matemática: Una publicación del ICM – ESPOL
2009, Vol. 7, No. 2

INTERVALOS DE CONFIANZA E INTERVALOS DE CREDIBILIDAD PARA UNA PROPORCIÓN ESTUDIO COMPARATIVO

Cervantes Víctor[1], Díaz Iván[2], Rodríguez Diana[3], Cepeda Cuervo Edilberto[4]

Resumen. *En este artículo, basado en Cepeda-Cuervo et al. (2008), se evalúa y se compara el comportamiento de diferentes metodologías utilizadas para la obtención de intervalos de confianza y/o de credibilidad, analizando sus probabilidades de cobertura estimada, su longitud esperada y la varianza de su longitud. Definidos estos tres conceptos, la comparación entre los intervalos considerados se desarrolla mediante procesos computacionales utilizando el paquete estadístico R. En este proceso, además de la verificación de conclusiones conocidas, como el mal comportamiento del Intervalo de Wald y la sobre-cobertura del intervalo exacto, se determinan, entre otros aspectos, características de los intervalos relacionadas con la variabilidad de su longitud.*

Palabras claves: proporciones, intervalos de confianza, métodos bayesianos, intervalos de credibilidad, cobertura.

1. INTRODUCCIÓN

En varias ocasiones se han comparado las propiedades frecuentistas de diferentes métodos para la obtención de estos intervalos para una proporción p [Agresti & Coull, 1998, Agresti & Caffo, 2000, Brown et al., 2002]. En diversos textos de enseñanza estadística recomiendan el intervalo de Wald cuando npq es mayor que 5 o 10 [Canavos, 1988]; sin embargo, es manifiesto el mal comportamiento del intervalo de Wald aún cuando estas condiciones son satisfechas [Agresti & Coull, 1998, Brown et al., 2001]. Sobre este intervalo, además, aún existe la creencia errónea de que presenta problemas sólo cuando p está cerca de 0 ó 1, o cuando el tamaño de la muestra n es bastante pequeño. En este trabajo se comparan diferentes propuestas de intervalos de confianza y de intervalos de credibilidad para una proporción p. Para cada uno de los intervalos se analiza el valor esperado de la probabilidad de cobertura, la longitud del intervalo y la varianza de ésta para distintos valores de p y del tamaño de la muestra n.

Después de una breve introducción, en la sección 2 se incluyen aspectos teóricos de los intervalos a comparar. En la sección 3 se hace una breve reseña de la metodología usada en la comparación de los intervalos. En la sección 4 se incluyen los resultados relacionados con probabilidad de cobertura, longitud esperada de los intervalos y variabilidad de la longitud de los intervalos. En la sección 5 se consideran distribuciones a priori informativas y se analizan los intervalos a posteriori para diferentes valores de p y de tamaño de la muestra. Finalmente, en la sección 6 se incluyen algunas conclusiones y recomendaciones.

2. ALGUNOS MÉTODOS PARA LA OBTENCIÓN DE INTERVALOS DE CONFIANZA Y DE CREDIBILIDAD PARA UNA PROPORCIÓN

En esta sección se exponen aspectos teóricos de intervalos de confianza y de credibilidad para una proporción p. En 2.1 se consideran los intervalos de confianza (IC) de *Wald*, *Exacto*, *Score* y *Wald Ajustado*. En 2.2 se incluye el concepto de intervalos de credibilidad y se definen estos intervalos para una proporción p cuando no existe información a priori acerca de la misma y cuando la información a priori se puede expresar a través de la distribución *Beta*.

2.1 INTERVALOS CLÁSICOS

El *intervalo de Wald* es el presentado por la mayoría de los textos estadísticos para la estimación de IC para una proporción. Se basa en la distribución asintótica del estimador de la proporción muestral $\hat{p} = x/n$, donde x representa el número de éxitos en n ensayos, y está definido por:

$$IC_{Wald} = \hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \qquad (1)$$

[1] *Cervantes Víctor H., Departamento de Estadística - Universidad Nacional de Colombia. (e_mail: vhcervantesb@unal.edu.co)*
[2] *Díaz M. Iván, Departamento de Estadística - Universidad Nacional de Colombia. (e_mail: ildiazm@unal.edu.co)*
[3] *Rodríguez Diana J., Departamento de Estadística - Universidad Nacional de Colombia. (e_mail: dprodriguezc@unal.edu.co)*
[4] *Cepeda Cuervo Edilberto, Departamento de Estadística - Universidad Nacional de Colombia. (e_mail: ecepedac@unal.edu.co)*

Donde $z_{\alpha/2}$ es el cuantil $(1 - z_{\alpha/2})$ de la distribución normal estándar. El *intervalo de confianza exacto* de Clopper-Pearson se basa en la inversión de una prueba binomial a dos colas iguales de la hipótesis $H_0 : p = p_0$. Los límites inferior y superior son las soluciones en p_0 de las ecuaciones

$$\sum_{k=x}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2$$

y

$$\sum_{k=0}^{x} \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2$$

, respectivamente. Para este intervalo, el límite inferior es 0 cuando $x = 0$ y el límite superior es 1 cuando $x = n$. Es fácil mostrar que el límite inferior es el cuantil $\alpha/2$ de una distribución Beta con parámetros x y $x-n+1$, denotada por $B(x,x-n+1)$, y que el límite superior es el cuantil $1-\alpha/2$ de una distribución $B(x+1,n-x)$ [Agresti & Coull, 1998]. El *intervalo de Score*, también conocido como el *intervalo de Wilson*, tiene la forma:

$$IC_{Score} = \tilde{p} \pm \frac{z_{\alpha/2}}{\tilde{n}} \sqrt{(\hat{p}\hat{q} + z_{\alpha/2}^2/4n)/n} \tag{2}$$

donde

$$\tilde{x} = \hat{p} + \frac{z_{\alpha/2}}{2n}, \tilde{n} = 1 + z_{\alpha/2}/n, \tilde{p} = \frac{\tilde{x}}{\tilde{n}}, \hat{p} = \frac{x}{n}$$

y $\hat{q} = 1 - \hat{p}$.

El *intervalo de Wald ajustado* [Agresti & Coull, 1998] está definido como:

$$IC_{Wald-Adj} = \tilde{p} \pm z_{\alpha/2}\sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}} \tag{3}$$

Donde $\tilde{x} = x + \frac{z_{\alpha/2}^2}{2}, \tilde{n} = n + z_{\alpha/2}^2$ y $\tilde{p} = \frac{\tilde{x}}{\tilde{n}}$.

Para el caso particular del intervalo de $(1-\alpha)100\% = 95\%$ de confianza, $z_{0.025}^2 = (1.96)^2 \approx 4$. Esto equivale a añadir 4 pseudo-observaciones: dos éxitos y dos fracasos al cálculo de la proporción muestral. [Agresti & Caffo, 2000] argumentan que la adición de pseudo-observaciones está motivada por el sesgo que presenta la distribución binomial cuando p se aproxima a 0 o a 1, por lo cual $\hat{p}$ no debe ser el punto medio del intervalo.

2.2 INTERVALOS DE CREDIBILIDAD

En la aproximación Bayesiana la estimación por *intervalos* se define por una evaluación simple de las distribuciones a posteriori de los parámetros. Así, si $\theta \in \Theta$ es una cantidad desconocida, $C \subset \Theta$ es una región de $100(1-\alpha)\%$ de credibilidad para θ si $P(\theta \in C \mid x) \geq 1-\alpha$. En este caso $(1-\alpha)$ es llamado el nivel de credibilidad. Si θ es un escalar, la región C está dada usualmente por un intervalo $[c_1,c_2]$ [Bernardo & Smith, 2000].

2.2.1 INTERVALOS SIN INFORMACIÓN PREVIA

En esta sección se consideran regiones de confianza para una proporción considerando como a priori no informativa una distribución $B(1,1)$ y la priori de Jeffreys $B(1/2,1/2)$. En el primer caso, el *intervalo bayesiano* de dos colas iguales y de $1-\alpha$ de credibilidad está dado por:

$$IC_{no\ informativo} = [B(x+1, n-x+1, \alpha/2), B(x+1, 1-\alpha/2)] \tag{4}$$

donde $B(m_1,m_2,\alpha)$ denota el cuantil α de una distribución $B(m_1,m_2)$. En el segundo caso, el *intervalo de colas iguales* y $1-\alpha$ de credibilidad está dado por:

$$IC_{Jeffreys} = [B(x+1/2, n-x+1/2, \alpha/2), B(x+1/2, n-x+1/2, 1-\alpha/2] \tag{5}$$

2.2.2 INTERVALOS CON INFORMACIÓN PREVIA

En algunos estudios se tiene información a priori acerca de la proporción en consideración. Este conocimiento puede ser incorporado en la obtención del intervalo de confianza para la proporción p a través de una distribución a priori informativa que haga uso de este conocimiento. Si la información a priori se incorpora a través de una distribución beta $B(a,b)$, el intervalo de confianza para p está definido por los cuantiles de la distribución a posteriori como en (6)

$$IC_{INformativo(a,b)} = [B(\alpha+x, n+\beta-x, \alpha/2), B(a+x, n+b-x, 1-\alpha/2)] \tag{6}$$

En este trabajo se analizan dos distribuciones a priori posibles cuando se conoce que el parámetro p es pequeño. La primera usando una distribución $B(0.5,2)$, y la otra, una $B(1.5,5)$.

3. METODOLOGÍA

Al evaluar intervalos de confianza (o de credibilidad) se consideran la longitud del intervalo, la probabilidad de cobertura (PC) y la varianza de la longitud del intervalo, el cual es incluido en este artículo. Un buen método debe proponer intervalos con probabilidades de cobertura muy *cercanas* a los niveles de confianza nominal y con valores pequeños del valor esperado y de la varianza de su longitud. La PC para una proporción p es $PC = \sum_{x=0}^{n} I(x,p)P_p(x)$, con $I(x, p)$ igual a 1 si el intervalo contiene a p cuando $X = x$ e igual a 0 si este no contiene a p, y $P_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$. En todos los casos, la PC, la longitud esperada del intervalo y su varianza, son calculados mediante $E_p(f(X)) = \sum_x f(x)P_p(x)$.

Para obtener estas esperanzas, se obtuvieron los n+1 posibles intervalos para cada tamaño de muestra desde 1 hasta 500 y se calcularon los valores esperados y varianzas bajo los valores de $p = 0(0.01)1$[1]. Esta metodología fue implementada a través del software estadístico R [R Development Core Team, 2008].

4. COMPARACIÓN DE INTERVALOS PARA UNA PROPORCIÓN

4.1 COMPARACIÓN DE LA PROBABILIDAD DE COBERTURA

En la figura 1 se presenta la PC para $p = 0.01, 0.1, 0.3, 0.5$ y $n = 1,2,3, \ldots, 500$, y en la figura 2 la PC como una función de p para $n = 10, 50$ y 100, a un nivel nominal del 95%. Puede observarse también, a partir de las figuras 1 y 2, que el intervalo de *Score* presenta una probabilidad de cobertura que fluctúan alrededor del valor nominal para todos los tamaños de muestra y que el intervalo *bayesiano* con distribución a priori uniforme presenta caídas muy fuertes en la PC respecto al valor nominal cuando el tamaño de muestra es pequeño, $n = 10$.

El intervalo *bayesiano* con a priori de Jeffreys presenta caídas más fuertes que el intervalo *bayesiano* con distribución a priori uniforme en tamaños de muestra pequeños, como $n = 10$ y $n = 50$. Para tamaños grandes de muestra, todos los intervalos presentan probabilidades de cobertura muy cercanas al valor nominal, excepto para valores de p muy próximos a cero y a uno.

Sin embargo, las tendencias señaladas previamente para cada uno de los intervalos se conservan.

4.2 COMPARACIÓN DE LA LONGITUD DE LOS INTERVALOS

En la figura 3 se representa la longitud esperada de los intervalos analizados, como función del parámetro p, para distintos tamaño de muestra. Se observa que los intervalos de *Wald*, *Exacto* y *Wald ajustado* presentan un comportamiento diferente al de los intervalos de *Score* y *bayesianos*. Los intervalos de *Wald ajustado* y *Exacto* tienen longitud mayor que la de todos los demás intervalos, y el intervalo de *Wald* tiene mayor longitud que la de los intervalos de *Score*, *Wald ajustado* y *bayesianos* para valores de p próximos a 0.5 y longitudes mucho menores que las de los demás intervalos para valores de p próximos a los extremos del intervalo (0,1). Los intervalos *Score* y *bayesianos* presentan longitudes muy similares entre sí para cualquier valor de p y de n. En la figura 4 se muestra la varianza de la longitud de los intervalos. Puede verse que el intervalo de *Wald ajustado* presenta la menor varianza de todos los intervalos para cualquier tamaño de muestra y para cualquier valor de p. Los intervalos *Score* y *bayesiano con a priori uniforme* presentan varianzas muy similares para los distintos tamaños de muestra; igualmente ocurre con los intervalos *Exacto* y *bayesiano con a priori de Jeffreys*. El intervalo de *Wald*, por su parte, presenta la mayor variablilidad entre todos los intervalos presentados.

1 Es decir desde 0 hasta uno en pasos de 0.01.

FIGURA 1

Intervalos de confianza e intervalos de credibilidad para una proporción

Probabilidad de cobertura de los IC para distintos tamaños de muestra y valores de p a un nivel nominal del 95%

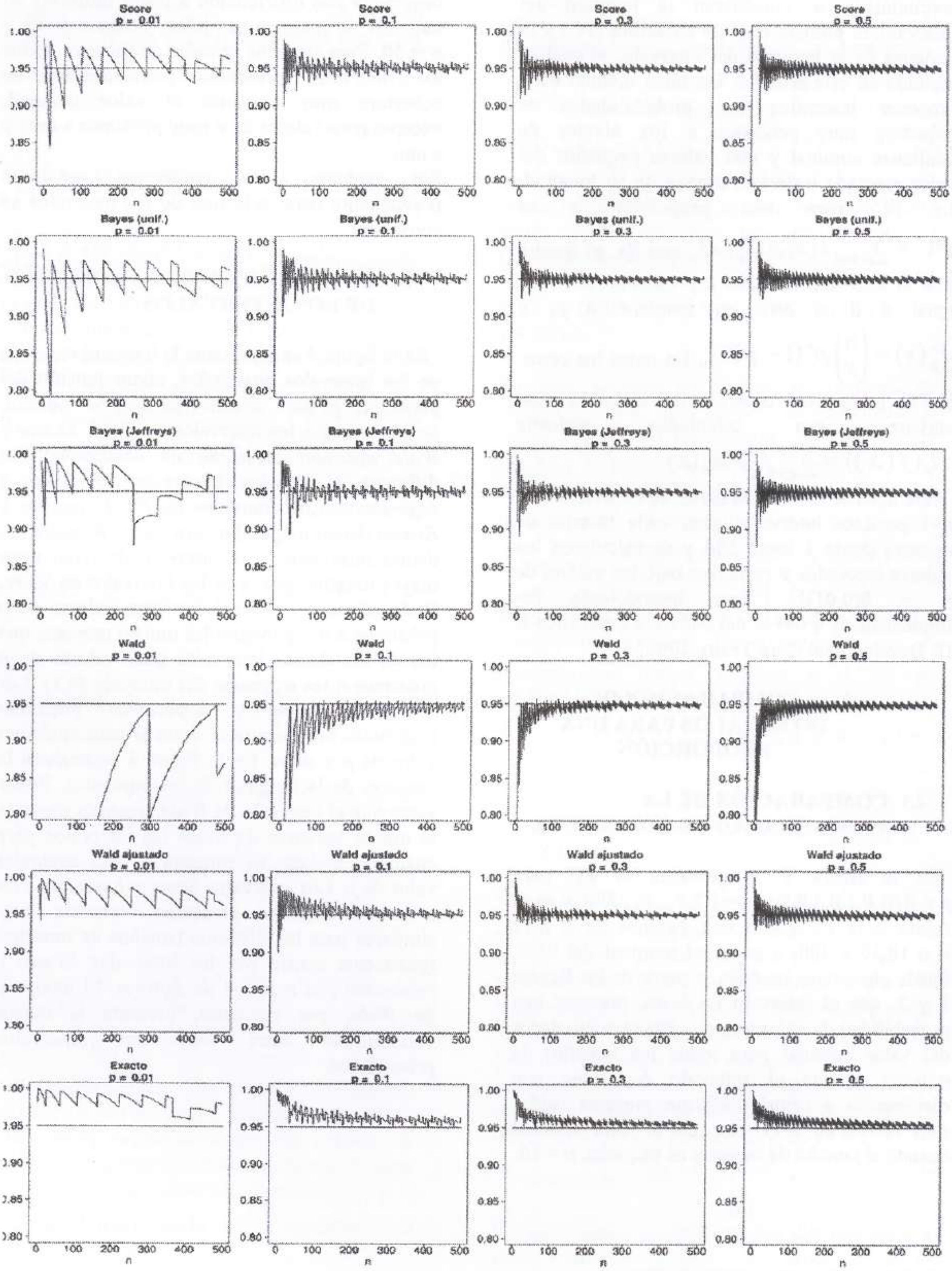


FIGURA 2

Intervalos de confianza e intervalos de credibilidad para una proporción

Probabilidad de cobertura de los IC para diferentes tamaños de muestra a un nivel nominal del 95%

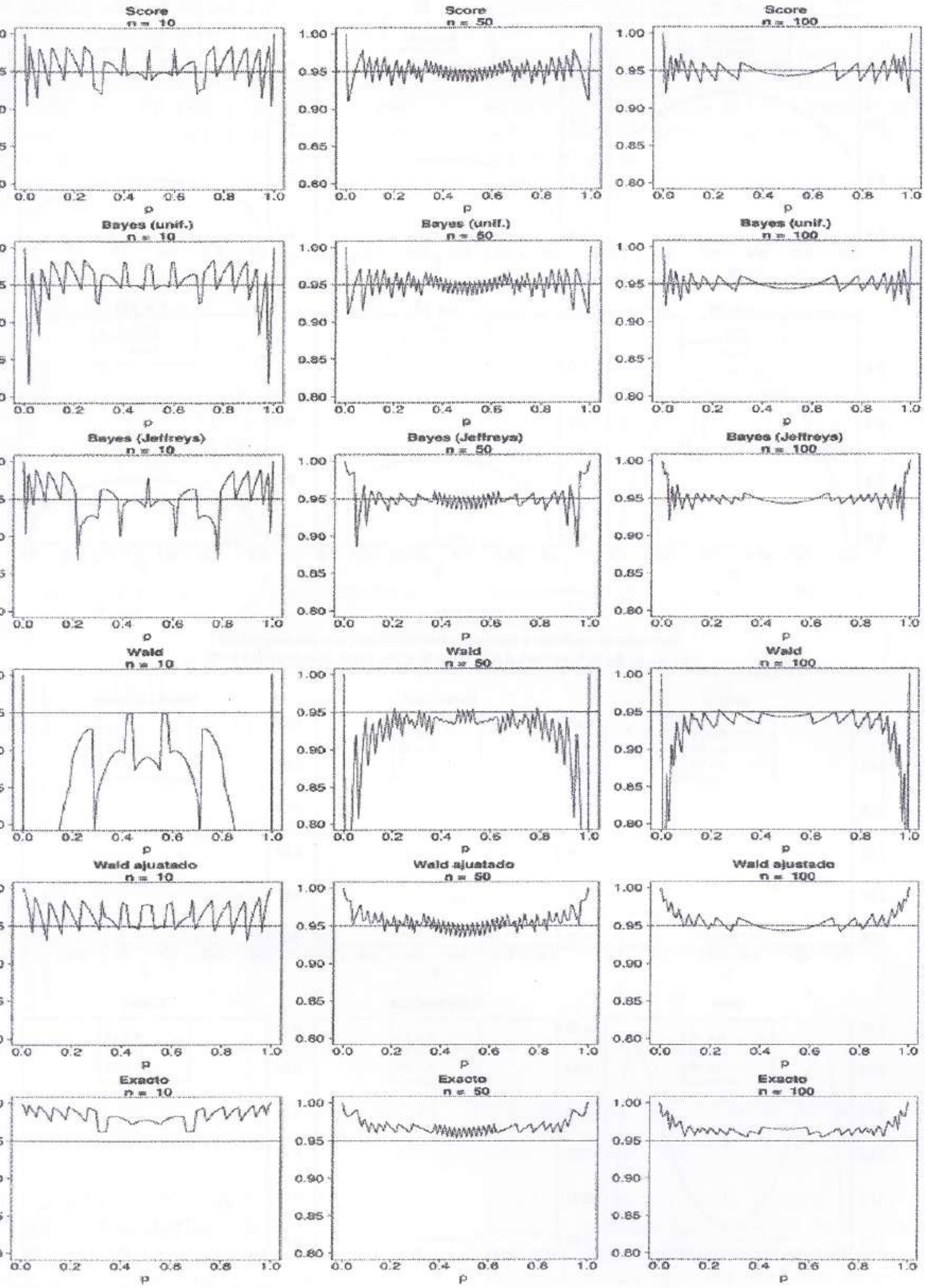


FIGURA 3

Intervalos de confianza e intervalos de credibilidad para una proporción

Longitud esperada de los intervalos para distintos tamaños de muestra y distintos valores de p a un nivel nominal del 95%

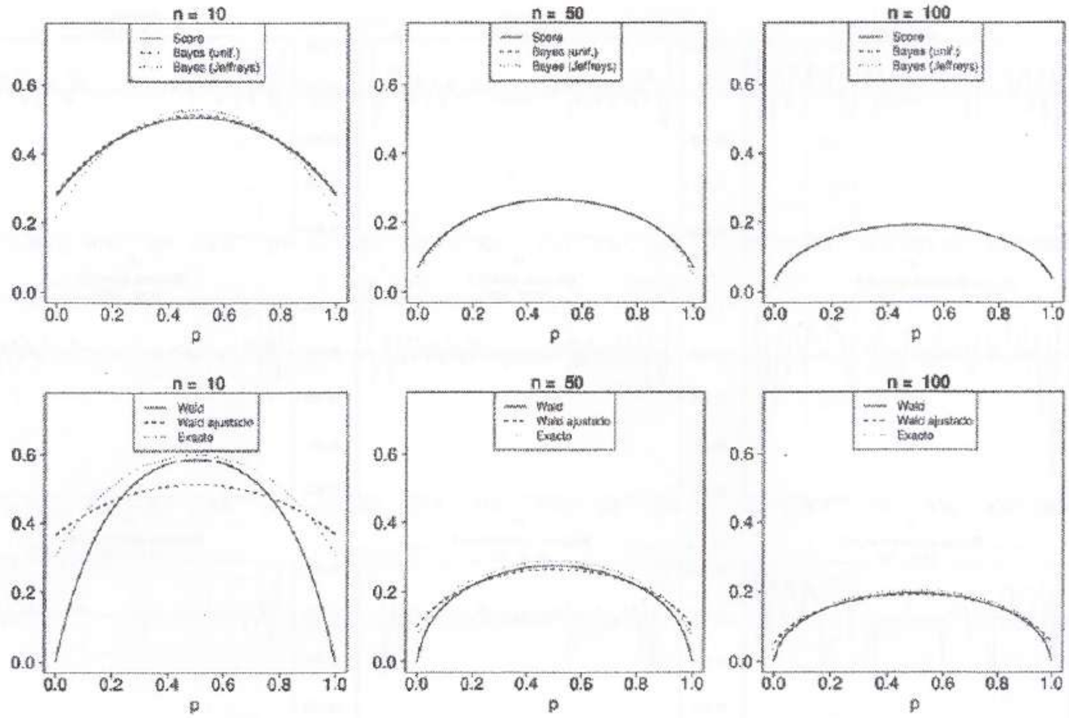
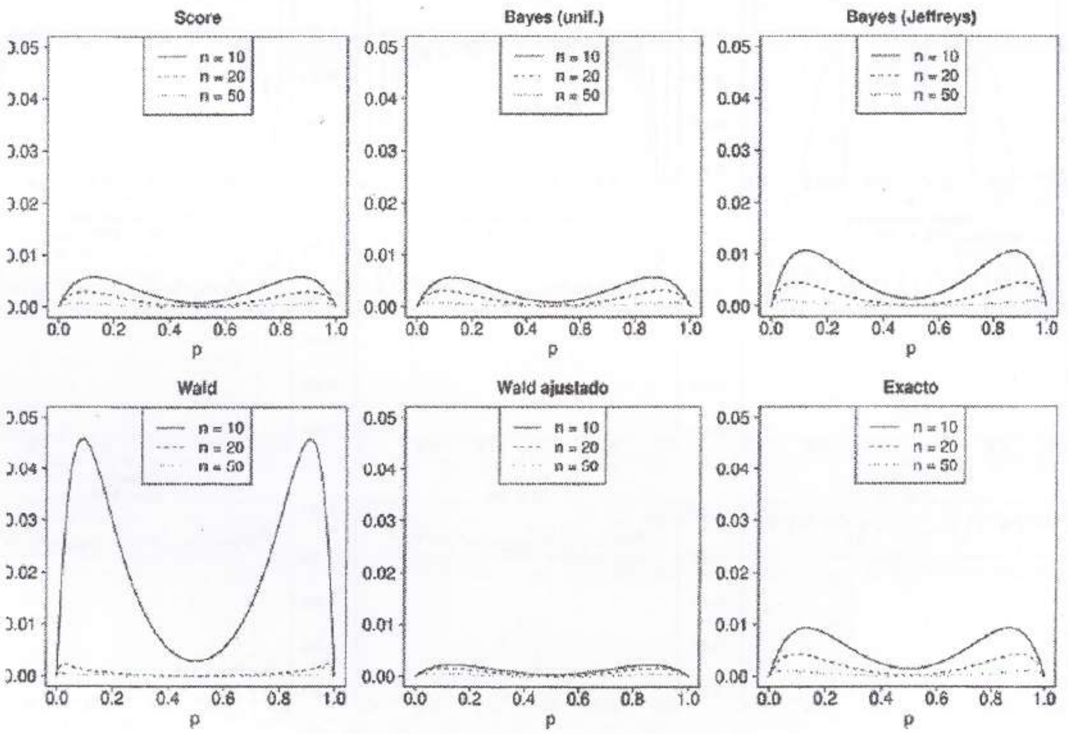


FIGURA 4

Intervalos de confianza e intervalos de credibilidad para una proporción

Varianza de la longitud del intervalo a un nivel nominal del 95%



5. INTERVALOS DE CREDIBILIDAD CON INFORMACIÓN PREVIA

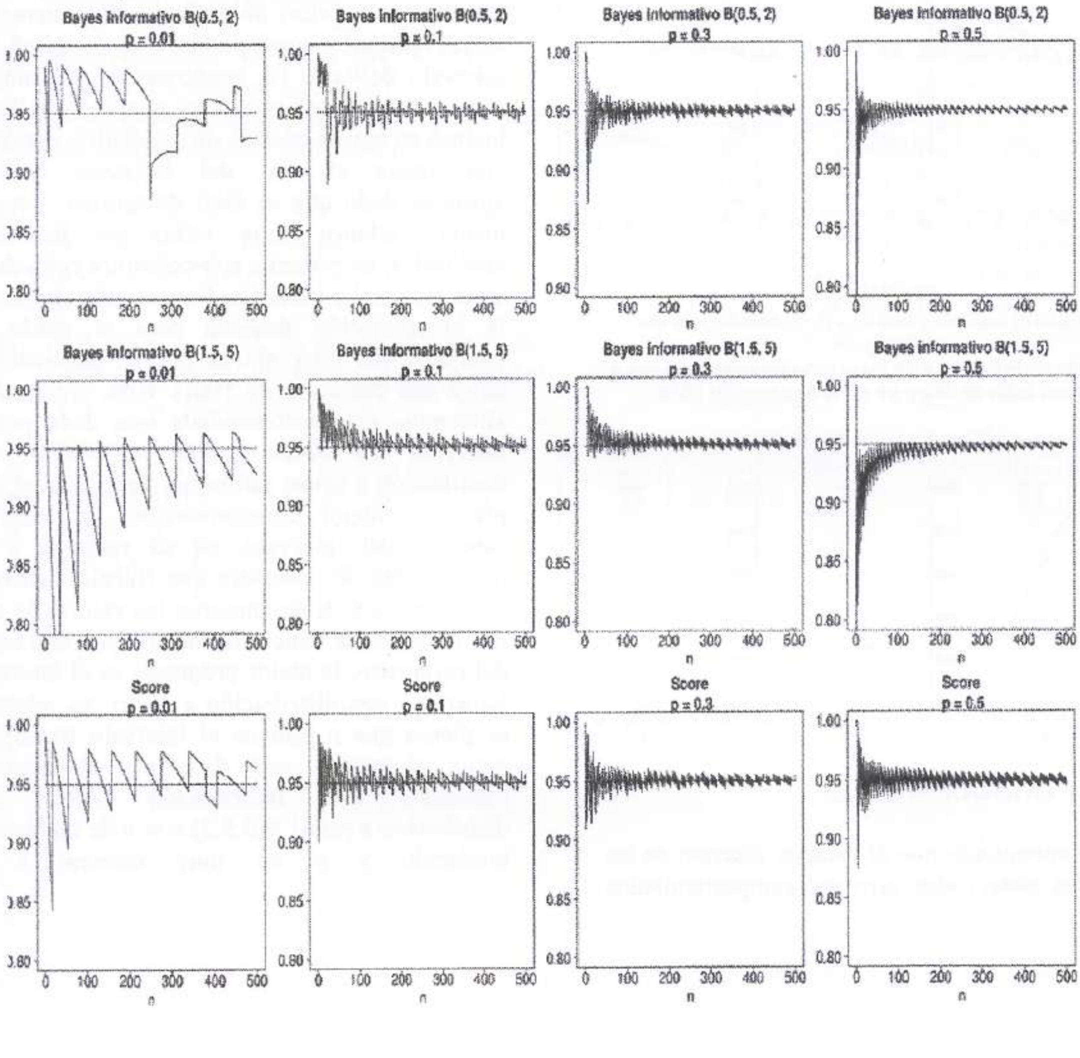
5.1 COMPARACIÓN DE LA PROBABILIDAD DE COBERTURA

En la figura 5 se presenta el comportamiento de la PC de los intervalos *bayesianos con información previa* dada por las distribuciones $B(0.5,2)$ y $B(1.5,5)$, junto con el del intervalo *Score*. Puede verse que la fluctuación del valor de cobertura es menor en los intervalos *bayesianos con información previa* siempre que p no sea muy próximo de cero, $p \sim 0.01$. El intervalo con distribución a priori $B(0.5,2)$ fluctúa alrededor del valor nominal, mientras que el intervalo con a priori $B(1.5,5)$ presenta sub y sobre coberturas sistemáticas para los diferentes valores de p y n.

FIGURA 5

Intervalos de confianza e intervalos de credibilidad para una proporción

PC del intervalo bayesiano con a priori informativa para diferentes tamaños de muestra a un nivel nominal del 95%



5.2 COMPARACIÓN DE LA LONGITUD DE LOS INTERVALOS

La figura 6 muestra las longitudes de los intervalos *bayesianos con información previa*, como se consideró en la sección 5.1, y del intervalo de *Score*. Los intervalos de *bayesianos con información previa* tienen una longitud de intervalo menor que el intervalo *Score*. El intervalo con a priori $B(0.5,2)$ tiene la menor longitud. En la figura 7 se presenta la varianza de la longitud de los intervalos *bayesianos con información previa* y del intervalo de *Score*. El intervalo con a priori $B(0.5,2)$ presenta la mayor varianza de los tres y el intervalo con a priori $B(1.5,5)$, la menor.

FIGURA 6

Intervalos de confianza e intervalos de credibilidad para una proporción

Longitud esperada de los intervalos bayesianos con a priori informativa a un nivel nominal del 95%

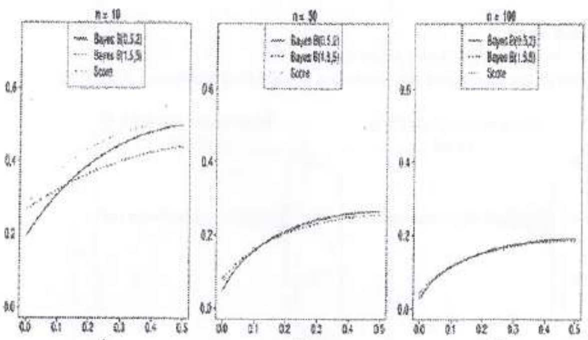
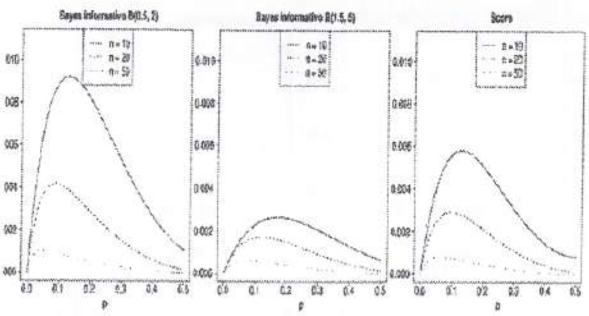


FIGURA 7

Intervalos de confianza e intervalos de credibilidad para una proporción

Varianza de la longitud de los intervalos bayesiano con a priori informativa a un nivel nominal del 95%



6. CONCLUSIONES

Se ha encontrado que el carácter discreto de las variables observadas provoca comportamientos inesperados en la obtención de IC de una proporción binomial [Agresti & Min, 2001] independientemente del método empleado para construirlos. Con respecto a la PC y a la longitud esperada de los intervalos, los resultados obtenidos en esta investigación coinciden con los reportados por [Agresti & Coull, 1998, Brown et al., 2001, Brown et al., 2002, Henderson & Meyer, 2001]. En cuanto a la varianza de longitud de los intervalos, la menor corresponde al intervalo de Wald ajustado. Luego, las correspondientes a los intervalos Score y bayesiano con distribución a priori uniforme. Después se encuentran las varianzas de la longitud de los intervalos Exacto y de Bayes con distribución a priori de Jeffreys. Finalmente, la mayor varianza corresponde a la longitud del intervalo deWald. En consecuencia, en ningún caso se recomienda el uso del intervalo de Wald, incluso en cursos básicos de estadística, donde es conveniente el uso del intervalo deWald ajustado, dado que es fácil de estimar, tiene la menor varianza entre todos los intervalos analizados, no presenta sub-cobertura respecto al valor nominal y tiene una formulación simple. Si la interpretación deseada para el grado de confianza del intervalo es que en promedio se tenga una cobertura de $100(1-\alpha)\%$, entonces la alternativa más recomendada está dada por el intervalo de *Score* y de *bayesiano* con distribución a priori uniforme, ya que es el que presentan mejor comportamiento en longitud esperada del intervalo, en su varianza y en probabilidad de cobertura que fluctúa alrededor del valor nominal sin importar los valores de n y p. Si además se tiene información previa a cerca del parámetro, la mejor propuesta es el intervalo *bayesiano* con distribución a priori. Si, además, se piensa que p está en el intervalo $(0,0.5]$, la mejor alternativa está dada por el intervalo *bayesiano con información previa* con distribución a priori $B(0.5,2)$ con n de un tamaño moderado y p no muy cercano a 0.

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

[1]. **AGRESTI, A. & COULL, B.** (1998). *"Approximate is better than "exact" for intervalo estimation of binomial proportions. The American Statistician".* 52:119–126.

[2]. **AGRESTI, A. & CAFFO, B.** (2000). *"Simple and effective confidence intervals for proportion and differences of proportions result from adding two successes and two failures. The American Statistician".* 54:280–288.

[3]. **BROWN, L., CAI, D. & DASGUPTA, A.** (2002). *"Confidence intervals for a binomial proportion and asymptotic expansions. The Annals of Statistics".* 30:160–201.

[4]. **CANAVOS, G.** (1988). *"Probabilidad y Estadística: Aplicaciones y métodos".* McGraw Hill, México, D.F.

[5]. **BROWN, L., CAI, D. & DASGUPTA, A.** (2001). *"Interval estimation for a binomial proportion. Statistical Science".* 16:101–133.

[6]. **BERNARDO, J. M. & SMITH, A. F.** (2000). *Bayesian theory.* John Wiley & Sons, Ltd., Chichester, England.

[7]. **R DEVELOPMENT CORE TEAM.** (2008). *"R: A language and environment for statistical computing".* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org. ISBN 3-900051-07-0.

[8]. **AGRESTI, A. & MIN, Y.** (2001). *"On small-sample confidence intervals for parameters in discrete distribution. Bometrics".* 57:963–971.

[9]. **HENDERSON, M. & MEYER, M.** (2001). *"Exploring the confidence interval for a binomial parameter in a first course in statistical computing. The American Statistician".* 55:337–344.

[10]. **CEPEDA, E., AGUILAR, W., CERVANTES, V., CORRALES, M., DÍAZ, I. & RODRÍGUEZ, D.** (2008). *"Intervalos de confianza e intervalos de confiabilidad para una proporción. Revista Colombiana de Estadística".* 30:211-228. http://www.estadistica.unal.edu.co/publicaciones/estadistica/rce/V31/bodyv31n2/v31n2a06.html.