

Intervalos de confianza e intervalos de credibilidad para una proporción

Confidence Intervals and Credibility Intervals for a Proportion

EDILBERTO CEPEDA-CUERVO^a, WILSON AGUILAR^b, VÍCTOR CERVANTES^c,
MARTHA CORRALES^d, IVÁN DÍAZ^e, DIANA RODRÍGUEZ^f

DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE
COLOMBIA, BOGOTÁ, COLOMBIA

Resumen

En este artículo se evalúa y se compara el comportamiento de diferentes metodologías empleadas para la obtención de intervalos de confianza 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 sobrecobertura del intervalo exacto, se determinan, entre otros aspectos, características de los intervalos relacionadas con la variabilidad de su longitud.

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

Abstract

Different methodologies for obtaining confidence and credibility intervals for a proportion are studied in this paper. Expected coverage, length and length variance of the interval are defined and used as a means for comparing the intervals produced by each methodology presented. These indicators were calculated using the statistical package R, used to characterize each interval; furthermore, some known properties, such as Wald interval's proportion undercoverage and Exact interval's overcoverage are verified in this study.

Key words: Proportions, Confidence intervals, Bayesian methods, Credibility interval, Coverage.

^aProfesor asociado. E-mail: ecepedac@unal.edu.co

^bEstudiante de maestría. E-mail: weaguilarl@unal.edu.co

^cEstudiante de maestría. E-mail: vhcervantesb@unal.edu.co

^dEstudiante de maestría. E-mail: mlcorralesb@unal.edu.co

^eEstudiante de maestría. E-mail: ildiazm@unal.edu.co

^fEstudiante de maestría. E-mail: dprodriguez@unal.edu.co

1. Introducción

Un problema estadístico habitual consiste en estimar los parámetros que ayudan a caracterizar una variable. El cálculo de intervalos de confianza para la estimación de parámetros es una técnica que permite hacer declaraciones sobre qué valores se pueden esperar de ese parámetro, lo que resulta útil para expresar la incertidumbre vinculada con los hallazgos sobre una estimación. Los intervalos de confianza dependen de la estimación del parámetro, obtenida a partir de la muestra, del tamaño muestral y del nivel de confianza seleccionado.

La estimación de la probabilidad de éxito de un experimento es un problema que ha sido tratado por muchos autores en diversos escritos, al igual que el cuestionable desempeño del intervalo clásico de Wald. De igual forma, muchos autores han planteado métodos alternativos para obtener estos intervalos de confianza. Agresti & Coull (1998) compararon el intervalo de *score* (Wilson) con el intervalo exacto y con los intervalos clásicos aproximados y argumentan que el intervalo de *score* (Wilson) es el mejor entre ellos. Agresti & Caffo (2000), basados en el artículo de Agresti & Coull (1998), muestran que el intervalo de Wald se mejora notablemente al adicionar pseudo observaciones a las muestras. Henderson & Meyer (2001) presentan un conjunto de intervalos de confianza frecuentistas e intervalos de credibilidad bayesianos con algunas elecciones a priori. Brown et al. (2002) mostraron que las propiedades del intervalo de Wald son pobres y realizan una comparación teórica de este con cuatro intervalos alternativos adicionales por expansiones asintóticas de sus probabilidades de cobertura: el intervalo de *score*, el intervalo de prueba de razón de verosimilitud, el bayesiano con a priori no informativa de Jeffreys y el intervalo corregido de Agresti y Coull. Ellos recomiendan usar el intervalo de Wilson o el intervalo de Jeffreys si el tamaño de muestra es pequeño; el de Agresti y Coull, si el tamaño de muestra es grande. También muestran que la probabilidad de cobertura de intervalo de Wald tiene un sesgo pronunciado y sistemático. Newcombe & Merino (2006) comparan diversos métodos para la construcción de intervalos de confianza para la estimación de proporciones y de las diferencias entre proporciones independientes.

A pesar de todo, aún existe la errónea creencia de que el intervalo de Wald presenta problemas solo cuando p está cerca de 0 ó 1, o cuando el tamaño de la muestra n es bastante pequeño. En diversos textos de enseñanza estadística recomiendan el intervalo de Wald cuando npq es mayor que 5 ó 10 (Canavos 1988). Sin embargo, en los artículos de Agresti & Coull (1998) y Brown et al. (2001), es manifiesto el mal comportamiento del intervalo de Wald, aun cuando estas condiciones son satisfechas. Ellos concluyeron que el intervalo de Wald debería ser usado únicamente cuando npq sea bastante grande.

El objetivo de este artículo es comparar 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 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 que van a compararse. 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 de confianza 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 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 intervalos de confianza 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

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

donde $z_{\alpha/2}$ es el cuantil $(1-\alpha/2)$ de la distribución normal estándar. Teóricamente, este intervalo tiene, para valores grandes de n , un nivel de confianza aproximado de $(1-\alpha)100\%$.

A pesar de lo simple que resulta su cálculo y de su gran difusión, muchos estudios han mostrado que este intervalo tiene un mal desempeño (Agresti & Coull 1998, Brown et al. 2002, Newcombe & Merino 2006). Newcombe & Merino (2006) muestran que si se sustituye $\hat{p}$ por cero en (1), se obtiene un error estándar igual a cero y el intervalo se degenera, ya que los límites inferior y superior son cero. De igual forma, cuando $\hat{p}$ es uno, los límites superior e inferior son 1. Aún más, cuando x es pequeño, el límite inferior puede ser menor que cero; similarmente, cuando $n-x$ es pequeño, el límite superior puede exceder de 1, en cuyo caso los intervalos obtenidos no tienen sentido. Correa & Sierra (2003), entre otros, señalan que los intervalos de Wald tienen una probabilidad de cobertura muy por debajo del nivel de confianza nominal, en especial cuando el tamaño de muestra es pequeño.

El intervalo de confianza exacto de Clopper-Pearson para una proporción p se basa en la inversión de una prueba binomial a dos colas iguales de la hipótesis $H_0 : p = p_0$ (Agresti & Coull 1998). 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

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

respectivamente. Para este intervalo, el límite inferior es 0 cuando $x = 0$ y el límite superior es 1 cuando $x = n$.

Es fácil mostrar (Brown et al. 2002) 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 $B(x+1, n-x)$. Este intervalo tiene probabilidad de cobertura mayor que o igual al nivel de confianza nominal; sin embargo, es muy conservador y su uso práctico solo es útil cuando esta característica es necesaria (Agresti & Coull 1998).

El intervalo de *score* (Agresti & Coull 1998, Newcombe & Merino 2006), también conocido como el intervalo de Wilson, se forma al invertir las aproximaciones del teorema del límite central para la familia de pruebas a dos colas de $H_0 : p = p_0$. Entonces, no se rechaza H_0 , con base en la aproximación por el teorema de límite central, si y solo si p_0 está en el intervalo. El intervalo de score tiene la forma

$$IC_{score} = \tilde{p} \pm \frac{z_{\alpha/2}}{\tilde{n}} \sqrt{\left(\tilde{p}\tilde{q} + \frac{z_{\alpha/2}^2}{4n} \right) / n} \quad (2)$$

donde $\tilde{x} = \hat{p} + \frac{z_{\alpha/2}^2}{2n}$, $\tilde{n} = 1 + z_{\alpha/2}^2/n$, $\tilde{p} = \frac{\tilde{x}}{\tilde{n}}$, $\hat{p} = \frac{x}{n}$ y $\hat{q} = 1 - \hat{p}$.

La probabilidad de cobertura del parámetro p por este intervalo está mucho más cercana al nivel de confianza nominal que la del intervalo exacto y que el intervalo de Wald (Agresti & Coull 1998), aun sin considerar el tamaño de la muestra ni las frecuencias o proporciones observadas (Newcombe & Merino 2006).

Agresti & Coull (1998) presentan una nueva propuesta de intervalo basada en una corrección del intervalo de confianza de Wald, conocido como el intervalo de Wald ajustado. El ajuste propuesto para estos intervalos está basado en la aproximación al intervalo *score* por el intervalo de Wald. En (1), el valor medio del intervalo se reemplaza por $\tilde{p}$; así, el intervalo queda definido como

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

donde $\tilde{x} = x + z_{\alpha/2}^2/2$, $\tilde{n} = n + z_{\alpha/2}^2$ 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 & Coull 1998). Agresti & Coull (1998) recomendaron la implementación de este intervalo ajustado en los cursos elementales de estadística debido a su sencillez y a que se comporta mejor que el intervalo de Wald (1) para los niveles usuales de confianza. Además, 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. El número de pseudo observaciones añadidas, algunas veces denotado por k^2 , varía de acuerdo con el nivel de confianza deseado para el intervalo. El valor k^2 empleado para el nivel de confianza de 90 % es $2.7 \approx 3$ pseudo observaciones; para 95 % es 4; y para 99 % es $5.4 \approx 5$.

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 $[c_1, c_2]$ (Migon & Gamerman 1999, Bernardo & Smith 2000).

2.2.1. Intervalos sin información previa

Históricamente, el uso de distribuciones a priori no informativas ha mostrado propiedades frecuentistas buenas (Wasserman 1991, citado por Newcombe & Merino 2006). En esta sección se consideran regiones de confianza para una proporción, considerando primero, como información a priori, una distribución $B(1, 1)$ y, luego, una a priori no informativa de Jeffreys.

En el primer caso, el intervalo bayesiano de dos colas iguales y de $1 - \alpha$ de credibilidad está dado por

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

donde $B(m_1, m_2, \alpha)$ denota el cuantil α de una distribución $B(m_1, m_2)$.

En el segundo caso, considerando la distribución a priori no informativa de Jeffreys, $B(1/2, 1/2)$, el intervalo de colas iguales y $1 - \alpha$ de credibilidad está dado por

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

Tanto el intervalo de Wilson (2) como el intervalo de Jeffreys tienen un excelente funcionamiento en términos de la probabilidad de cobertura media; la del intervalo de Jeffreys es ligeramente superior. La cobertura media del intervalo de Jeffreys está muy cercana al nivel nominal, aun para valores bastante pequeños de n (Brown et al. 2002).

2.2.2. Intervalos con información previa

En algunos estudios se tiene información a priori acerca de la proporción en consideración. Por ejemplo, existen líneas de producción automatizadas donde se tiene conocimiento experto que el margen de error en la línea de producción no supera 0.5 %. 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):

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

En la sección 5 se analizan dos propuestas de distribuciones a priori posibles cuando se conoce que el parámetro p es pequeño. La primera propone una distribución $B(0.5, 2)$; la segunda, una $B(1.5, 5)$. Estas dos distribuciones pertenecen a la clase conjugada de distribuciones a priori de la distribución binomial (Migon & Gamerman 1999, Bernardo & Smith 2000, Henderson & Meyer 2001).

3. Metodología

Al evaluar intervalos de confianza (o de credibilidad), se consideran dos conceptos importantes para determinar cuáles métodos son más eficaces: la longitud del intervalo, que indica su precisión, y la probabilidad de cobertura, definida teóricamente como $P(L_{inf} \leq p \leq L_{sup})$, donde L_{inf} y L_{sup} son variables aleatorias que indican, respectivamente, el límite inferior y el límite superior del intervalo. Un tercer elemento importante en la selección de estos métodos es la varianza de la longitud del intervalo, el cual es incluido en este artículo.

Tanto la probabilidad de cobertura (PC) como los límites de un intervalo son funciones del número x de éxitos observados en la muestra a través de los límites inferior y superior del intervalo. La probabilidad de cobertura para una proporción p es

$$PC = \sum_{x=0}^n I(x, p) P_p(x) \quad (7)$$

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 si $P_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$. Los límites del intervalo, presentados en la sección 2.2, están definidos de acuerdo con el método elegido y su longitud está dada por $L_{sup} - L_{inf}$. En todos los casos, la PC, los límites del intervalo y la longitud esperada del intervalo y su varianza son calculados mediante la ecuación

$$E_p(f(X)) = \sum_x f(x) P_p(x) \quad (8)$$

Con el fin de comparar los valores de la PC y de la longitud de cada uno de los intervalos expuestos, así como sus varianzas, se obtuvieron los $n + 1$ posibles

intervalos para cada tamaño de muestra desde 1 hasta 500 y se calcularon los valores esperados y las varianzas bajo los valores de $p = 0(0.01)1$. Esta metodología fue implementada a través del software estadístico R (R Development Core Team 2007).

4. Comparación de intervalos de confianza y de intervalos de credibilidad para una proporción

Para la comparación de los métodos usados para la estimación de intervalos, se determina el comportamiento de los intervalos propuestos, estableciendo en la sección 4.1 su probabilidad de cobertura y en la sección 4.2 el valor esperado y la varianza de su longitud. 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.

4.1. Comparación de la probabilidad de cobertura

Con el fin de observar el comportamiento de la probabilidad de cobertura de cada uno de los intervalos relacionados en la sección 2, se obtuvo su valor esperado para diferentes valores de p y de n , como se indica en la sección 3. En la figura 1 se representa esta probabilidad para $p = 0.01, 0.1, 0.3, 0.5$ y $n = 1, 2, 3, \dots, 500$; en la figura 2, la probabilidad de cobertura, como una función de p para $n = 10, 50$ y 100 en un nivel nominal de 95 %. Se observa que la probabilidad de cobertura del intervalo de Wald está muy lejana de la nominal para todos los tamaños de muestra y que, aun con tamaños muy grandes de la muestra, esta probabilidad es siempre menor que el valor nominal.

De la figura 1, se obtienen las siguientes conclusiones. El intervalo exacto tiene probabilidades de cobertura por encima del nivel nominal para todos los tamaños de muestra. El intervalo de Wald ajustado presenta probabilidades de cobertura mucho más cercanas al valor nominal que el intervalo de Wald y que el exacto. Para valores de p cercanos a 0 o a 1, esta probabilidad de cobertura es mayor que el valor nominal.

Puede observarse también, en las figuras 1 y 2, que el intervalo de *score* presenta una probabilidad de cobertura que fluctúa 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 probabilidad de cobertura respecto al valor nominal cuando el tamaño de muestra es pequeño, $n = 10$. Para tamaños de muestra mayores, las fluctuaciones en la probabilidad de cobertura son muy similares a las presentadas por el intervalo de *score*. 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 $n = 50$. En tamaños de muestra grandes, presenta fluctuaciones más cercanas al valor nominal que todos los demás intervalos.

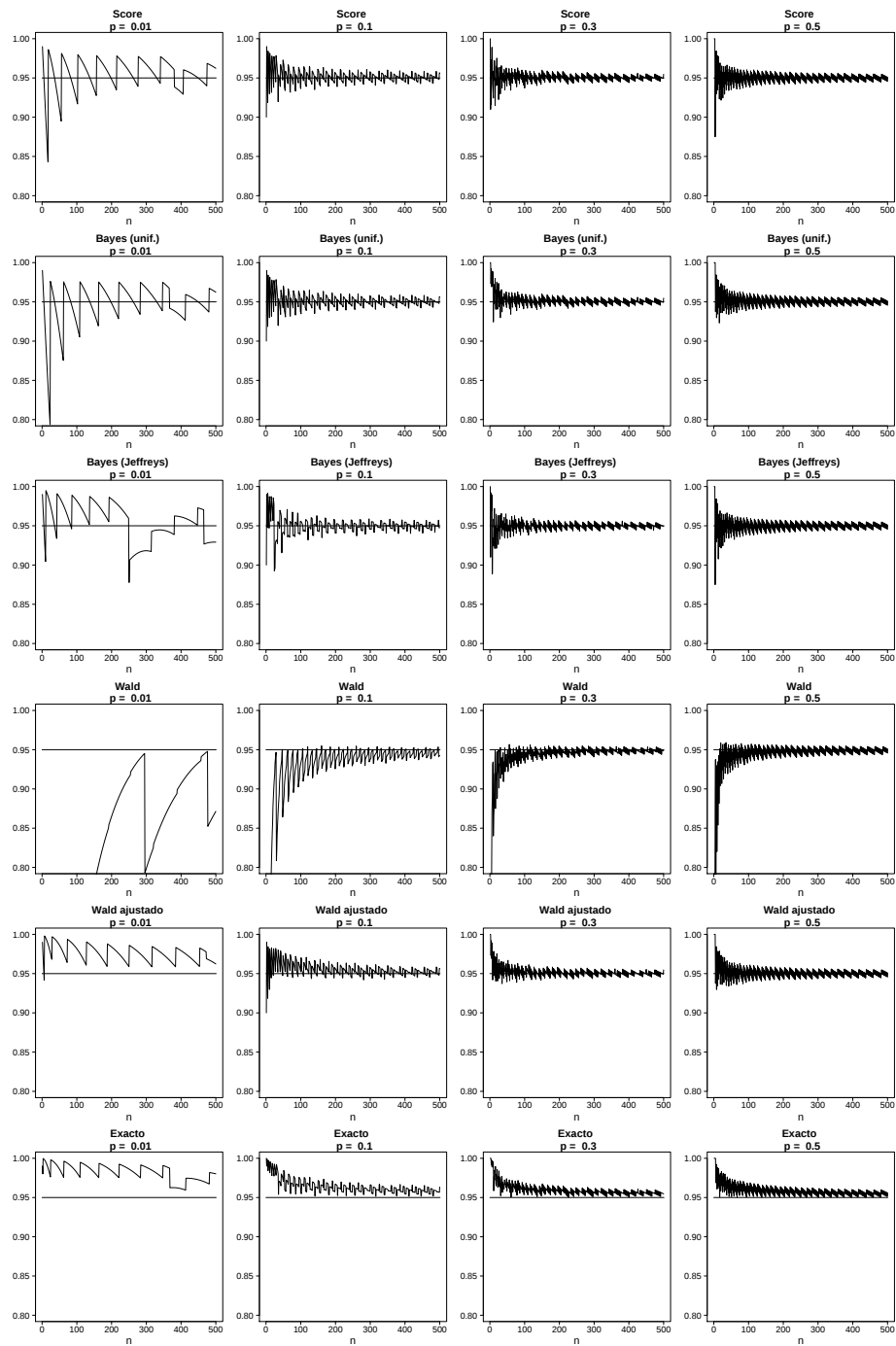


FIGURA 1: Probabilidad de cobertura de los intervalos de confianza para distintos tamaños de muestra y valores de p en un nivel nominal de 95 %.

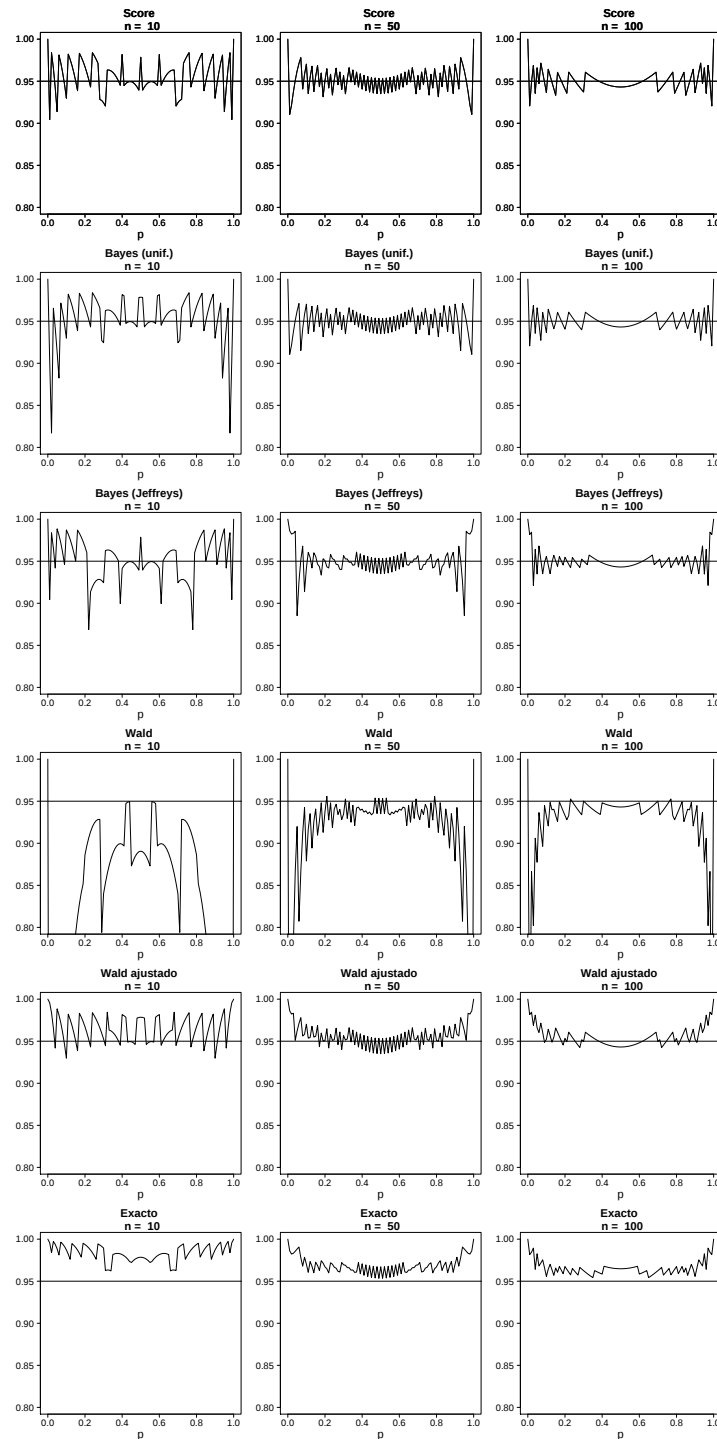


FIGURA 2: Probabilidad de cobertura de los intervalos de confianza para diferentes tamaños de muestra en un nivel nominal de 95 %.

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 0 y a 1. Sin embargo, las tendencias señaladas previamente para cada uno de los intervalos se conservan.

La figura 3 representa la probabilidad de cobertura como una función de p y del tamaño de la muestra. En ella se resume y complementa, para cada uno de los intervalos analizados, el comportamiento de la probabilidad de cobertura observado en las figuras 1 y 2.

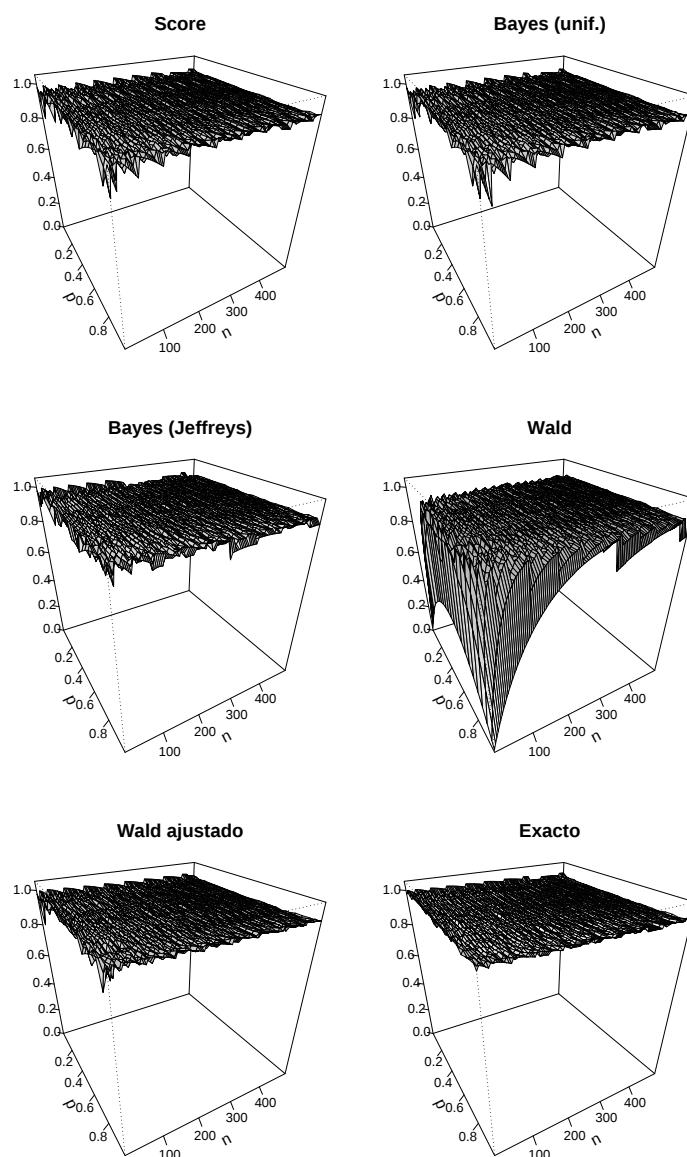


FIGURA 3: Probabilidad de cobertura de los intervalos analizados variando p y n .

4.2. Comparación de la longitud de los intervalos de confianza y de los intervalos de confiabilidad para una proporción

Lo ideal para obtener estimaciones de intervalos de confianza (o de credibilidad) es determinarlas a partir de propuestas que tengan probabilidad de cobertura próxima a la nominal, longitud esperada más corta y varianza de la longitud del intervalo más pequeña.

En la figura 4 la longitud esperada de los intervalos analizados se representa como función del parámetro p para distintos tamaño de muestra. Los intervalos de Wald, exacto y Wald ajustado presentan un comportamiento diferente al de los intervalos de *score* y bayesianos. Esta diferencia es clara para tamaños de muestra pequeños, $n = 10$ y $n = 50$, donde se observa que los intervalos de Wald ajustado y exacto tienen longitud mayor que la de los demás intervalos, y que 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 $(0, 1)$.

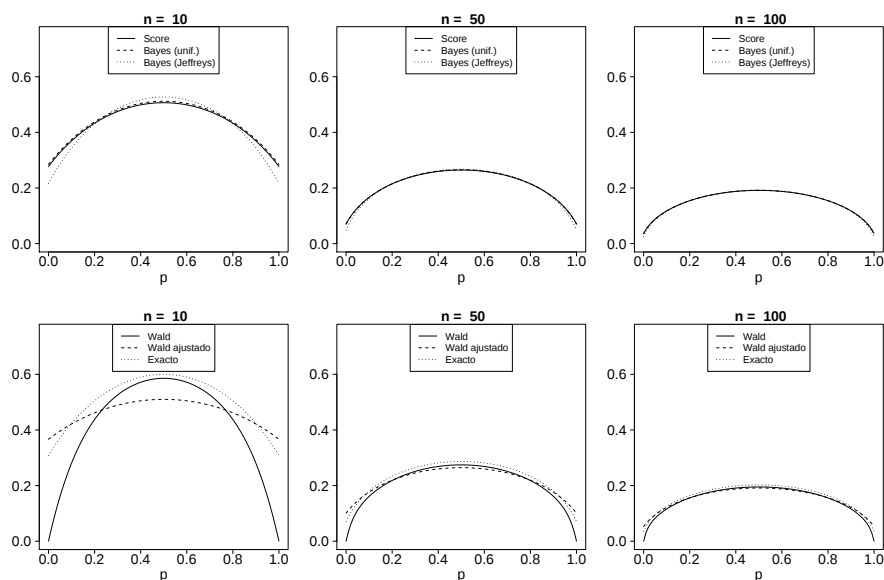


FIGURA 4: Longitud esperada de los intervalos para distintos tamaños de muestra y distintos valores de p en un nivel nominal de 95 %.

Los intervalos *score* y bayesianos presentan longitudes muy similares entre sí para cualquier valor de p y de n . Cuando n es igual o mayor que 50, los valores esperados de las longitudes de todos los intervalos se hacen similares, excepto para el intervalo de Wald, cuya longitud tiende a 0 cuando p tiende a 0 o a 1.

La figura 5 muestra el intervalo esperado, como una función de p , para cada uno de los seis casos considerados en este estudio. En todos los casos, se consideraron tamaños de muestra $n = 10, 50, 100$. Se observa con claridad el comportamiento del

intervalo esperado a través de los valores posibles de p y a medida que el tamaño de la muestra se hace mayor.

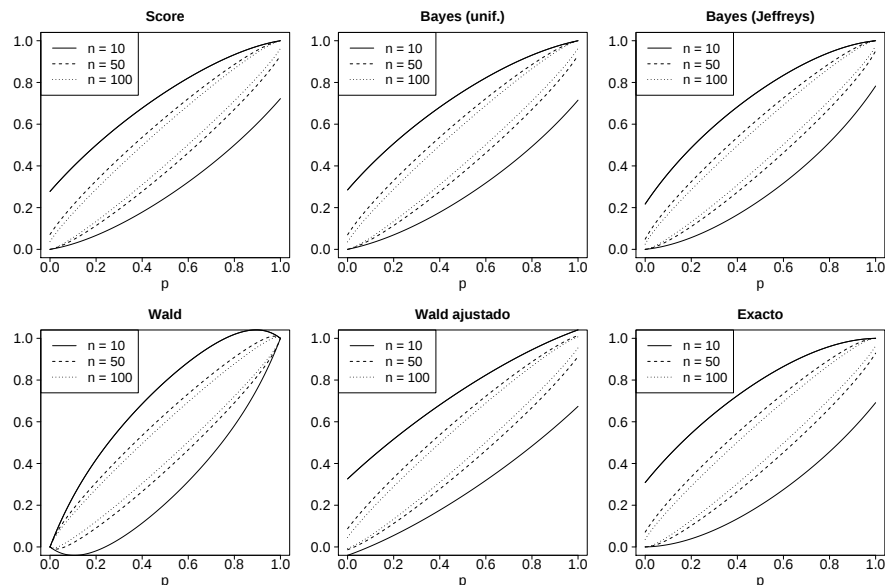


FIGURA 5: Intervalos de confianza esperados en un nivel nominal de 95 %.

En la figura 6 se muestra la varianza de la longitud de los intervalos. 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; lo mismo ocurre a los intervalos exacto y bayesiano con a priori de Jeffreys. El intervalo de Wald presenta la mayor variabilidad entre todos los intervalos presentados.

5. Intervalos de credibilidad con información previa

5.1. Comparación de la probabilidad de cobertura

En la figura 7 se presenta el comportamiento de la probabilidad de cobertura 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*, para $n = 1, 2, \dots, 500$ y $p = 0.01, 0.1, 0.3, 0.5$; en la figura 8, para distintos valores de $p \in (0, 0.5)$ y $n = 10, 50, 100, 500$. Puede verse que la fluctuación del valor de cobertura es menor en los intervalos bayesianos con información previa siempre que p no esté muy próximo a 0, $p \sim 0.01$, o que n sea grande, $n \sim 500$. Sin embargo, para 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 .

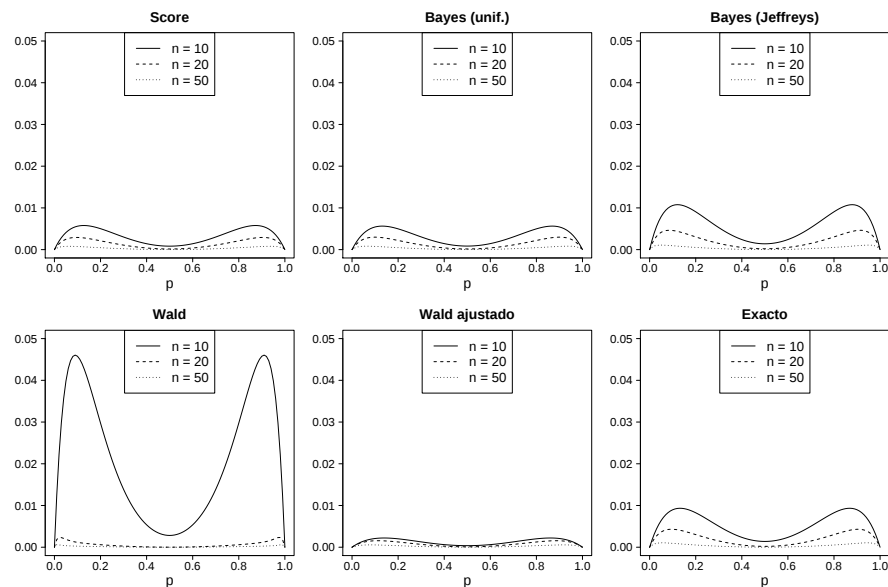


FIGURA 6: Varianza de la longitud del intervalo en un nivel nominal de 95 %.

5.2. Comparación de la longitud de los intervalos

Las figuras 9 y 10 muestran las longitudes de los intervalos bayesianos con información previa, como se consideró en la sección 5.1, y del intervalo de *score*. Los intervalos 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 11 se presenta la varianza de la longitud de los intervalos bayesianos con información previa, como se consideró en la sección 5.1, y del intervalo de *score*. El intervalo con a priori $B(0.5, 2)$ presenta la mayor varianza de los tres; el intervalo con a priori $B(1.5, 5)$, la menor.

6. Otros intervalos de confianza para una proporción

6.1. Intervalo logit

El intervalo *logit* para una proporción p está definido por la expresión

$$IC_{Logit} = \left[\frac{e^{\lambda_I}}{1 + e^{\lambda_I}}, \frac{e^{\lambda_S}}{1 + e^{\lambda_S}} \right]$$

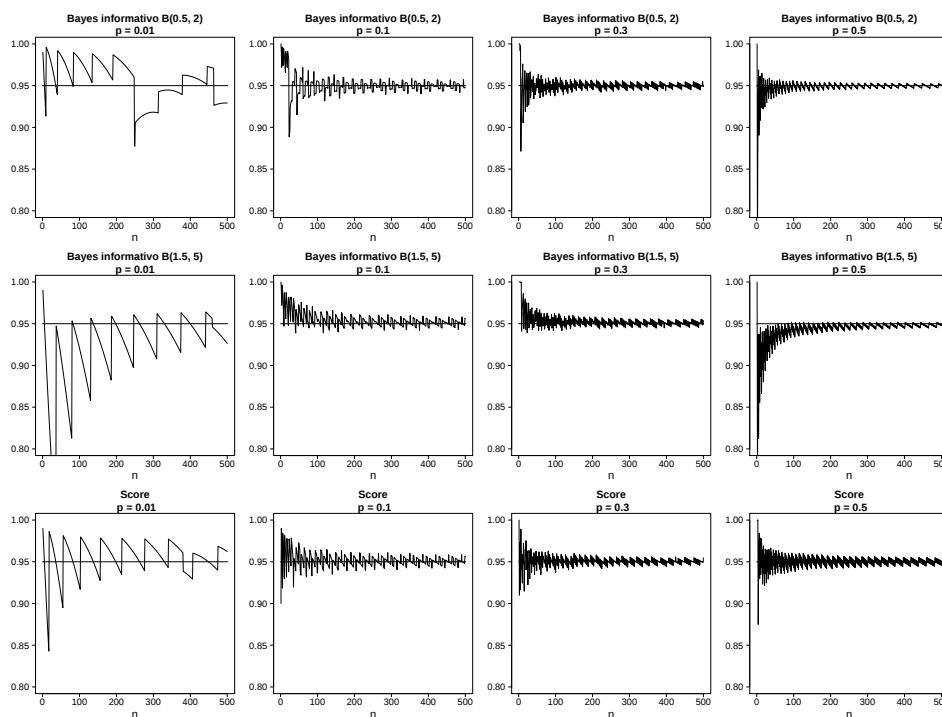


FIGURA 7: Probabilidad de cobertura del intervalo de confianza bayesiano con a priori informativa para diferentes tamaños de muestra en un nivel nominal de 95 %

donde $\lambda_I = \hat{\lambda} - Z_{1-\alpha/2} \hat{V}^{\frac{1}{2}}$ y $\lambda_I = \hat{\lambda} + Z_{1-\alpha/2} \hat{V}^{\frac{1}{2}}$. Se obtiene invirtiendo el intervalo de confianza IC_λ del logaritmo de la razón de probabilidades $\lambda = \log\left(\frac{p}{1-p}\right)$,

$$IC_\lambda = \left[\hat{\lambda} - Z_{1-\alpha/2} \hat{V}^{\frac{1}{2}}, \hat{\lambda} + Z_{1-\alpha/2} \hat{V}^{\frac{1}{2}} \right]$$

donde

$$\begin{aligned} \hat{\lambda} &= \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \log\left(\frac{X}{n-X}\right) \\ \hat{V} &= Var(\hat{\lambda}) = \frac{n}{X(n-X)} \end{aligned}$$

Este intervalo se comporta muy bien en términos de cobertura de p lejos de 0 y de 1, pero su longitud esperada es aún mayor que la longitud del intervalo de Clopper-Pearson (Brown et al. 2002), la cual se muestra en la figura 4.

6.2. Intervalo arc sen

El intervalo arc sen para una proporción p está definido por la expresión

$$IC_{arcsine} = \left[\sin^2\left(\arcsin\left(\hat{p}^{\frac{1}{2}}\right) - \frac{1}{2}Z_{\alpha/2}n^{-\frac{1}{2}}\right), \sin^2\left(\arcsin\left(\hat{p}^{\frac{1}{2}}\right) + \frac{1}{2}Z_{\alpha/2}n^{-\frac{1}{2}}\right) \right]$$

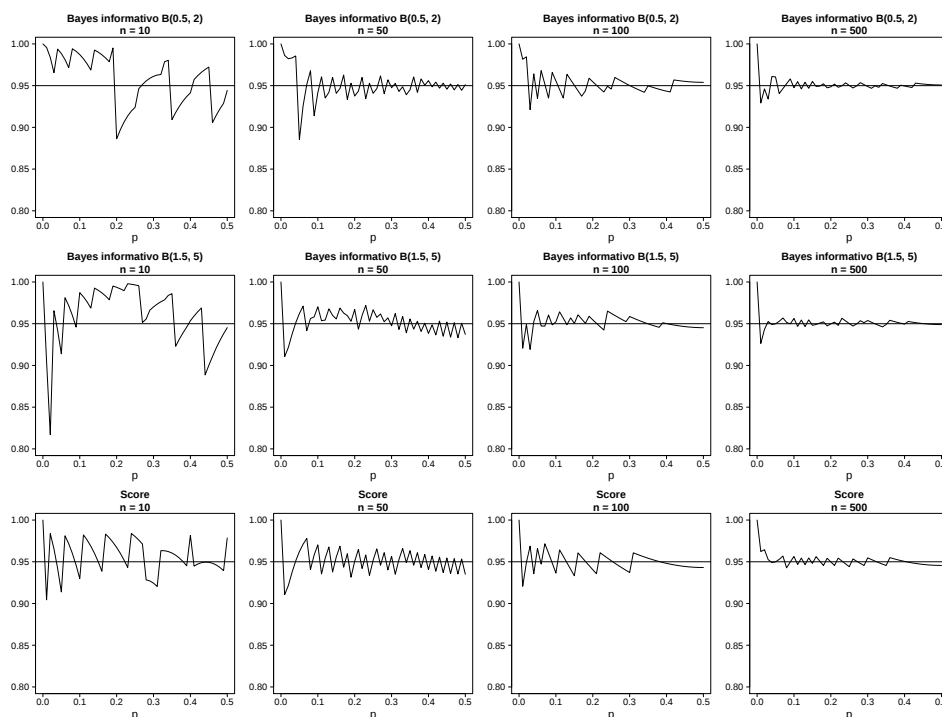


FIGURA 8: Probabilidad de cobertura del intervalo de confianza bayesiano con a priori informativa a lo largo de $(0, 0.5]$ en un nivel nominal de 95 %.

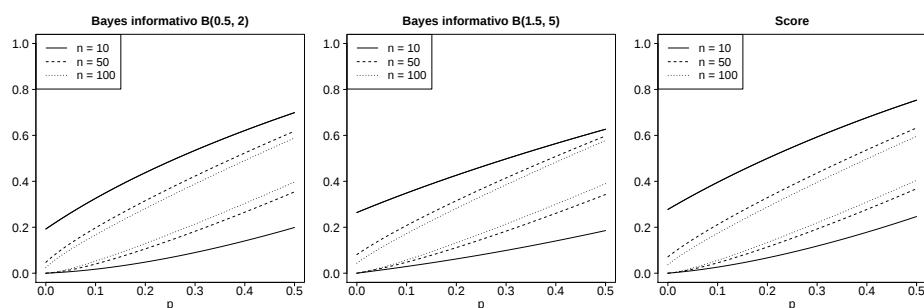


FIGURA 9: Intervalos bayesianos esperados con a priori informativa en un nivel nominal de 95 %.

donde $\hat{p} = X/n$. Se obtiene invirtiendo el intervalo de confianza IC_δ de $\delta = g(p) = \arcsen\left(p^{\frac{1}{2}}\right)$:

$$IC_\delta = \left[\arcsen\left(\hat{p}^{\frac{1}{2}}\right) - \frac{1}{2}Z_{\alpha/2}n^{-\frac{1}{2}}, \quad \arcsen\left(\hat{p}^{\frac{1}{2}}\right) + \frac{1}{2}Z_{\alpha/2}n^{-\frac{1}{2}} \right]$$

donde $\hat{\delta} = g(\hat{p}) = \arcsen\left(\hat{p}^{\frac{1}{2}}\right)$ y $Var(\hat{p}) = [g'(\hat{p})]^2 Var(p) = 1/4n$.

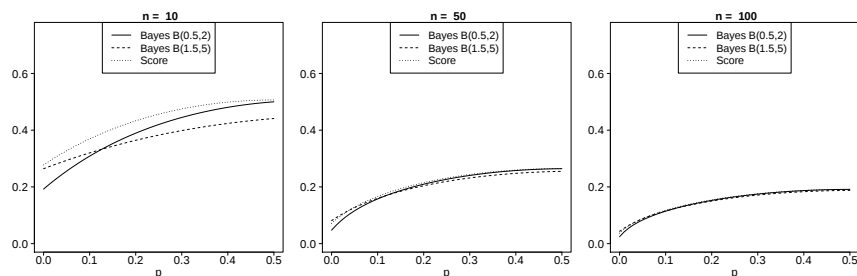


FIGURA 10: Longitud esperada de los intervalos bayesianos con a priori informativa en un nivel nominal de 95 %.

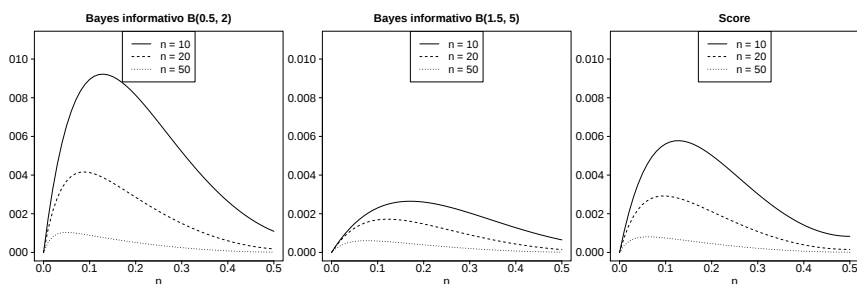


FIGURA 11: Varianza de la longitud de los intervalos bayesianos con a priori informativa en un nivel nominal de 95 %.

Este intervalo se comporta razonablemente bien para valores de p no próximos a 0 o a 1, pero la probabilidad de cobertura está próxima a 0 para valores de p próximos a estos valores; presenta un comportamiento peor que el del intervalo de Jeffreys.

6.3. Intervalo basado en verosimilitud relativa

Este intervalo se construye a partir de la prueba de razón de verosimilitud que acepta $H_0 : p = p_0$ si $-2 \log(\Lambda) \leq k^2$, donde

$$\Lambda_n = \frac{L(p_0)}{\sup_p L(p)} = \frac{p_0^X (1 - p_0)^{n-X}}{(X/n)^X (1 - X/n)^{n-X}}$$

donde L es la función de verosimilitud.

Brown et al. (2002) muestran que este intervalo tiene buenas propiedades. Sin embargo, no es fácil de calcular.

7. Conclusiones

Se ha encontrado que el carácter discreto de las variables observadas provoca comportamientos inesperados en la obtención de intervalos de confianza de una

proporción binomial, independientemente del método empleado para construirlos (Agresti & Min 2001).

Con respecto a la probabilidad de cobertura 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) y 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 de Wald. 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 de Wald ajustado, dado que es fácil de estimar, tiene la menor varianza entre todos los intervalos analizados, no presenta subcobertura 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 bayesiano con distribución a priori uniforme, ya 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 acerca de los parámetros, la mejor propuesta es el intervalo bayesiano con distribución a priori. Si 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 $B(0.5, 2)$, n de un tamaño moderado y p no muy cercano a 0.

[Recibido: enero de 2008 — Aceptado: septiembre de 2008]

Referencias

- 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.
- Agresti, A. & Coull, B. (1998), 'Approximate is Better than Exact for Interval Estimation of Binomial Proportions', *The American Statistician* **52**, 119–126.
- Agresti, A. & Min, Y. (2001), 'On Small-Sample Confidence Intervals for Parameters in Discrete Distribution', *Biometrics* **57**, 963–971.
- Bernardo, J. M. & Smith, A. F. M. (2000), *Bayesian Theory*, John Wiley & Sons, Ltd., Chichester, England.
- Brown, L., Cai, D. & DasGupta, A. (2001), 'Interval Estimation for a Binomial Proportion', *Statistical Science* **16**, 101–133.

- Brown, L., Cai, D. & DasGupta, A. (2002), 'Confidence Intervals for a Binomial Proportion and Asymptotic Expansions', *The Annals of Statistics* **30**, 160–201.
- Canavos, G. (1988), *Probabilidad y Estadística: Aplicaciones y métodos*, McGraw Hill, México D. F., México.
- Correa, J. & Sierra, E. (2003), 'Intervalos de confianza para la comparación de dos proporciones', *Revista Colombiana de Estadística* **26**, 61–75.
- 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.
- Migon, H. & Gamerman, D. (1999), *Statistical Inference: An Integrated Approach*, Oxford University Press Inc., Madison Avenue, United States.
- Newcombe, R. & Merino, C. (2006), 'Intervalos de confianza para las estimaciones de proporciones y las diferencias entre ellas', *Interdisciplinaria* **23**, 141–154.
- R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org>
- Wasserman, L. (1991), An Inferential Interpretation of Default Priors, Technical report, Dept. Statistics, Carnegie Mellon Univ.