Algunas observaciones sobre las inconsistencias del uso de la aproximación normal en intervalos de confianza cuando las observaciones son Bernoullis

Marcos Morales Cortes[a], Hortensia J. Reyes Cervantes[a], Félix Almendra Arao[b]

[a]Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla,
18 sur y Av. Sn. Claudio, Col. Sn. Manuel, C.P. 72570, Puebla, Pue.
averandmeph@gmail.com, hreyes@fcfm.buap.mx

[b]UPIITA del Instituto Politécnico Nacional. Av. IPN 2580, Col Laguna Ticomán, 07340,
México DF.
falmendra@ipn.mx.

RESUMEN

En la mayoría de las disciplinas del conocimiento es muy frecuente realizar experimentos Bernoulli, en donde dos posibles resultados aleatorios mutuamente excluyentes denominados éxito (E) y fracaso (F). Sus resultados son generalmente dados a través de intervalos de confianza para la estimación del parámetro p (0<p<1), de la distribución Binomial. En la literatura estadística se encuentran recomendaciones para su uso, las más conocidas son: **1)** np, n(1-p) son ≥ 5 (o 10). **2)** np(1-p) ≥ 5 (o 10). **3)** n$\hat{p}$, n(1-$\hat{p}$) son ≥5 (o 10). **4)** $\hat{p}\pm 3\sqrt{\hat{p}(1-\hat{p})/n}$ no contiene al 0 o al 1. **5)** n es suficientemente grande. **6)** n al menos ≥ 50 a menos que p sea muy pequeño. En el presente trabajo se analiza el comportamiento del intervalo de Wald también conocido como intervalo estándar (IE), por medio de sus probabilidades de cobertura, se analizan además algunos de los criterios más conocidos para su aplicación. En conclusión tenemos que mediante la prueba de Wald se construyen intervalos de confianza inestables y poco confiables. Nuestro interés en este trabajo fue mostrar estas inconsistencias ya que son intervalos de confianza muy usados en la práctica.

INTRODUCCIÓN

Un problema estadístico muy común es la estimación de los parámetros que ayudan a caracterizar una variable aleatoria. El uso de intervalos de confianza para la estimación de parámetros, nos permite hacer investigaciones sobre qué valores se pueden esperar para ese parámetro. Los intervalos de confianza dependen de la muestra aleatoria, del tamaño muestral y del nivel de confianza seleccionado. Así por tanto debemos familiarizarnos con el concepto de *Intervalo de confianza*, presentamos las siguientes definiciones relacionadas con este concepto.

Definición: Una estimación por intervalo de un parámetro θ es algún par de funciones de la muestra aleatoria ***X***, es decir, L($X_1, X_2, \ldots, X_n$) y U($X_1, X_2, \ldots, X_n$), que satisfacen L(x)≤U(x) para todo x∈ **Ω**. El intervalo aleatorio [L(***X***),U(**X**)] es llamado un estimador por intervalo.

Definición: Sea [L(***X***),U(***X***)] un estimador por intervalo para θ, la probabilidad de cobertura de [L(***X***),U(***X***)] es la probabilidad de que el intervalo aleatorio contenga al parámetro θ, es decir, Pr_θ(θ ∈[L(***X***),U(***X***)]).

Denotaremos a la probabilidad de cobertura como PC.

Definición: Sea [L(***X***),U(***X***)] un estimador por intervalo para θ, el coeficiente de confianza *α* es el ínfimo de las probabilidades de cobertura, es decir, $\alpha = inf_{\theta}\ Pr_{\theta}(\theta \in [L(\boldsymbol{X}),U(\boldsymbol{X})])$.

Finalmente, los estimadores por intervalo junto con una medida de confianza, usualmente un coeficiente de confianza, son conocidos como intervalos de confianza.
Existen diversas formas de construir intervalos de confianza tales como:

- Invirtiendo una prueba estadística.
- Usando una variable aleatoria pivotal.
- Usando el método estadístico.
- Usando intervalos bayesianos.

En la mayoría de las disciplinas del conocimiento cuando se trabaja con individuos (objetos, animales, plantas o personas) es muy frecuente realizar experimentos Bernoulli, es decir experimentos en donde la variable aleatoria tiene únicamente dos posibles resultados aleatorios los cuales son mutuamente excluyentes denominados, éxito (E) y fracaso (F) tal que P(X=E)=p y P(X=F)=1-p, 0<p<1. Y decimos que Y es una variable aleatoria Bernoulli con parámetro *p* y denotamos este hecho mediante *Y~Bernoulli (p)* la variable aleatoria Y tiene función de densidad,

(1) $$f_Y(y) = p^y(1-p)^{1-y}, \text{ para } y=0,1 \text{ y } 0<p<1.$$

Sin embargo en la repetición sucesiva de estos experimentos, los investigadores están interesados en una nueva variable aleatoria definida por el número de ensayos favorables Y_i E de eventos Bernoulli, denotaremos a esta variable aleatoria por X y de esta forma X es la suma de variables aleatorias Bernoulli independientes e idénticamente distribuidas por tanto X tiene una distribución binomial con parámetro n y p, donde n es el número de ensayos y 0<p<1, y así X tiene funciones de densidad y de distribución respectivamente dadas por,

(2) $$f_{\boldsymbol{X}}(\boldsymbol{x}) = \binom{n}{x}p^x(1-p)^{n-x}, \text{ para } x=0, 1,\ldots, n \text{ y } 0<p<1.$$

(3) $$F_X(x) = P(\boldsymbol{X} \le x) = \sum_{i=0}^{x}\binom{n}{i}p^i(1-p)^{n-i}, \text{ para } x=0,1,\ldots, n \text{ y } 0<p<1.$$

La construcción de intervalos de confianza para estimar el parámetro p, de la distribución binomial es un problema muy frecuente en el trabajo estadístico, en la gran mayoría de los libros y textos de los cursos de estadística es presentado el intervalo de Wald obtenido por la aplicación del teorema central del límite, conocido también como intervalo estándar IE.

(4) $$\text{IE}= \left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Donde $z_{\frac{\alpha}{2}}$ denota el 1-$\frac{\alpha}{2}$ cuantil de la distribución normal estándar y $\hat{p}$ es el estimador de máxima verosimilitud para p dado por,

(5) $$\hat{p} = \frac{\boldsymbol{X}}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

El anterior es un intervalo de confianza para *p* al 100(1-α)%.

Revisando la literatura estadística podemos ver que existen varias recomendaciones para el uso del intervalo de Wald. Algunas de estas recomendaciones son:

1. np, n(1-p) son ≥ 5 (o 10).
2. np(1-p) ≥ 5 (o 10).
3. n$\hat{p}$, n(1-$\hat{p}$) son ≥5 (o 10).
4. $\hat{p}\pm 3\sqrt{\hat{p}(1-\hat{p})/n}$ no contenga 0 o 1.
5. n suficientemente grande
6. n ≥ 50 a menos que p sea muy pequeño.

De esta manera, se piensa que si solamente tomamos en cuenta recomendaciones como las anteriores, el intervalo de Wald será un buen estimador por intervalo.

METODOLOGÍA

El objetivo de este trabajo es mostrar las inconsistencias que tiene el uso del intervalo de Wald aún cuando se cumplen las condiciones anteriormente mencionadas. Para esto analizaremos su probabilidad de cobertura PC, definida por $Pr_{\theta}(\theta \in [L(\boldsymbol{X}),U(\boldsymbol{X})])$, para distintos valores de p y distintos tamaños de muestra n, a un nivel de confianza del 95%. En el presente trabajo se analizaron ejemplos usando el intervalo estándar, observamos que se tienen estimaciones poco precisas cuando utilizamos valores de p que están cercanos a 0 o al 1 o cuando n es pequeña (<20), sin embargo se observará además que tal falta de precisión acontece también incluso para valores grandes de n y valores de p cercanos a 0.5.

Se realizó un programa en R y se presentan las gráficas con los resultados, tales gráficas muestran las PC de cálculos con valores fijos de p, p=0.2 y p=0.4 y n tomando valores de 10 a 200.

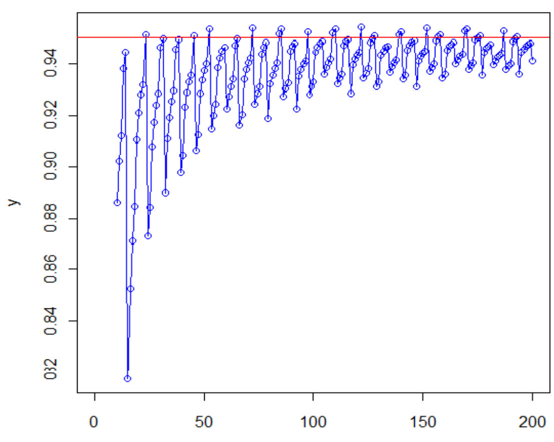

FIG.1; p=0.2 n=10,11,...,200

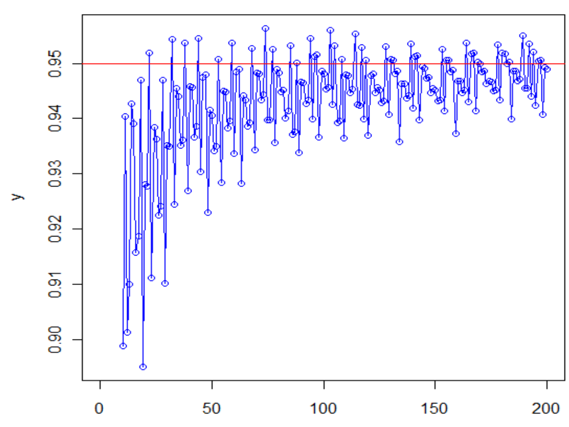

FIG.2; p=0.4 n=10,11,...,200

En la FIG.1 se observa que la PC para la mayoría de los resultados obtenidos es menor que 0.95 (el nivel nominal), de hecho solamente un pequeño porcentaje tiene PC≥0.95, esto sucede aún cuando se cumplen las condiciones mencionadas. En la FIG.2 los cálculos con PC≥0.95 ha aumentado, sin embargo, el porcentaje tal que PC<0.95 es muy grande cuando p=0.4, aún también cuando se cumplen las condiciones anteriores.

En la mayoría de los textos de estadística se plantea que el intervalo de Wald tiene un mejor desempeño cuando p=0.5, veamos este caso:

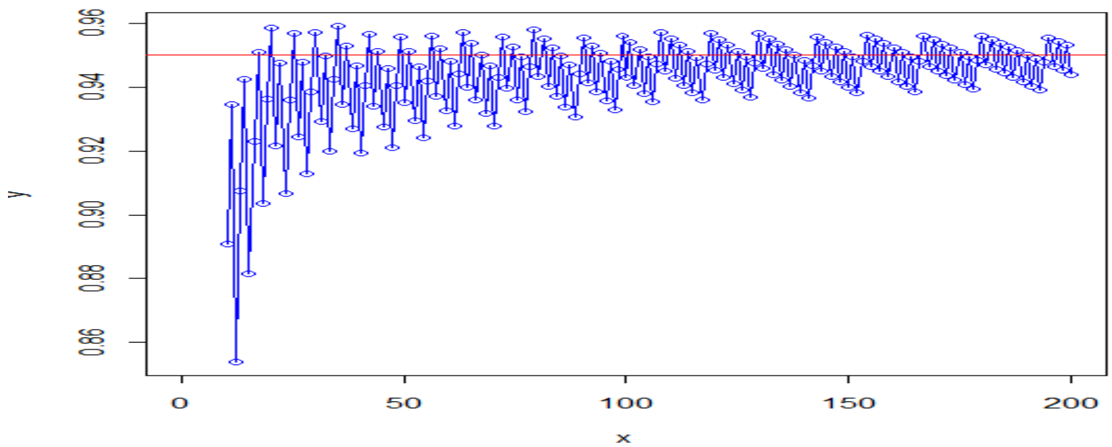

FIG.3; p=0.5 n=10,11,....,200

Puede verse que el número de resultados que tienen una probabilidad de cobertura menor a 0.95 disminuye pero sigue siendo muy grande.

El criterio 5 (ver abajo) trata acerca de n suficientemente grande, ¿qué tan grande?, consideremos que grande significa, n≥100, y construimos gráficas para n=200 FIG.4 y n=1000 FIG 5, y p toma los valores de 0.01, 0.02, …, 0.99.

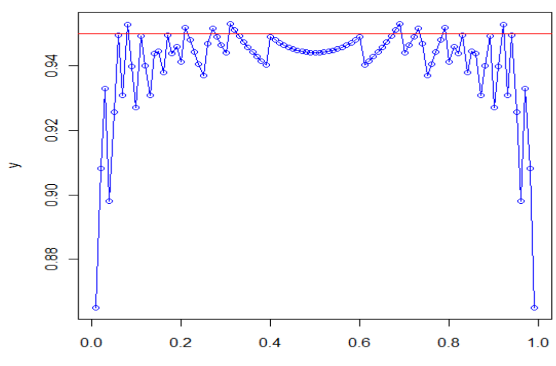

FIG.4; p=0.01,0.02,...,0.99

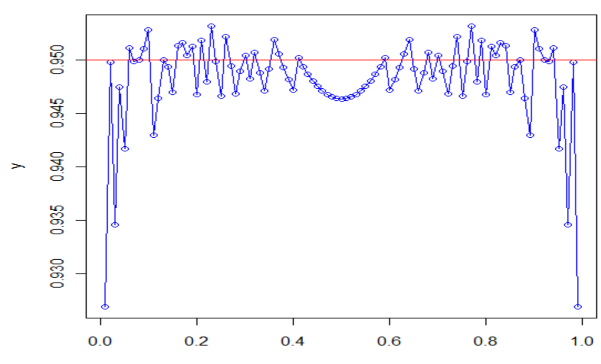

FIG.5; n=1000, p=0.01,0.02,...,0.99

De las FIG.4 y FIG.5 vemos que el número de resultados con PC<0.95 es menor cuando n=1000 que cuando n=200, también vemos que en ambos casos se cumplen las condiciones arriba mencionadas y el número de resultados que cumple que PC<0.95 aún es muy grande.

Podría pensarse que si aumentamos n a valores muy grandes el porcentaje de resultados con PC<0.95 disminuirá. Analizamos esto de la siguiente forma: para los siguientes intervalos se calculará el porcentaje de resultados con PC<0.95 cuando cumplen cada uno de los siguientes criterios:

Criterio 1: np≥5 y n(1-p) ≥5.

Criterio 2: np≥10 y n(1-p) ≥10.

Criterio 3: np(1-p)≥5.

Criterio 4: $np(1p) \geq 10$.
Criterio 5: Diremos que n es grande si $n \geq 100$.
Criterio 6: $n \geq 50$ y p no pequeño suponga que p es pequeño si $p \leq 2$, por simetría también si $p \geq 0.8$, entonces p no es pequeña si $0.2 < p < 0.8$.

La siguiente tabla muestra el porcentaje de resultados con PC<0.95 para los dos siguientes casos:

I. p toma los valores 0.01, 0,02, …, 0.99.
II. p toma los valores 0.001, 0.002, …, 0.999.

	CRITERIO 1		CRITERIO 2		CRITERIO 3		CRITERIO 4		CRITERIO 5		CRITERIO 6	
INTERVALO	I	II	I	II	I	II	I	II	I	II	I	I
[1,500]	79.10	79.30	77.55	77.75	78.81	79.03	77.07	77.28	78.06	78.45	72.95	73.24
[501,1000]	70.50	70.39	69.89	69.97	70.49	70.39	69.88	69.96	70.50	70.78	64.28	64.39
[1001,1500]	67.15	67.10	67.15	66.83	67.15	67.10	67.14	66.82	67.15	67.33	61.55	61.54
[1501,2000]	64.92	65.11	64.92	64.90	64.92	65.11	64.92	64.90	64.92	65.27	59.93	59.93
[2001,2500]	63.53	63.55	63.53	63.40	63.53	63.55	63.53	63.40	63.53	63.70	59.05	58.62
[2501,3000]	62.90	62.79	62.90	62.64	62.90	62.79	62.90	62.64	62.90	62.86	58.42	58.15
[3001,3500]	61.42	61.81	61.42	61.68	61.42	61.81	61.42	61.68	61.42	61.89	57.20	57.35
[3501,4000]	60.94	61.24	60.94	61.16	60.94	61.24	60.94	61.16	60.94	61.31	56.63	56.96
[4001,4500]	60.62	60.59	60.62	60.51	60.62	60.59	60.62	60.51	60.62	60.67	56.71	56.49
[4501,5000]	59.50	60.10	59.50	60.02	59.50	60.10	59.50	60.02	59.50	60.18	55.68	56.18
[5001,5500]	59.53	59.75	59.53	59.67	59.53	59.75	59.53	59.67	59.53	59.75	56.11	55.83
[5501,6000]	59.3	59.45	59.3	59.37	59.3	59.45	59.3	59.37	59.3	59.45	55.7	55.65
[6001,6500]	58.45	59.12	58.45	59.04	58.45	59.12	58.45	59.04	58.45	59.12	55.24	55.40
[6501,7000]	58.62	58.88	58.62	58.79	58.62	58.88	58.62	58.79	58.62	58.88	55.09	55.29
[7500,7501]	58.20	58.66	58.20	58.58	58.20	58.66	58.20	58.58	58.20	58.66	54.85	55.17
[7501,8000]	58.12	58.22	58.12	58.13	58.12	58.22	58.12	58.13	58.12	58.22	54.95	54.83
[8001,8500]	57.75	58.12	57.75	58.03	57.75	58.12	57.75	58.03	57.75	58.12	54.98	54.76
[8501,9000]	57.32	57.95	57.32	57.86	57.32	57.95	57.32	57.86	57.32	57.95	54.27	54.58
[9001,9500]	57.59	57.70	57.59	57.62	57.59	57.70	57.59	57.62	57.59	57.70	54.55	54.44
[9501,10000]	56.80	57.39	56.80	57.30	56.80	57.39	56.80	57.30	56.80	57.39	53.88	54.23

De esta forma se observa en el primer caso, que el porcentaje de resultados con PC<0.95 es muy grande y en algunas ocasiones es mayor cuando se toman los valores del caso II y como puede apreciarse estos porcentajes sobrepasan el 50%, esto también puede apreciarse en FIG1-FIG5.

Veamos qué pasa con las probabilidades de cobertura para n muy grandes. Para esto tomemos valores grandes para n, digamos n=1000001 y n=2000001 y p toma los valores de 0.01, 0.02 ,…, 0.99.

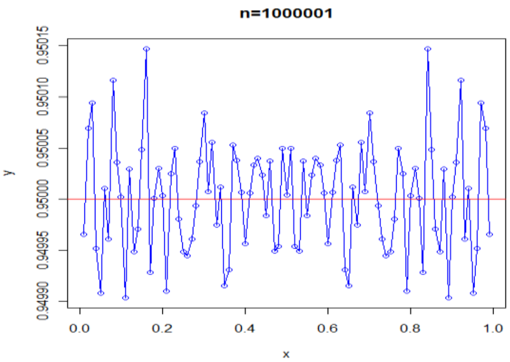

FIG.6, p=0.01,0.02, ..., 0.99

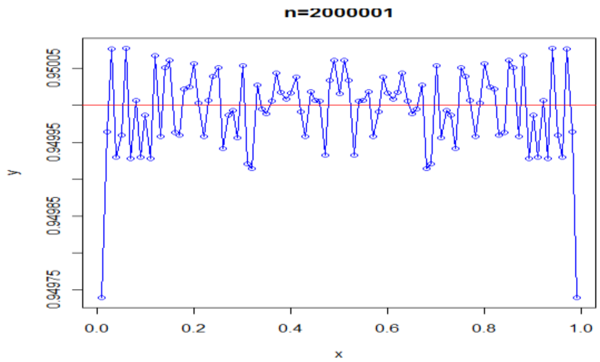

FIG.7, p=0.01,0.02, ..., 0.99

Como podemos observar el porcentaje de resultados con PC<0.95 sigue siendo muy alto, sin embargo, la probabilidad de cobertura para estos resultados está muy cercana a 0.95. Podemos considerar un buen estimador por intervalo cuando se cumple que PC≥0.93.

Veamos esto con las siguientes gráficas que muestran los resultados con PC<0.93 para n=100 y p=0.01, 0.02, …, 0.99 FIG.8 y n=1000 y p=0.001, 0.002, …, 0.999 FIG.9.

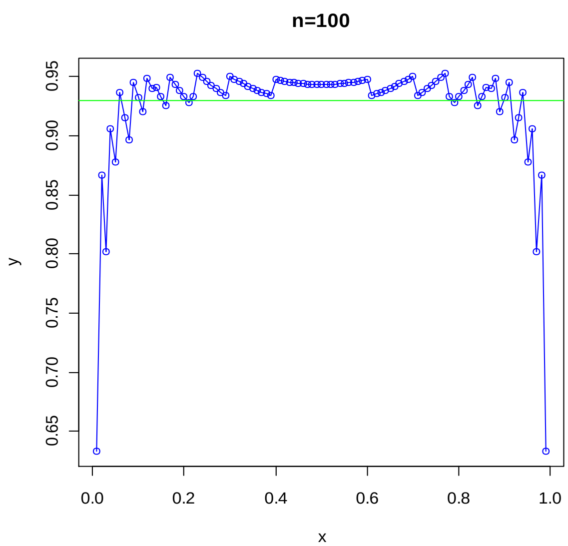

FIG.8, p=0.01, 0.02, ..., 0.99

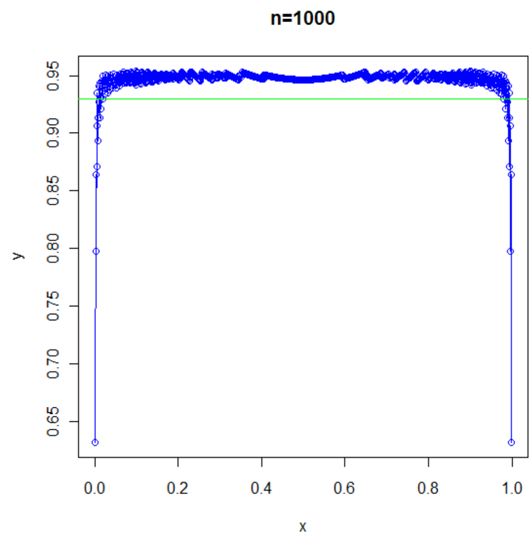

FIG.9, p=0.001, 0.002, ..., 0.999

La tabla siguiente muestra los porcentajes de los resultados con PC<0.93 para cada intervalo, donde p toma los valores 0.01, 0.02, …, 0.99

INTERVALO	CRITERIO 1	CRITERIO 2	CRITERIO 3	CRITERIO 4	CRITERIO 5	CRITERIO 6
[1,500]	8.99	4.36	7.99	3.45	7.99	0.81
[501,1000]	2.40	0.99	2.38	0.97	2.40	0.00
[1001,1500]	0.93	0.93	0.93	0.89	0.93	0.00

[1501,2000]	0.49	0.49	0.49	0.49	0.49	0.00
[2001,2500]	0.19	0.19	0.19	0.19	0.19	0.00
[2501,3000]	0.06	0.06	0.06	0.06	0.06	0.00
[3001,3500]	0	0	0	0	0	0

Como puede observarse si tomamos como un buen estimador al intervalo de Wald a un nivel nominal del 95% cuando la $PC \geq 0.93$, el porcentaje de las resultados con $PC < 0.93$ es muy pequeño, haciendo estas aclaraciones, el intervalo de Wald es un buen estimador por intervalo para el nivel de confianza del 95%.

CONCLUSIONES

El hecho de que la variable aleatoria es discreta provoca intervalos de confianzas inestables, a pesar de cumplir las suposiciones para su uso. Nuestro interés en este trabajo fue mostrar estas inconsistencias ya que son intervalos de confianza muy usados en la práctica.

BIBLIOGRAFÍA

1. Brown, L., Cai, D. & DasGupta, A. (2001), ‘Interval Estimation for a Binomial Proportion’, Statistical Science 16, 101–133.
2. Intervalos de confianza e intervalos de credibilidad para una proporción, (2008) Revista Colombiana de Estadística, volumen 31, no. 2
3. George Casella, Roger L. Berger (2002), Statistical Inference, Duxbury Thomson Learning.
4. Navidi W. (2006). Estadística para ingenieros y científicos, Mc Graw Hill.
5. Hogg, R.V. y Craig, A.T. (1978). Introduction to Mathematical Statistics. Cuarta Edición. Collier MacMillan International: New York.