Probabilidad y Estadística

Grado en Ingeniería Informática

Tema 6
Teoremas límite

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. La Ley de grandes números.
2. El teorema central del límite.

Objetivos principales

- Familiarizarse con los distintos modos de convergencia.
- Entender la importancia y aplicabilidad del TCL.

Introducción

Los resultados más célebres e importantes de la Teoría de la Probabilidad son los conocidos como **leyes de los grandes números**. Tales leyes no son otra cosa que teoremas que aseguran cierta convergencia de variables aleatorias bajo unas condiciones determinadas. Ahora bien, hay varios modos de convergencia de variables aleatorias

En este tema se analizan esos modos de convergencia y se muestran las principales leyes de los grandes números.

Ejemplo introductorio

Realizamos lanzamientos sucesivos e independientes de una moneda equilibrada.

$$X_i = \begin{cases} 1, & \text{si en el } i\text{-ésimo lanzamiento sale cara,} \\ 0, & \text{si en el } i\text{-ésimo lanzamiento sale cruz.} \end{cases}$$

$X_1, X_2, \ldots$ sucesión de v.a. independientes de Bernoulli, $\mathrm{B}(1; 1/2)$.

$$S_n = X_1 + \cdots + X_n.$$

($S_n \equiv$ número de caras en los n primeros lanzamientos)

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}.$$

($S_n/n \equiv$ proporción de caras en los n primeros lanzamientos)

Pregunta: ¿Qué ocurre con S_n/n cuando n es grande ($n \to \infty$)?

Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 1/2)$

Simulación: Realizamos $n = 5000$ lanzamientos de la moneda.

```
lista = RandomInteger[BinomialDistribution[1, 0.5], 5000];
Take[lista, 25]
```

{0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1}

`e[k_Integer] := Extract[lista, k]; p[n_] :=` $\frac{1}{n}\sum_{i=1}^{n}$ `e[i];`

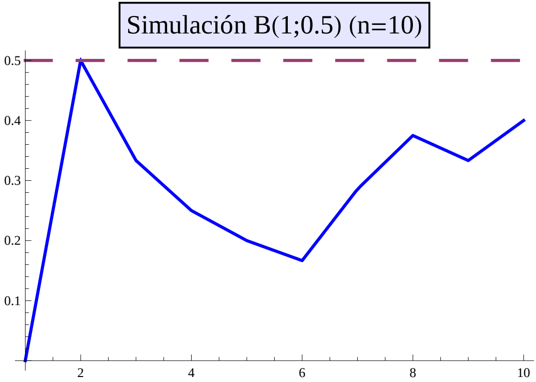
```
lista2 = {}; Do[lista2 = Append[lista2, {n, p[n]}], {n, 1, 5000}];
TableForm[Take[N[lista2], 10], TableHeadings → {None, {"n", "p[n]"}}]
```

n	p[n]
1.	0.
2.	0.5
3.	0.333333
4.	0.25
5.	0.2
6.	0.166667
7.	0.285714
8.	0.375
9.	0.333333
10.	0.4

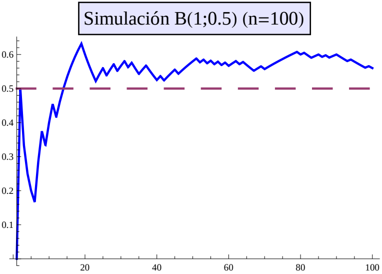
Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 1/2)$

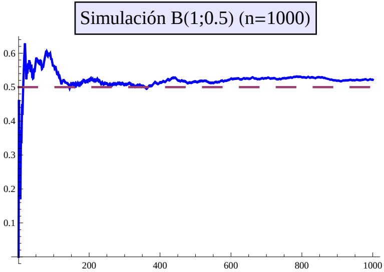
```
int = Interpolation[lista2, InterpolationOrder → 1];
Plot[{int[x], 0.5}, {x, 1, 10}, PlotRange → All,
 PlotStyle → {{Blue, Thickness[.006]}, {Thickness[.006], Dashing[{0.05, 0.05}]}}]
```

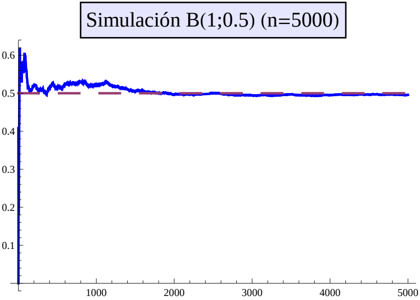
n	p[n]
1.	0.
2.	0.5
3.	0.333333
4.	0.25
5.	0.2
6.	0.166667
7.	0.285714
8.	0.375
9.	0.333333
10.	0.4

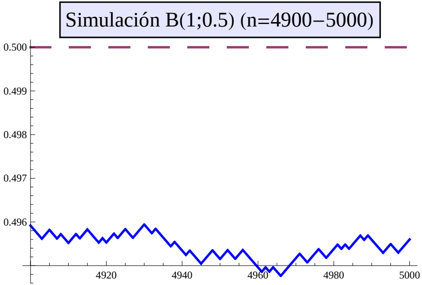


{0, 1, 0, 0, 0, 0, 1, 1, 0, 1}


Simulación B(1;0.5) (n=100)
0.6
0.5
0.4
0.3
0.2
0.1
20
40
60
80
100


Simulación B(1;0.5) (n=1000)
0.6
0.5
0.4
0.3
0.2
0.1
200
400
600
800
1000


Simulación B(1;0.5) (n=5000)
0.6
0.5
0.4
0.3
0.2
0.1
1000
2000
3000
4000
5000


Simulación B(1;0.5) (n=4900−5000)
0.500
0.499
0.498
0.497
0.496
4920
4940
4960
4980
5000

Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 0{,}75)$

Simulación: Realizamos $n = 5000$ lanzamientos de la moneda trucada con probabilidad de cara 0,75..

```
lista3 = RandomInteger[BinomialDistribution[1, 0.75], 5000];
Take[lista3, 25]
```

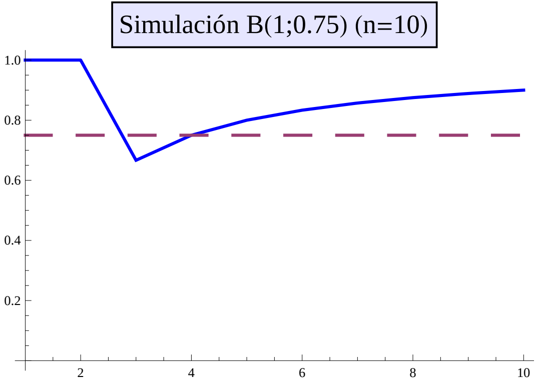
{1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0}

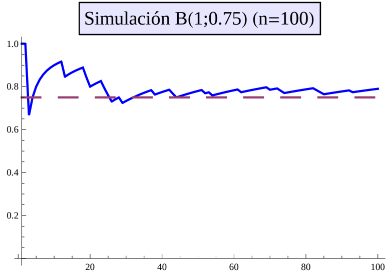
$$\texttt{e2[k\_Integer] := Extract[lista3, k]; p2[n\_] :=} \frac{1}{n}\sum_{i=1}^{n} \texttt{e2[i];}$$

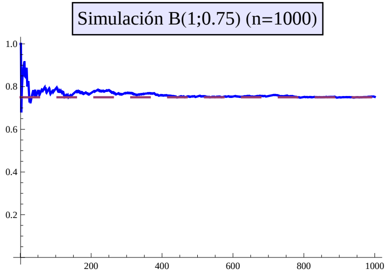
n	p2[n]
1.	1.
2.	1.
3.	0.666667
4.	0.75
5.	0.8
6.	0.833333
7.	0.857143
8.	0.875
9.	0.888889
10.	0.9

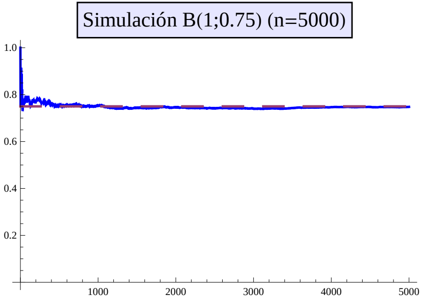
Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 0{,}75)$

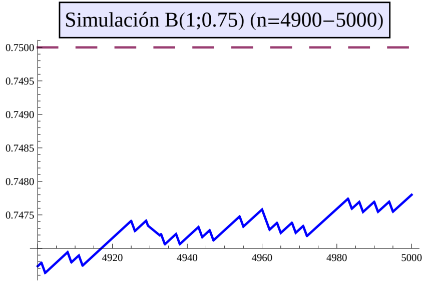
n	p2[n]
1.	1.
2.	1.
3.	0.666667
4.	0.75
5.	0.8
6.	0.833333
7.	0.857143
8.	0.875
9.	0.888889
10.	0.9




Simulación B(1;0.75) (n=100)
1.0
0.8
0.6
0.4
0.2
20
40
60
80
100


Simulación B(1;0.75) (n=1000)
1.0
0.8
0.6
0.4
0.2
200
400
600
800
1000


Simulación B(1;0.75) (n=5000)
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
5000


Simulación B(1;0.75) (n=4900−5000)
0.7500
0.7495
0.7490
0.7485
0.7480
0.7475
4920
4940
4960
4980
5000

Conclusiones de las simulaciones

- Si $X \sim \mathrm{B}(1; p)$, parece claro que, en algún sentido,

$$S_n/n \to p, \qquad \text{cuando } n \to \infty.$$

- $p = \mathrm{E}X$, cuando $X \sim \mathrm{B}(1; p)$.
- Como S_n/n es la media de la muestra $X_1, \ldots, X_n$ (**media muestral**), parece razonable que, en general,

$$S_n/n \to \mu, \qquad \text{cuando } n \to \infty \text{ (en algún sentido).}$$

Objetivo del tema: Estudiar la convergencia de una sucesión de variables aleatorias. Especialmente nos interesa estudiar la convergencia de la media muestral.

Nota: Hay diferentes formas de convergencia de variables aleatorias.

Convergencia en media cuadrática

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X en media cuadrática**, y escribimos $X_n \xrightarrow{\text{m-2}} X$, si

$$\mathrm{E}(X_n - X)^2 \to 0, \qquad n \to \infty.$$

Idea: Si $X_n \xrightarrow{\text{m-2}} X$, el valor esperado de la distancia de las variables de la sucesión a la variable límite tiende a cero, cuando n crece.

Teorema: Ley de grandes números en media cuadrática

Sean $X_1, X_2, \ldots,$ variables independientes con igual media μ y varianza σ^2. Se tiene

$$\frac{S_n}{n} \xrightarrow{\text{m-2}} \mu, \quad n \to \infty.$$

Pregunta: Se pueden rebajar las condiciones del teorema anterior.

Convergencia en probabilidad

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X en probabilidad**, y escribimos $X_n \xrightarrow{\mathrm{P}} X$, si

$$\text{para cada } \epsilon > 0,\ \mathrm{P}(|X_n - X| > \epsilon) \to 0, \quad n \to \infty.$$

Idea: Si $X_n \xrightarrow{\mathrm{P}} X$, a largo plazo, cada variable de la sucesión está muy concentrada entorno a la variable límite.

Nota: Relación entre los modos de convergencia

- $X_n \xrightarrow{\text{m-2}} X \Longrightarrow X_n \xrightarrow{\mathrm{P}} X$.
- $X_n \xrightarrow{\mathrm{P}} X \not\Longrightarrow X_n \xrightarrow{\text{m-2}} X$.

Teorema: Ley débil de grandes números

Sean $X_1, X_2, \ldots,$ variables incorreladas con igual media μ y varianza σ^2. Se tiene

$$\frac{S_n}{n} \xrightarrow{\mathrm{P}} \mu, \quad n \to \infty.$$

Nota histórica: La ley débil de Bernoulli

Históricamente, la primera ley de grandes números fue obtenida por Bernoulli en el siglo XVIII. La ley se refiere a la proporción de caras que se obtienen en n lanzamientos de una p-moneda (moneda que cae de cara con probabilidad p) y al modo en que esa proporción se aproxima a p cuando el número de lanzamientos es grande. En términos más o menos coloquiales dice lo siguiente:

Haciendo un número suficientemente grande de lanzamientos, podemos conseguir que la probabilidad de que dicha proporción se diferencie de p en menos de una cantidad prefijada esté tan próxima a 1 como queramos.

Teorema: Ley débil de Bernoulli

Si $X_1, X_2, \ldots$ son independientes y tienen la misma distribución de Bernoulli de parámetro p, entonces

$$\frac{S_n}{n} \xrightarrow{\mathrm{P}} p.$$

Convergencia casi segura

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X casi seguramente** o **con probabilidad 1**, y escribimos $X_n \xrightarrow{\text{c.s.}} X$, si

$$\mathrm{P}(X_n \to X) = \mathrm{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Idea: Si $X_n \xrightarrow{\text{c.s.}} X$, salvo en un conjunto de probabilidad nula se da la convergencia puntual de X_n a X.

Nota: Relación entre los modos de convergencia

- $X_n \xrightarrow{\text{c.s.}} X \Longrightarrow X_n \xrightarrow{\mathrm{P}} X$.
- $X_n \xrightarrow{\mathrm{P}} X \not\Longrightarrow X_n \xrightarrow{\text{c.s.}} X$.
- $X_n \xrightarrow{\text{c.s.}} X \not\Longrightarrow X_n \xrightarrow{\text{m-2}} X$.
- $X_n \xrightarrow{\text{m-2}} X \not\Longrightarrow X_n \xrightarrow{\text{c.s.}} X$.

Convergencia casi segura

Sea $X, X_1, X_2, \dots$ variables aleatorias. Decimos que X_n **converge a X casi seguramente** o **con probabilidad 1**, y escribimos $X_n \xrightarrow{\text{c.s.}} X$, si

$$\mathrm{P}(X_n \to X) = \mathrm{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Idea: Si $X_n \xrightarrow{\text{c.s.}} X$, salvo en un conjunto de probabilidad nula se da la convergencia puntual de X_n a X.

Teorema: Ley fuerte de grandes números (Kolmogorov)

Sean $X_1, X_2, \dots,$ variables independientes con igual distribución y media μ. Se tiene

$$\frac{S_n}{n} \xrightarrow{\text{c.s.}} \mu, \quad n \to \infty.$$

Nota: La Ley fuerte de grandes números es uno de los resultados más importantes de la Teoría de la probabilidad.

Resumen ideas principales

- Existen diferentes modos de convergencia de una sucesión de v.a.
- Los **modos de convergencia fuertes** son: convergencia casi segura (principalmente) y convergencia en media cuadrática. Los **modos de convergencia débiles** son: convergencia en probabilidad y sobre todo convergencia en distribución (siguiente punto).
- La **ley fuerte de grandes números** asegura que la media (aritmética) de una sucesión de variables independientes y con igual distribución (iid) converge con probabilidad 1 a μ (media poblacional).
- Esto resuelve el problema introductorio y confirma las simulaciones. La ley fuerte de grandes números para variables de Bernoulli afirma que si $X_1, X_2, \ldots$ son independientes y $\mathrm{B}(1; p)$, entones
$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \longrightarrow p, \quad \text{casi seguramente.}$$
- Al realizar experimentos sucesivos e independientes, el porcentaje de veces que acaece un suceso de probabilidad p será aproximadamente el $p \times 100\,\%$. De esta forma se recupera la definición frecuentista de la probabilidad.

Convergencia en distribución

Sea $X, X_1, X_2, \ldots$ variables aleatorias y sean F y F_n las funciones de distribución de X y X_n, respectivamente. Decimos que X_n **converge a** X **en distribución**, y escribimos $X_n \xrightarrow{\mathcal{D}} X$, si

$$F_n(x) \to F(x), \quad \text{para todo } x \in C_F, \quad n \to \infty,$$

donde C_F es el conjunto de puntos de continuidad de F.

Idea: Si $X_n \xrightarrow{\mathcal{D}} X$, la probabilidad de $\{X_n \leq a\}$ (a constante) es parecida a la probabilidad de $\{X \leq a\}$ (cuando n es grande).

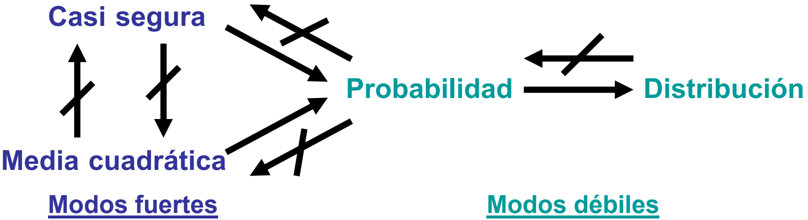
Pregunta: ¿Por qué no $\forall x \in \mathbb{R}$ en la definición anterior?

Observación: Si $X_n = 1/n$ cte. y $X = 0$ cte. Es razonable esperar que $X_n \to X$ en cualquier modo de convergencia.

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0, \end{cases} \quad \text{y} \quad F_n(x) \to \begin{cases} 0 & \text{si } x \leq 0, \\ 1 & \text{si } x > 0, \end{cases} \quad \textbf{no es f.d.}$$

Nota: La convergencia casi segura y en media cuadrática se llaman **modos de convergencia fuertes** (sobre todo la convergencia c.s.). La convergencia en probabilidad y en distribución se denominan **modos de convergencia débiles** (sobre todo la convergencia en distribución).

Esquema de las relaciones mutuas



Nota: Los recíprocos de las anteriores implicaciones no son ciertos en general.

Idea: *Una variable que es el resultado de la suma de muchos efectos independientes entre sí sin que ninguno domine al total es aproximadamente normal.*

El Teorema Central del Límite (TCL)

Sean $X_1, X_2, \ldots,$ variables independientes y con la misma distribución de probabilidad tal que $EX_i = \mu$, $\mathrm{Var}X_i = \sigma^2 > 0$ $(i \geq 1)$. Entonces, si $S_n = X_1 + \cdots + X_n$, se tiene

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0;1) \quad (n \uparrow \infty).$$

Nota: El TCL es uno de los resultados más importantes de la Teoría de la probabilidad.

Observaciones importantes sobre el TCL

- Cuando n es grande, el TCL asegura que

$$X_1 + \cdots + X_n \approx \mathrm{N}\left(n\mu; \sigma\sqrt{n}\right) \quad \text{(en distribución).}$$

- En general, n se considera "grande" a partir de 30, $n \geq 30$. Sin embargo, esta regla genérica para una "buena aproximación" se debe tomar con cautela ya que depende de la distribución de partida (como veremos enseguida).
- **Importante:** No se impone ninguna *hipótesis distribucional* sobre las X_i salvo que tengan segundo momento finito. **El resultado es válido para <u>cualquier distribución</u>** (discreta o continua) verificando estas condiciones.
- Para calcular probabilidades de *cualquier* suma (grande) de v.a. i.i.d., es suficiente aproximar la suma por una variable normal y utilizar las tablas de probabilidad de la normal.

Ejemplo: La demanda diaria de un producto tiene media 30 y desviación típica 6. Supuesta la independencia de la demanda de cada día respecto de los restantes:

(a) ¿Cuál es la distribución (aproximada) de la demanda en un periodo de 182 días?

(b) ¿Cuál es la probabilidad de que en 182 días, el número de unidades demandadas supere 6370 unidades?

Observaciones importantes:

- La variable aleatoria *demanda* es una variable discreta que toma los valores $0, 1, 2, \ldots$ (Muy distinta de la normal.)
- ¡¡¡¡¡Ni siquiera conocemos la distribución de probabilidad de la variable demanda!!!!! (Sólo sabemos su media y desviación.)

El **Teorema de De Moivre-Laplace** establece una aproximación normal de la distribución binomial. Es un caso particular del TCL. Históricamente, primero se mostró este resultado antes de obtener el TCL en su versión general.

El Teorema de De Moivre-Laplace

Sean $X_1, \ldots, X_n, \ldots$ v.a. independientes con la misma distribución de probabilidad de Bernoulli $\mathrm{B}(1; p)$. Entonces:

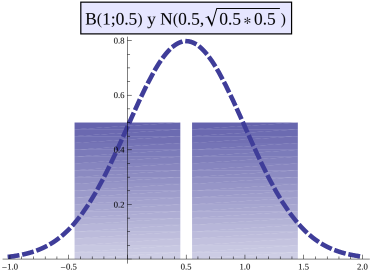
$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0; 1) \quad (n \uparrow \infty),$$

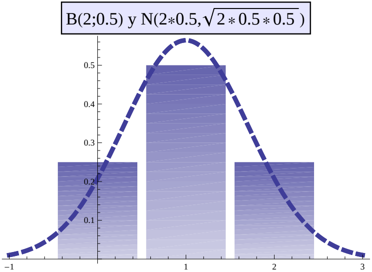
donde $S_n = X_1 + \cdots + X_n \sim \mathrm{B}(n; p)$ (Binomial).

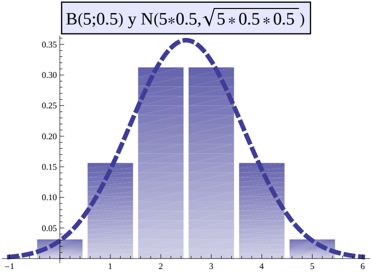
Observaciones:

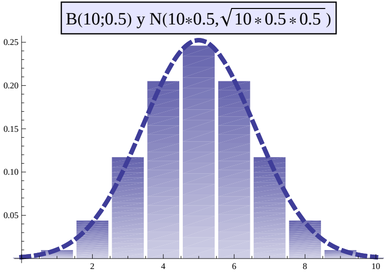
- Cuando n es grande, $\mathrm{B}(n; p) \approx \mathrm{N}\left(np; \sqrt{np(1-p)}\right)$ (en distribución).
- Cuando n es grande, podemos calcular probabilidades binomiales mediante la aproximación a la normal.

El Teorema de De Moivre-Laplace

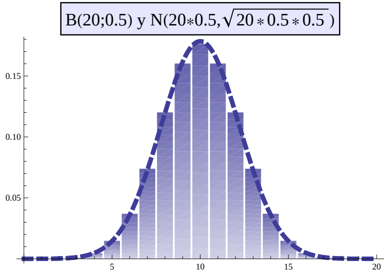


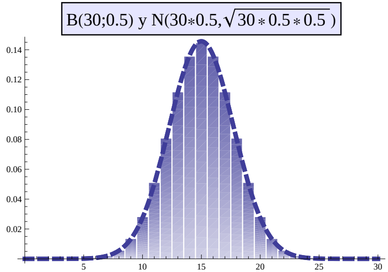


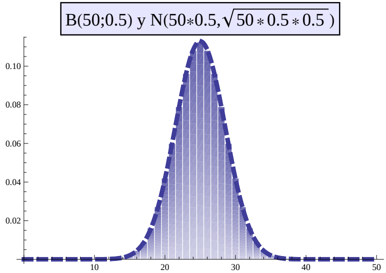


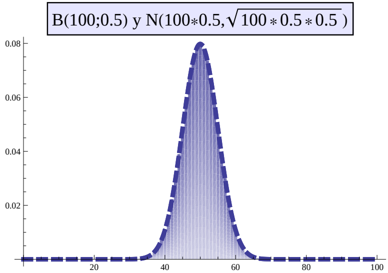


El Teorema de De Moivre-Laplace

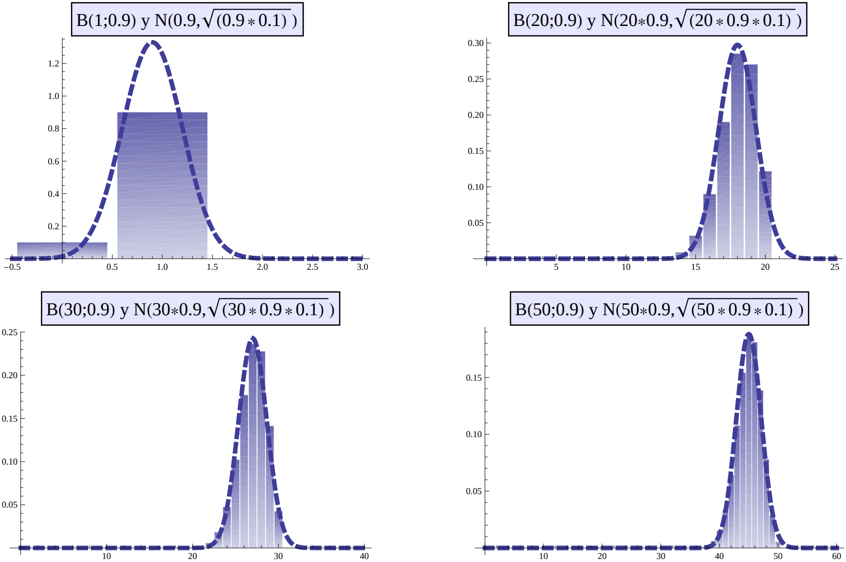




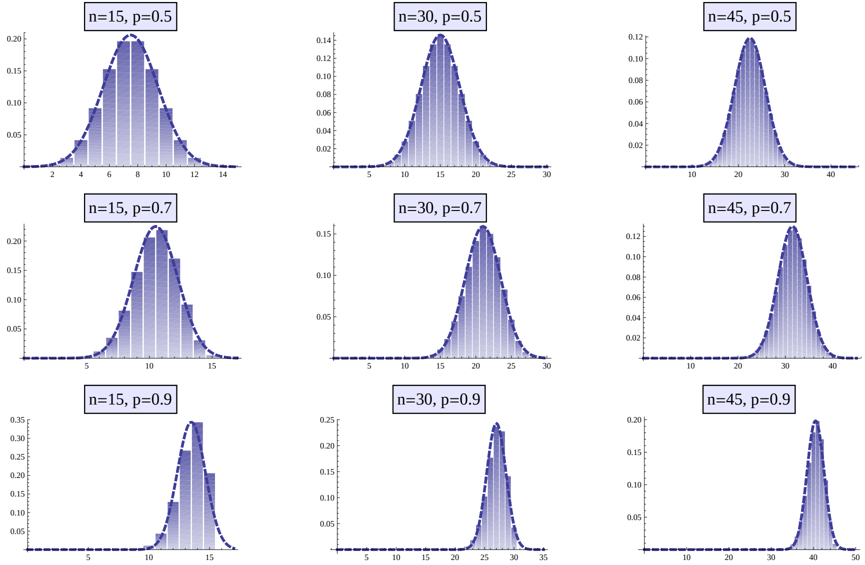




El Teorema de De Moivre-Laplace



El Teorema de De Moivre-Laplace



Aplicaciones del TCL

(1) **Caso general**: En general, n se considera *grande* a partir de 30. Es decir, si $n \geq 30$ y $X_1, \ldots, X_n$ son v.a. independientes y con igual distribución con media μ y desviación típica $\sigma > 0$, entonces

$$X_1 + \cdots + X_n \approx \mathrm{N}\left(n\mu; \sigma\sqrt{n}\right).$$

(2) **Caso binomial**: Si $S_n \sim \mathrm{B}(n; p)$ y queremos calcular $\mathrm{P}(S_n \leq k)$ ($k \in \{0, \ldots, n\}$).

(2.1) Si n y p están en las tablas o la probabilidad es fácil de calcular computacionalmente, se calcula directamente.

(2.2) Si n y p *no* están en las tablas o su cálculo es delicado:

- Si $n \geq 30$, $\pi \leq 0,1$ y $n\pi < 18$ (con π : $\text{mín}\{p, 1-p\}$), entonces se utiliza la aproximación de Poisson

$$S_n \approx \mathrm{P}(np).$$

- Si $n \geq 30$ y $0,1 < p < 0,9$, se utiliza el TCL

$$S_n \approx \mathrm{N}\left(np; \sqrt{np(1-p)}\right).$$

Aplicaciones del TCL: Caso binomial

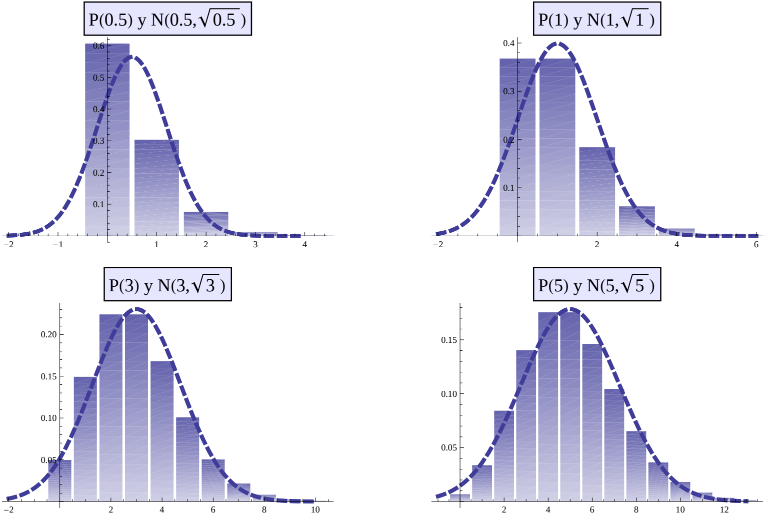
Ejemplo: Conocemos que el 70 % de los usuarios del metro utilizan bonometro.

(a) Si cogemos una muestra de 15 viajeros, ¿cuál es la probabilidad de que al menos 8 de ellos tengan este tipo de billete?

(b) Si tomamos una muestra de 10 personas en 15 estaciones, ¿cuál es la probabilidad de que sean menos de 95 los viajeros que no disponen de bonometro?

Aplicaciones del TCL: Distribución de Poisson

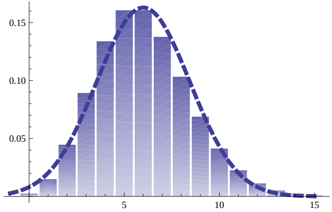
Si $Y \sim \mathrm{P}(\lambda)$ con $\lambda > 5$, entonces $Y \sim \mathrm{P}(\lambda) \approx \mathrm{N}\left(\lambda; \sqrt{\lambda}\right)$.



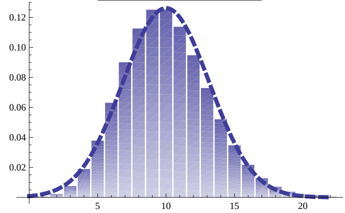
Aplicaciones del TCL: Distribución de Poisson

Si $Y \sim \mathrm{P}(\lambda)$ con $\lambda > 5$, entonces $Y \sim \mathrm{P}(\lambda) \approx \mathrm{N}\left(\lambda; \sqrt{\lambda}\right)$.

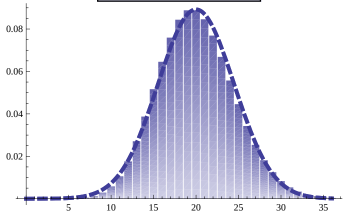
P(6) y N(6,$\sqrt{6}$)



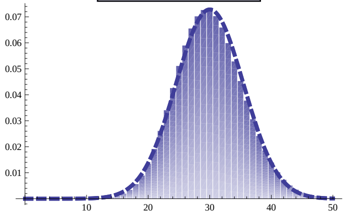
P(10) y N(10,$\sqrt{10}$)



P(20) y N(20,$\sqrt{20}$)



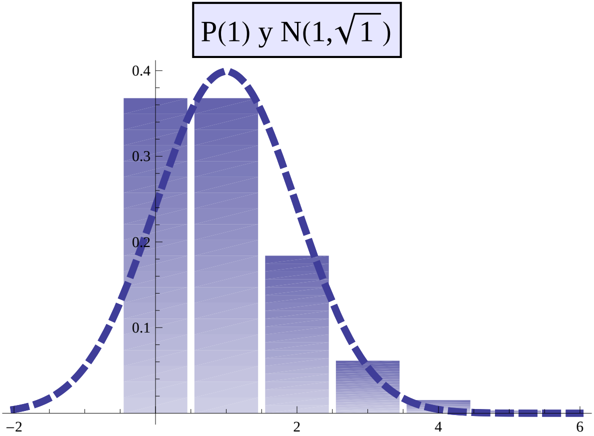
P(20) y N(20,$\sqrt{20}$)



Aplicaciones del TCL: Distribución de Poisson

Pregunta: ¿Por qué $X_1 \sim \mathrm{P}(\lambda) \approx \mathrm{N}(1;1)$?

$$X_1 \sim Y_1 + \cdots + Y_{100}, \text{ con } Y_i \sim \mathrm{P}(1/100) \text{ ind.}$$



Aplicaciones del TCL: Distribución de Poisson

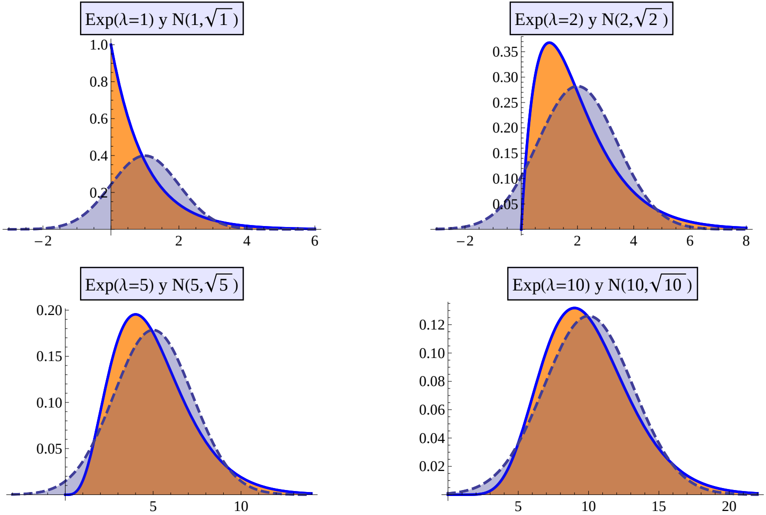
Ejemplo: Un supermercado tiene tres puertas de entrada. Se supone que el número de personas que acuden diariamente por cada una de las puestas son independientes y siguen una distribución de Poisson de medias 200, 150 y 50, respectivamente.

(a) ¿Cuál es la distribución del número total de personas que entra diariamente?

(b) ¿Cuál es la probabilidad de que en 200 días la afluencia de personas supere la cifra de 40260?

Aplicaciones del TCL: Distribución exponencial

Si $X_1, X_2 \cdots \sim \mathrm{Exp}(1)$ independientes, $S_n \sim \mathrm{N}(n; \sqrt{n})$ (n grande).



Aplicaciones del TCL: Distribución exponencial

Si $X_1, X_2 \cdots \sim \text{Exp}(1)$ independientes, $S_n \sim \text{N}(n; \sqrt{n})$ (n grande).