Suma de variables i.i.d.

i.i.d.

Independientes
Idénticamente
Distribuidas

Suma de variables i.i.d.

- Dos v.a. continuas son independientes sii
- Dos v.a. discretas son independientes sii
- Si dos v.a. (continuas o discretas) son independientes, entonces

Suma de variables i.i.d.

v.a. i.i.d. con y

$$\mathrm{E}\left[X+Y\right]=\mathrm{E}\left[X\right]+\mathrm{E}\left[Y\right]=2\mu$$

$$\mathrm{Var}\left[X+Y\right]=\mathrm{Var}\left[X\right]+\mathrm{Var}\left[Y\right]+2\mathrm{COV}\left[X,Y\right]=2\sigma^2$$

Suma de variables i.i.d.

v.a. i.i.d. con y

$$\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = ?$$

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = ?$$

Suma de variables i.i.d.

v.a. i.i.d. con y

$$\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu$$

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = n\sigma^2$$

Suma de variables i.i.d.

v.a. i.i.d. con y

Promedio
:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\mathrm{E}\left[\overline{X}_n\right] = ?$

$\mathrm{Var}\left[\overline{X}_n\right] = ?$

Suma de variables i.i.d.

v.a. i.i.d. con y

Promedio:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathrm{E}\left[\overline{X}_n\right] = \mu$$

$$\mathrm{Var}\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$$

Suma de v.a. independientes

Casos especiales

Suma de v.a. independientes

¡Ojo!

- Esos eran casos **especiales**, no siempre se da
- No ejemplos: suma de uniformes, suma de exponenciales, etc.

Suma de v.a. independientes

¡Ojo!

- Un no-ejemplo de interés:

v.a. i.i.d., con

$$\sum_{i=1}^{n} X_i \quad ?$$

Suma de v.a. independientes

¡Ojo!

- Un no-ejemplo de interés:

v.a. i.i.d., con

$$\sum_{i=1}^{n} X_i \quad Bino(n,p)$$

Suma de v.a. independientes

¡Ojo!

- Otro no-ejemplo de interés:

v.a. i.i.d., con

$$S_n = \sum_{i=1}^{n} X_i$$

Suma de v.a. independientes

v.a. i.i.d., con

Tiempo hasta el -ésimo evento de un proceso de Poisson .

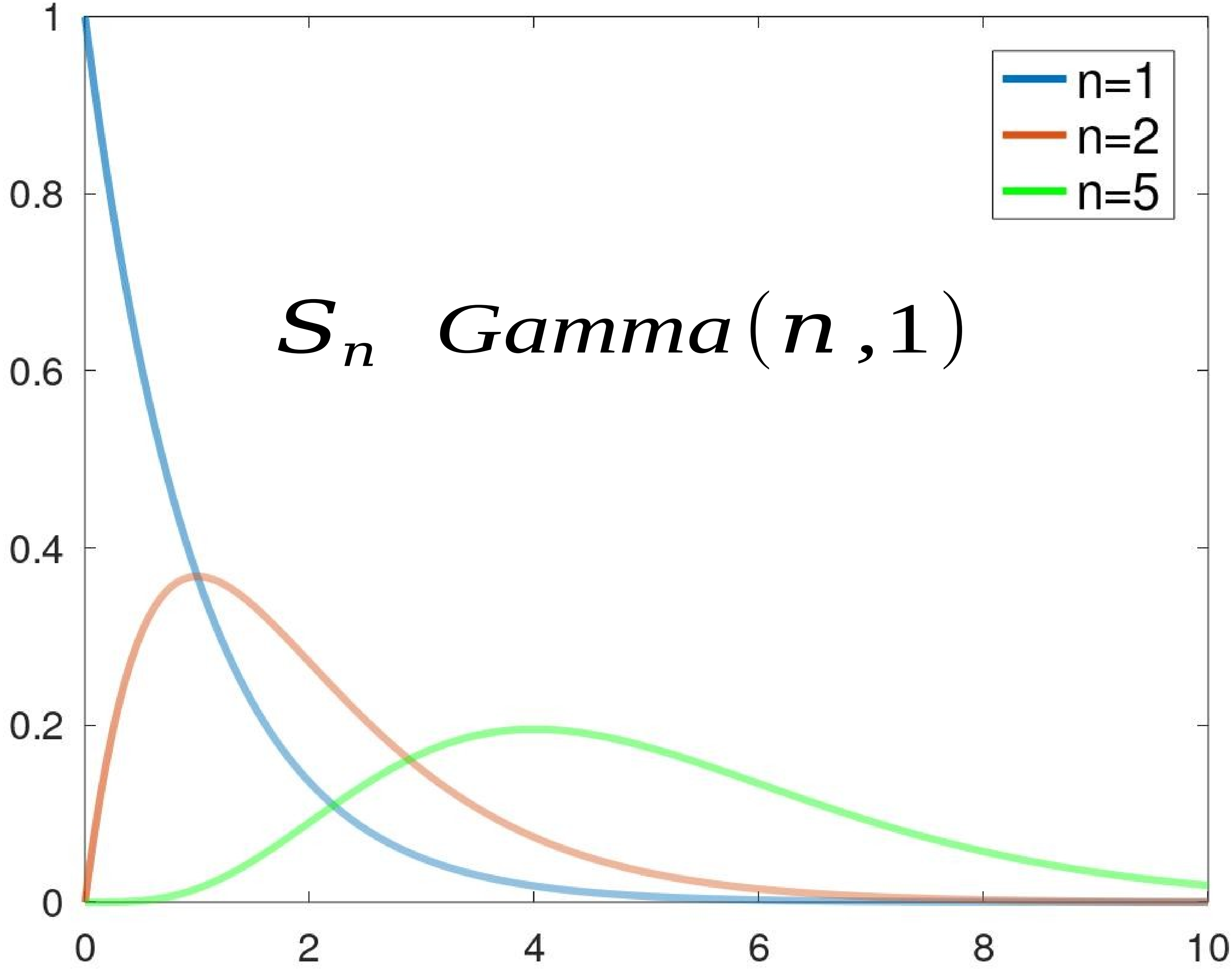
$$P\left(S_n > t\right) = P\left(N\left(t\right) < n\right) = \sum_{i=0}^{n-1} \frac{\left(\lambda t\right)^i}{i!} e^{-\lambda t}$$

Suma de v.a. independientes

v.a. i.i.d., con

$$S_n \quad Gamma(n,\lambda)$$

$$f_{S_n}(t)=\frac{\lambda(\lambda t)^{n-1}e^{-\lambda t}}{(n-1)!}\quad t>0$$



$S_n \quad Gamma(n, 1)$
n=1
n=2
n=5
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10

Preguntas

Desigualdad de Markov

Sea   una v.a. que sólo toma valores no-negativos. Entonces

$$P\left(X \geq \alpha\right) \leq \frac{\mathrm{E}[X]}{\alpha}$$

Desigualdad de Markov

Idea de la demostración para v.a. continua

Desigualdad de Chebychev

Sea una v.a. con y . Entonces

$$P\left(¿\, X - \mu \vee \geq \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2}$$

Desigualdad de Chebychev

Demostración usando desigualdad de Markov

Desigualdad de Chebychev

Ejemplo 1: v.a. normal

$$P(\text{¿}X-\mu\vee\geq k\sigma)\leq\frac{1}{k^2}$$

	Probabilidad	**Cota**
1	0.3173	1.0000
2	0.0455	0.2500
3	0.0027	0.1111
4	0.0001	0.0625

Desigualdad de Chebychev

Ejemplo 2: promedio

v.a. i.i.d. con y

$$\mathrm{E}\left[\overline{X}_n\right]=\mu \qquad \mathrm{Var}\left[\overline{X}_n\right]=\frac{\sigma^2}{n}$$

$$P\left(¿\,\overline{X}_n-\mu \vee \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Ley Débil de los Grandes Númer

v.a. i.i.d. con y . Luego, para todo

$$\lim_{n \to \infty} P\left(¿\,\overline{X}_n - \mu \vee \geq \varepsilon \right) = 0$$

Decimos que converge a en probabilidad

Ley Débil de los Grandes Númer

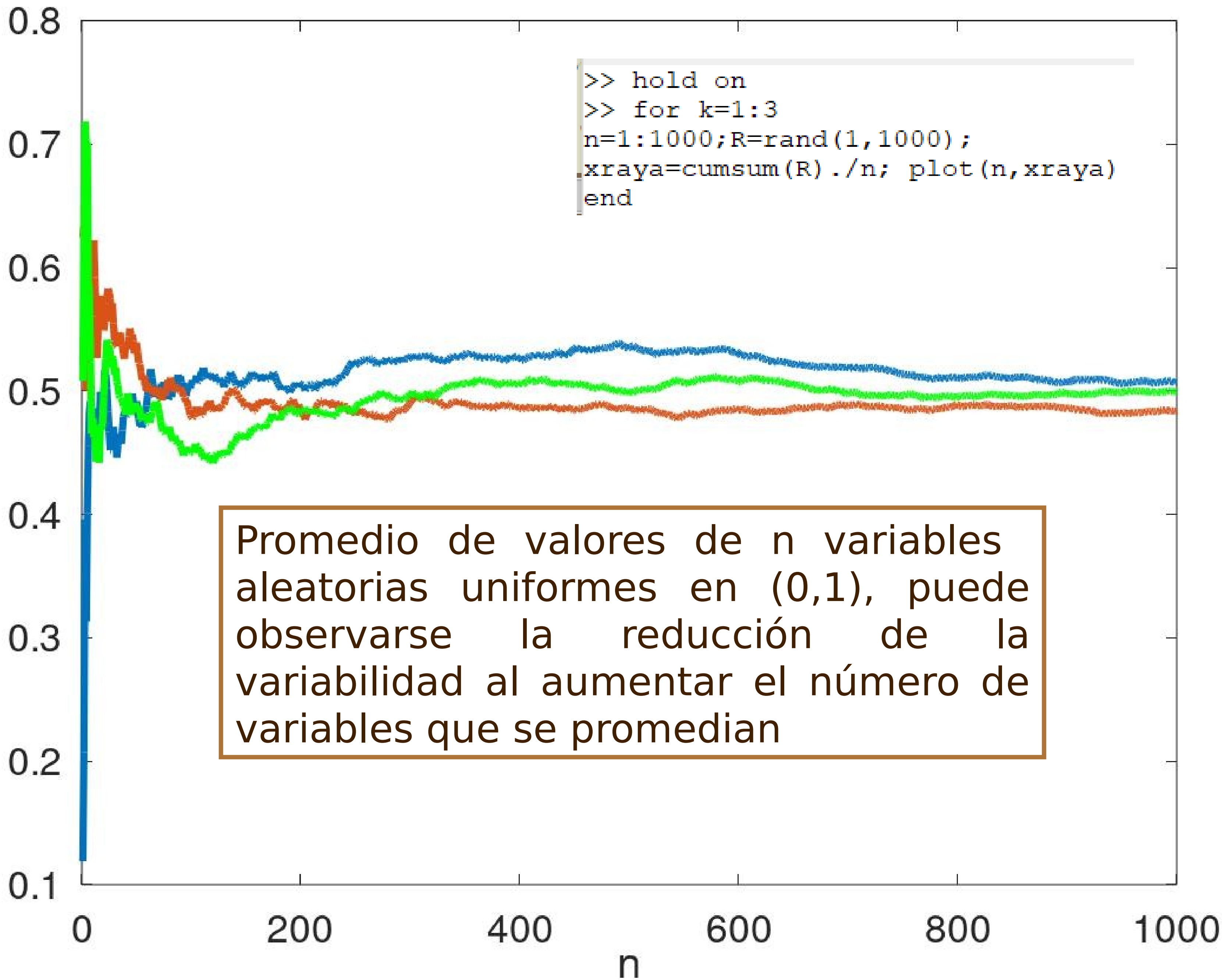
Implicancia práctica:

Podemos estimar las probabilidades repitiendo experimentos independientes un gran número de veces

Ley Débil de los Grandes Númer

Ejemplo en Octave

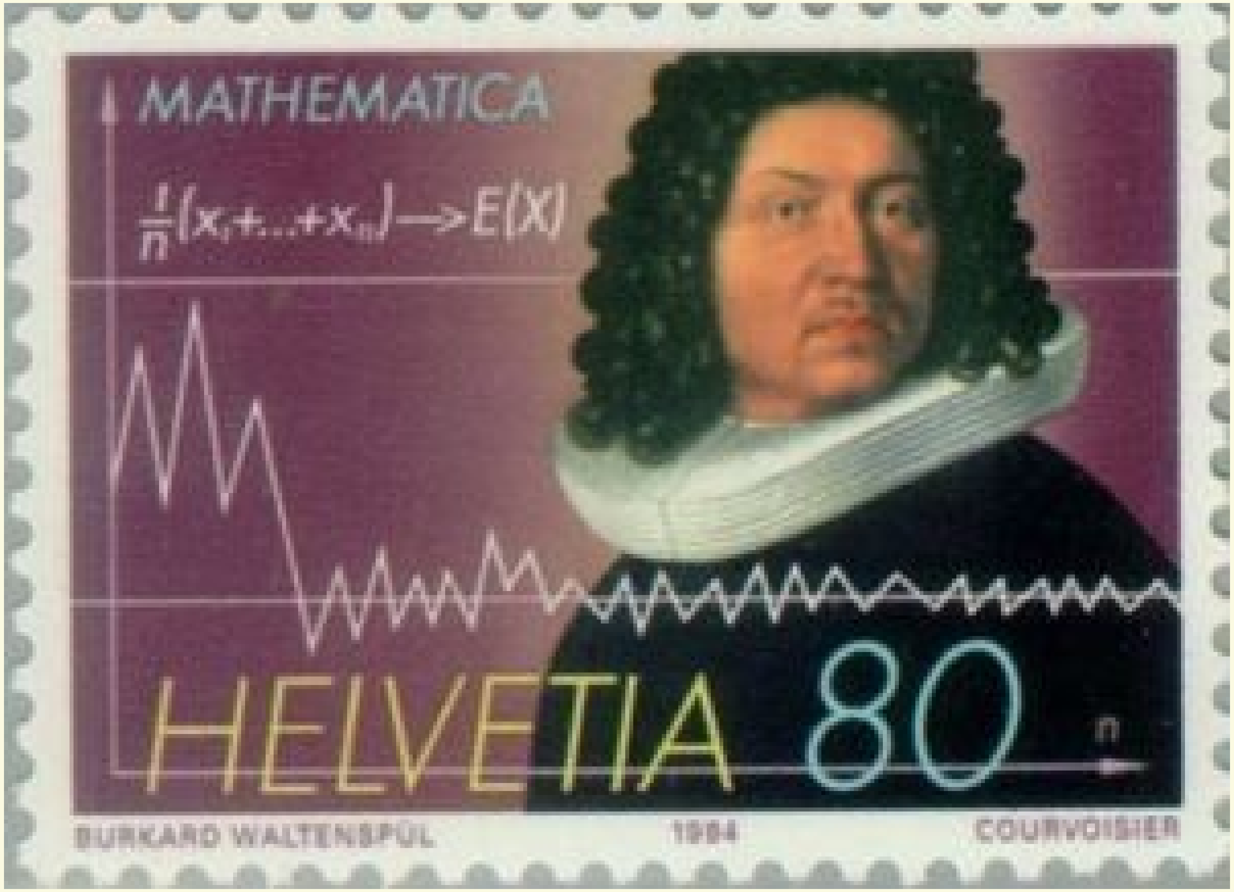
```
x = rand(1000,100);
pro = cumsum(x)./(1:1000)';
n = 1:1000;
figure(1)
plot(n,pro(:,1),n,pro(:,2),n,pro(:,3))
figure(2)
p = sum((abs(pro-0.5)>0.01)');
plot(n,p)
```



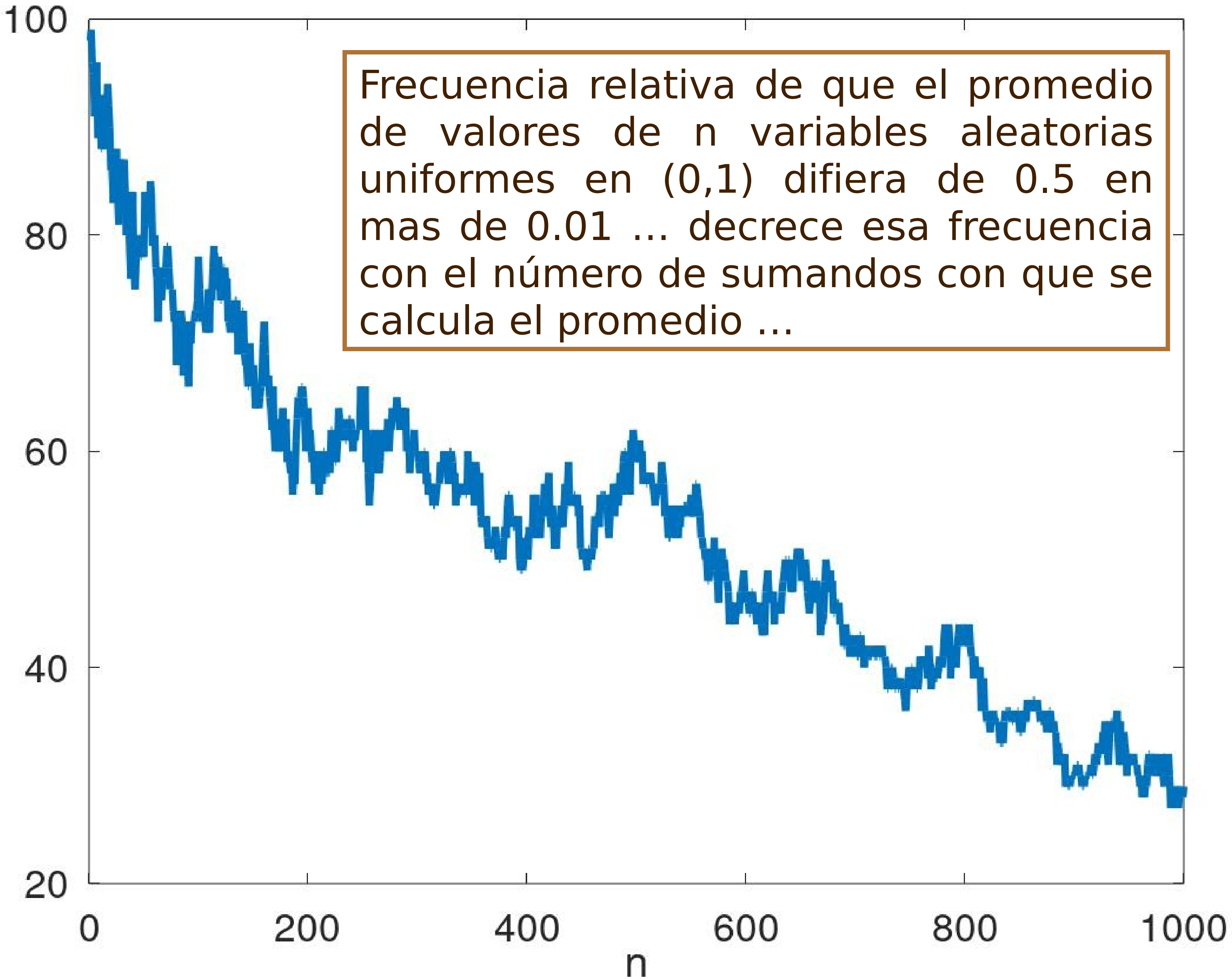
Promedio de valores de n variables aleatorias uniformes en (0,1), puede observarse la reducción de la variabilidad al aumentar el número de variables que se promedian

Ley Débil de los **Grandes Números**:

Jacob *Bernoulli* (1655-1705) en una estampilla suiza de 1994, creador de la ley de los grandes números



Observen el gráfico de la tendencia del promedio hacia la media y la noción de convergencia en probabilidad del promedio hacia la media en el ángulo superior izquierdo.



100
80
60
40
20
0
200
400
600
800
1000
n
Frecuencia relativa de que el promedio de valores de n variables aleatorias uniformes en (0,1) difiera de 0.5 en mas de 0.01 ... decrece esa frecuencia con el número de sumandos con que se calcula el promedio ...

Preguntas

Teorema Central del Límite

v.a. i.i.d. con , . Luego, para grande

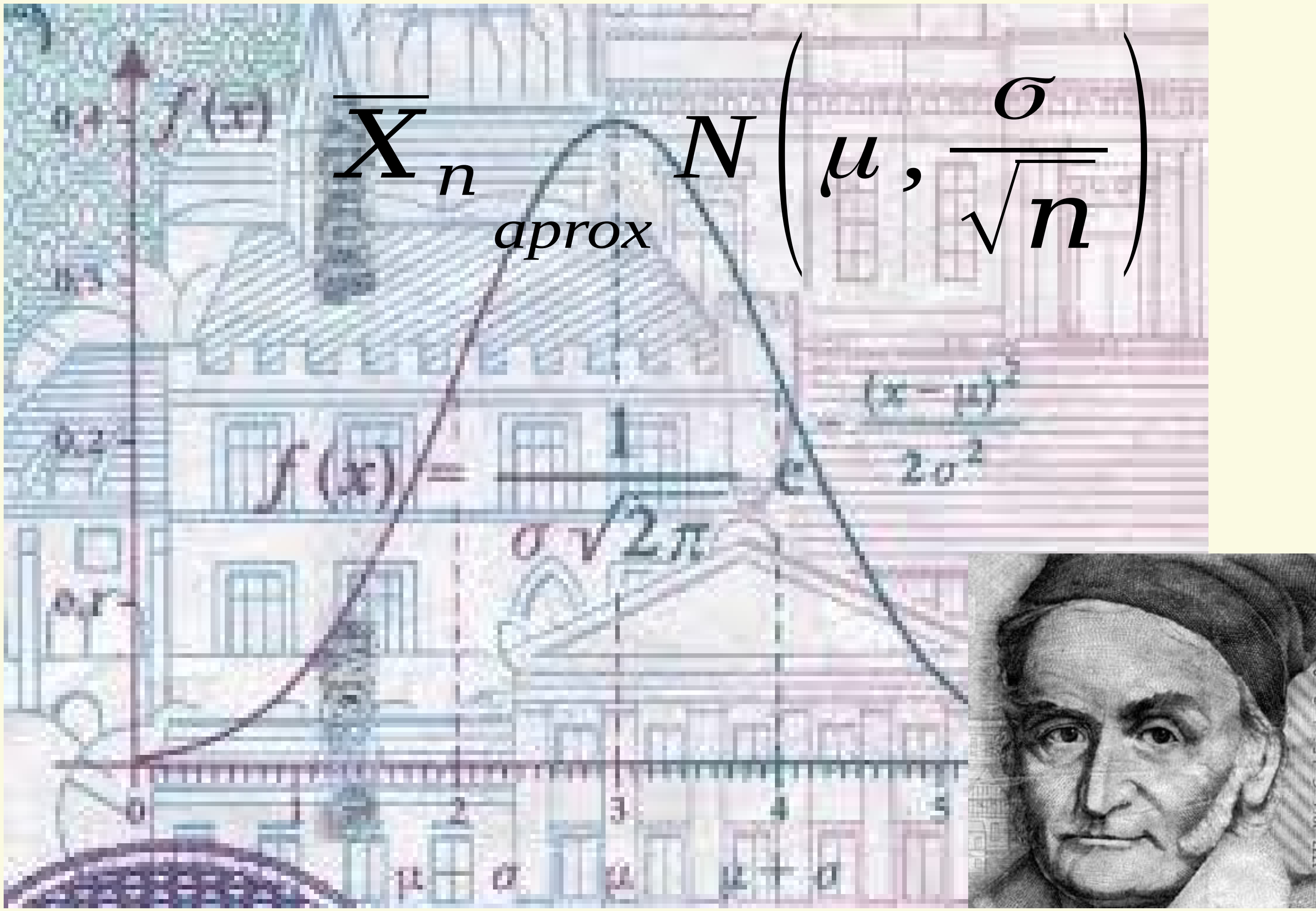
$$\overline{X}_n \underset{aprox}{} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\sum_{i=1}^{n} X_i \underset{aprox}{} N\left(n\mu, \sqrt{n}\sigma\right)$$

Teorema Central del Límite

v.a. i.i.d. con , . Luego, para grande

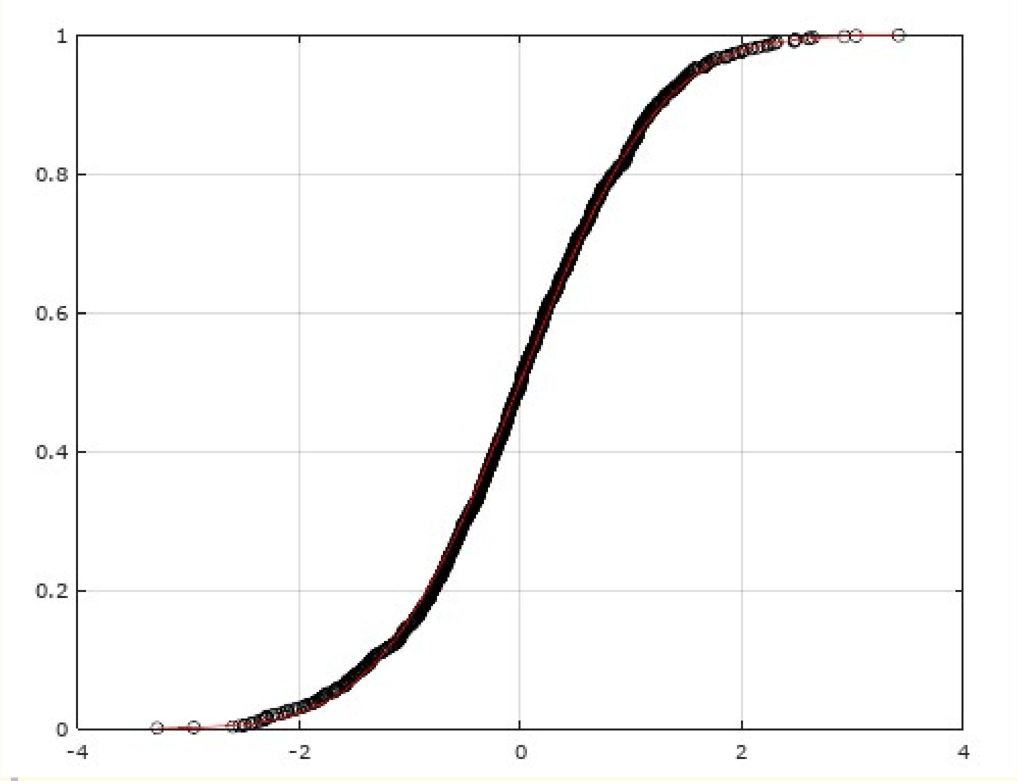
$$\overline{X}_n \underset{aprox}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

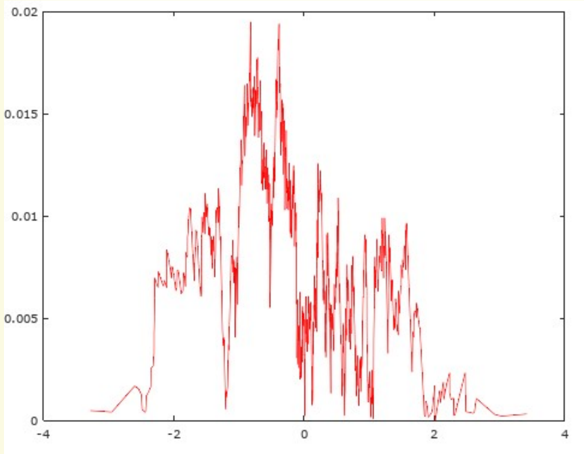


Teorema Central del Límite

v.a. i.i.d. con , . Luego, para grande

$$\overline{X}_n \underset{aprox}{} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$






```
R=rand(100,1000);
M=mean(R);
MO=sort(M);I=(1:1000)/1000;
m=mean(MO);s=std(MO);
z=(MO-m)/s;Fz=normcdf(z);
plot(z,I,'ko',z,Fz,'r-')
plot(z,abs(I-Fz),'r-')
```

Superposición de la distribución empírica de 1000 valores estandarizados de la media de 100 números aleatorios con distribución uniforme en (0,1) y la función de distribución normal estándar. En cuadro aparte la distancia entre ambas (para este caso fue menor que 0.02).

Teorema Central del Límite

Ejemplo 1

$$X_i \quad U(0,1)$$

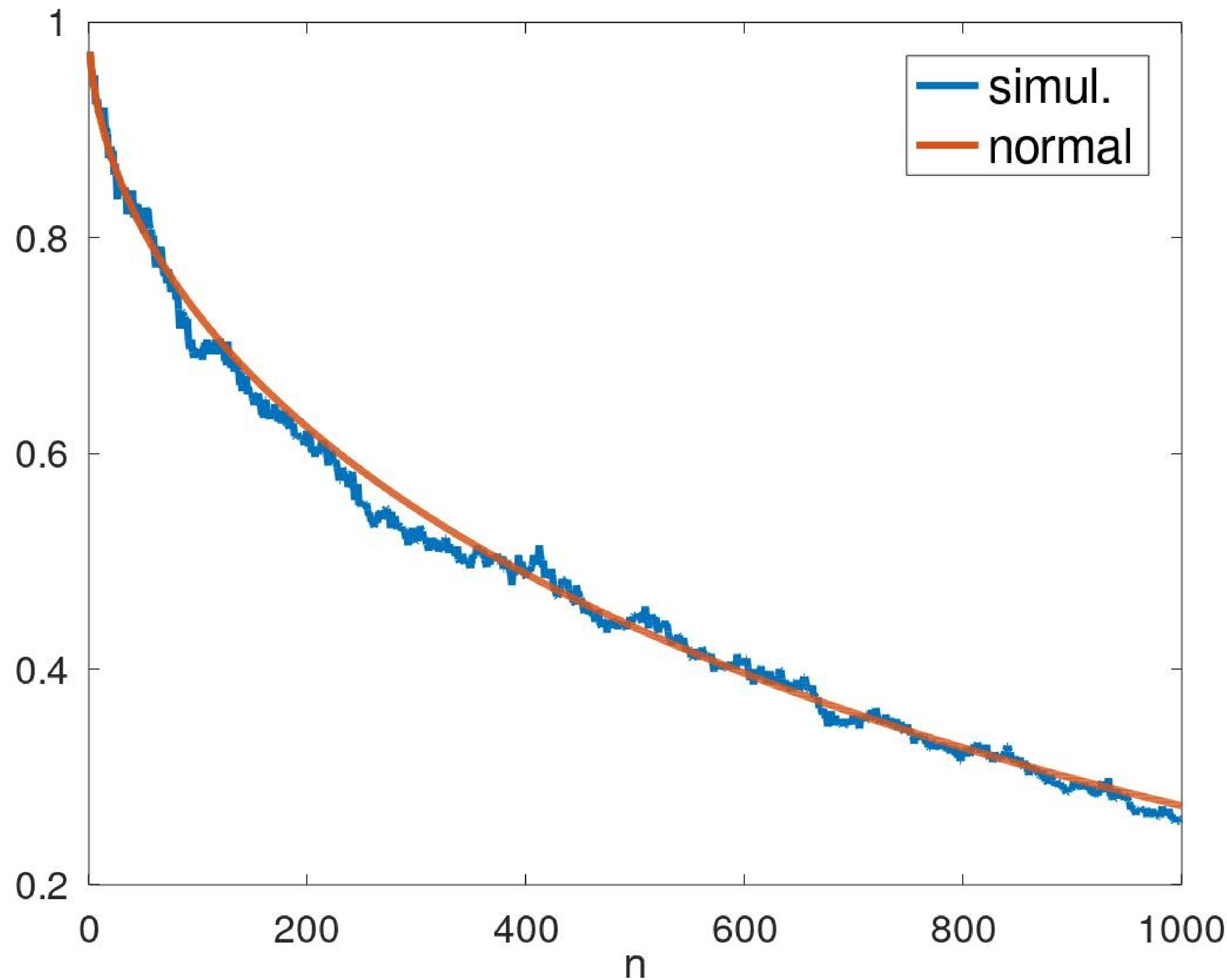
$$\overline{X}_n \;{}_{aprox}\; N\left(0.5, \frac{1}{\sqrt{12n}}\right)$$

$$P\left(¿\,\overline{X}_n - 0.5 \vee \geq \varepsilon\right) \approx 2\left(1 - \Phi\left(\varepsilon\sqrt{12n}\right)\right)$$

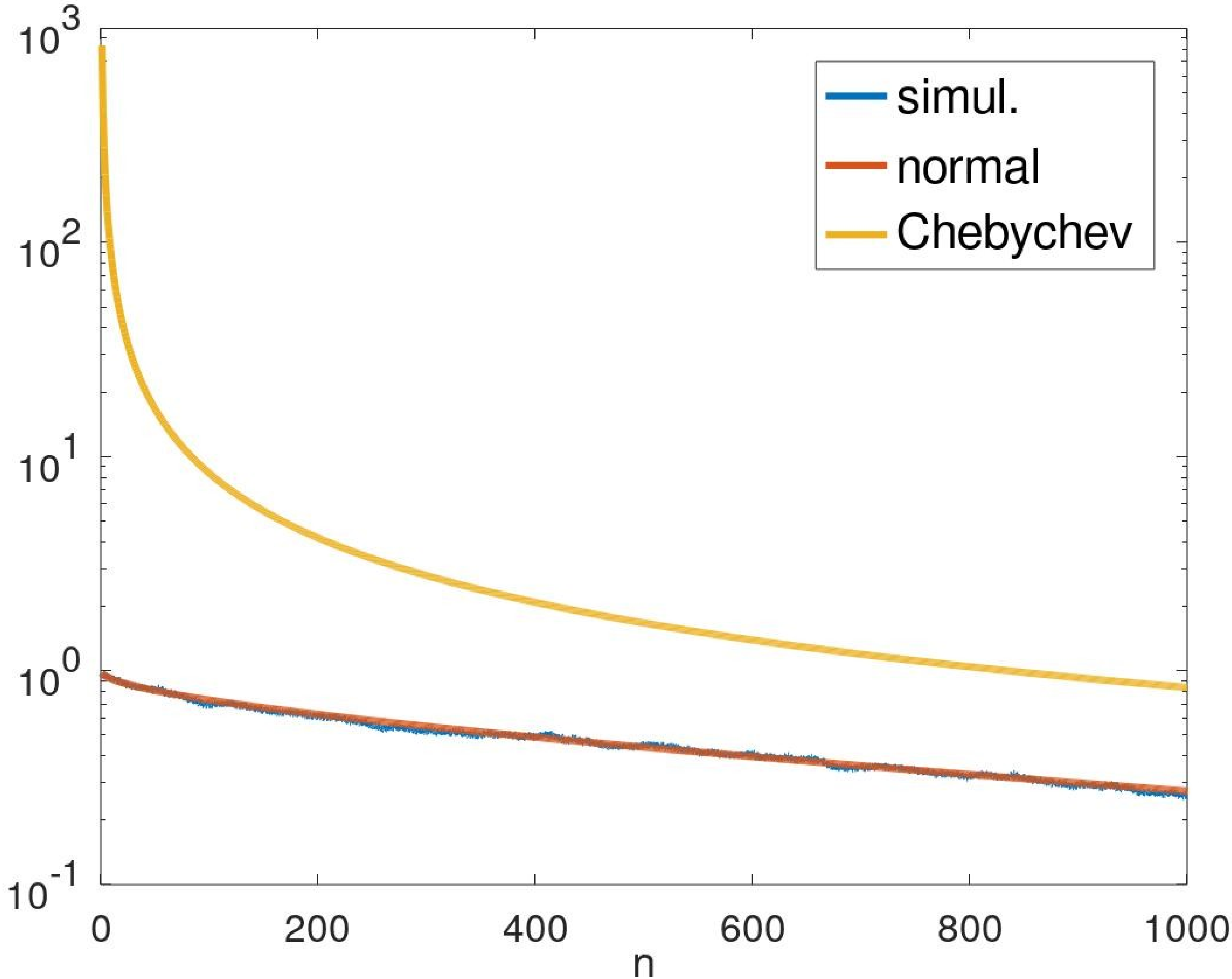
Teorema Central del Límite

Ejemplo 1 en Octave

```
x = rand(1000,1000);
pro = cumsum(x)./(1:1000)';
n = 1:1000;
p = sum((abs(pro-0.5)>0.01)')/1000;
figure(1)
plot(n,p,n,2*(1-normcdf(0.01*sqrt(12*n))))
figure(2)
semilogy(n,p,n,2*(1-normcdf(0.01*sqrt(12*n))))
hold on
semilogy(n,1./(12*0.01^2*n))
```



simul.
normal
1
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
n



10^3
10^2
10^1
10^0
10^{-1}
0
200
400
600
800
1000
n
simul.
normal
Chebychev

Teorema Central del Límite

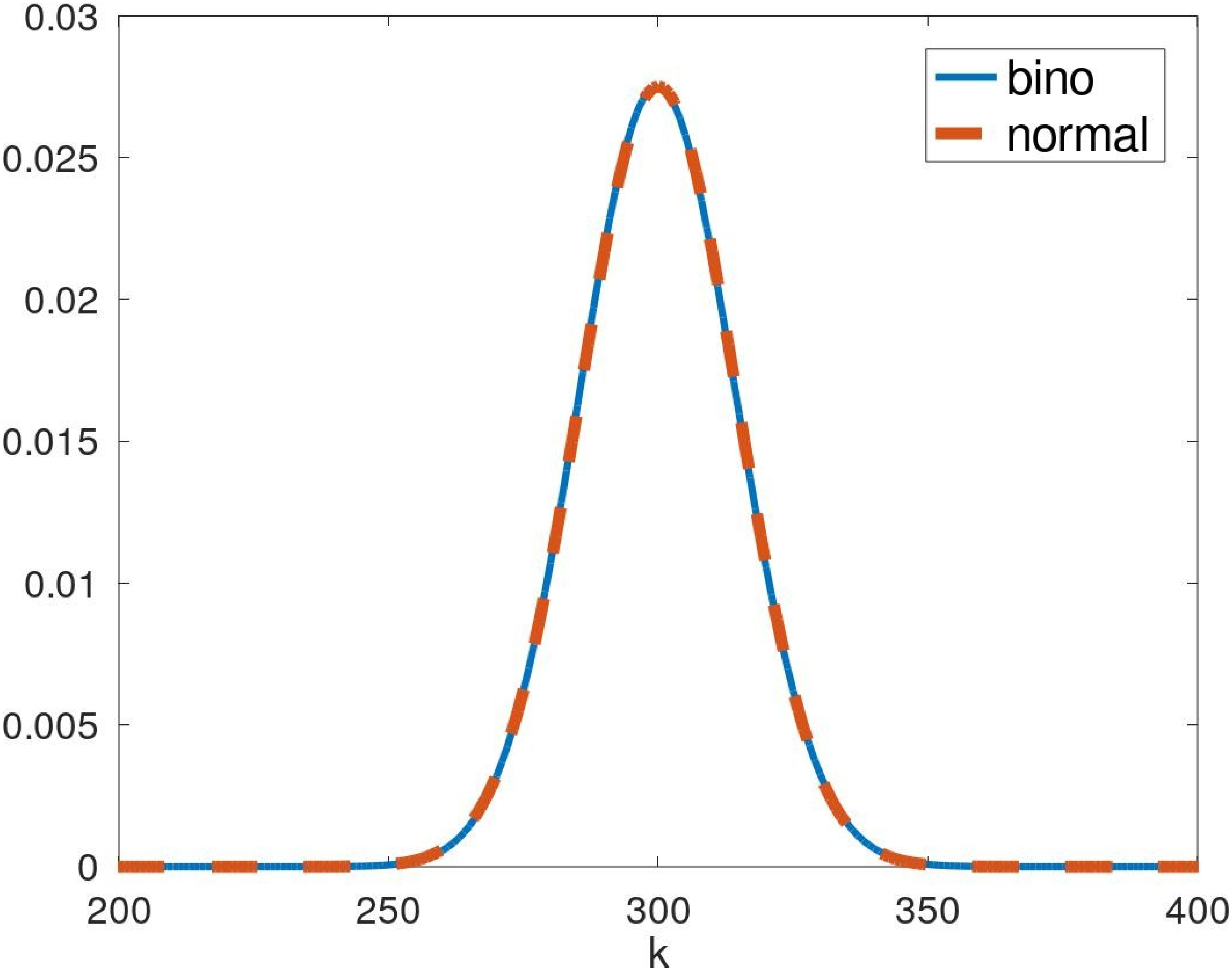
Ejemplo 2

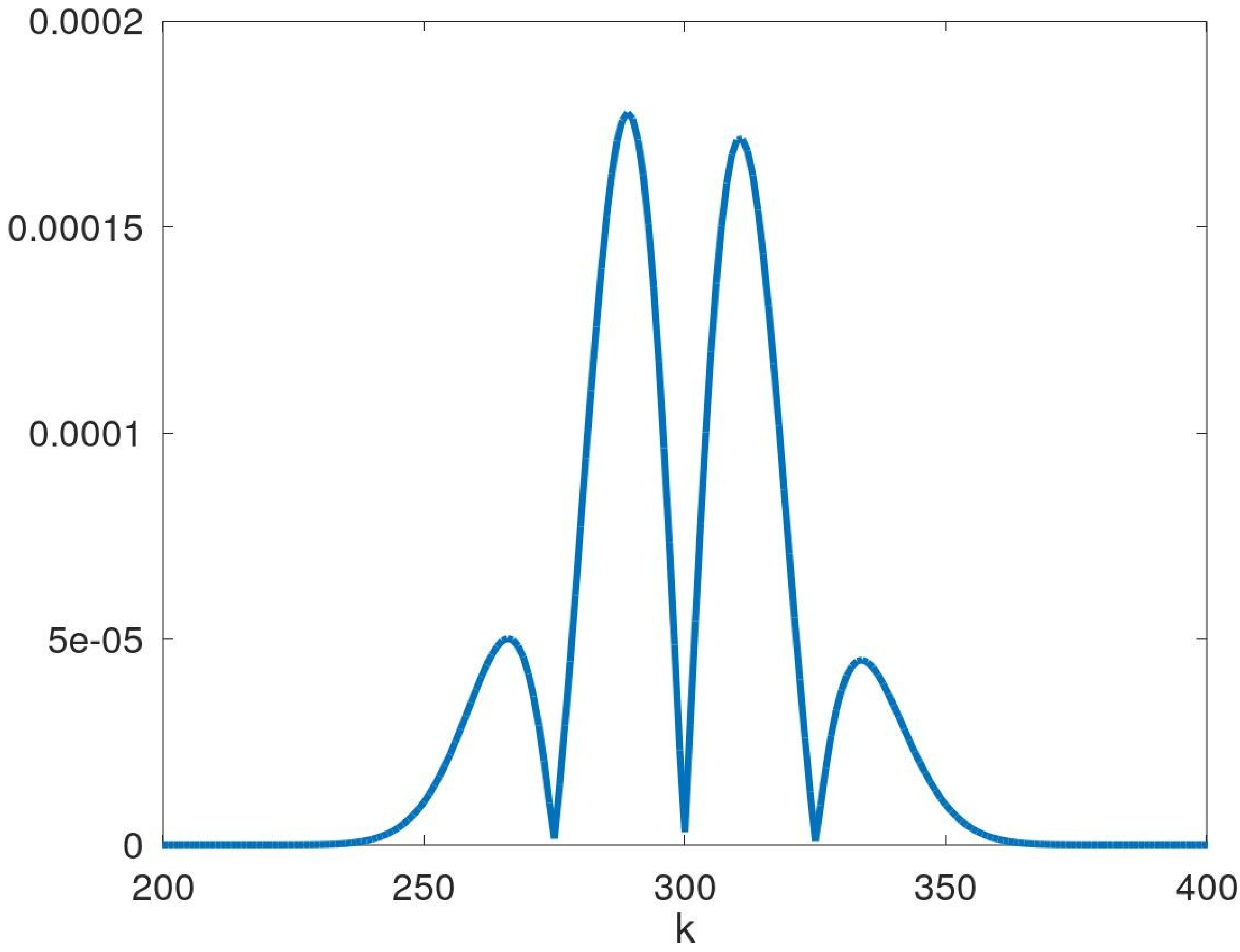
$$S_n \quad Bino(n,p) \underset{aprox}{,} N\left(np, \sqrt{np(1-p)}\right)$$

Teorema Central del Límite

Ejemplo 2 Octave

```
n = 1000; p = 0.3;
k = 0:n;
pe= binopdf(k,n,p);
pa= normcdf((k+0.5-n*p)/sqrt(n*p*(1-p)));
pa= pa-normcdf((k-0.5-n*p)/sqrt(n*p*(1-p)));
figure(1)
plot(k,pe,k,pa)
figure(2)
plot(k,abs(pe-pa))
```


0.03
0.025
0.02
0.015
0.01
0.005
0
200
250
300
350
400
k
bino
normal

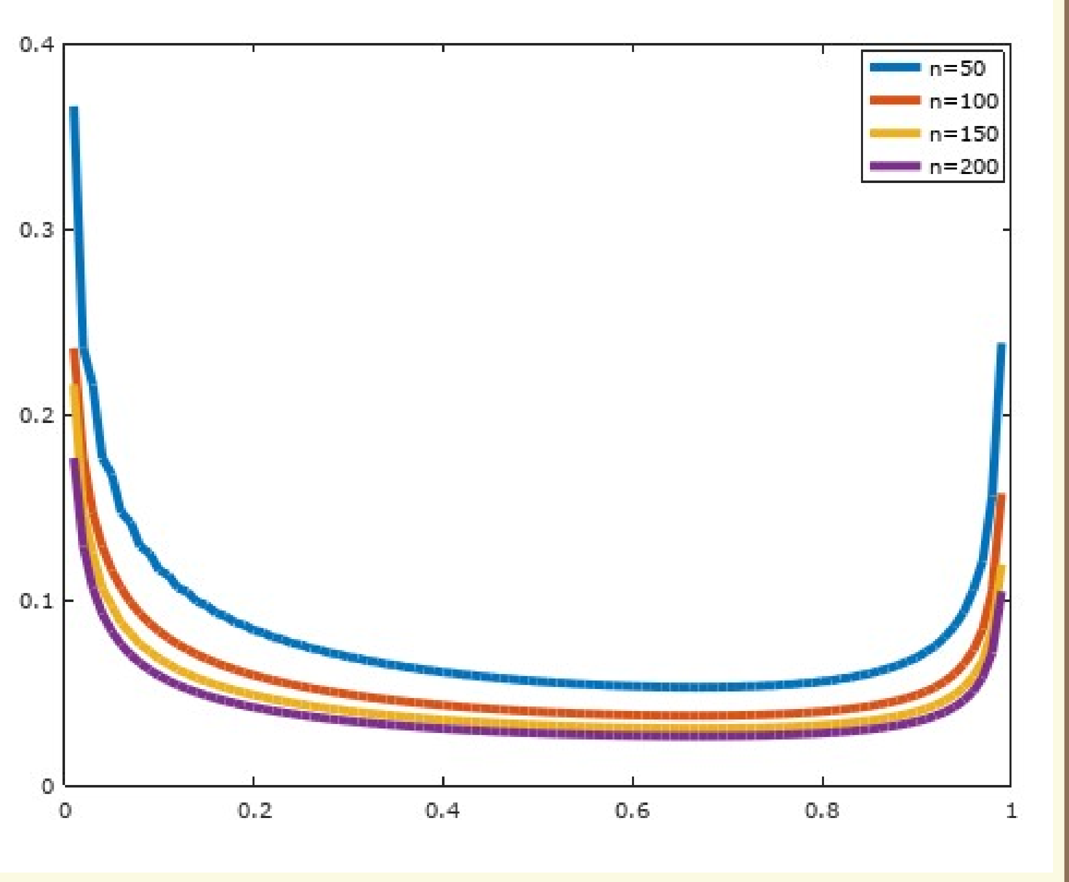


0.0002
0.00015
0.0001
5e-05
0
200
250
300
350
400
k

Aproximación normal de la binomial

Máxima diferencia entre los valores de la función de distribución binomial y la aproximación normal en función de p para varios valores de n. Puede observarse que en regiones intermedias de valores de p y para varios valores crecientes de n el error es menor que 0.1

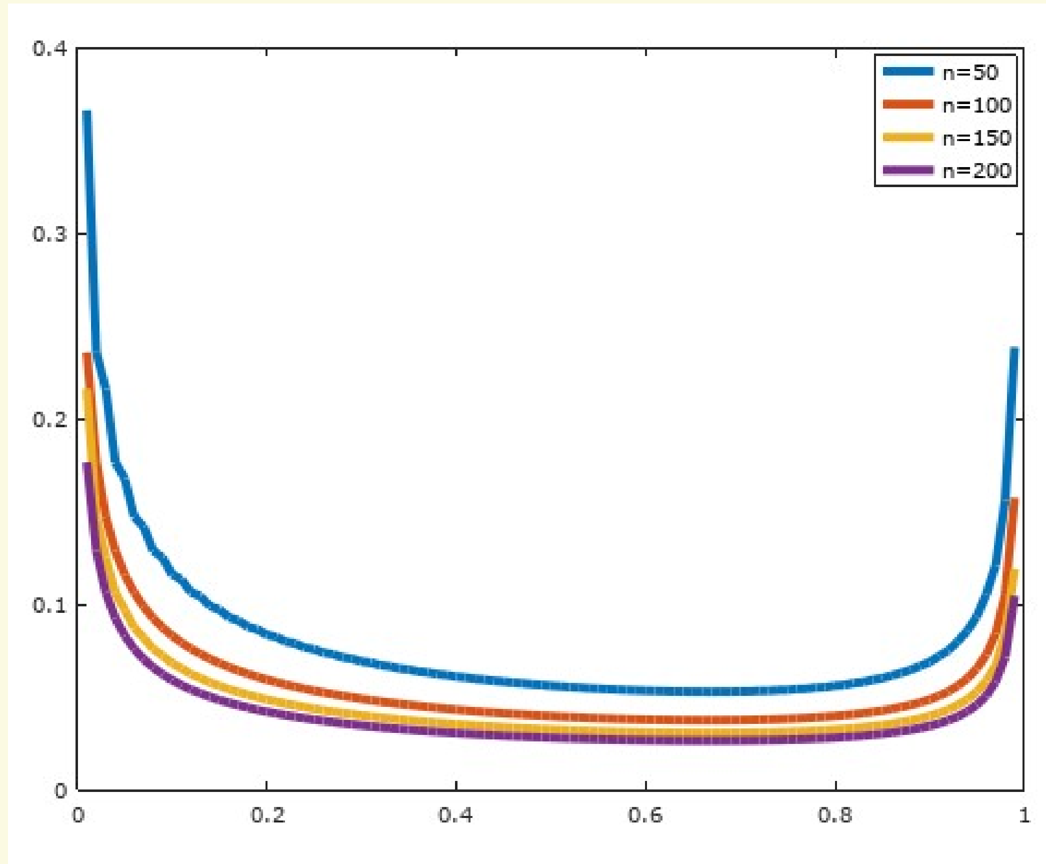
```
p=0.01:0.01:0.99;
for i=1:length(p)
for j=1:length(n)
P=binocdf(0:n(j),n(j),p(i));
m=n(j)*p(i);s=sqrt(n(j)*p(i)*(1-p(i)));
N=normcdf(0:n(j),m,s);
E(i,j)=max(abs(P-N));
end
end
plot(p',E,'linewidth',4)
legend('n=50','n=100','n=150','n=200')
```



Aproximación normal de la binomial

Máxima diferencia entre los valores de la función de distribución binomial y la aproximación normal en función de p para varios valores de n. Puede observarse que en regiones intermedias de valores de p y para varios valores crecientes de n el error es menor que 0.1

```
p=0.01:0.01:0.99;
for i=1:length(p)
for j=1:length(n)
P=binocdf(0:n(j),n(j),p(i));
m=n(j)*p(i);s=sqrt(n(j)*p(i)*(1-p(i)));
N=normcdf(0:n(j),m,s);
E(i,j)=max(abs(P-N));
end
end
plot(p',E,'linewidth',4)
legend('n=50','n=100','n=150','n=200')
```

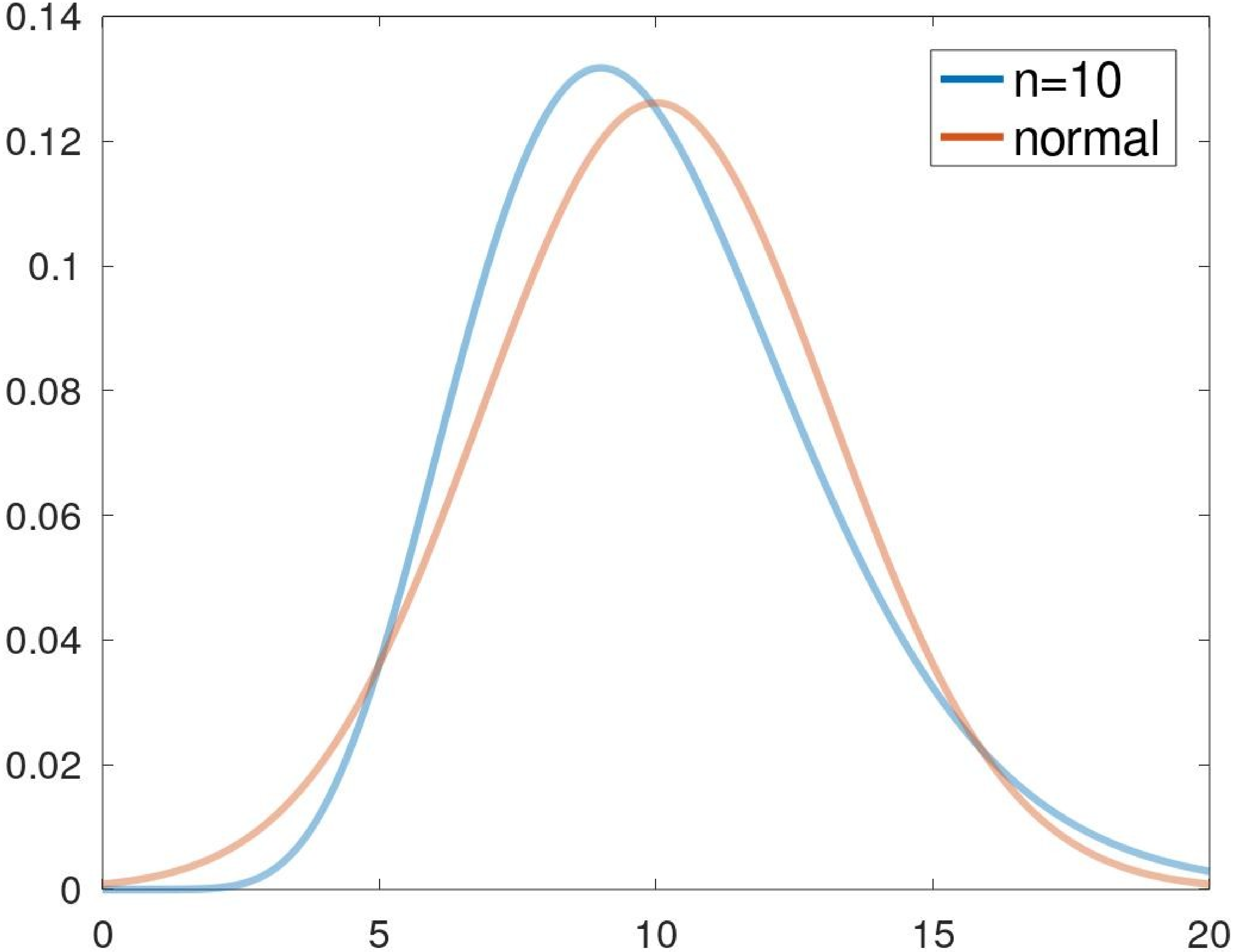


Teorema Central del Límite

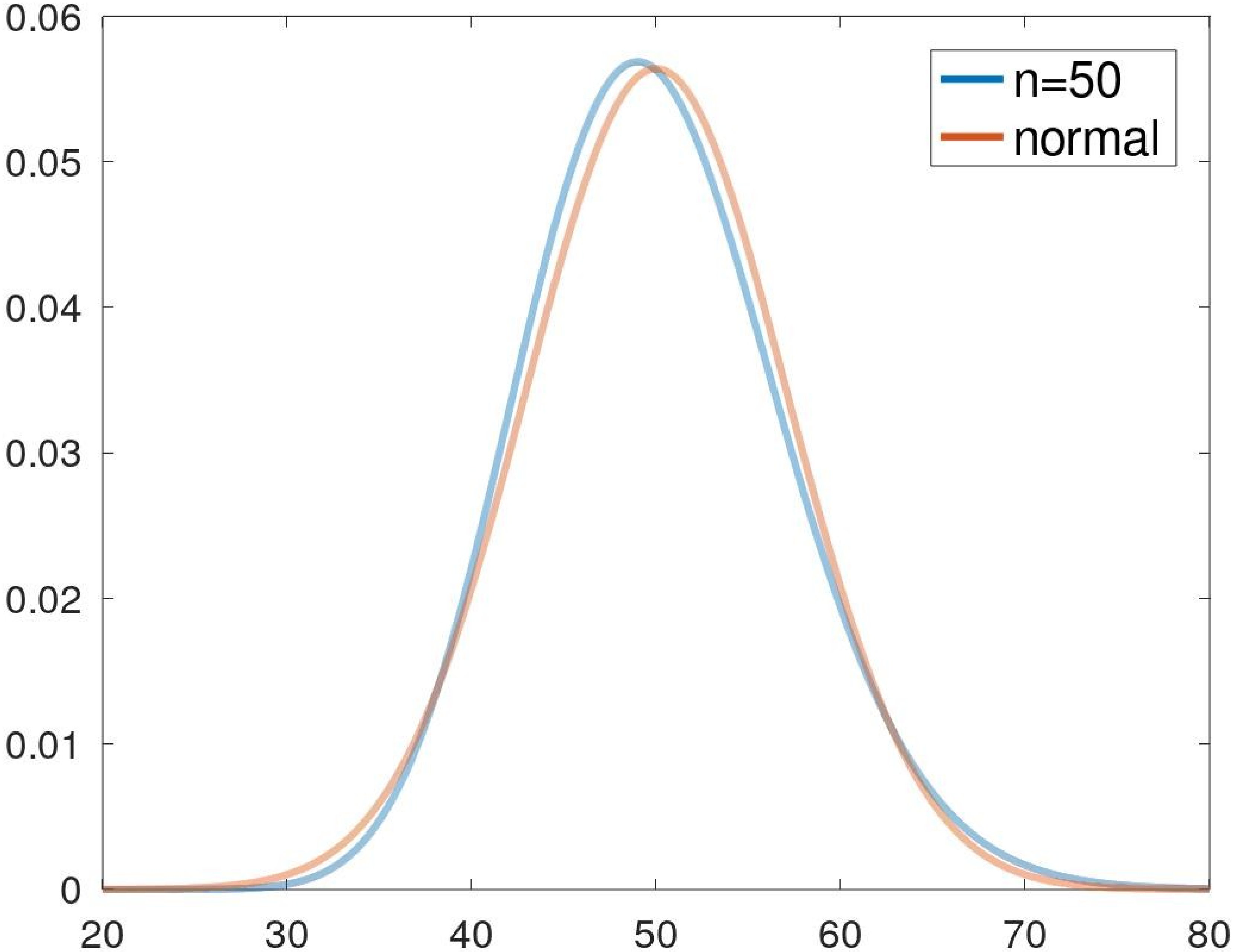
Ejemplo 3

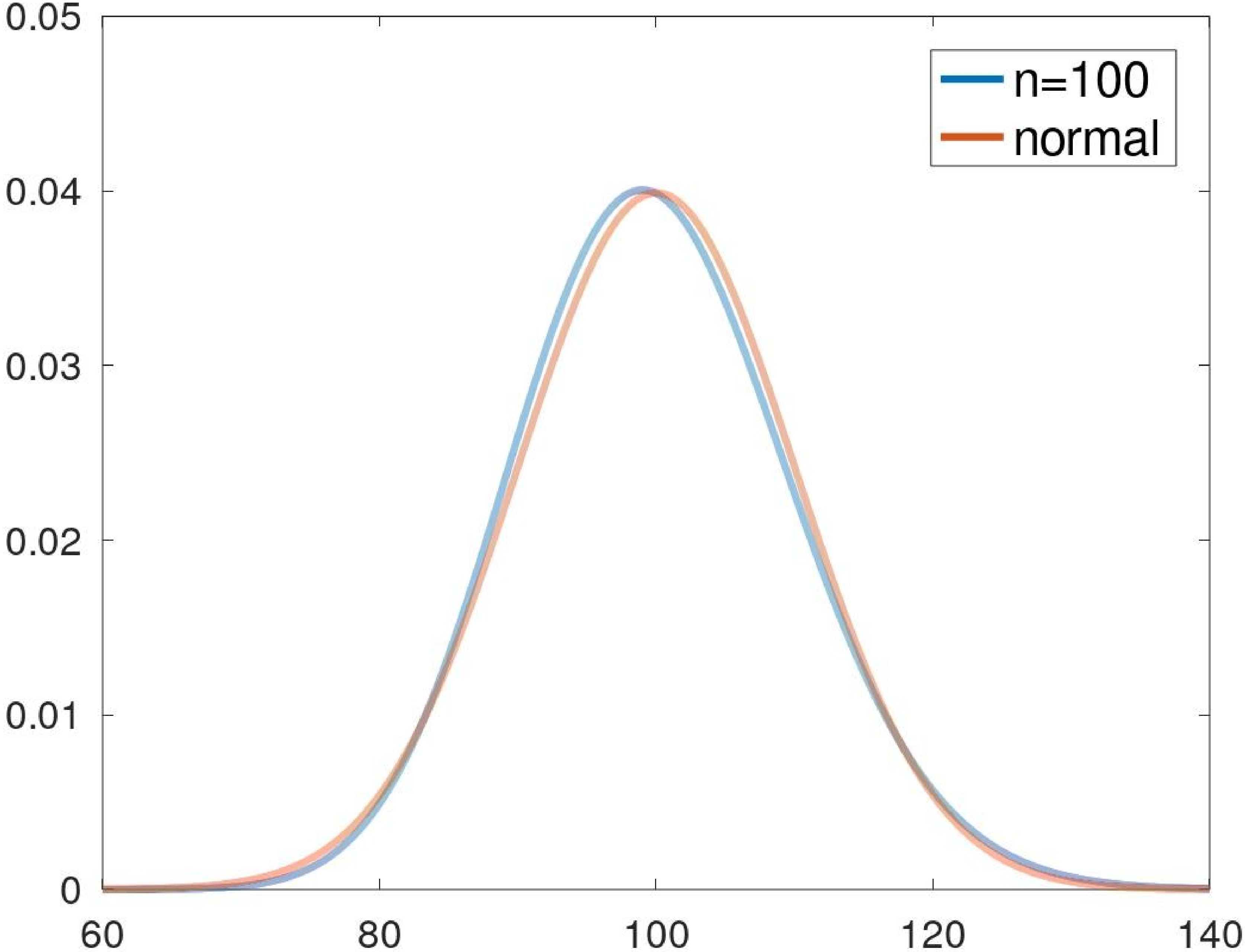
,

$$S_n \underset{aprox}{} N\left(n\lambda, \sqrt{n}\,\lambda\right)$$



n=10
normal
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20


0.06
0.05
0.04
0.03
0.02
0.01
0
20
30
40
50
60
70
80
n=50
normal



0.05
0.04
0.03
0.02
0.01
0
60
80
100
120
140
n=100
normal

Teorema Central del Límite

10. Los tiempos de servicio para los clientes que llegan a una caja de un supermercado pueden suponerse variables aleatorias independientes con un promedio de 1.5 minutos y una dispersión de 1 minuto.

   a) Calcular la probabilidad aproximada de que pueda atender a 100 clientes en menos de 2 horas de tiempo total de servicio de esta caja.

   b) Calcular el número de clientes n tal que la probabilidad de dar servicio a todos en menos de 2 horas sea aproximadamente 0.1.

Supongamos que se presentan n clientes y sea S_n el tiempo necesario para atenderlos. Sea T_k el tiempo de atención del cliente k con k entre 1 y n. Supongamos que las variables aleatorias T_k son independientes y claramente idénticamente distribuidas. Se tiene:

E(**Tk**) = 100 1.5 = 150 ;

V(**Mk**) = 100 1^2= 100

Teorema Central del Límite

Se desea calcular p = $P(S_{100} < 120)$. Se puede aproximar el valor de p usando el teorema central del límite dado que puede suponerse que "**n es grande**", … Para ello es necesario conocer el valor esperado y la varianza de S_{100}

Se tiene $E(S_{100}) = 150$ y $V(S_{100}) = 100 \; 1 = 100$ ya que los tiempos por cliente son variables aleatorias independientes e idénticamente distribuidas.

Entonces usando el teorema central del límite:

$$p = P\left(S_{100} < 120\right) \approx \Phi ¿$$

$$p \approx \Phi\left(-3\right) \approx 0.00135$$

Teorema Central del Límite

Se desea calcular *n* tal que p = P(S_n < 120) = 0.1. Se puede aproximar el cálculo usando el teorema central del límite suponiendo que "**n es grande**", … Para ello es necesario conocer el valor esperado y la varianza de S_n

Se tiene E(S_n) = 1.50 *n* y V(S_n) = n 1 = *n* ya que los tiempos por cliente son variables aleatorias independientes e idénticamente distribuidas.

Entonces usando el teorema central del límite:

$$P(S_n < 120) \approx \Phi ¿$$

(

Teorema Central del Límite

En un negocio de comida rápida los clientes pueden elegir entre tres combos: el pequeño a 70 pesos, el mediano a 140 pesos y el grande a 210 pesos. Cada cliente que llega, en forma independiente, elige el pequeño con probabilidad 0.3, el mediano con probabilidad 0.6 o el grande con probabilidad 0.1. Calcule en forma aproximada la probabilidad de que 100 clientes en conjunto gasten más de 13000 pesos.

Supongamos que se presentan n clientes y sea S_n el gasto realizado por los n clientes y M_k lo gastado por el cliente k con k entre 1 y n. Supongamos que las variables aleatorias M_k son independientes y claramente idénticamente distribuidas. Se tiene:

m	70	140	210
$P(M_k = m)$	0.3	0.6	0.1

E(**Mk**) = (70 0.3 + 140 0.6 + 210 0.1) = 126 ;

V(**Mk**) = 17640- 126^2 = 1764

Teorema Central del Límite

Se desea calcular p = $P(S_{100} > 13000)$. Se puede aproximar el valor de p usando el teorema central del límite dado que puede suponerse que "**n es grande**", … Para ello es necesario conocer el valor esperado y la varianza de S_{100}

Se tiene $E(S_{100}) = 100\ \ 126 = 12600$ y $V(S_{100}) = 100\ \ 1764 = 176400$ ya que los gastos por cliente son variables aleatorias independientes e idénticamente distribuidas.

Entonces usando el teorema central del límite:

$$p = P\left(S_{100} > 13000\right) = 1 - P\left(S_{100} \leq 13000\right) \approx 1 - \Phi\text{¿}$$

$$p \approx 1 - \Phi\left(0.9524\right) = \Phi\left(-0.9524\right) \approx 0.1705$$

Preguntas