

**PEDECIBA Informática**  
**Instituto de Computación – Facultad de Ingeniería**  
**Universidad de la República**  
**Montevideo, Uruguay**

---

---

## **Reporte Técnico RT 06-19**

---

---

**Sobre el correcto uso de intervalos de confianza  
para la estimación de proporciones binomiales**

**Javier Cohenar González**

**2006**

Sobre el correcto uso de intervalos de confianza para la estimación de proporciones binomiales

Cohenar González, Javier

ISSN 0797-6410

Reporte Técnico RT 06-19

PEDECIBA

Instituto de Computación – Facultad de Ingeniería

Universidad de la República

Montevideo, Uruguay, 2006

# SOBRE EL CORRECTO USO DE INTERVALOS DE CONFIANZA PARA LA ESTIMACIÓN DE PROPORCIONES BINOMIALES

Javier Cohenar González

Departamento de Investigación Operativa -INCO  
Facultad de Ingeniería  
Universidad de la República  
Uruguay

## ABSTRACT

*En este breve reporte técnico se presentan valores numéricos del desempeño del método clásico de estimación de proporciones binomiales aplicados al problema de la confiabilidad diámetro acotada, conjuntamente con resultados similares y análisis teóricos obtenidos de una búsqueda bibliográfica. Dicha búsqueda mostró que la estimación de las proporciones antes mencionadas fue retomado recientemente debido a resultados experimentales que indicaban un desempeño del método estándar inferior con respecto a lo usualmente pensado. Dado lo extendido de la aplicación del método estándar, tanto en el cálculo científico, la enseñanza o en la práctica profesional, se incluyen brevemente dos alternativas al intervalo clásico para cubrir las distintas necesidades existentes.*

## 1. INTRODUCCIÓN

En los últimos años las redes peer-to-peer (P2P) han ido ganando relevancia debido a su popularidad en ambientes como Internet. Tanto a nivel de diseño como de operación de estas redes, un objetivo fundamental es alcanzar una alta probabilidad de conexión entre los nodos manteniendo la carga de la red controlada.

En el contexto de estas redes se define el problema de la medida de la Confiabilidad Diámetro Acotada (CDA) [12], esto es, dada una red representada por un grafo  $\mathbf{G}$  cuyas aristas no dirigidas están sujetas a fallas con probabilidad  $r_u$  (siendo  $u$  una arista del grafo), un subconjunto  $\mathbf{K}$  de nodos del grafo (también llamado conjunto de nodos terminales) y un número entero positivo  $\mathbf{D}$  o diámetro, se denomina Confiabilidad Diámetro Acotada a la probabilidad de que haya al menos un camino de largo máximo  $\mathbf{D}$  para cada par de nodos terminales.

Siendo la CDA nuestra medida principal a estudiar, es de fundamental importancia conocer el comportamiento de las distintas herramientas disponibles para calcularla o estimarla. Una dificultad que surge al momento de estudiar este problema es la ausencia de algoritmos que puedan calcular la CDA de manera exacta en tiempos prácticos razonables cuando la cantidad de nodos o aristas es grande, es por ello que se recurre a métodos estadísticos aproximados para su evaluación. En particular, se pueden utilizar técnicas como simulación por el método MonteCarlo Crudo (MCC) e Intervalos de Confianza (IC) para la estimación. Sin embargo, dichos IC son válidos bajo ciertas hipótesis que pueden o no cumplirse en la práctica y por tanto es de interés conocer cómo es afectada la capacidad de predicción de los IC en distintas instancias de este problema y con distintos parámetros de la simulación (por ejemplo, largo de una réplica).

Para evaluar el desempeño de los IC se emplean sobre grafos cuya CDA se conoce a priori las llamadas Funciones de Cobertura (FC)[8] que, cuando las hipótesis de los IC se cumplen, deberían devolver valores distribuidos uniformemente. Por lo tanto, evaluando el grado o medida con que la FC se ajusta a dicha distribución a través del p-valor de un test de ajuste (por ejemplo el test de Kolmogorov-Smirnov) se obtendría una caracterización del desempeño de los IC escogidos para distintas condiciones de la simulación e instancias del problema. Esto a su vez, brindará información sobre los parámetros adecuados para una simulación en este tipo de contextos.

En el caso de la CDA específicamente, valores pequeños de ésta son esperables. A su vez, diversos textos advierten (a veces de manera poco precisa) sobre lo inadecuado del método clásico de estimación (intervalo de Wald) para ciertos valores o ciertas condiciones de la CDA, por lo que resulta de interés entender exactamente el alcance de esta forma de estimación.

Los resultados obtenidos mostraron un comportamiento poco satisfactorio del método clásico de estimación (intervalo de Wald) lo cual motivó una búsqueda en la literatura especializada de resultados similares o IC alternativos. Al parecer, el problema de la estimación de una proporción binomial fue retomado en años recientes debido a la constatación del comportamiento insatisfactorio antes mencionado del IC de Wald, en especial [10] muestra que dicho comportamiento es peor de lo que hasta el momento se había anticipado proponiendo una serie de IC alternativos.

Este reporte está organizado de la siguiente manera: en la sección 2. se introducen los conceptos utilizados a lo largo del reporte, en especial aquellos relacionados con los métodos de estimación de proporciones binomiales y la función de cobertura; en la sección 3. se describen los experimentos realizados (3.1) y los resultados obtenidos para el intervalo estándar (3.2, 3.3 y 3.4); en 3.5 se discuten las propiedades de algunos intervalos alternativos ante la constatación del mal desempeño del intervalo estándar; finalmente, en 4 se brindan las conclusiones y un resumen final.

## 2. CONCEPTOS PRELIMINARES Y CONTEXTO DEL TRABAJO

### 2.1 Método de MonteCarlo

El algoritmo básico de Monte Carlo crudo aplicado al problema de la confiabilidad de redes, consiste en generar aleatoriamente distintos estados de una red dada verificando en cada iteración si la condición de conectividad entre los nodos terminales se cumple. En caso afirmativo, se registra el caso como un experimento exitoso. Al finalizar la ejecución se tendrá información sobre la cantidad de experimentos exitosos en relación a la cantidad total de realizados y por lo tanto, una estimación de la confiabilidad diámetro acotada.

El esquema básico del algoritmo empleado es el sugerido en [3], y que aquí se presenta utilizando una notación similar.

1. Inicialización :  $X = 0$
2. Para cada experimento  $i = 1, \dots, n$ 
  - 2.1 Para cada arista  $u = 1, \dots, |E|$   
sortear  $U$  de Uniforme(0,1)  
Si  $(U \in [0, r_u])$  entonces  $x_u = 1$  sino  $x_u = 0$
  - 2.2 Evaluar la función de estructura  $\Phi(\chi)$  y hacer  $X = X + \Phi(\chi)$
3. Calcular el estimado de la media  $\hat{X} = X/n$
4. Calcular el estimado de la varianza  $\hat{V} = \hat{X}(1-\hat{X})/(n-1)$

donde

$|E|$  representa el cardinal del conjunto de ejes  $E$

$x_u$  indica si la arista  $u$  se encuentra activa (1) o no (0)

$\Phi(\chi)$  evalúa si se cumple la condición de conectividad en el grafo  $\chi$  (vale 1 o 0)

## 2.2 Intervalos de confianza y la función de cobertura.

Para el caso de la confiabilidad de redes, las variables aleatorias calculadas siguen una distribución de Bernoulli con parámetro CDA. Por lo tanto la media y la varianza estimadas en cada replicación son

Por razones de practicidad y coherencia con la notación habitual en estadística, la CDA también se denotará  $p$ , siendo  $\hat{CDA}$  o  $\hat{p}$  sus estimaciones empíricas.

Obsérvese que los resultados  $X_i$  están distribuidos de acuerdo a una distribución Bernoulli de parámetro CDA por lo que la distribución de  $p$  es binomial de parámetro CDA.

Cuando las muestras son iid se pueden obtener intervalos de confianza según [7]:

$$I = \left[ \hat{p} - z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}, \quad \hat{p} + z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right]$$

donde  $z_{\alpha/2}$  representa al valor de la función normal inversa evaluada en  $\alpha/2$ . En diversos libros de texto introductorios en estadística se brindan algunas condiciones que deben cumplirse para que el intervalo tenga validez. Aquí se resumen algunas de ellas, según son tratadas en [10]:

- i.  $np, (1-p)n \geq 5$  (o 10)*
- ii.  $np(1-p) \geq 5$  (o 10)*
- iii.  $\hat{p}, (1-\hat{p})n \geq 5$  (o 10)*
- iv.  $\hat{p} \pm 3\sqrt{\hat{p}(1-\hat{p})/n}$  no contenga ni 0 ni 1*
- v.  $n$  grande*
- vi.  $n \geq 50$  pero  $p$  no muy pequeño*

Como se muestra en [10] estas condiciones “heurísticas” no aseguran el correcto desempeño del intervalo de Wald (más aún, alguna de ellas, por ejemplo la condición v, resultan ambiguas o no verificables como es el caso de *i* y *ii*. Como se verá más adelante los experimentos realizados para este informe tienden a verificar las apreciaciones de [10].

En este informe se utilizan además las llamadas funciones de cobertura [8]. Siguiendo la presentación hecha en [8], todo intervalo de confianza se puede representar mediante una región del espacio que depende de los valores experimentales a partir de los cuales fue calculada y un nivel de confianza  $\eta$ , esto es,  $R(\eta, X)$  donde  $X$  representa el conjunto de datos experimentales.

Entonces se define  $\eta^*$  como el mínimo valor de  $\eta$  que asegura que el parámetro a estimar  $\theta$  está contenido en la región  $R(\eta, X)$ . Formalmente,

$$\eta^* = \inf\{\eta \in [0,1]; \theta \in R(\eta, X)\}.$$

Cuando la estimación se realiza bajo hipótesis verdaderas,  $\eta^*$  estará uniformemente distribuido, esto es,

$$\text{Prob}(\eta^* \leq \eta) = \text{Prob}\{\theta \in R(\eta, X)\} = \eta.$$

De forma similar pero a nivel empírico se puede definir  $G(\eta^*)$  como

$$G(\eta^*) = \dots$$

la cual puede ser testeada con un test de ajuste para verificar su uniformidad. En este contexto  $\eta_j^*, j=1,2,\dots,m$  son los valores observados de  $\eta^*$  para cada intervalo  $j$  dados una muestra de  $m$  intervalos. Para este reporte se utilizó el test de Kolmogorov-Smirnov (KS) [9] para la prueba de uniformidad.

El test de Kolmogorov-Smirnov es un test estadístico que, en su versión original, se aplica para determinar si un conjunto de variables aleatorias provienen de una distribución continua determinada. Dado el resultado de un test estadístico, una medida del grado de seguridad en las conclusiones del mismo viene dado por el llamado p-valor, el cual indica en valores de confianza el punto en el cual se comienza a rechazar la hipótesis nula (en nuestro caso la hipótesis nula es la suposición de que los datos provienen de una distribución uniforme) conforme se incrementa el nivel de significancia del test. Por lo tanto, cuanto más bajo el p-valor más seguridad tenemos de las conclusiones extraídas del test.

### 3. ANALISIS DE LOS DATOS OBTENIDOS

#### 3.1 NOTACION EMPLEADA Y DESCRIPCIÓN DE LOS EXPERIMENTOS REALIZADOS

##### 3.1.1 Notación y Experimentos realizados

La etapa de experimentación consistió en generar  $m$  simulaciones de largo  $n$ .

Todos los experimentos realizados están basados en simulaciones de Monte Carlo crudo de  $m$  réplicas (o intervalos) de largo  $n$ . Cada replicación tiene asociado un valor  $\eta_j^*$  y para cada CDA existe un p-valor del test KS que denominaremos  $p_v(CDA, m, n)$ . La CDA se calculó utilizando un diámetro ( $D$ ) de 2 y el conjunto de nodos terminales igual al conjunto de nodos del grafo.

Para este reporte se utilizaron dos tipos de grafos los cuales son ilustrados en la Figura 3.1, Figura 3.2 y los valores de su CDA dada una confiabilidad de arista y diámetro. Cabe recalcar que los resultados de los experimentos no dependen de la forma del grafo ni del diámetro elegido, sino de su CDA puesto que nuestro problema se reduce siempre a la estimación de la media de una variable Binomial de parámetro CDA.

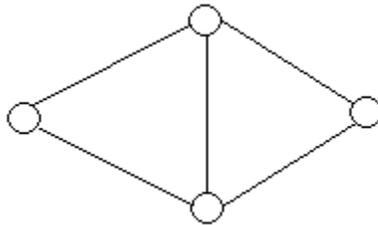


Figura 3.1. Puente



Figura 3.2. Arista simple

ra	CDA
0,05	0,000256
0,1	0,00208
0,2	0,01696
0,3	0,05724
0,4	0,13312
0,5	0,25
0,6	0,40608
0,7	0,58996
0,8	0,77824
0,83	0,830349
0,85	0,862846
0,9	0,93312
0,95	0,981694

Tabla 3.1. Equivalencias entre confiabilidad de arista y CDA para grafo de Figura 3.1

El algoritmo para la evaluación de la CDA fue implementado en lenguaje C y la evaluación de la existencia de caminos que comuniquen a todos los puntos del conjunto de terminales se realizó mediante una adaptación del algoritmo de Dijkstra. El generador de números aleatorios utilizado fue el Mersenne Twister [5][6]. Para el cálculo de la función de cobertura se utilizó la biblioteca DCDFLIB [4].

La información de los grafos sobre los cuales se realizan los cálculos es leída en el formato GXL mediante la biblioteca Libxml2 [1]. Dicha biblioteca fue desarrollada para el Proyecto Genoma Humano y permite –entre otros- la lectura y edición de archivos XML. El formato de los grafos GXL (Graph eXchange Language, subformato de XML) fue diseñado para el intercambio de información a través de grafos de manera estándar [2].

### 3.1.2 Verificaciones y validaciones efectuadas

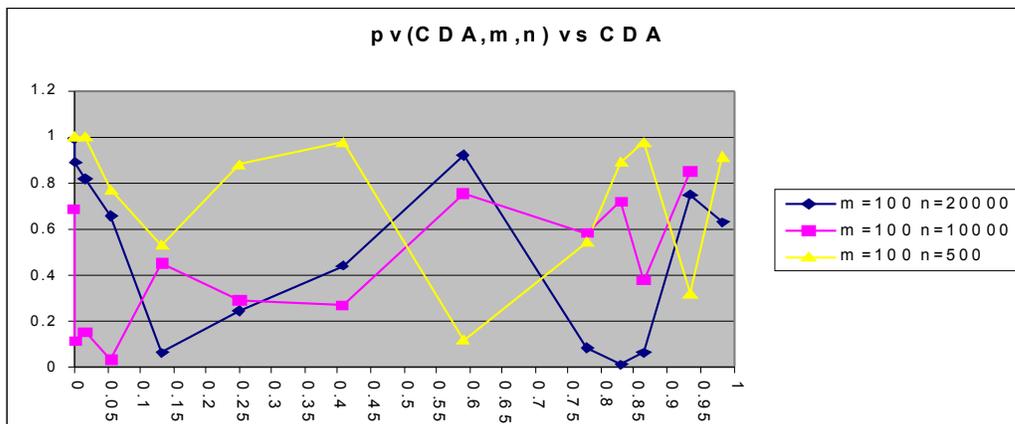
Para un número reducido de experimentos se realizó el test de rachas (“up and down runs test”)[7] al conjunto de datos conformados por la CDA estimada en cada replicación del método para una confiabilidad dada, no pudiéndose encontrar evidencia que sugiera que dichas muestras no constituyen una muestra aleatoria simple. Este test fue hecho con el ánimo de encontrar posibles errores en el manejo del generador de números aleatorios, por ejemplo, semillas repetidas o solapamiento de los streams.

### 3.2 ESTUDIO DE LA UNIFORMIDAD DE LA FUNCIÓN DE COBERTURA

De los resultados obtenidos se desprende que

- a) los p-valores hallados en el test de KS muestran un comportamiento oscilatorio conforme varía la confiabilidad de arista;
- b) contrariamente a lo que cabría esperar, el aumentar el número de réplicas o la longitud de las mismas a partir de determinado punto no brinda un mejor desempeño del intervalo;
- c) las distribuciones obtenidas para la función de cobertura se apartan fuertemente de la distribución uniforme;

Los resultados obtenidos se ilustran en la Gráfica 3.1 y Gráfica 3.2



GRAFICA 3.1 Pvalor obtenido del test de Kolmogorov-Smimov para varios experimentos

En este gráfico se aprecia las oscilaciones del p-valor del test de KS. A pesar de esto, algunas características de los gráficos se conservan; parece haber una tendencia en la disminución del p-valor para confiabilidades menores a 0.35, luego aumenta bruscamente, desciende y vuelve a aumentar, lo cual sugiere algún tipo de correlación positiva.

La tabla 3.2.1 se muestran los valores de los gráficos 3.3 y 3.2. Si bien existe una mejora en el p-valor promedio al aumentar el largo desde 500 a 10000, no es claro que exista una mejora al hacerlo desde 10000 a 20000. De hecho, en este último caso los p-valores promedio son de 0.4833 y 0.5042 respectivamente. Más aún, si se considera la desviación estándar de los p-valores tenemos que para un largo de 10000 la desviación es 0.3066 mientras que para un largo de 20000 es de 0.3831, lo cual indica que las oscilaciones vistas se agravan con el aumento del largo de la réplica.

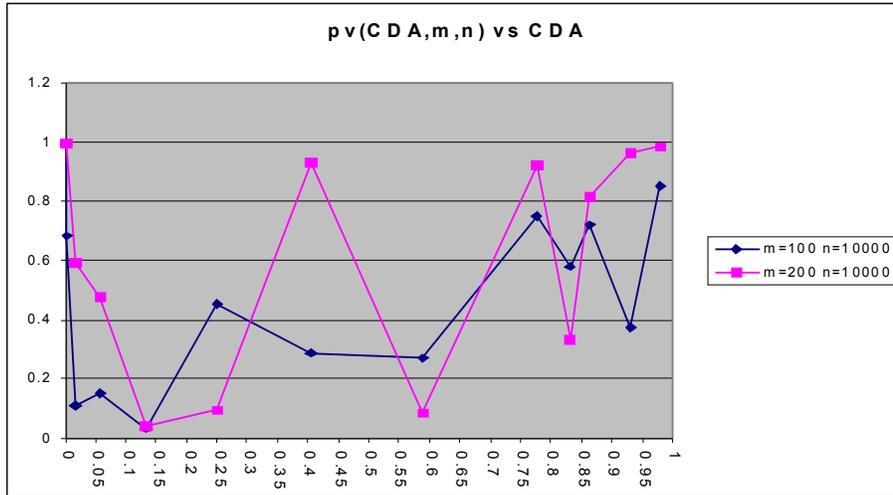
n	10000	10000	500	20000
m	100	200	100	100
CDA				
0.00025				0.98786
6	1	1	1	8
	0.68473	0.99636		0.88954
0.00208	4	7	1	2
	0.11273	0.59493	0.99991	0.81984
0.01696	8	5	3	1
	0.15340	0.47671		0.65438
0.05724	8	2	0.77243	3
	0.03287	0.04018	0.53278	0.06204
0.13312	9	6	2	7
	0.45200	0.09589		
0.25	4	8	0.88214	0.24368
	0.28998	0.93035	0.97690	0.44034
0.40608	5	1	3	9
	0.27089	0.08626	0.12073	
0.58996	4	1	1	0.91828
	0.75180	0.92312	0.54515	0.08730
0.77824	1	5	7	2
0.83034	0.58336	0.33410	0.89296	0.00993
9	9	6	1	8
0.86284	0.72284	0.81553	0.97519	0.06334
6	6	9	1	1
	0.37552	0.96357	0.31984	0.75112
0.93312	4	3	7	3
0.98169	0.85300	0.98708	0.91567	0.62723
4	2	4	8	4

Tabla 3.2.1 P-valores de las gráficas 3.2 y 3.3

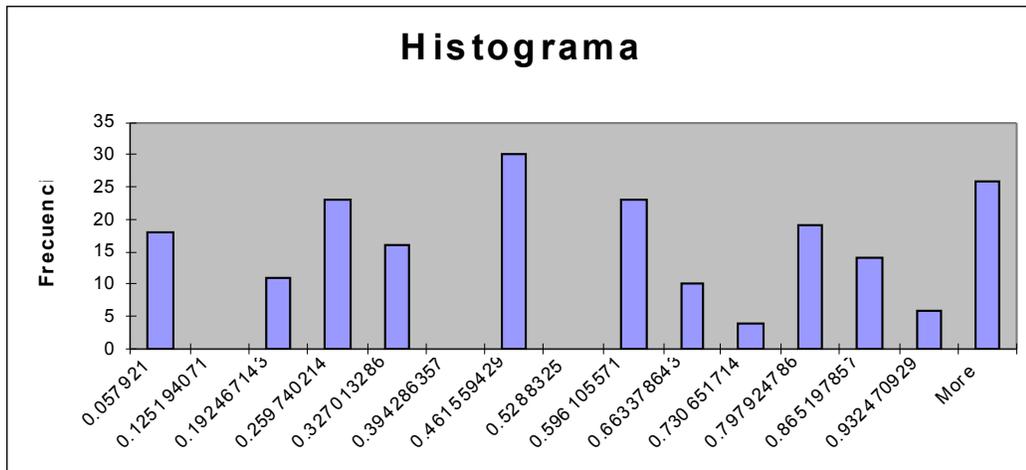
Si se mantiene el largo de réplica fijo y se aumentan la cantidad de réplicas, los resultados son similares, esto es, se aprecian oscilaciones conforme aumenta la confiabilidad de la arista, existe una correlación entre ambas series de datos, no se produce una mejora en cuanto al p-valor promedio y la magnitud de la oscilaciones medida a través de la desviación estándar de los p-valores aumenta.

Un punto llamativo es que mas allá de las oscilaciones, no es raro que los p-valores alcancen magnitudes de más de 0.9 (inclusive 1) para valores de la confiabilidad de arista medios (en el

entorno de 0.5, por ejemplo), donde a priori se podrían suponer valores bajos. Un histograma típico de la función de cobertura calculada podría ser el mostrado en la Gráfica 3.3.



GRAFICA 3.2. Pv(CDA,m,n) vs CDA



GRAFICA 3.3.  $m=200$ ,  $n=1000$  y una  $CDA=0.981694$ .

Un chequeo rápido nos permite concluir que se verifican las condiciones de “validez” del intervalo de Wald *iii* y *iv* (en este último caso, de los doscientos intervalos calculados solo uno violó *iv*, por lo que se considera como aceptable), sin embargo el desempeño del intervalo es cuestionable.

El apartamiento con respecto a la distribución uniforme en la Gráfica 3.3 es notorio.

La razón por la cual se perciben oscilaciones en los p-valores calculados está relacionada con la naturaleza discreta y oblicuidad (*skewness*) de la distribución binomial [10][11]. Este punto será tratado con más detalle en 3.2 cuando se discuta el problema de las oscilaciones en las coberturas observadas.

### 3.3 COBERTURA DEL INTERVALO DE WALD PARA NIVELES DE SIGNIFICANCIA MENORES A 0.1

Los niveles de significancia de interés práctico son aquellos menores a 0.1, por lo tanto es de particular interés saber cómo el intervalo de Wald se comporta en dicha región. Anteriormente se mostró que los p-valores del test de KS sugieren un apartamiento de la distribución uniforme importante, pero eso no descarta la posibilidad de que el intervalo de Wald se comporte de forma aceptable para niveles de significación menores a 0.1, mostrando cierta “robustez” en dicha zona. Sin embargo, el análisis de los datos muestra que

- a) los niveles de cobertura observada no necesariamente tienden a mejorar conforme aumenta el largo de la réplica.
- b) la cobertura observada es en general, menor a la deseada para los valores experimentados de largo de réplica para niveles de significancia menores a 0.1.

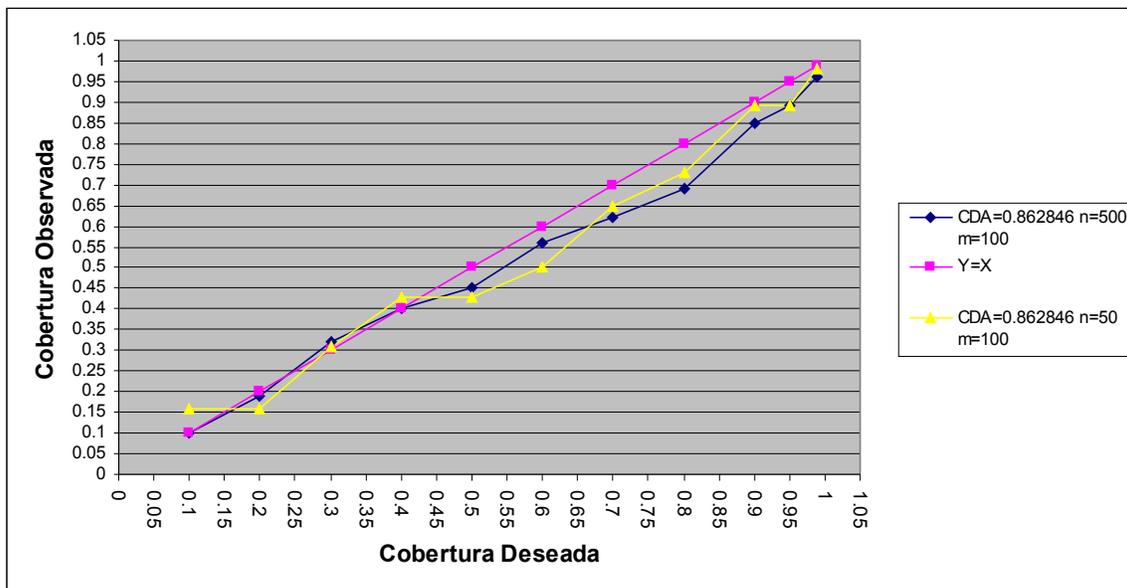


GRAFICO 3.4 Cobertura Observada vs Cobertura Deseada

El Gráfico 3.4 ilustra los puntos anteriores para un caso particular ( $r_a=0.85$   $n=500$  y  $50$ ). Además muestra lo inadecuado de algunas de las condiciones en 2. Por ejemplo, las condiciones *iii*, *iv*, *v* y *vi* se verifican (en realidad, la condición *iii* no se verifica totalmente para el caso de largo 50 cuando se fuerza a la desigualdad a ser mayor a 10), sin embargo las predicciones del intervalo distan de lo aceptable.

En el Gráfico 3.5 se aprecia el comportamiento obtenido para  $n=500$ . Si se compara este gráfico con la Figura 5 de [10], resulta claro que para una correcta caracterización del intervalo, una grilla muy fina de valores de CDA se debe aplicar debido a la presencia de oscilaciones (ver [10] y [11]). Un desarrollo teórico para explicar las oscilaciones puede ser encontrado en [11] y por lo tanto las conclusiones que se extraigan del Gráfico 3.5 deben ser hechas con cautela.

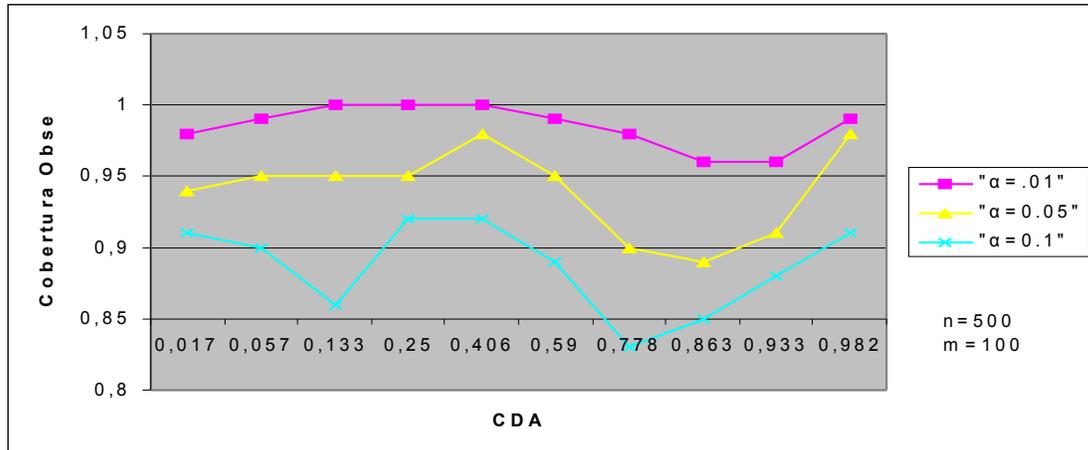


GRAFICO 3.5. Cobertura Observada vs CDA (sin grandes efectos oscilatorios)

El Gráfico 3.6 pretende ilustrar la magnitud de las oscilaciones en los valores de la cobertura en función de la CDA. Por ejemplo, para una cobertura deseada de 0.9 y CDA de 0.44, la cobertura observada es de 0.89 mientras que para una CDA de 0.46 la cobertura observada es de 0.95.

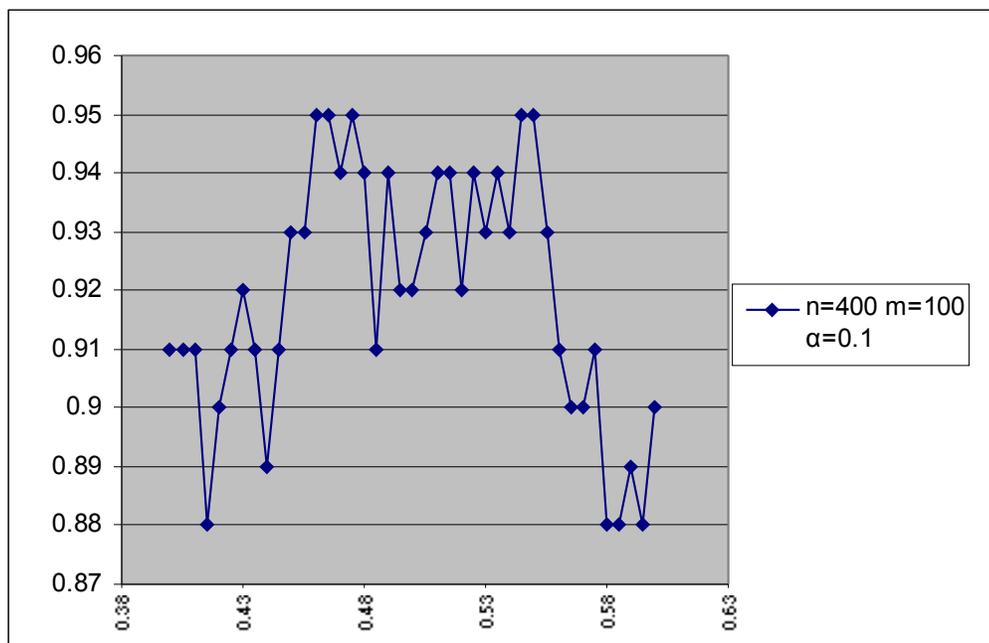


GRAFICO 3.6. Cobertura observada vs CDA

Obsérvese que también existe cierto sesgo en la cobertura media. El tema del sesgo será retomado en 3.4.

Debido a las oscilaciones en la cobertura observada es que se pueden explicar las oscilaciones en los p-valores obtenidos. El test de KS tiene como estadístico una magnitud proporcional a la mayor diferencia entre el valor de la distribución de probabilidades empírica y la teórica. Por lo tanto, ante variaciones importantes en la cobertura observada para dos niveles de CDA

próximos, el estadístico tiende a variar de forma significativa y de igual manera el p-valor. Esto explica el comportamiento observado en las gráficas 3.2 y 3.3.

Otro aspecto importante a destacar es que la cobertura del intervalo de Wald se deteriora de forma más significativa que la inicialmente prevista cuando  $p$  se acerca a 0 o 1 (ver [10]).

### 3.4 INFLUENCIA DEL SESGO Y LAS OSCILACIONES EN LOS RESULTADOS OBTENIDOS

En [11] el lector interesado podrá encontrar una fundamentación analítica de los fenómenos descritos anteriormente. Dicha exposición se basa en propiedades asintóticas de la distribución binomial cuando se la aproxima mediante lo que se conoce como expansiones de Edgeworth de uno y dos términos. Dada la complejidad técnica de dicha exposición, aquí se resumen los resultados más importantes sin entrar en detalles sobre las derivaciones específicas de estos.

Existen dos razones por las cuales el desempeño del IC de Wald no es el deseado:

- a) sesgo;
- b) oscilaciones dependientes de la CDA y  $n$  en la cobertura.

#### 3.4.1 Sesgo

La presencia de sesgo en el intervalo estándar se debe principalmente a que se encuentra centrado en el punto “equivocado”. La derivación del intervalo se basa en que

$$W_n \equiv \frac{n^{1/2}(\hat{p} - p)}{\sqrt{\hat{p}\hat{q}}} \xrightarrow{n} N(0,1)$$

Como se muestra en [11], a pesar de que asintóticamente  $W_n$  tiene sesgo 0, varianza 1, oblicuidad 0 y kurtosis 3, para  $n$  moderados estos valores difieren significativamente. Por ejemplo, para  $p=0.25$  y  $n=50$ ,  $E(W_n)$  vale aproximadamente -0.1 y para  $n=150$   $E(W_n)$  alcanza valores en el entorno de -0.05. Además, es posible demostrar que el signo del sesgo es negativo si  $p < 0.5$  y positivo cuando  $p > 0.5$ .

Un estudio de los valores de la varianza, oblicuidad y kurtosis muestra una convergencia lenta hacia los valores asintóticos, por lo que concluimos que  $W_n$  presenta un comportamiento alejado de la distribución normal para varios valores de interés práctico. Nuevamente se recomienda consultar [11] por más ejemplos numéricos.

#### 3.4.2 Oscilaciones

La presencia de oscilaciones proviene de dos fuentes: la oblicuidad de la distribución binomial y el error en el redondeo producto de su naturaleza discreta.

Para el intervalo bilateral estudiado en este reporte el error de redondeo resulta asintóticamente dominante y es de orden  $n^{-1/2}$  mientras que el error por oblicuidad es de orden  $n^{-1}$  (estos resultados se obtienen mediante expansiones de Edgeworth ver [10] y [11]).

De acuerdo a [11], la probabilidad de cobertura se puede escribir como  $P_{n,p}(L_{n,p} \leq X \leq U_{n,p})$  donde  $L_{n,p}$  y  $U_{n,p}$  son enteros que dependen tanto de  $n$  como de  $p$  tal que pequeños cambios en estos valores provocan grandes cambios en  $L_{n,p}$  y  $U_{n,p}$ .

Para ilustrar esto último aquí se repite un ejemplo brindado en [11]. Si se supone que  $p=0.5$ ,  $\alpha=0.05$  y  $n=39$ , entonces  $L_{n,p} = 14$  y  $U_{n,p} = 25$ . Sin embargo cuando se aumenta  $n$  a 40 se obtiene  $L_{n,p} = 15$  y  $U_{n,p} = 25$ . Por lo tanto, al aumentar  $n$  se pierde el punto  $X=14$  y esto crea en  $n=40$  una cobertura menor que en  $n=39$ . De acuerdo a lo mostrado en [10], estos puntos “desafortunados”

de  $n$  se pueden encontrar para varios valores de  $n$  (inclusive mayores de 1000). Existe un comportamiento similar cuando se deja fijo  $n$  y se hace variar  $p$ . Esto es lo que genera el desempeño errático o caótico del intervalo de Wald.

Como resumen, la expansión de Edgeworth de la cobertura del intervalo de Wald puede escribirse como:

$$C(n, \alpha, p) = (1 - \alpha) + O(n^{-1/2}) \text{oscilación} + O(n^{-1}) \text{sesgo} + O(n^{-1}) \text{oscilación} + O(n^{-3/2})$$

El primer término del miembro derecho lo podríamos interpretar como la influencia del teorema central del límite, el segundo debido al error de redondeo, el tercero el sesgo por convergencia lenta y el cuarto es un término oscilatorio por efecto de la oblicuidad de la distribución binomial.

Esto es perfectamente coherente con la deducción del intervalo de Wald, que como se ha mencionado anteriormente, está basada en una convergencia asintótica a la distribución normal. Lo que la expansión de EdgeWorth indica es que esta convergencia existe, pero se ve "demorada" por otros efectos.

### 3.5 INTERVALOS ALTERNATIVOS

Dada la importancia del problema de estimar una proporción binomial y el mal desempeño del método clásico (intervalo de Wald) es de interés encontrar una forma alternativa de lograr resolver este problema. Para este reporte se estudian dos intervalos tal cual son propuestos en [10], a saber, el intervalo de Agresti-Coull y el intervalo de Wilson (en [10] y [11] se presentan otras alternativas).

Antes de presentar en detalle estos intervalos, es de interés señalar algunas propiedades deseables a cumplir por cualquier intervalo alternativo:

- a) Un intervalo de  $1-\alpha\%$  debería cubrir de forma aproximada el valor correcto en  $1-\alpha\%$  de las veces, en contraste con el enfoque conservador que dice que la cobertura debería ser mayor o igual a  $1-\alpha\%$  (por una interesante discusión sobre este punto ver los comentarios agregados en [10], especialmente las reflexiones de los profesores Casella, Corcoran y Mehta y las respuestas de Brown, Cai y daGupta).
- b) El largo del intervalo es importante. Es claro que a efectos de estimar un parámetro para un nivel dado de confianza se desea que el largo del intervalo sea lo menor posible.
- c) Dada la necesidad de presentar en cursos de nivel introductorios en estadística un intervalo para estimar una proporción binomial, la simplicidad de la forma del intervalo es importante.

#### 3.5.1 Intervalo de Agresti-Coull (AC)

El intervalo AC tiene la siguiente forma:

$$CI_{AC} = \tilde{p} \pm z_{1-\alpha/2} (\tilde{p}\tilde{q})^{1/2} \tilde{n}^{-1/2}$$

donde  $\tilde{X} = X + \frac{(z_{1-\alpha/2})^2}{2}$ ,  $\tilde{n} = n + z_{1-\alpha/2}^2$

$$\tilde{p} = \tilde{X} / \tilde{n}, \quad \tilde{q} = 1 - \tilde{p}$$

Este intervalo es el más simple de los dos presentados en este reporte y mantiene una forma

similar a la de Wald corrigiendo sus características negativas. De hecho, cuando  $n$  tiende a infinito AC tiende al intervalo de Wald. Otra característica importante y tendiente a mejorar el desempeño del intervalo de Wald viene dada por el hecho de que el intervalo tiene un centro distinto al intervalo de Wald (tendiente a corregir el sesgo mencionado anteriormente en 3.4.1) y el término que reemplaza a la varianza con una “varianza corregida”. Estas características lo hacen adecuado para la enseñanza a nivel de introducción al tema.

### 3.5.2 Intervalo de Wilson (W)

Este intervalo tiene la siguiente forma:

$$CI_w = \tilde{p} \pm \frac{z_{1-\alpha/2} n^{1/2}}{\tilde{n}} \left[ \hat{p}\hat{q} + \frac{(z_{1-\alpha/2})^2}{4n} \right]^{1/2}$$

A pesar de la mayor complejidad de la forma del intervalo, cuando  $n$  tiende a infinito el intervalo W converge al valor del intervalo de Wald. Obsérvese que tiene el mismo centro que el intervalo AC pero difiere en su longitud.

## 3.6 PROPIEDADES DE LOS INTERVALOS ALTERNATIVOS

### 3.6.1 Cobertura

Tal cual se mencionó en 3.5 los intervalos que son de interés son aquellos que en promedio contienen al valor verdadero  $1-\alpha\%$ . Sin embargo, a la hora de evaluar el desempeño de los intervalos es de interés conocer a partir de qué valor de  $n$  es posible obtener una mínima cobertura. En la Tabla 1 de [11] se muestran dichos valores para el caso de  $1-\alpha=0.95$  y una cobertura mínima de 0.93, y aquí se reproduce parcialmente en la Tabla 3.5.1:

$p/n_{min}$	0.01	0.03	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
	275	100										
<b>Wald</b>	7	5	526	286	141	118	89	82	56	64	54	71
<b>W</b>	407	42	111	36	24	41	13	31	18	7	13	32
<b>AC</b>	1	5	10	11	8	4	1	1	10	1	1	32

Tabla 3.5.1.

Mínimo valor de  $n$  que asegura una cobertura de 0.93 para una cobertura deseada o nominal de 0.95

Resulta llamativo la no monotonía de  $n_{min}$  en  $p$  de los valores obtenidos y el hecho de que el intervalo AC tenga valores de  $n_{min}$  significativamente menores que los del resto de los intervalos. Existen modificaciones para el intervalo de Wilson de forma de mejorar su desempeño cerca de 0 y 1 (ver [10]). En dichos casos, el intervalo AC tiende a comportarse de manera un tanto conservadora en el sentido de brindar una cobertura un tanto superior a la nominal. En el contexto de confiabilidad en redes, es de interés trabajar con intervalos de buen desempeño en dichos casos, por ejemplo, es de esperar valores de  $p$  pequeños en el contexto de la CDA.

### 3.6.2 Largo de los intervalos

Las longitudes de los intervalos estudiados presentan diferencias significativas (ver [10] por gráficos ilustrativos de este punto y [11] por resultados analíticos utilizando expansiones de Edgeworth). Como ya se señaló, el intervalo de Wald tiene propiedades de cobertura muy

inferiores en las zonas de  $p$  cercanas a 0 o 1, por lo tanto en el estudio del largo no tiene sentido realizar comparaciones entre el intervalo de Wald y los intervalos alternativos en dicha zona. Es de interés destacar que el intervalo de Wilson posee un menor largo que el resto de los intervalos. Por lo tanto, esta es otra característica en la cual el intervalo de Wald se ve superado.

Expansiones de Edgeworth para los intervalos W y AC muestran que también el error de redondeo resulta asintóticamente dominante y es de orden  $n^{-1/2}$  mientras que el error por oblicuidad es de orden  $n^{-1}$ .

#### 4. CONCLUSIONES Y RESUMEN

En este breve reporte se han tratado de resumir algunos resultados numéricos obtenidos en el estudio de la confiabilidad diámetro acotada en relación al desempeño del método estándar para la estimación de una media con intervalos de confianza. Dados los resultados obtenidos, una búsqueda bibliográfica fue comenzada con el propósito de hallar resultados similares. En esta búsqueda se destacaron los artículos [10] y [11], en donde se reportan experimentos exhaustivos y desarrollos teóricos que demuestran lo inadecuado del uso del método estándar (intervalo de Wald) para la estimación de proporciones binomiales. En este reporte se han transcrito las explicaciones y análisis brindados en dichos artículos los cuales deben ser consultados en caso de requerir mayor profundidad.

También se describen brevemente dos intervalos de confianza alternativos. La elección de estos es en base a las recomendaciones en [10] y [11], y sobretodo considerando que dicho intervalo debe ser tanto utilizado para el cálculo científico y/o práctica profesional como para la enseñanza.

#### Referencias

- [1] Libxml 2: The XML C parser and toolkit. <http://www.xmlsoft.org/> (Fecha de consulta 20/6/05)
- [2] Graph eXchange Language. <http://www.gupro.de/GXL/> (Fecha de consulta 20/6/05)
- [3] Cancela, H. Urquhart, M. "Adapting RVR simulation techniques for residual connectedness network reliability models." Reporte técnico INCO 00.02. Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay (2000).
- [4] Brown, B.W., Lovato, J., Russel, K., DCDFLIB: Library of C Routines for Cumulative Distribution Functions, Inverses, and Other Parameters. Version 1.1 Noviembre 1997 Disponible en [http://www.csit.fsu.edu/~burkardt/cpp\\_src/dcdfliib/dcdfliib.html](http://www.csit.fsu.edu/~burkardt/cpp_src/dcdfliib/dcdfliib.html) (Fecha de consulta 2/8/05)
- [5] Matsumoto, M., Nishimura, T., "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", ACM Trans. on Modeling and Computer Simulation Vol. 8, No. 1, January pp.3-30 (1998)
- [6] Cohenar-González, J., "Comparación del desempeño de generadores de números aleatorios". Reporte técnico INCO . Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay (2005)
- [7] Perera, G. *Probabilidad y Estadística Matemática. Un primer encuentro*. Centro de Estudiantes de Ingeniería, Facultad de Ingeniería, Universidad de la República, Uruguay (2002)
- [8] Schruben, L.W., "A coverage function for interval estimators of simulation response" Management Science Vol 26. No. 1, 18-27, January 1980
- [9] DeGroot, M., *Probabilidad y Estadística*. 2da Edición Addison-Wesley Iberoamericana (1988)
- [10] Brown, L.D., Cai, T.T., DasGupta, A., "Interval estimation for a binomial proportion" Statistical Science Vol 21, No. 2, 101-133, 2001. Disponible en <http://projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.ss/1009213286> (Fecha de consulta 20/6/05)
- [11] Brown, L.D., Cai, T.T., DasGupta, A., "Confidence Intervals for binomial proportion and asymptotic expansions" Annals of Statistics Vol 30, No. 1, 160-201, 2002 Disponible en <http://projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.aos/1015362189> (Fecha de consulta 20/6/05)
- [12] Cancela, H., Petingi, L., "Diameter constrained network reliability: exact evaluation by factorization and bounds" Reporte Técnico INCO 01.03 Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay (2001)