12), 2381–2384.

Horn, P. S. (1988). A biweight prediction interval for random samples. *Journal of the American Statistical Association*, *83*(401), 249–256.

Horn, P. S. (1990). Robust quantile estimators for skewed populations. *Biometrika*, *77*(3), 631–636.

Horn, P. S. & Pesce, A. J. (2003). Reference intervals: an update. *Clinica Chimica Acta.*, *334*, 5–23.

Horn, P. S., Pesce, A. J., & Copeland, B. E. (1998). A robust approach to reference interval. *Clinical Chemistry*, *44*(3), 622–631.

Hyndman, R. J. & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, *50*(4), 361–365.

Jellinek, E. M. & Looney, J. M. (1939). Statistics of some biochemical variables on healthy men in the age range of twenty to forty–five years. *The Journal of biological chemistry*, 621–630.

Jiménez Díaz, M. & Schosinsky Nevermann, K. (2000). Valores de referencia de Colinesterasa plasmática y eritrocítica en población costarricense. Comparación del desempeño clínico de ambas enzimas. *Revista Costarricense de Ciencias Médicas.*, *21*(3-4), 117–126.

John, J. A. & Draper, N. R. (1980). An alternative family of transformations. *Applied Statistics*, *29*(2), 190–197.

Koduah, M., Iles, T. C., & Nix, B. J. (2004). Centile charts I: New method of assessment for univariate reference intervals. *Clinical Chemistry*, *50*(5), 901–906.

Linnet, K. (1987). Two-stage transformation systems for normalization of reference distributions evaluated. *Clinical Chemistry*, *33*(3), 381–386.

Linnet, K. (2000). Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clinical Chemistry*, *46*(6), 867–869.

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology.* Londres: Chapman & Hall.

Molinero, L. M. (2007). Estimación de intervalos de referencia de variables biológicas. Descargado el 17/Marzo/2007 de la URL: http://www.seh-lelha.org/intervalref.htm.

Muiño, J., Voit, E., & Sorribas, A. (2006). GS-distributions: A new family of distributions for continuous unimodal variables. *Computational Statistics & Data Analysis*, *50*, 2769–2798.

NCCLS (1992). *How to define and determine reference intervals in the Clinical Laboratory: Proposed guideline.* 12(2). National Committee for Clinical Laboratory Standards. Document C–28–P.

Nordøy, E. S. & Thoresen, S. I. (2002). Reference values for serum biochemical. parameters in free-ranging harp seals. *Veterinary Clinical Pathology*, *31*(3), 98–105.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, *74*(365), 105–121.

Paul Murphy, J., Work, T., Hunter, D., McFie, E., & Fjelline, D. (1994). Serologic survey and serum biochemical reference ranges of the free-ranging mountain lion (*Felis concolor*) in california. *Journal of Wildlife Diseases*, *30*(2), 205–215.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Royston, P. & Wright, E. M. (1998). A method for estimating age–specific reference intervals ("normal ranges") based on fractional polynomials and exponential transformation. *J. R. Statist. Soc. A*, *161*(Part 1), 79–101.

Solberg, H. E. (1987). Approved recommendation (1987) on the theory of reference values. part 5. statistical treatment of collected reference values. determination of reference limits. *Clinica Chimica Acta*, *170*(2-3), S13 – S32.

Solberg, H. E. (2004). The IFCC recommendation on estimation of reference intervals. The RefVal Program. *Clin. Chem. Lab. Med.*, *42*(7), 710–714.

Tanyalcin, T., Aslan, D., Kurtulmus, Y., Gökalp, N., & Kumanlioglu, K. (2000). Reference intervals of serum folate and vitamin B12 developed from data of healthy subjects. *Accreditation and Quality Assurance*, *5*, 383–387.

Thompson, M. L. & Theron, G. B. (1990). Maximum likelihood estimation of reference centiles. *Statistics In Medicine*, *9*(5), 539 – 548.

Tirosh, A., Shai, I., Tekes-Manova, D., Israeli, E., Pereg, D., Shochat, T., Kochba, I., & Rudich, A. (2005). Normal fasting plasma glucose levels and type 2 diabetes in young men. *New England Journal of Medicine*, *353*(14), 1454–1462.

Trujillano, J., Muiño, J. M., March, J., & Sorribas, A. (2007). A more flexible parametric estimation of univariate reference intervals: A new method based on the GS-distribution. *Clinica Chimica Acta*, *379*(1–2), 71–80.

Virtanen, A., Kairisto, V., & Uusipaikka, E. (1998). Regression-based reference limits: determination of sufficient sample size. *Clinical Chemistry*, *44*(11), 2353?–2358.

Wright, E. M. & Royston, P. (1997). A comparison of statistical methods for age-related reference intervals. *J. R. Statist. Soc. A*, *160*(Part 1), 47–69.

Wright, E. M. & Royston, P. (1999). Calculating reference intervals for laboratory measurements. *Statistical Methods in Medical Research*, *8*, 93–112.

Yip, P. M., Chan, M. K., Nelken, J., Lepage, N., Brotea, G., & Adeli, K. (2006). Pediatric reference intervals for lipids and apolipoproteins on the VITROS 5,1 FS Chemistry System. *Clinical Biochemistry*, *39*, 978–983.