REVISTA INVESTIGACIÓN OPERACIONAL Vol., 31, No.1, 80-92, 2010

EDUCATIONAL ISSUES/
ASPECTOS EDUCACIONALES

ALGORITMOS EN R PARA LA ENSEÑANZA DE LA APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL A LA NORMAL

Mariana Silvia Santamaría[1] y Marta Susana Malla [2]
Departamento de Matemática, Universidad Nacional del Sur, Bahía Blanca, Argentina

RESUMEN
Una de las dificultades que suelen presentar los alumnos universitarios de los cursos básicos en el aprendizaje de probabilidad y de estadística es la comprensión de que la distribución de probabilidad discreta binomial pueda aproximarse mediante la distribución de probabilidad continua normal. Se presenta en este trabajo una propuesta didáctica consistente en dos algoritmos de muy fácil manejo, desarrollados con el paquete informático R de distribución libre. Estos algoritmos introducen al alumno mediante representaciones gráficas y tablas en las ideas básicas de aproximación. Permiten una participación más activa del alumno en el proceso de aprendizaje.

ABSTRACT
One of the difficulties students from basic courses often show in learning probability and statistics at the university is understanding how the continuous normal probability distribution approximates the discrete binomial probability distribution. In this paper a didactic proposal is presented consisting in two algorithms, very easy to use, developed with software R which is completely free. These algorithms aim to introduce students in basic concepts of normal approximation to the binomial distribution by means of graphical representations and tables. They allow an active participation of student in the learning process.

KEY WORDS: didactic proposal, graphical representations, tables, central limit theorem.

MSC 97U40, 62P99

1. INTRODUCCIÓN

Existe una variedad de situaciones que pueden ser modeladas por una distribución de probabilidad binomial y cuyo cálculo resulta complicado si no se recurre a la aproximación normal para valores grandes del parámetro *n*. Si bien la expresión de la función de densidad de la distribución normal estándar no resulta sencilla dado que involucra raíces, exponentes y números como π y e, existen tablas de la distribución normal estándar para el cálculo de probabilidades ampliamente difundidas.

Una de las dificultades que suelen presentar los alumnos universitarios de los cursos básicos en el aprendizaje de probabilidad y de estadística (correspondiente a las carreras de pregrado de universidades nacionales argentinas)[3], es

[1]msanta@criba.edu.ar
[2] msmalla@criba.edu.ar
[3] En Argentina, las asignaturas de Probabilidad y de Estadística forman parte de los cursos básicas iniciales (18 a 21 años de edad de los alumnos) de las currículas de las carreras de Ingeniería, Matemática, Bioquímica y otras.

la comprensión de que una distribución discreta, binomial, pueda aproximarse mediante una distribución continua, normal.

Dado que la computadora constituye en la actualidad una herramienta didáctica útil en el proceso de enseñanza/aprendizaje de diversas disciplinas, se presenta en este trabajo una propuesta didáctica consistente en dos algoritmos de muy fácil manejo desarrollados con el paquete informático R. Dichos algoritmos introducen al alumno mediante representaciones gráficas y tablas en las ideas básicas de aproximación. El paquete R está disponible en internet a través del Comprehensive R Archive Network (CRAN) cuya dirección es http://cran.r-project.org/. Uno de los algoritmos permite variar los valores de los parámetros de la distribución binomial a fin de evaluar en forma visual, a través de gráficos de barras, su efecto sobre la forma de la distribución y la bondad de la aproximación. El otro, proporciona en forma de tabla las probabilidades binomiales y su aproximaciones a la normal.

El concepto de aproximación presentado en este trabajo se basa en el enfoque propuesto por Mosteller y colaboradores en el libro *Probability with Statistical Applications* (1961). Dicho enfoque describe, cómo los gráficos de barras para distribuciones binomiales ajustadas tienden a la forma de una distribución normal y además compara numéricamente las probabilidades de las distribuciones binomiales con sus correspondientes aproximaciones a la normal estándar.

2. METODOLOGÍA

Algunas propiedades de la distribución binomial, tales como su simetría, desplazamiento, achatamiento y dispersión fueron expuestas en el trabajo de M. Santamaría y M. Malla (2007) como así también una breve introducción sobre la definición y notación de una variable aleatoria binomial, B(n, p).

Para una demostración gráfica de la forma en que la distribución binomial puede ser aproximada por la distribución normal estándar, $N(0,1)$, deben considerarse dos aspectos fundamentales:

1. La distribución binomial es discreta y la distribución normal es continua, por este motivo las probabilidades representadas por las ordenadas binomiales requieren ser reemplazadas por áreas, ya que éstas representan las probabilidades en las distribuciones continuas.
2. A medida que n crece se produce un desplazamiento y achatamiento de la distribución binomial, para evitar esto es necesario un cambio de escala para la variable binomial.

El siguiente procedimiento muestra una forma sencilla de ajustar una distribución binomial por una función de probabilidad continua y utilizar áreas para representar las probabilidades dadas por las ordenadas de la distribución discreta (F. Mosteller y colaboradores, 1961; R. Johnson y P. Kuby, 2004).

La distribución de probabilidad de una variable aleatoria discreta, en particular de una variable binomial, suele representarse por medio de un conjunto de segmentos de recta trazados en los valores x correspondientes al rango de la variable X, cuyas longitudes (ordenadas) representan las probabilidades de cada x. El método consiste en reemplazar cada ordenada de la distribución binomial por un rectángulo centrado en x cuyo ancho es la unidad y cuya altura es igual a la ordenada binomial original. De este modo el área de cada rectángulo tiene la misma medida numérica que el alto de la ordenada. Si bien el objetivo del trabajo es considerar valores de n grandes, a fin de mostrar este concepto mediante un ejemplo, **Figura 1**, se considera la distribución binomial con n = 8 y p = 0.5.

En la **Figura 1** se observa que el área entre 3-0.5 y 3+0.5 tiene el mismo valor, $1 \times 0.2187 = 0.2187$, que la altura de la ordenada en 3 que es 0.2187. En general, el área correspondiente al intervalo que va de x-0.5 a x+0.5, **Figura 1** (b), coincide con la altura de la ordenada x, **Figura 1** (a). Es decir, las probabilidades dadas como ordenadas se representan como áreas.

Para comprender el cambio de escala en la distribución binomial, necesario para evitar su desplazamiento y achatamiento a medida que n crece, se define una sucesión de distribuciones binomiales para un p fijo. Para conservar el valor de n, se adoptará en la notación de la sucesión de variables aleatorias X el subíndice n, es decir X_n.

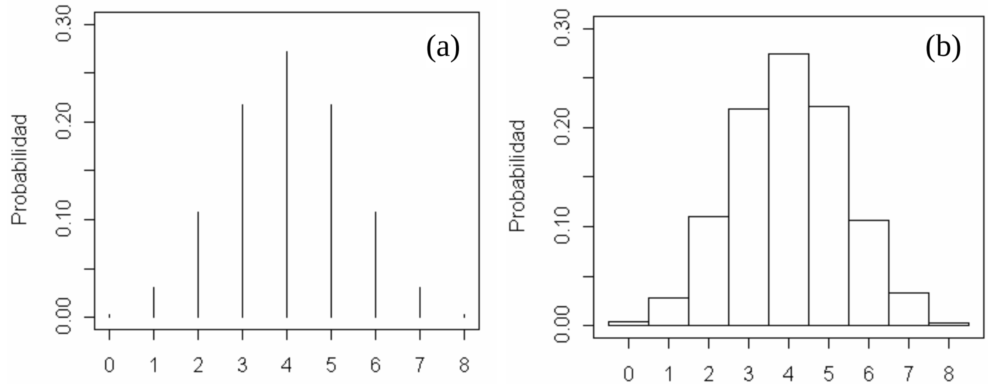

Figura 1. (a) *B(8,0.5)* con probabilidades como ordenadas. (b) *B(8,0.5)* con probabilidades como áreas.

Para evitar el desplazamiento de las sucesivas variables binomiales X_n se introduce el conjunto de variables $Y_n = X_n - np$. Dado que la media de la distribución de X_n es np y su desviación estándar $\sqrt{npq}$, $npq > 0$, la variable Y_n tiene media 0 y desviación estándar $\sqrt{npq}$. De este modo cada Y_n tiene la misma media que la distribución normal estándar pero, en general, no tiene la misma desviación estándar. La sucesión de distribuciones Y_n aún presenta achatamiento a medida que *n* crece. Para evitar esto, se crea un nuevo conjunto de variables $Z_n = Y_n / \sqrt{npq}$. La distribución de cada Z_n tiene media 0 y desviación estándar 1, tal como la distribución normal estándar. Al cambiar a las variables Z_n se ajusta la escala de las X_n produciendo un cambio en el ancho de los rectángulos que se utilizan en la representación de probabilidades mediante áreas. Para comprender este concepto se expone el siguiente ejemplo para una distribución binomial de parámetros *n* = 16 y *p* = 0.5, **Figura 2**.

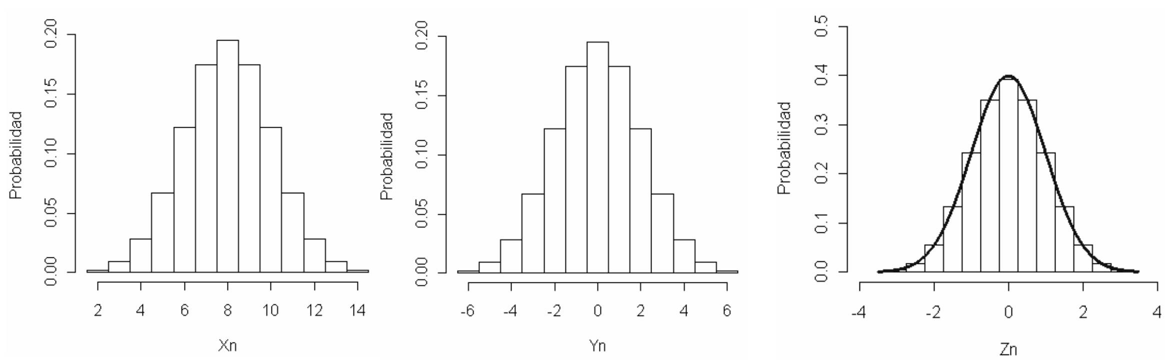

Figura 2. Gráficos de barras para la aproximación de *B(16,0.5)* a la normal.
Las escalas horizontal y vertical de Z_n difieren de X_n y Y_n.

En general para cualquier distribución binomial el rectángulo centrado en *x* en la escala de X_n tiene por área:

base x altura = $1 \times b(x;n,p) = b(x;n,p)$

Mientras que el correspondiente rectángulo en la escala de Z_n tiene por área:

base x altura = $\left(\dfrac{1}{\sqrt{npq}}\right) \times \left[\sqrt{npq}\; b(x;n,p)\right] = b(x;n,p)$

Los resultados gráficos y numéricos que se presentan en este trabajo pretenden ilustrar como una sucesión de gráficos de barras ajustados de distribuciones binomiales para un valor fijo de *p* y *n* creciente se aproxima a una distribución normal. Este resultado se expresa más formalmente mediante el Teorema Central del Límite (J. Blaiotta, P. Delieutraz, 2004, J. Gubner 2006).

2.1 Teorema Central del Límite (de Lindeberg – Lévy)

Sea $\{W_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E(W_n)=\mu$ y $0<Var(W_n)=\sigma^2<\infty$. Sea $S_n = \sum_{j=1}^{n} W_j$ con $n=1, 2, \ldots$

Entonces

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{L} N(0,1).$$

De Moivre, en 1733, fue quien primero proporcionó una versión del teorema sólo para variables Bernoulli para el caso especial p = 1/2, mientras que Laplace lo generalizó al caso p arbitrario cuyo resultado se enuncia a continuación:

Sea $\{W_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas con distribución Bernoulli con parámetro p, $p\in(0,1)$.

Entonces,

$$\frac{S_n - n\,p}{\sqrt{n\,p\,q}} \xrightarrow{L} N(0,1).$$

Dado que la suma de n variables aleatorias Bernoulli independientes con probabilidad p es una variable aleatoria por lo tanto por el teorema resulta que para n grande la distribución binomial B(n, p) tiende a $N(n\,p, n\,p\,q)$.

En la práctica para el cálculo de una probabilidad binomial entre a y b, $a < b$, valores pertenecientes al rango de la variable, el resultado del teorema establece, para valores grandes de n, que:

$$\mathrm{P}(\mathrm{a}\leq \mathrm{X_n}\leq \mathrm{b}) = \mathrm{P}(\mathrm{a}-0.5\leq \mathrm{X_n}\leq \mathrm{b}+0.5) =$$

$$\mathrm{P}\left(\frac{\mathrm{a}-0.5-\mathrm{np}}{\sqrt{\mathrm{npq}}}\leq\frac{\mathrm{X_n}-\mathrm{np}}{\sqrt{\mathrm{npq}}}\leq\frac{\mathrm{b}-0.5-\mathrm{np}}{\sqrt{\mathrm{npq}}}\right)\approx \mathrm{P}\left(\frac{\mathrm{a}-0.5-\mathrm{np}}{\sqrt{\mathrm{npq}}}\leq \mathrm{Z_n}\leq\frac{\mathrm{b}-0.5-\mathrm{np}}{\sqrt{\mathrm{npq}}}\right)$$

Si bien la aproximación Normal a la distribución binomial mejora al aumentar n, otro factor a tener en cuenta es la posición del intervalo [a, b]. La aproximación es mejor si el intervalo contiene a la media $n\,p$ y empeora a medida que el intervalo se desplaza hacia la izquierda o la derecha de $n\,p$. Por otra parte, la aproximación es mejor en las ordenadas centrales que en las ordenadas correspondientes a las colas de la distribución. Los distintos textos sobre probabilidad no siempre están de acuerdo acerca de cuan grande debe ser n para lograr una buena aproximación. En general se acepta que n sea mayor o igual que 30 para poder aplicar el Teorema Central del Límite. Sin embargo una buena aproximación también depende de p y una posible guía para determinar cuando se puede utilizar la aproximación normal es tener en cuenta el cálculo de np y nq. Si ambos "*np y nq son mayores o iguales a 5*" la aproximación será buena (R. Walpole y R. Myers, 1992, pp. 165). Otro criterio para una buena aproximación es el sugerido por Rohatgi (1984, pp. 587) que establece $np(1-p) > 10$.

2.2 Descripción y Funcionamiento de los Algoritmos

A continuación se describen los algoritmos desarrolladas e implementados en el paquete informático R (V. Ricci, 2005, R Development Core Team, 2005) que constituyen el aspecto original del presente trabajo. Los dos algoritmos o funciones (como se los denomina en lenguaje R) dependen de dos argumentos: *n* y *p*, que constituyen los parámetros de la distribución binomial.

La función **BinomialAproxNormal** permite evaluar para cada par de valores *n* y *p* mediante gráficos de barras, la forma de la distribución binomial original, centrada y estandarizada junto con la superposición de la curva de distribución normal estándar. De este modo, variando los argumentos de la función, a medida que *n* aumenta y bajo las condiciones mencionadas con anterioridad para el parámetro *p*, puede apreciarse en forma gráfica como la distribución binomial tiende a la normal.

Por otra parte, la función **TablaProbabilidades** permite calcular para cada par valores de *n* y *p* las probabilidades, tanto puntuales como acumuladas, de la distribución binomial original y su aproximación normal. Mediante la presentación de estos valores en forma de tablas es posible evaluar la bondad de la aproximación a través de la comparación de las áreas de ambas distribuciones. Los valores permitidos para los argumentos de ambas funciones son: para *n*, enteros entre 2 y 200, y para *p*, reales entre 0.01 y 0.99. Si bien entre estos valores permitidos para los parámetros algunos no cumplen las condiciones requeridas para una buena aproximación a la normal, fueron incluidos en la programación de los algoritmos a fin de que el educando pueda apreciar que en esos casos la aproximación no es buena.

```
BinomialAproxNormal<-function(n,p){
cociente=n/trunc(n)
if (n<2 | n>200 | cociente>1) stop("n debe ser un número natural entre 2 y 200" )
if (p <0.01| p>0.99) stop("p debe variar entre 0.01 y 0.99" )
binom1=dbinom(0:n,n,p)
binom1Red3=round(binom1,3)
matrizBin<-matrix(0:n,ncol= (n+1))
matrizFrec=matrix(binom1Red3,ncol= (n+1))
matrizTabla=rbind(matrizBin, matrizFrec)
TablaTrunc=matrizTabla[,matrizTabla[2,]>=0.001]
fila1=TablaTrunc[1,]
fila2=TablaTrunc[2,]*1000
Xn=rep(fila1,fila2)
hist(Xn,breaks=seq(min(fila1)-0.5,    max(fila1)+0.5,by=1),freq=F,ylab=c("Probabilidad"),main=paste("    "))
#histograma variable X
Yn=Xn-n*p
windows()
hist(Yn,breaks=seq(min(fila1-n*p)-0.5,max(fila1-n*p)+0.5,by=1),freq=F, ylab=c("Probabilidad"), main=paste(" "))
#histograma variable Y=X-n*p
Zn=Yn/sqrt(n*p*(1-p))
windows()
inversadesvio=1/(sqrt(n*p*(1-p)))
hist(Zn,freq=F,breaks=seq(min(Zn)-1/2* inversadesvio, max(Zn)+ 1/2* inversadesvio, by=1/(sqrt(n*p*(1-
p)))),xlim=c(-4,4), ylim=c(0,0.5), ylab=c("Probabilidad"), main=paste(" ")) #histograma variable Z=X-n*p/
qrt(n*p*(1-p))
xaleatorios=seq(-3.5,3.5,length=100000)
ynormales=dnorm(xaleatorios, 0,1)
lines(xaleatorios, ynormales,col="black",lwd=3)
}
TablaProbabilidades<- function(n,p){
cociente=n/trunc(n)
if (n<2 | n>200 | cociente>1) stop("n debe ser un número natural entre 2 y 200" )
if (p <0.01| p>0.99) stop("p debe variar entre 0.01 y 0.99" )
datosx=0:n
binomialprob<-dbinom(0,n,p)
binomialprobacum<-pbinom(0,n,p)
normalprob<- pnorm((0+0.5 - n*p)/sqrt(n*p*(1-p)),0,1)
normalprobacum<- pnorm((0+0.5 - n*p)/sqrt(n*p*(1-p)),0,1)
for (i in 1:(n-1)){
```

```
binomialprob<-cbind(binomialprob,dbinom(i,n,p))
normalprob<-cbind(normalprob,pnorm((i+0.5 - n*p)/sqrt(n*p*(1-p)),0,1)- pnorm((i-0.5 - n*p)/sqrt(n*p*(1-p)),0,1))
binomialprobacum<-cbind(binomialprobacum,pbinom(i,n,p))
normalprobacum<-cbind(normalprobacum,pnorm((i+0.5 - n*p)/sqrt(n*p*(1-p)),0,1))
}
binomialprob<-cbind(binomialprob,dbinom(n,n,p))
normalprob<-cbind(normalprob,1-pnorm((n-0.5 - n*p)/sqrt(n*p*(1-p)),0,1))
binomialprobacum<-cbind(binomialprobacum,pbinom(n,n,p))
normalprobacum<-cbind(normalprobacum, pnorm(((n-1)+0.5 - n*p)/sqrt(n*p*(1-p)),0,1)+1-pnorm((n-0.5 - n*p)/sqrt(n*p*(1-p)),0,1))
probpuntual<-matrix(c(datosx, round(binomialprob,digits=4),
round(normalprob,digits=4)), ncol=3, dimnames=list(rep(“ “,(n+1)), rep(“      “,3)))
probacum<-matrix(c(round(binomialprobacum,digits=4), round(normalprobacum,digits=4)), ncol=2, dimnames=list(rep(“ “,(n+1)), c(“          “,“      “) ))
tabla.probab<-cbind(probpuntual,probacum)
cat(rep("",11),"x",rep("",3),"b(x;n,p)", rep(" ",1),"Aprox. normal",rep("",3)," P(X <= x)", rep("",3),"Aprox. normal","\n", rep(" ",13),"a b(x;n,p)", rep(" ",10), “a P(X <= x)”, "\n")
print(tabla.probab)
}
```

Al final del trabajo se incluye un Apéndice que contiene una breve descripción de la obtención e instalación del software R para Windows y puesta en funcionamiento de los algoritmos.

3. RESULTADOS

Para exhibir el funcionamiento de ambos algoritmos se eligieron valores de n = 5, 35 y 100 y de p = 0.2 y 0.5.

Para obtener los gráficos y tablas respectivas en R se ejecutaron las funciones **BinomialAproxNormal** y **TablaProbabilidades** escribiendo en la consola de R a continuación del símbolo > su nombre y entre paréntesis los argumentos correspondientes, de la siguiente manera:

```
BinomialAproxNormal(5,0.2);       TablaProbabilidades(5,0.2)
BinomialAproxNormal(35,0.2);      TablaProbabilidades(35,0.2)
BinomialAproxNormal(100,0.2);     TablaProbabilidades(100,0.2)
BinomialAproxNormal(5,0.5);       TablaProbabilidades(5,0.5)
BinomialAproxNormal(35,0.5);      TablaProbabilidades(35,0.5)
BinomialAproxNormal(100,0.5);     TablaProbabilidades(100,0.5)
```

Se sugiere que el alumno ejecute los dos algoritmos utilizando otros pares de valores de n y p, a fin de involucrarse de manera activa en el proceso de comprensión y aprendizaje de la aproximación binomial a la normal.

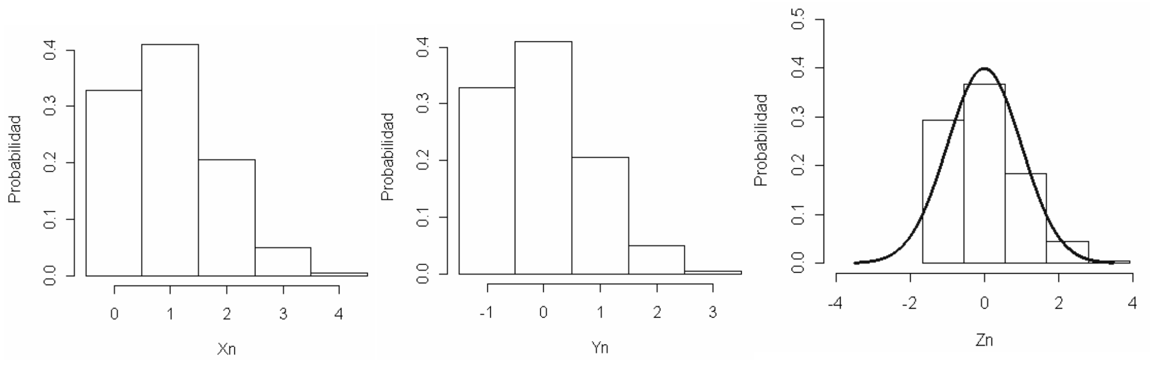

Figura 3. Gráficos de barras para la aproximación de *B(5,0.2)* a la normal.
Las escalas horizontal y vertical de Z_n difieren de X_n y Y_n.

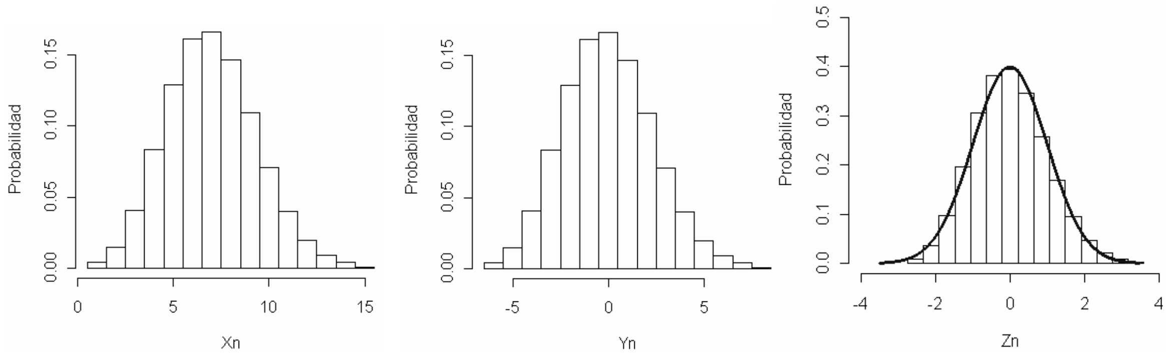

Figura 4. Gráficos de barras para la aproximación de *B(35,0.2)* a la normal.
Las escalas horizontal y vertical de Z_n difieren de X_n y Y_n.

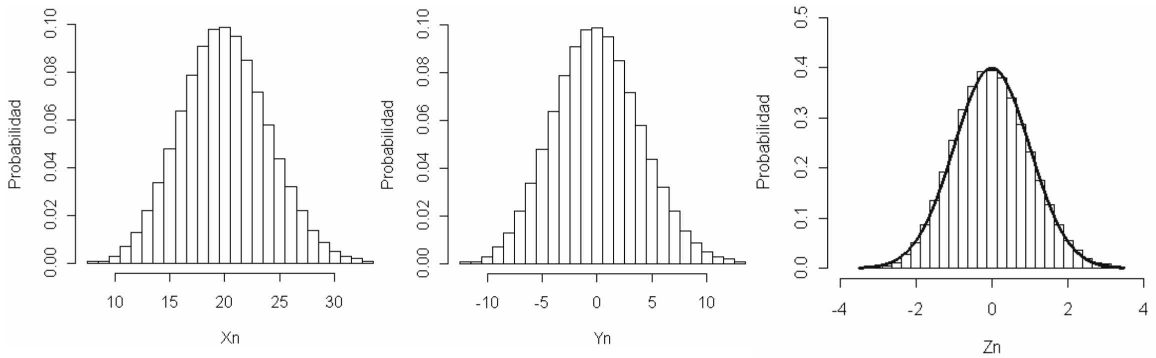

Figura 5. Gráficos de barras para la aproximación de *B(100,0.2)* a la normal.
Las escalas horizontal y vertical de Z_n difieren de X_n y Y_n.

Las representaciones gráficas, **Figuras 3-5**, pretenden ilustrar que el límite de una sucesión de gráficos de barras de una distribución binomial ajustada para un p fijo, p=0.2, y n creciente, n = 5, 35, 100, es una distribución normal. A partir de las figuras puede observarse como al cambiar de las escalas X_n, Y_n a la escala Z_n el ancho de los rectángulos deja de ser una unidad.

La **Tablas 1–3** contienen en su primera columna los valores del recorrido de la variable binomial. La segunda y tercera columna muestran respectivamente las probabilidades puntuales de la distribución binomial y su aproximación a la normal calculada sobre cada intervalo [x-0.5, x+0.5]. Las dos últimas columnas de las tablas contienen las probabilidades acumuladas de la distribución binomial y de su correspondiente aproximación a la normal estándar.

Tabla 1. Valores puntuales de la distribución *B(5,0.2)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
0	0.3277	0.2881	0.3277	0.2881
1	0.4096	0.4238	0.7373	0.7119
2	0.2048	0.2413	0.9421	0.9532
3	0.0512	0.0442	0.9933	0.9974
4	0.0064	0.0025	0.9997	1.0000
5	0.0003	0.0000	1.0000	1.0000

Tabla 2. Valores puntuales de la distribución *B(35,0.2)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
0	0.0004	0.0030	0.0004	0.0030
1	0.0035	0.0070	0.0040	0.0101
2	0.0151	0.0186	0.0190	0.0286
3	0.0415	0.0410	0.0605	0.0696
4	0.0830	0.0758	0.1435	0.1454
5	0.1286	0.1177	0.2721	0.2631
6	0.1607	0.1532	0.4328	0.4163
7	0.1665	0.1673	0.5993	0.5837
8	0.1457	0.1532	0.7450	0.7369
9	0.1093	0.1177	0.8543	0.8546
10	0.0710	0.0758	0.9253	0.9304
11	0.0404	0.0410	0.9656	0.9714
12	0.0202	0.0186	0.9858	0.9899
13	0.0089	0.0070	0.9947	0.9970
14	0.0035	0.0022	0.9982	0.9992
15	0.0012	0.0006	0.9995	0.9998
16	0.0004	0.0001	0.9999	1.0000
17	0.0001	0.0000	1.0000	1.0000
18	0.0000	0.0000	1.0000	1.0000
19	0.0000	0.0000	1.0000	1.0000

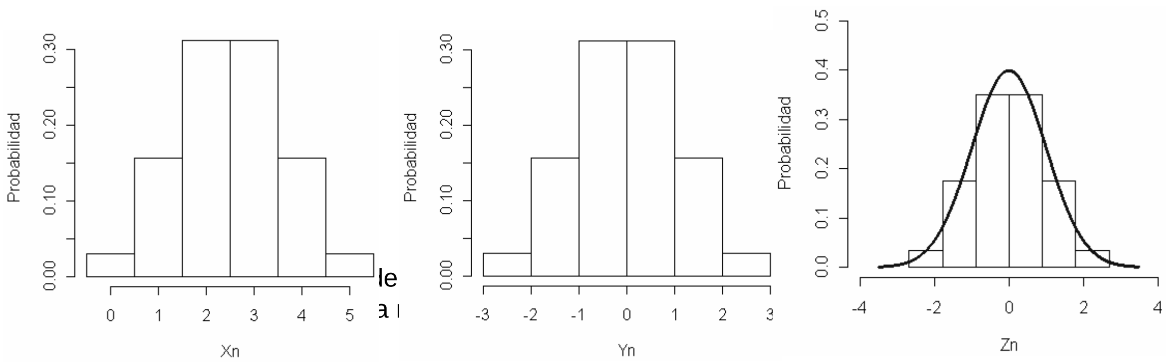

Figura 6. Gráficos de barras para la aproximación de *B(5,0.5)* a la normal. Las escalas horizontal y vertical de *Zn* difieren de *Xn* y *Yn*.

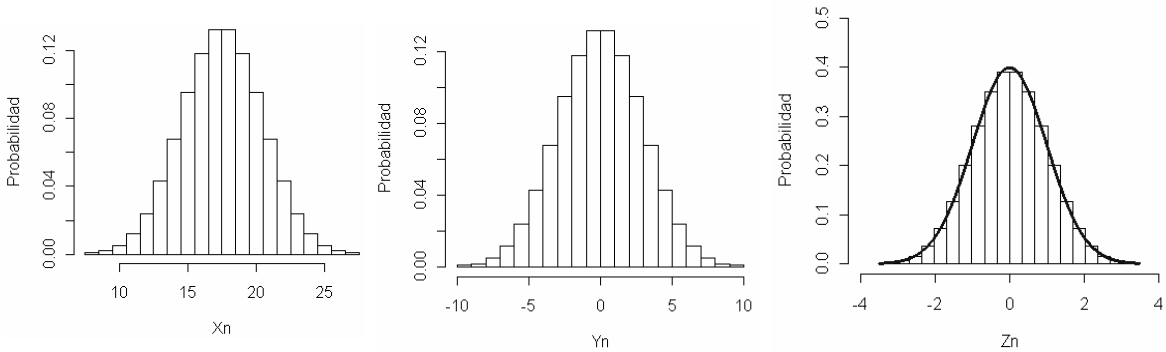

Figura 7. Gráficos de barras para la aproximación de *B(35,0.5*) a la normal. Las escalas horizontal y vertical de *Zn* difieren de *Xn* y *Yn*.

Tabla 3. Valores puntuales de la distribución *B(100,0.2)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
4	0.0000	0.0000	0.0000	0.0001
5	0.0000	0.0001	0.0000	0.0001
6	0.0001	0.0002	0.0001	0.0004
7	0.0002	0.0005	0.0003	0.0009
8	0.0006	0.0011	0.0009	0.0020
9	0.0015	0.0023	0.0023	0.0043
10	0.0034	0.0044	0.0057	0.0088
11	0.0069	0.0080	0.0126	0.0168
12	0.0128	0.0136	0.0253	0.0304
13	0.0216	0.0217	0.0469	0.0521
14	0.0335	0.0325	0.0804	0.0846
15	0.0481	0.0457	0.1285	0.1303
16	0.0638	0.0605	0.1923	0.1908
17	0.0789	0.0752	0.2712	0.2660
18	0.0909	0.0878	0.3621	0.3538
19	0.0981	0.0964	0.4602	0.4503
20	0.0993	0.0995	0.5595	0.5497
21	0.0946	0.0964	0.6540	0.6462
22	0.0849	0.0878	0.7389	0.7340
23	0.0720	0.0752	0.8109	0.8092
24	0.0577	0.0605	0.8686	0.8697
25	0.0439	0.0457	0.9125	0.9154
26	0.0316	0.0325	0.9442	0.9479
27	0.0217	0.0217	0.9658	0.9696
28	0.0141	0.0136	0.9800	0.9832
29	0.0088	0.0080	0.9888	0.9912
30	0.0052	0.0044	0.9939	0.9957
31	0.0029	0.0023	0.9969	0.9980
32	0.0016	0.0011	0.9984	0.9991
33	0.0008	0.0005	0.9993	0.9996
34	0.0004	0.0002	0.9997	0.9999
35	0.0002	0.0001	0.9999	0.9999
36	0.0001	0.0000	0.9999	1.0000
37	0.0000	0.0000	1.0000	1.0000

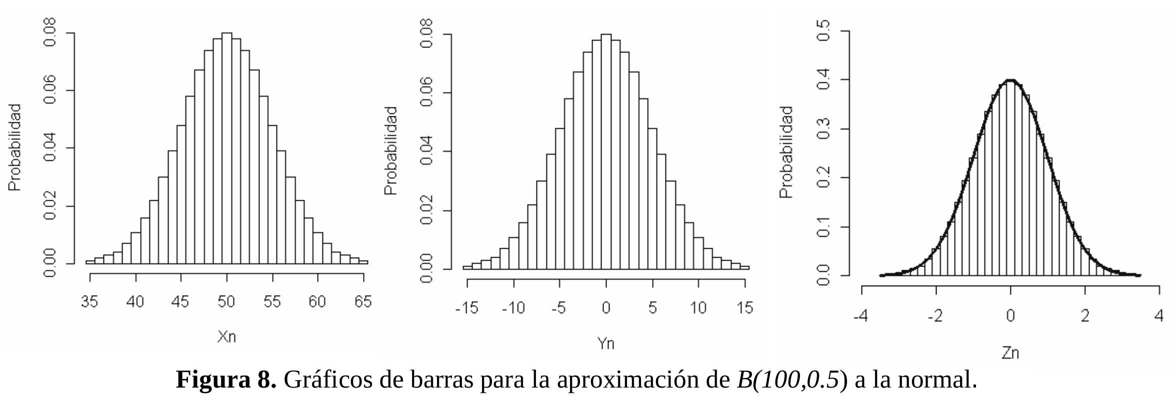

Figura 8. Gráficos de barras para la aproximación de *B(100,0.5)* a la normal. Las escalas horizontal y vertical de *Zn* difieren de *Xn* y *Yn*.

Se puede observar al comparar las **Figuras 3–5** con las **Figuras 6–8**, que la aproximación a la normal es mejor cuando la distribución binomial es simétrica, es decir para p=0.5, que cuando la distribución binomial es asimétrica, caso p=0.2. El alumno también puede ejecutar los algoritmos para distintos valores de n y p=0.15 y 0.5 con el fin de observar que para el caso p=0.15 se requiere de un n más grande que para el caso p=0.5 para obtener una buena aproximación.

La **Tablas 4–6** contienen las probabilidades puntuales y acumuladas de la distribución binomial y su aproximación a la normal estándar.

Tabla 4. Valores puntuales de la distribución *B(5,0.5)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
0	0.0312	0.0368	0.0312	0.0368
1	0.1562	0.1487	0.1875	0.1855
2	0.3125	0.3145	0.5000	0.5000
3	0.3125	0.3145	0.8125	0.8145
4	0.1562	0.1487	0.9688	0.9632
5	0.0312	0.0368	1.0000	1.0000

Tabla 5. Valores puntuales de la distribución *B(35,0.5)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
5	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0001	0.0001	0.0001
7	0.0002	0.0003	0.0003	0.0004
8	0.0007	0.0008	0.0009	0.0012
9	0.0021	0.0022	0.0030	0.0034
10	0.0053	0.0056	0.0083	0.0090
11	0.0121	0.0123	0.0205	0.0213
12	0.0243	0.0242	0.0448	0.0455
13	0.0430	0.0427	0.0877	0.0881
14	0.0675	0.0671	0.1553	0.1552
15	0.0945	0.0942	0.2498	0.2495
16	0.1182	0.1182	0.3679	0.3677
17	0.1321	0.1323	0.5000	0.5000
18	0.1321	0.1323	0.6321	0.6323
19	0.1182	0.1182	0.7502	0.7505
20	0.0945	0.0942	0.8447	0.8448
21	0.0675	0.0671	0.9123	0.9119
22	0.0430	0.0427	0.9552	0.9545
23	0.0243	0.0242	0.9795	0.9787
24	0.0121	0.0123	0.9917	0.9910
25	0.0053	0.0056	0.9970	0.9966
26	0.0021	0.0022	0.9991	0.9988
27	0.0007	0.0008	0.9997	0.9996
28	0.0002	0.0003	0.9999	0.9999
29	0.0000	0.0001	1.0000	1.0000

Tabla 6. Valores puntuales de la distribución *B(100,0.5)* y su acumulada junto con las aproximaciones a la normal.

x	b(x;n,p)	Aprox. normal a b(x;n,p)	P(X <= x)	Aprox. normal a P(X <= x)
30	0.0000	0.0000	0.0000	0.0000
31	0.0001	0.0001	0.0001	0.0001
32	0.0001	0.0001	0.0002	0.0002
33	0.0002	0.0003	0.0004	0.0005
34	0.0005	0.0005	0.0009	0.0010
35	0.0009	0.0009	0.0018	0.0019
36	0.0016	0.0016	0.0033	0.0035
37	0.0027	0.0027	0.0060	0.0062
38	0.0045	0.0045	0.0105	0.0107
39	0.0071	0.0071	0.0176	0.0179
40	0.0108	0.0109	0.0284	0.0287
41	0.0159	0.0158	0.0443	0.0446
42	0.0223	0.0222	0.0666	0.0668
43	0.0301	0.0300	0.0967	0.0968
44	0.0390	0.0389	0.1356	0.1357
45	0.0485	0.0484	0.1841	0.1841
46	0.0580	0.0579	0.2421	0.2420
47	0.0666	0.0666	0.3086	0.3085
48	0.0735	0.0736	0.3822	0.3821
49	0.0780	0.0781	0.4602	0.4602
50	0.0796	0.0797	0.5398	0.5398
51	0.0780	0.0781	0.6178	0.6179
52	0.0735	0.0736	0.6914	0.6915
53	0.0666	0.0666	0.7579	0.7580
54	0.0580	0.0579	0.8159	0.8159
55	0.0485	0.0484	0.8644	0.8643
56	0.0390	0.0389	0.9033	0.9032
57	0.0301	0.0300	0.9334	0.9332
58	0.0223	0.0222	0.9557	0.9554
59	0.0159	0.0158	0.9716	0.9713
60	0.0108	0.0109	0.9824	0.9821
61	0.0071	0.0071	0.9895	0.9893
62	0.0045	0.0045	0.9940	0.9938
63	0.0027	0.0027	0.9967	0.9965
64	0.0016	0.0016	0.9982	0.9981
65	0.0009	0.0009	0.9991	0.9990
66	0.0005	0.0005	0.9996	0.9995
67	0.0002	0.0003	0.9998	0.9998
68	0.0001	0.0001	0.9999	0.9999
69	0.0001	0.0001	1.0000	1.0000
70	0.0000	0.0000	1.0000	1.0000

4. CONCLUSIONES

Los dos algoritmos que se presentan son flexibles debido a que permiten variar los valores de sus argumentos (parámetros de la distribución binomial) posibilitando una participación más activa del alumno en el proceso de aprendizaje del tema propuesto. Su implementación además de ser sencilla es accesible tanto al profesor como al educando en virtud de que R es un software libre, fácil de instalar, mantener actualizado y utilizar por no expertos en informática. Los gráficos y tablas que resultan de la ejecución de los algoritmos, ilustran de una manera didáctica y clara, como el límite de una sucesión de gráficos de barras de una distribución binomial ajustada es una distribución normal.

REFERENCIAS

[1] BLAIOTTA J. y P. DELIEUTRAZ (2004): Teorema Central del Límite. www.union-**matematica.org.ar/reunion_anual/anteriores/monografiatcl41.pdf**

[2] GUBNER J. (2006): **Probability and Random Processes for Electrical and Computer Engineers**. Cambridge University Press, Cambridge.

[3] JOHNSON, R. y P. KUBY (2004): **Estadística Elemental**. Thomson Learning, México.

[4] MOSTELLER, F., ROURKE R. y THOMAS, G.B. Jr., (1961): **Probability with Statistical Application**.. Addison Wesley, Massachusetts.

[5] R DEVELOPMENT CORE TEAM (2005): **Introducción a R. Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos.** http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf

[6] RICCI V. (2005): **Fitting Distributions with R.** http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

[7] ROHATGI V. (1984): **Statistical Inference**. John Wiley & Sons, New York.

[8] SANTAMARÍA M. y MALLA M. (2007): Comprensión de las Propiedades de la Distribución Binomial mediante Gráficos con R". **XXXV Coloquio Argentina de Estadística**, Mar del Plata, Argentina.

[9] SEEFELD K. y LINDER E. (2007): **Statistics Using R with Biological Examples**. University of Hampshire, Durham, NH. http://cran.r-project.org/doc/contrib/

[10] WALPOLE R. y MYERS R. (1992): **Probabilidad y Estadística**. McGraw-Hill, México.

RECEIVED FEBRUARY 2009
REVISED JUNE 2009

APÉNDICE

Obtención e instalación de R para WIndows

R consiste en un proyecto, un lenguaje y un entorno de software (K. Seefeld y E. Linder, 2007). Como proyecto R es parte del GNU, un proyecto libre (www.gnu.org) con licencia sin restricciones. Como lenguaje, R es un dialecto del lenguaje S. Como software, R es interactivo, está diseñado para ejecutar cálculos, manipular datos y producir gráficos. Aunque es un entorno de líneas de comando no está diseñado exclusivamente para programadores y es además, de muy fácil uso.

Para obtener un archivo ejecutable de R para windows ingresar al siguiente sitio de internet: http://cran.es.r-project.org/bin/windows/base/

y descargar el archivo R- 2.6.2-win32.exe (última versión a la fecha Abril 2008) o su actualizado.

Una vez descargado este archivo, hacer “clic” dos veces en su ícono y seguir las instrucciones de instalación. Para ejecutar el programa se hace “clic” dos veces en el ícono de R que se encuentra en el escritorio de la computadora. Esto abre el programa y crea una ventana llamada “R console”, consola de R, donde el usuario escribe las instrucciones que ejecutará el programa. El símbolo > en la consola es el apuntador (“prompt”) e indica que el programa está listo para recibir instrucciones (comandos).

Una forma de incorporar a R y poner en funcionamiento los dos algoritmos consiste en copiar cada uno de ellos en un archivo de texto, por ejemplo Word, y salvar el documento. Mediante el uso del mouse copiar cada algoritmo y pegar en la consola de R. De esa manera ya se pueden ejecutar variando los parámetros según se explicó con anterioridad.

El comando q( ) cierra el programa R. Al salir de la sesión de R e intentar cerrar el programa aparece en la consola la siguiente pregunta: “Salvar imagen de área de trabajo?” Al aceptar, todos los objetos creados en R se guardan en el directorio de trabajo en un archivo con extensión “.Rdata”. De este modo los algoritmos son almacenados en forma permanente como nuevas funciones de R, y pueden ejecutarse directamente escribiendo sus nombres y parámetros en la consola. Una forma de obtener ayuda en R consiste en seleccionar el menú Ayuda y buscar en las distintas opciones que éste ofrece.