Universidad Rey Juan Carlos

GRADO EN MATEMÁTICAS

PROBABILIDAD

TEMA 5

CONVERGENCIAS ESTOCÁSTICAS

Sonia Hernández Alonso
Área de Estadística e Investigación Operativa (URJC)

Esquema

- Sucesiones de variables aleatorias
- Convergencia en probabilidad, en L^p y casi seguro
- Combinaciones lineales de variables aleatorias
- Leyes de los grandes números
- Teorema del límite central
- Ejemplos de aplicación del TCL
- Otras versiones del TCL

Sucesiones de variables aleatorias

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Colecciones numerables de variables aleatorias

- En el tema anterior hemos examinado vectores aleatorios n-dimensionales, con n finito.
- En este tema vamos a analizar **secuencias infinitas numerables** de variables aleatorias.
- Dado un **espacio probabilístico** $(\Omega, \mathscr{A}, P)$, consideraremos una **sucesión de variables aleatorias,** $\{X_n\}_{n\in\mathbb{N}}$.
- En especial nos plantearemos qué ocurre con las probabilidades sobre X_n cuando n se hace grande, es decir. qué es lo que pasa en el **límite**.

Ejemplos: sucesiones de variables aleatorias

- **Ejemplo 1:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P(X_n = 1) = \frac{n^2 - 1}{n^2} \qquad \text{y} \qquad P(X_n = 5) = \frac{1}{n^2}.$$

Nos planteamos ¿a dónde converge la sucesión $\{X_n\}_{n\in\mathbb{N}}$?

¿En qué sentidos?

Ejemplos: sucesiones de variables aleatorias

- **Ejemplo 1:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P(X_n = 1) = \frac{n^2-1}{n^2} \qquad \text{y} \qquad P(X_n = 5) = \frac{1}{n^2}.$$

Nos planteamos ¿a dónde converge la sucesión $\{X_n\}_{n\in\mathbb{N}}$?

¿En qué sentidos?

- **Ejemplo 2:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P(X_n = 0) = \frac{n^2-1}{n^2} \qquad \text{y} \qquad P\left(X_n = n^2\right) = \frac{1}{n^2}.$$

Nos planteamos ¿a dónde converge la sucesión $\{X_n\}_{n\in\mathbb{N}}$?

¿En qué sentidos...?

Ejemplos: sucesiones de variables aleatorias

- **Ejemplo 3:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$X_n \sim Exp(n).$$

Nos planteamos ¿a dónde converge la sucesión $\{X_n\}_{n\in\mathbb{N}}$?

¿En qué sentidos?

¿Cómo convergen las variables aleatorias?

- Los ejemplos anteriores ponen de manifiesto que hay diferentes formas de considerar la convergencia de variables aleatorias.
- Tengamos en cuenta que una sucesión de variables aleatorias no es una sucesión de números reales.
- Cada una de las variables aleatorias define una distribución de probabilidades... y esto permite diferentes puntos de vista para definir la convergencia.

Convergencia en probabilidad, en L^p y casi seguro

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Varios tipos de convergencia

- Dado un **espacio probabilístico** $(\Omega, \mathscr{A}, P)$, consideraremos una **sucesión de variables aleatorias,**

$$\{X_n\}_{n\in\mathbf{N}},$$

 y otra variable X definida obre el mismo espacio.

- Vamos a definir diferentes formas de que la secuencia $\{X_n\}_{n\in\mathbf{N}}$ pueda converger a la variable X.

- Por el momento definiremos la **convergencia en probabilidad, en L^p y casi seguro**.

- Más adelante introduciremos el concepto de **convergencia en distribución**.

Convergencia en probabilidad

- **Definición:** Se dice que la sucesión de variable aleatorias $\{X_n\}_{n\in\mathbf{N}}$ **converge en probabilidad a** X si para todo $\epsilon > 0$ se verifica

$$\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0.$$

  es decir, si

$$\lim_{n\to\infty} P\left(\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\right) = 0.$$

- Esto se denota abreviadamente como

$$X_n \xrightarrow{P} X.$$

- **Observación:** en muchos casos este límite en probabilidad es una **variable degenerada**, es decir, una **constante**.

Ejemplos: convergencia en probabilidad

- **Ejercicio 1:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P(X_n = 1) = \frac{n^2 - 1}{n^2} \quad \text{y} \quad P(X_n = 5) = \frac{1}{n^2}.$$

Demostrar que

$$X_n \xrightarrow{P} 1.$$

resolución:.........pizarra

Convergencia en media cuadrática

- **Definición:** Se dice que la sucesión de variable aleatorias $\{X_n\}_{n\in\mathbf{N}}$ **converge en media cuadrática a** X si

$$\lim_{n\to\infty} E\left|X_n - X\right|^2 = 0.$$

- Nótese que esto implica que la dispersión de X_n con respecto al *límite* X, se aproxima a 0 al crecer n.

- Abreviadamente lo denotaremos

$$X_n \xrightarrow{L^2} X.$$

Ejemplos: convergencia en media cuadrática

- **Ejercicio 2:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P\left(X_n = 0\right) = \frac{n^2 - 1}{n^2} \qquad \text{y} \qquad P\left(X_n = n^2\right) = \frac{1}{n^2}.$$

1. ¿Se verifica en este caso que $X_n \xrightarrow{P} 0$?
2. ¿Se verifica en este caso que $X_n \xrightarrow{L^2} 0$?

resolución:.........pizarra

Convergencia en L^p

- La convergencia en media cuadrática es un caso particular de la llamada **convergencia en L^p**, que puede definirse para cualquier p positivo.
- **Definición:** Dado $p \in \mathbb{R}^+$, se dice que la sucesión de variable aleatorias $\{X_n\}_{n\in\mathbf{N}}$ **converge en media cuadrática a X** si

$$\lim_{n\to\infty} E\,|X_n - X|^p = 0.$$

- Abreviadamente lo denotaremos

$$X_n \overset{L^p}{\to} X.$$

Ejemplos: convergencia en L^p

- **Ejercicio 3:** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tales que, para cada $n \in \mathbb{N}$,

$$P(X_n = 0) = \frac{n^2 - 1}{n^2} \qquad \text{y} \qquad P\left(X_n = n^2\right) = \frac{1}{n^2}.$$

1. ¿Se verifica en este caso que $X_n \xrightarrow{L^1} 0$?
2. ¿Para qué valores de $p > 0$ se verifica $X_n \xrightarrow{L^p} 0$?

resolución:.........pizarra

Convergencia casi seguro

- **Definición:** Se dice que la sucesión de variable aleatorias $\{X_n\}_{n\in\mathbf{N}}$ **converge casi seguro a X** si

$$P\left(X_n \to X\right) = 1.$$

  es decir si

$$P\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

- **Observación:** La convergencia casi seguro es una condición más fuerte que la convergencia en probabilidad.

- **Proposición 1:** Si

$$X_n \stackrel{c.s.}{\to} X,$$

  entonces

$$X_n \stackrel{P}{\to} X.$$

A partir de variables incorrelacionadas....

- Ahora vamos a centrarnos en sucesiones que se definen a partir de variables aleatorias que están **incorrelcionada entre sí**.
- Y muy especialmente en variables que son **independientes** y además siguen todas la **misma distribución**.
- Este caso es especialmente importante en **inferencia estadística**.

Combinaciones lineales

de variables aleatorias

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Transformaciones lineales de variables aleatorias

- Recordemos que, si X es una variable aleatoria, y $a, b \in \mathbb{R}$ son constantes, entonces

$$
\begin{aligned}
E(aX + b) &= aE(X) + b, \\
V(aX + b) &= a^2 V(X).
\end{aligned}
$$

Combinaciones lineales de variables aleatorias

- Consideremos ahora combinaciones lineales de dos variables aleatorias.

  Sean X_1, X_2 variables aleatorias y $a_1, a_2, b \in \mathbb{R}$ constantes.

  Entonces se verifica que

$$E(a_1X_1 + a_2X_2 + b) = a_1E(X_1) + a_2E(X_2) + b,$$

$$V(a_1X_1 + a_2X_2 + b) = a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2Cov(X_1, X_2).$$

  demostración:.........pizarra

Combinaciones lineales de v. al. (continuación)

- Lo anterior implica, por ejemplo, que
  - Para la suma de dos variables aleatorias:

$$\begin{aligned} E(X_1 + X_2) &= E(X_1) + E(X_2), \\ V(X_1 + X_2) &= V(X_1) + V(X_2) + 2 \times Cov(X_1, X_2), \end{aligned}$$

  - Para la diferencia de dos variables aleatorias:

$$\begin{aligned} E(X_1 - X_2) &= E(X_1) - E(X_2), \\ V(X_1 - X_2) &= V(X_1) + V(X_2) - 2 \times Cov(X_1, X_2), \end{aligned}$$

  - Para el promedio de dos variables aleatorias:

$$\begin{aligned} E\left(\frac{X_1 + X_2}{2}\right) &= \frac{E(X_1) + E(X_2)}{2}, \\ V\left(\frac{X_1 + X_2}{2}\right) &= \frac{V(X_1) + V(X_2) + 2 \times Cov(X_1, X_2)}{4}, \end{aligned}$$

  etcetera

Combinaciones lineales de v. al. incorrelacionadas

- Las expresiones anteriores se simplifican cuando, en particular, **las variables aleatorias X_1 e X_2 son incorrelacionadas**:

  - Para la suma de dos variables aleatorias incorrelacionadas:

$$
\begin{aligned}
E\left(X_1 + X_2\right) &= E\left(X_1\right) + E\left(X_2\right), \\
V\left(X_1 + X_2\right) &= V\left(X_1\right) + V\left(X_2\right),
\end{aligned}
$$

  - Para la diferencia de dos variables aleatorias incorrelacionadas:

$$
\begin{aligned}
E\left(X_1 - X_2\right) &= E\left(X_1\right) - E\left(X_2\right), \\
V\left(X_1 - X_2\right) &= V\left(X_1\right) + V\left(X_2\right),
\end{aligned}
$$

  - Para el promedio de dos variables aleatorias incorrelacionadas:

$$
\begin{aligned}
E\left(\frac{X_1 + X_2}{2}\right) &= \frac{E\left(X_1\right) + E\left(X_2\right)}{2}, \\
V\left(\frac{X_1 + X_2}{2}\right) &= \frac{V\left(X_1\right) + V\left(X_2\right)}{4}.
\end{aligned}
$$

Combinaciones lineales de v. al. independientes

- Recordemos que **las variables aleatorias independientes siempre están incorrelacionadas**.
- Evidentemente esto implica que las simplificaciones de la página anterior para la varianza de combinaciones lineales de variables aleatorias siempre **son válidas cuando las variables son independientes**.
- También debemos recordar que, en general la **incorrelación no implica independencia**, aunque en el caso de variables gaussianas sí son condiciones equivalentes.

Combinaciones de v.al. independ's (continuación)

- Cuando, además de ser independientes, **las variables X_1 y X_2 tienen la misma esperanza y la misma varianza,** es decir, verifican

$$\begin{aligned} E(X_1) &= E(X_2) = \mu, \\ V(X_1) &= V(X_2) = \sigma^2, \end{aligned}$$

entonces se pueden simplificar aún más las expresiones:

- Para la **suma de dos variables aleatorias independientes con media μ y varianza σ^2:**

$$\begin{aligned} E(X_1 + X_2) &= 2\mu \\ V(X_1 + X_2) &= 2\sigma^2 \end{aligned}$$

- Para el **promedio de dos variables aleatorias independientes con media μ y varianza σ^2:**

$$\begin{aligned} E\left(\frac{X_1 + X_2}{2}\right) &= \frac{\mu + \mu}{2} = \mu, \\ V\left(\frac{X_1 + X_2}{2}\right) &= \frac{\sigma^2 + \sigma^2}{4} = \frac{\sigma^2}{2}, \end{aligned}$$

Variables aleatorias i.i.d.

- En este último caso lo que tenemos son dos variables aleatorias **independientes** y **la misma esperanza μ y la misma varianza σ^2.**
- Esta situación aparece siempre que las variables X e Y son realizaciones independientes de una distribución común, es decir, cuando son variables **i**ndependientes e **i**dénticamente **d**istribuidas, lo cual se denota

$$X, Y \textbf{ i.i.d.}$$

- Cuando se extraen al azar **muestras de una población**, lo habitual es que las observaciones se escojan de manera **independiente** unas de otras.
- Además, como todas las observaciones proceden de la misma población, constituyen realizaciónes de una variable aleatoria común, es decir, tienen **idéntica distribución**.
- Por tanto, las observaciones de la muestra constituyen una colección de variables aleatorias **i.i.d**

Colecciones de cariables aleatorias i.i.d.

- Si llamamos n al **tamaño de la muestra** es decir, al número de elementos extraídos de la población, y denotamos a estos elementos por $X_1, X_2, \ldots, X_n$, entonces tenemos que

$$X_1, X_2, \ldots, X_n \ \textbf{i.i.d.}$$

- Habitualmente, la muestra $X_1, X_2, \ldots, X_n$ se ha extraído de la población a fin de obtener información sobre una característica X.

- Por tanto, la muestra $X_1, X_2, \ldots, X_n$ consiste en n realizaciones independientes de la distribución que tiene en la población la variable X que queremos estudiar.

- Evidentemente, puesto que $X_1, X_2, \ldots, X_n$ tienen una distribución común, tienen todas ellas la misma media (que denotaremos μ) y las misma varianza (que denotaremos σ^2):

$$\begin{aligned} E(X_1) &= \ldots = E(X_n) = \mu, \\ V(X_1) &= \ldots = V(X_n) = \sigma^2. \end{aligned}$$

Combinaciones lineales de variables i.i.d.

- **Proposición 2:** Si $X_1, \ldots, X_n$ **son v. al. independientes e idénticamente distribuídas con esperanza** μ **y varianza** σ^2**,** entonces

1. La **suma** de las n variables aleatorias verifica

$$E\left(X_1 + \ldots + X_n\right) = \mu + \ldots + \mu = n\mu,$$

$$V\left(X_1 + \ldots + X_n\right) = \sigma^2 + \ldots + \sigma^2 = n\sigma^2.$$

2. El **promedio** de las n variables aleatorias verifica

$$E\left(\overline{X}_n\right) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\mu + \ldots + \mu}{n} = \frac{n\mu}{n} = \mu,$$

$$V\left(\overline{X}_n\right) = V\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\sigma^2 + \ldots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

demostración:.........pizarra

Combinaciones lineales de v.al. incorrelacionadas

- **Observación:** El resultado anterior también se verifica cuando las variables $X_1, \ldots, X_n$ **son incorrelacionadas**, aunque no sean independientes.

- Además, para que se verifique la **Proposición 2** tampoco es necesario que las variables sean idénticamente distribuídas, sino que basta con que **todas tengan la misma esperanza y la misma varianza**.

- Esto nos lleva a la siguiente **extensión de la Proposición 2**.

Extensión de la Proposición 2

- **Proposición 3:** Si $X_1, \ldots, X_n$ **son v. al. incorrelacionadas, todas ellas con esperanza** μ **y varianza** σ^2**,** entonces

  1. La **suma** de las n variables aleatorias verifica

  $$\begin{aligned} E\left(X_1 + \ldots + X_n\right) &= \mu + \ldots + \mu = n\mu, \\ V\left(X_1 + \ldots + X_n\right) &= \sigma^2 + \ldots + \sigma^2 = n\sigma^2. \end{aligned}$$

  2. El **promedio** de las n variables aleatorias verifica

  $$\begin{aligned} E\left(\overline{X}_n\right) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) &= \frac{\mu + \ldots + \mu}{n} = \frac{n\mu}{n} = \mu, \\ V\left(\overline{X}_n\right) = V\left(\frac{X_1 + \ldots + X_n}{n}\right) &= \frac{\sigma^2 + \ldots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

- *demostración:*.........inmediata

Error estándar

- La desviación típica de la media muestral se denomina **error estándar** y viene dada por

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}.$$

- Observemos que **a medida que aumenta el tamaño de la muestra, n, el error estándar disminuye**.

- Esto se traduce en que, **cuanto mayor sea el tamaño muestral, más concentrada estará la probabilidad de $\overline{X}$ en torno a μ.**

¿Podemos decir algo más?

- Sabemos ya que, para variables incorrelacionadas con la misma distribución se verifica

$$E\left(\sum_{i=1}^{n} X_i\right) = n\mu,$$
$$V\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2,$$

y que

$$E\left(\overline{X}\right) = \mu,$$
$$V\left(\overline{X}\right) = \frac{\sigma^2}{n}.$$

- Pero, ¿puede decirse algo más acerca de la distribución de $\sum_{i=1}^{n} X_i$, o de $\overline{X}$?
- La respuesta es **sí**. Los **teoremas del límite central** nos dicen mucho más.

Leyes de los grandes números

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Ejemplo: combinaciones lineales de v. al. i.i.d.

- **Ejercicio 4:** Consideremos una colección $\{X_i\}_{i\in\mathbf{N}}$ de variables aleatorias independientes e idénticamente distribuidas con distribución común

$$X_i \sim Exp(0.5)$$

1. ¿Cuáles son la esperanza y la varianza de $\overline{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$?
2. ¿Qué ocurre con la varianza de $\overline{X}_n$ cuando n tiende a infinito?
3. Intuitivamente, ¿qué sucederá con la distribución de probabilidad de $\overline{X}_n$ cuando n tiende a infinito?

- Las **leyes de los grandes números** nos indican qué ocurre con la distribución límite de $\overline{X}_n$ en situaciones de este tipo.

Ley débil de los grandes números de Khintchin

- Existen varias versiones de la ley débil de los grandes números. La más sencilla, probada por *Aleksandr Khinchin*, es la siguiente.

- **Proposición 4:** Sea $\{X_i\}_{i\in\mathbf{N}}$ una sucesión de variables aleatorias independientes e idéticamente distribuidas (i.i.d.) con

$$E(X_i) = \mu, \qquad \text{y} \qquad V(X_i) = \sigma^2 < \infty.$$

Consideremos la sucesión de promedios

$$\overline{X}_n = \frac{\sum\limits_{i=1}^{n} X_i}{n}.$$

Entonces,

$$\overline{X}_n \xrightarrow{P} \mu.$$

demostración:.........pizarra

Ejemplo: aplicación de la ley débil

- **Ejercicio 5:** Consideremos una sucesión $\{X_i\}_{i\in\mathbb{N}}$ de variables aleatorias independientes e idénticamente distribuidas con distribución común

$$X_i \sim Exp(0.5)$$

Para cada $n \in \mathbb{N}$ definimos

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

Comprobar que

$$\overline{X}_n \xrightarrow{P} 2.$$

Ley fuerte de los grandes números de Kolmogórov

- La ley débil de Khinchin, supuso una aportación muy importante al la teoría de la probabilidad.

- Sin embargo, este resultado fue mejorado posteriormente por *Andréi Kolmogórov* con su **ley fuerte de los grandes números**.

- **Proposición 5:** Sea $\{X_i\}_{i\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con

$$E(X_i) = \mu, \qquad \text{y} \qquad V(X_i) = \sigma^2 < \infty.$$

Consideremos la sucesión de promedios

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

Entonces,

$$\overline{X}_n \stackrel{c.s.}{\rightarrow} \mu.$$

Ejemplo: aplicación de la ley fuerte

- **Ejercicio 6:** Consideremos una secuencia $\{X_i\}_{i\in\mathbb{N}}$ de variables aleatorias independientes e idénticamente distribuidas con distribución común

$$X_i \sim Ge(0.2)$$

Para cada $n \in \mathbb{N}$ definimos

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

Comprobar que

$$\overline{X}_n \overset{c.s.}{\to} 5.$$

Generalización ley fuerte de los grandes números

- La ley fuerte de Kolmogorov puede extenderse a casos más generales.
- No es necesario que las variables de partida ean independientes, sino que basta con que sean incorrelacionadas.
- Y tampoco es preciso que todas tengan la misma distribución. Basta con que todas ellas tengan varianza finita.
- El resultado siguiente **generaliza la ley fuerte de los grandes números**.

Ley fuerte de los grandes números: caso general

- **Proposición 6***: Sea $X_1, X_2, \ldots,$ una sucesión de variables aleatorias incorrelacionadas con

$$E(X_i) = \mu_i, \qquad \text{y} \qquad V(X_i) = \sigma_i^2 < \infty.$$

Consideremos la sucesión de promedios

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

y sea

$$\overline{\mu}_n = E\left(\overline{X}_n\right) = \frac{\sum_{i=1}^{n} \mu_i}{n},$$

Entonces,

$$\overline{X}_n - \overline{\mu}_n \stackrel{c.s.}{\rightarrow} 0.$$

Teoremas del límite central

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

¿Qué pasa con la distribución en el límite?

- **Ejercicio 7:** Consideremos una colección $\{X_i\}_{i\in\mathbf{N}}$ de variables aleatorias e idénticamente distribuidas con distribución común

$$X_i \sim Exp(0.5)$$

1. ¿Cuáles son la esperanza y la varianza de $\sum_{i=1}^{n} X_i$?
2. ¿Qué ocurre con la varianza de $\sum_{i=1}^{n} X_i$ cuando n tiende a infinito?
3. ¿Puede decirse algo sobre la distribución de probabilidad de $\sum_{i=1}^{n} X_i$ cuando n tiende a infinito?

- Los **teoremas del límite central** establecen cuál es la distribución límite de $\sum_{i=1}^{n} X_i$ (y también de $\overline{X}_n$), en situaciones de este tipo.

Combinaciones lineales de v.al. gaussianas i.i.d.

- Recordemos que **cualquier combinación lineal de distribuciones normales es sigue también una distribución normal.**

- Por tanto, si $X_1, \ldots, X_n$ son variables aleatorias independientes y con distribución normal de esperanza μ y varianza σ^2, es decir, si

$$X_1, \ldots, X_n \sim^{i.i.d} N(\mu, \sigma^2),$$

entonces:

  - La distribución de la suma de estas n variables, es

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n \sim N\left(n\mu, n\sigma^2\right),$$

  - **La distribución del promedio de estas n variables, es**

$$\boxed{\overline{X}_n = \frac{X_1 + \ldots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

¿Y si las variables no son gaussianas?

- Nos planteamos ahora, ¿que ocurre con el promedio $\overline{X}_n$ si $X_1, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas (i.i.d.) pero **su distribución común no es gaussiana**?

- Y qué sucede con la distribución de la suma $S_n = \sum_{i=1}^{n} X_i$?

- Vamos a ver un resultado que nos asegura que, si n es suficientemente grande, y bajo condiciones muy generales, las distribuciones de $\overline{X}_n$ y S_n. se asemejan mucho a una normal.

Convergencia en distribución

- Para poder enunciar correctamente el teorema del límite central (TCL) necesitamos introducir previamente otra forma de **convergencia estocástica**.

- Ya hemos definido algunas formas de convergencia de variables aleatorias: convergencia en probabilidad, convergencia en L^p, convergencia casi seguro... Ahora definiremos la **convergencia en distribución**.

- **Definición:** Se dice que una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbf{N}}$ **converge en distribución** a una variable aleatoria X si verifica

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

  en todos los puntos x donde $F_X(x)$ es continua.

- Esto se denota abreviadamente como

$$X_n \xrightarrow{d} X.$$

- A partir de esta definición pasamos a enunciar el **Teorema del Límite Central** en su versión más sencilla.

Teorema del Límite Central para v.al. i.i.d)

URJC

- **Teorema 1:** Sea $X_1, X_2, \ldots,$ una sucesión de variables aleatorias i.i.d. con $\mu = E(X_i)$ y $\sigma^2 = V(X_i) < \infty$.

  Definimos

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

  y sea $G_n(x)$ la función de distribución de la variable aleatoria $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$.

  Entonces, para todo $x \in \mathbb{R}$ se verifica

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \Phi(x),$$

  es decir, **la variable aleatoria**

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

  converge en distribución a la normal estándar, Z.

Entendiendo en Teorema del Límite Central (TCL)

- Lo que establece el teorema del límite central es que, si $X_1, X_2, ..., X_n$ son variables aleatorias independientes e idénticamente distribuídas con media μ y varianza σ^2, entonces, para n **suficientemente grande** se verifica

$$\boxed{\frac{X_1 + \ldots + X_n}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right)}$$

sin que importe cuál sea la distribución de las variables $X_1, X_2, ..., X_n$.

- Este resultado es válido tanto para variables discretas como continuas, sean simétricas o asimétricas, unimodales o multimodales... Y esto va resultar muy útil en inferencia estadística.

- Más exactamente, lo que asegura este teorema es que, si $X_1, X_2, ..., X_n$ son **i.i.d.** con $E(X_i) = \mu$, $V(X_i) = \sigma^2$, entonces la **función de distribución** de la variable aleatoria

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

tiende a Φ cuando $n \to \infty$.

Referencia sobre la demostración del TCL

- Es decir, esta versión para vaiables i.i.d. del teorema central del límite establece que, si la varianza común es finita, entonces,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z,$$

  donde Z denota, como es habitual, la distribución normal estándar.

- Una demostración de este resultado puede encontrarse, por ejemplo, en la **Sección 7.4** del libro de Wackerly, Mendenhall y Sheaffer, *Estadística Matemática con Aplicaciones*.

Ejemplo del TCL: lanzamiento dado

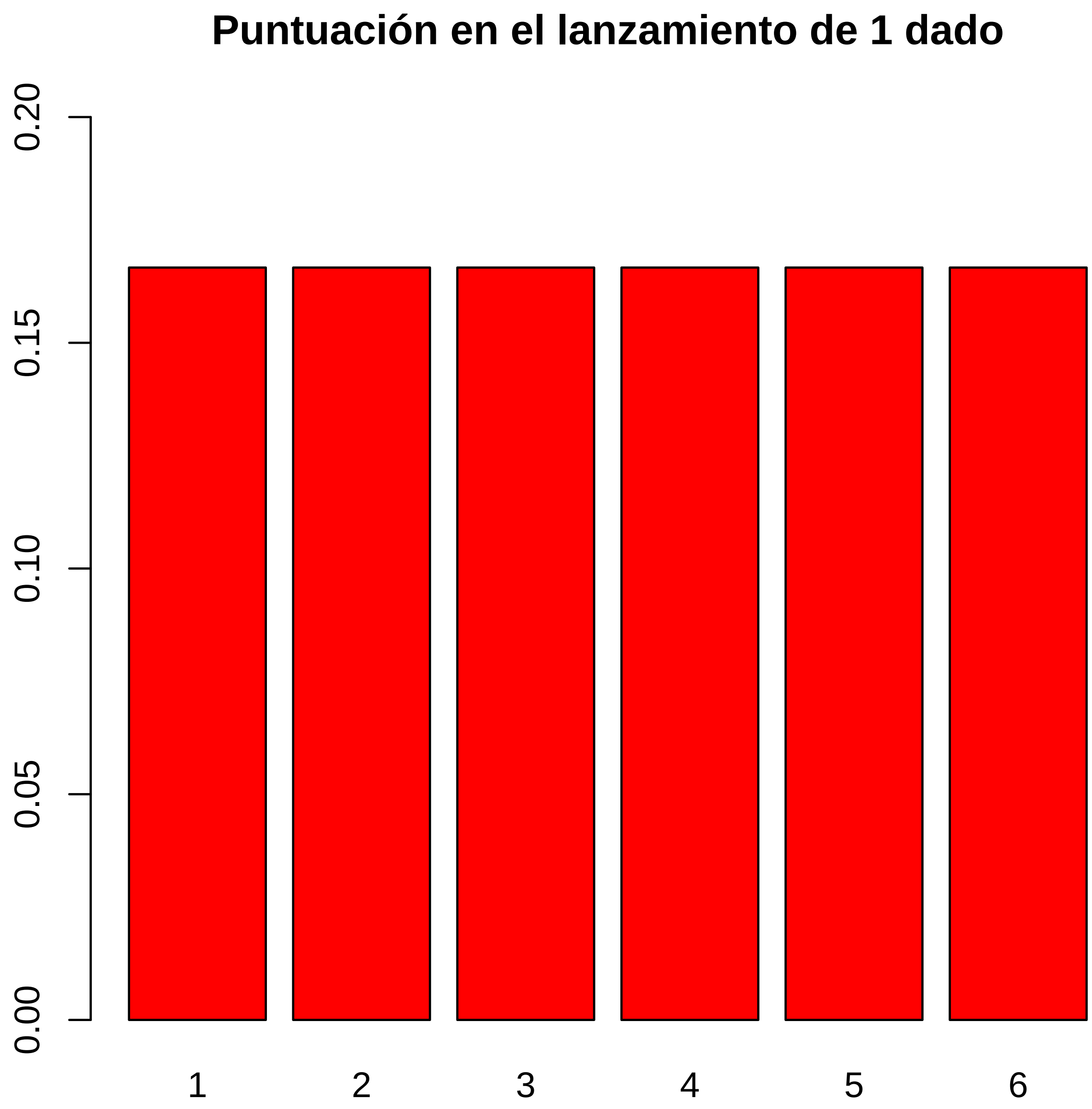
- Para ilustrar gráficamente el teorema central del límite, vamos a utilizar el caso del lanzamiento de un dado equilibrado.

- El resultado de cada lanzamiento es una variable aleatoria discreta, X, con función de masa

$$X \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

- En la transparencia siguiente aparece representada esta distribución mediante un barras. El gráfico evidencia que la distribución de X difiere mucho de una normal.

Ejemplo del TCL: lanzamiento dado (continuación)



Ejemplo del TCL: lanzamiento dado (continuación)

- Observemos que la esperanza de X es

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \ldots 6 \times \frac{1}{6} = \frac{21}{6} = 3.5$$

- Además

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \ldots 6^2 \times \frac{1}{6} = \frac{91}{6} = 15.17$$

- Por tanto la varianza de X es

$$V(X) = E(X^2) - E^2(X) = 15.17 - 3.5^2 = 2.92$$

Ejemplo del TCL: lanzamiento dado (continuación)

- Supongamos ahora que no lanzamos el dado una única vez, sino n veces. Los n lanzamientos son n variables aleatorias **independientes** y **todas con la misma distribución:**

$$X_1, X_2, X_3, ..., X_n \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

- El promedio de las n puntuaciones es otra variable aleatoria

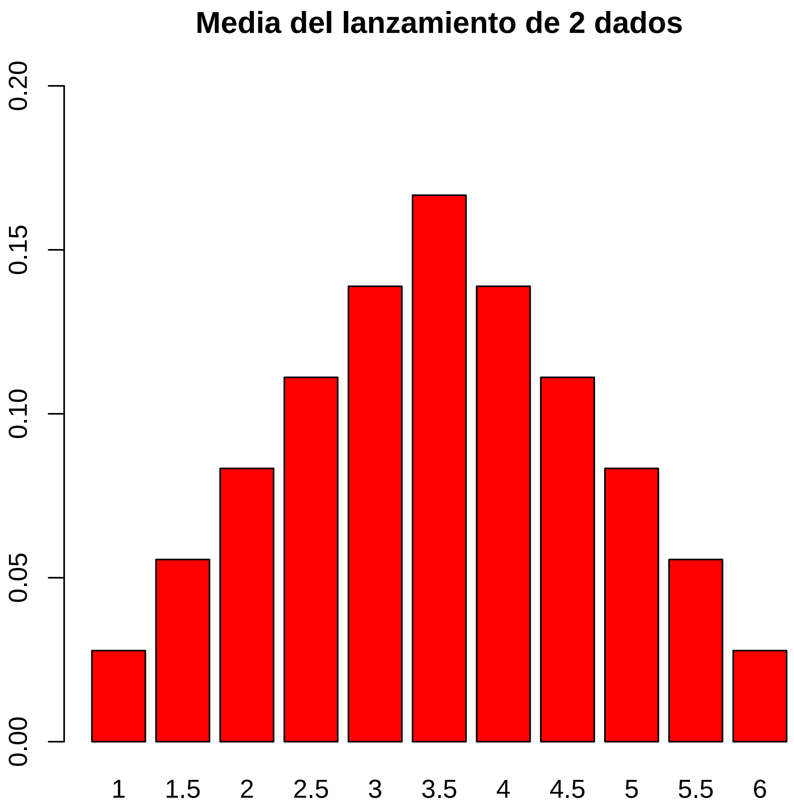
$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

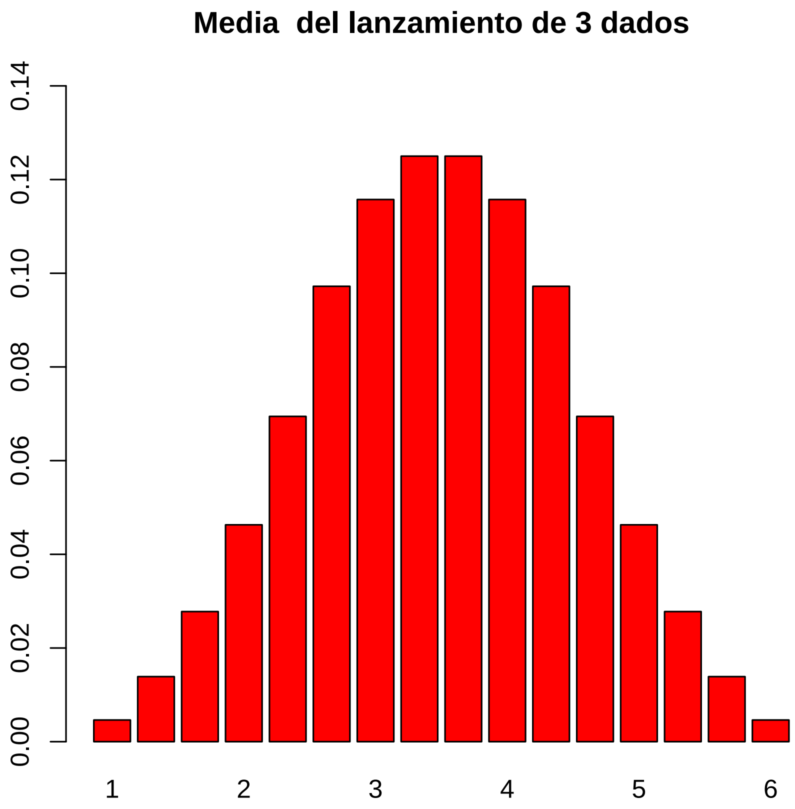
- El TCL implica que, cuando n es grande, se tiene

$$\overline{X} \simeq N\left(\mu = 3.5, \sigma^2 = \frac{2.92}{n}\right)$$

Ejemplo del TCL: lanzamiento dado (continuación)

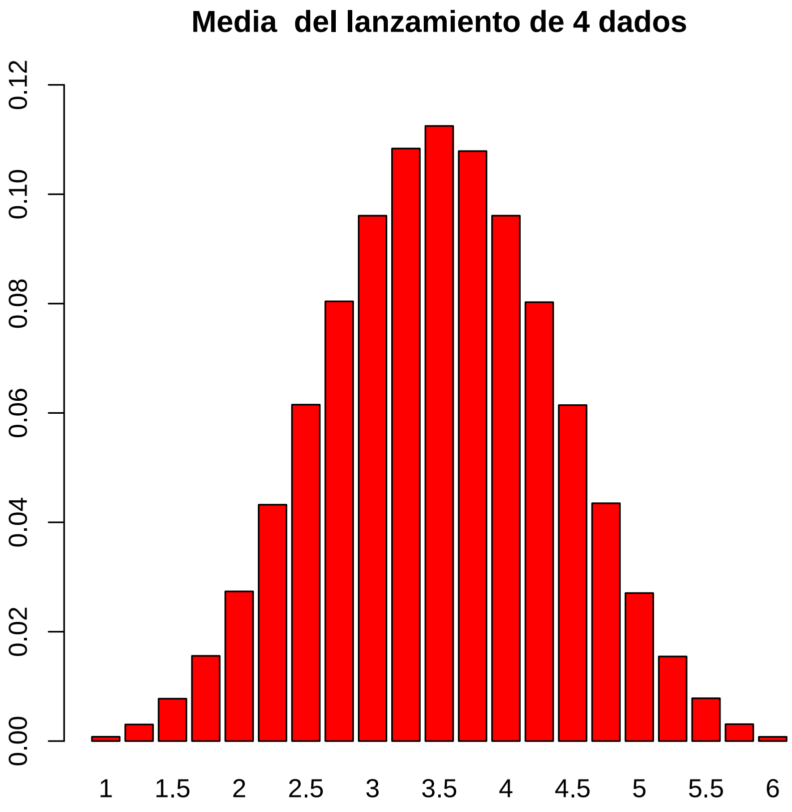
- Para comprobar lo que dice el TCL sobre este ejemplo, los gráficos siguientes muestran la distribución del promedio de 2 lanzamientos de un dado y de 3 lanzamientos, respectivamente:

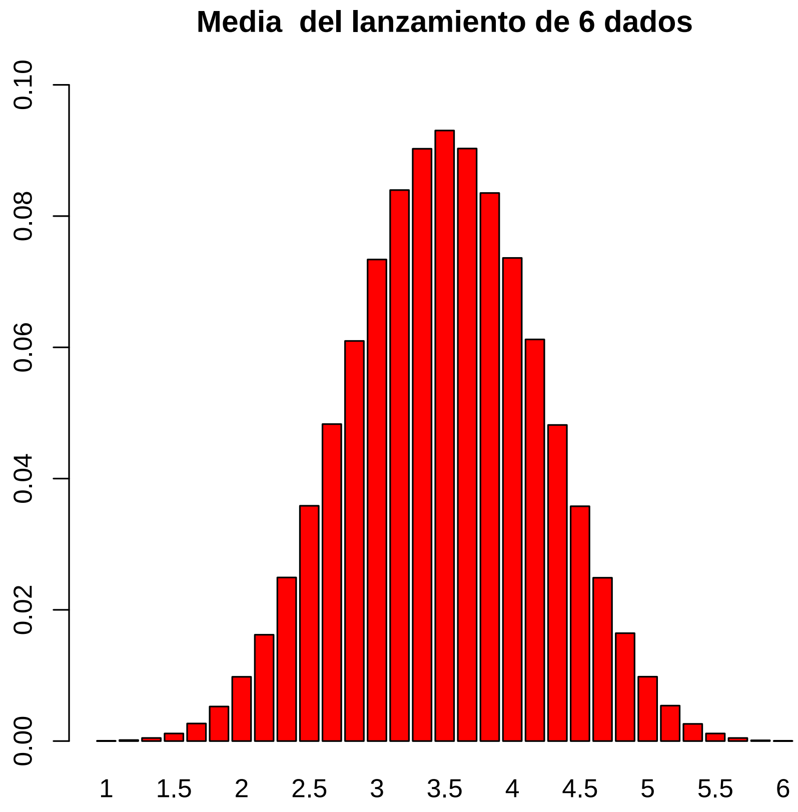




Ejemplo del TCL: lanzamiento dado (continuación)

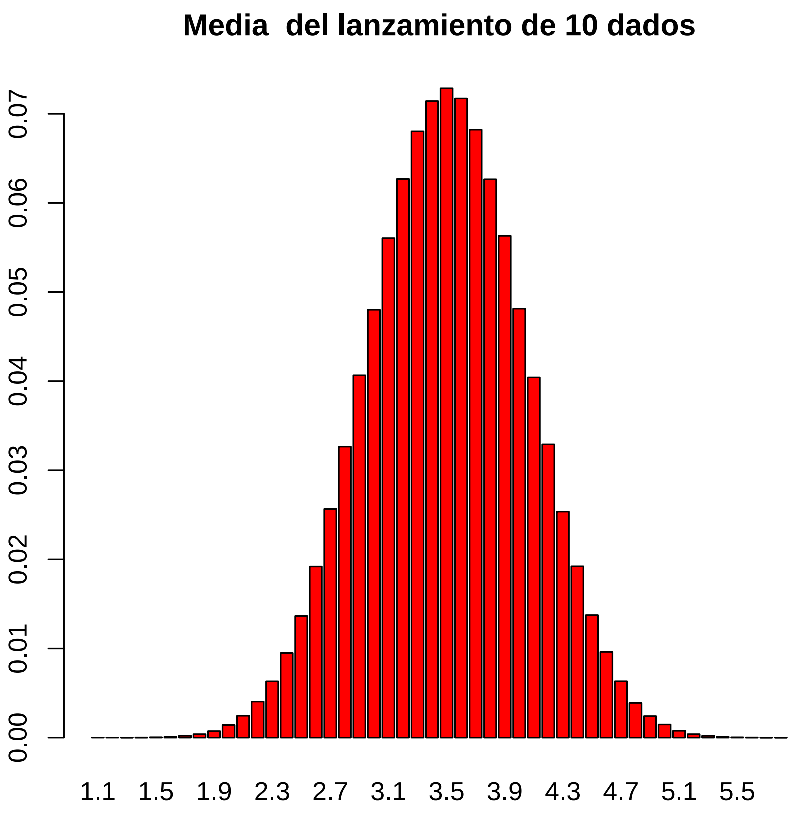
- Los gráficos siguientes muestran el promedio de 4 y 6 lanzamientos, respectivamente:

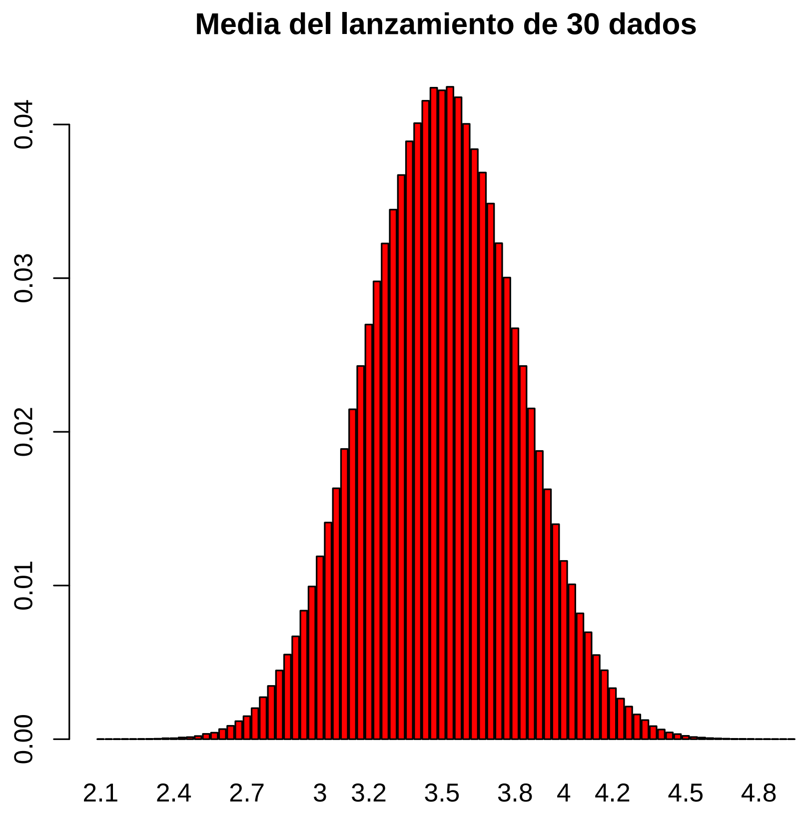




Ejemplo del TCL: lanzamiento dado (continuación)

- Representación del promedio de 10 y 30 lanzamientos:





- Es evidente que la distribución del promedio de las variables se va aproximando a una distribución normal centrada en $\mu = 3.5$.

Observaciones sobre el Teorema del Límite Central

- Se trata de un resultado muy importante (central) en estadística.
- A partir de una secuencia de variables aleatorias y sin asumir nada más que sean i.i.d., ¡¡obtenemos normalidad!!
- La clave de la demostración de este resultado es que la normalidad se obtiene por la **suma de cantidades "pequeñas" e independientes.**

  Nótese que es necesario asumir varianzas finitas para obtener normalidad.
- **En general no sabemos como de rápido se da esta convergencia a la normalidad. Es necesario comprobarlo para cada distribución.**

Ejemplo del TCL: aproximación media exponencial

- Supongamos que simulamos muestras aleatorias de tamaño n de una variable aleatoria X que sigue una distribución exponencial con media 10, es decir, cuya función de densidad es:

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{10}\, e^{-\frac{x}{10}} & \text{si } x \geq 0 \end{cases}$$

- El gráfico de la densidad anterior es

Ejemplo TCL: media exponencial (continuación)

- Se simularon muestras de la variable exponencial anterior, con tamaños muestrales $n = 5$ y $n = 25$.

- Esto se repitió 1000 veces, es decir, 1000 veces se generaron muestras de la v.a. exponencial con esos dos tamaños, y para cada muestra de las 1000 se calculó la media, es decir:

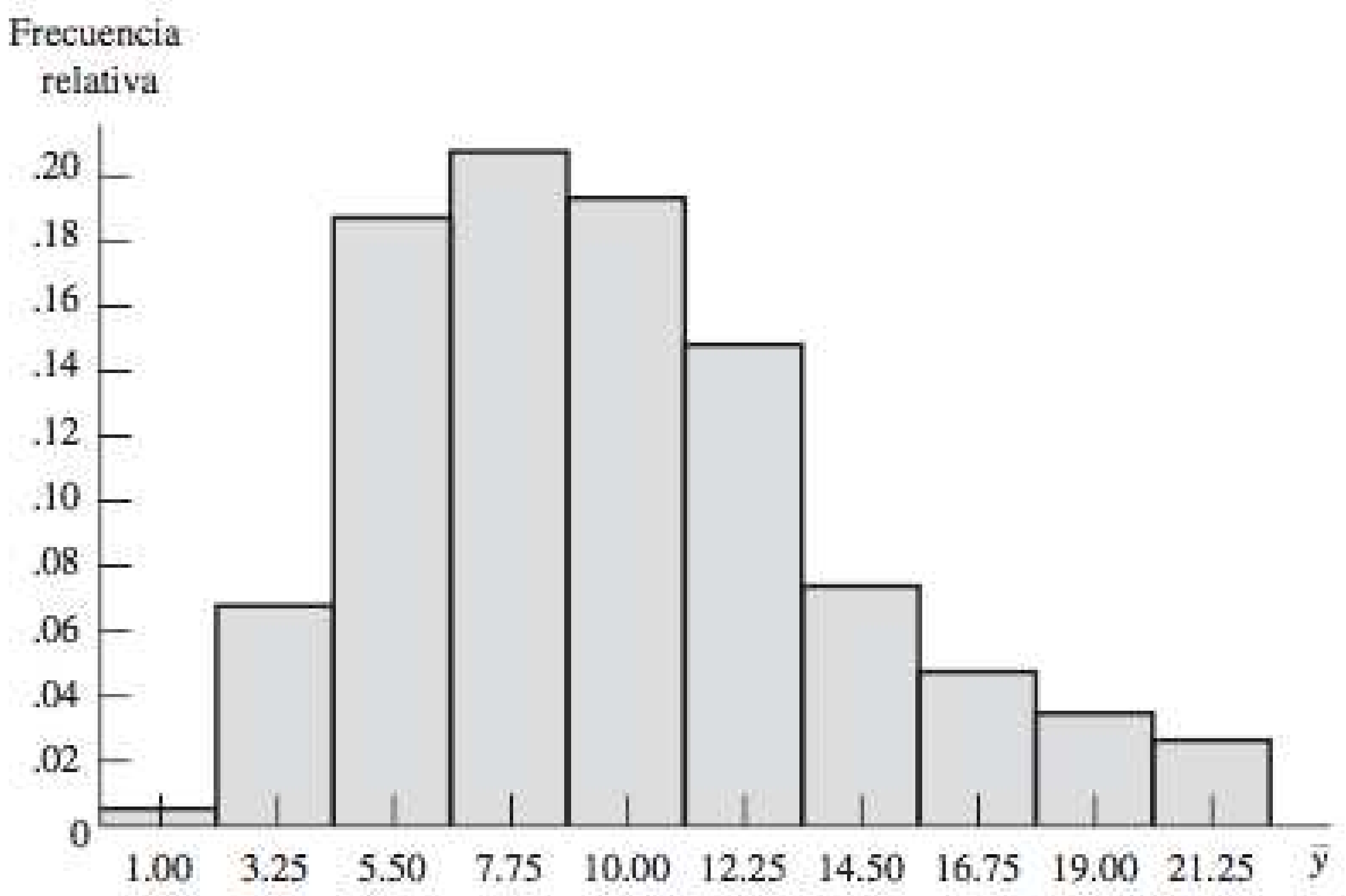
$$\overline{X}_5 = \frac{X_1 + X_2 + \ldots + X_5}{5},$$

  ó bien:

$$\overline{X}_{25} = \frac{X_1 + X_2 + \ldots + X_{25}}{25},$$

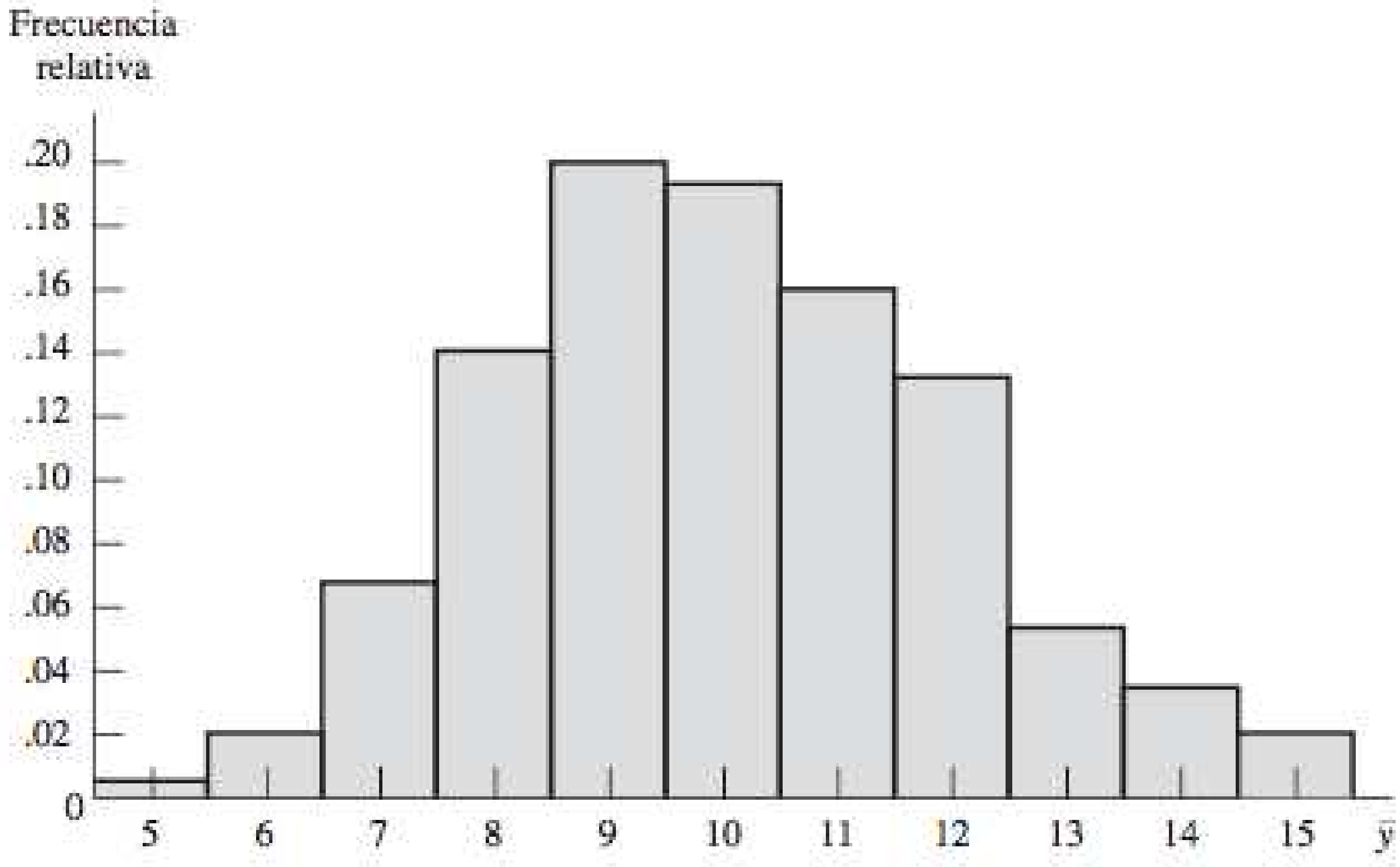
Ejemplo TCL: media exponencial (continuación)

- Este es el histograma de frecuencias relativas de las medias muestrales obtenidas en las 1000 repeticiones, para $n = 5$:



Ejemplo TCL: media exponencial (continuación)

- Y este es el histograma de la frecuencia relativa de las medias muestrales obtenidas en las 1000 repeticiones, para $n = 25$.



Ejemplo TCL: media exponencial (continuación)

- Por el TCL sabemos que

$$\overline{X} \approx N(\mu_{\overline{X}}, \sigma^2_{\overline{X}})$$

con

$$\mu_{\overline{X}} = E(\overline{X}) = \mu = 10 \quad \text{y} \quad \sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n} = \frac{100}{n},$$

- Podemos comparar estos números con los valores de la media y varianza muestrales observadas en el experimento de simulación para los casos $n = 5$ y $n = 25$:

Tamaño muestral	Promedio de 1000 medias muestrales	$\mu_{\overline{X}} = \mu$	Varianza de 1000 medias muestrales	$\sigma^2_{\overline{X}} = \frac{100}{n}$
n=5	9.86	10	19.63	20
n= 25	9.95	10	3.93	4

¿Cuál es la distribución asintótica de $S_n = \sum_{i=1}^{n} X_i$?

- El **Teorema 1** establece que, si $X_1, X_2, \ldots,$ son variables aleatorias i.i.d. con $\mu = E(X_i)$ y $\sigma^2 = V(X_i) < \infty$, entonces

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} X,$$

  donde Z es la distribición normal estándar.

- Nótese que esto implica de forma inmediata que

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{d} Z.$$

- En consecuencia, para n suficientemente grande, se puede utilizar la aproximación

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n \approx N\left(n\mu, n\sigma^2\right).$$

Ejemplos de aplicación del TCL

URJC

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Ejemplo TCL: peso de los perros

- **Ejercicio 8:** Se ha comprobado que el peso medio de los perros de la raza pitbull de cierta región es de 25 Kg, con una desviación típica de 20 Kg.

  Se selecciona al azar una muestra de 100 pitbulls de esta comarca.

  Calcular, de forma aproximada, la probabilidad de que el promedio de los pesos de esta muestra sea inferior a 20 Kg

Solución al Ejercicio 8

- Disponemos de una m.a.s. de pesos de los pitbulls, es decir, de una colección de variables i.i.d.,

$$P_1, P_2, \ldots, P_{100},$$

con $E(P_i) = 25$ y $V(P_i) = 20^2 = 400$, pero con distribución desconocida.

Tenemos que calcular una probabilidad sobre la variable aleatoria

$$\overline{P} = \frac{P_1 + P_2 + \ldots + P_{100}}{100}.$$

Aunque la distribución de las P_i es desconocida, como disponemos de una **muestra grande** el **Teorema del Límite Central** garantiza que la distribución de $\overline{P}$ es aproximadamente normal, con media la misma que la del pesos de cada pitbull, es decir

$$E(\overline{P}) = 25 \text{ kg},$$

y con varianza,

$$V(\overline{P}) = \frac{\sigma^2}{100} = \frac{400}{100} \text{ Kg}^2 = 4 \text{ kg}^2.$$

Solución al Ejercicio 8 (continuación)

Es decir que

$$\overline{P} \approx N(\mu = 25, \sigma^2 = 4)$$

Por tanto, la probabilidad de que el promedio de los pesos de esta muestra de tamaño 100 sea inferior a 20 Kg es, aproximadamente,

$$\begin{aligned} P\left[\overline{P} < 20\right] &\simeq P\left[\frac{\overline{P} - 25}{\sqrt{4}} < \frac{20 - 25}{\sqrt{4}}\right] = P\left[Z < \frac{-5}{2}\right] = P\left[Z < -2.5\right] \\ &= P\left[Z > 2.5\right] = 1 - \Phi(2.5) = 1 - 0.9930 = \mathbf{0.0062} \end{aligned}$$

Ejemplo TCL: turistas franceses e ingleses

- **Ejercicio 9:** Tras efectuar un amplio estudio sobre los turistas franceses e ingleses que pasan sus vacaciones en la Comunidad de Madrid, se ha comprobado que el gasto medio en alimentación de los franceses es de 21 euros por persona y día, con una varianza de 36 euros, mientras que para los ingleses el gasto medio es de de 18 euros persona y día, con una desviación típica de 3 euros.

  Se ha realizado una encuesta a 300 turistas franceses y a 200 ingleses.

  Suponiendo que cada grupo de turistas encuestados es una selección de variables **.,i.d.**, y que los gastos en ambos grupos con independientes:

  1. Calcular la probabilidad de que el **gasto total** de la muestra de los franceses supere los 7000 euros.
  2. ¿Cómo es de probable que los **gastos medios** diarios en alimentación de los turistas franceses no superen a los gastos medios de los ingleses en más de 2.5 euross?

Solución al Ejercicio 9

- Como es habitual, Z denotará la distribución normal estándar,

$$Z \sim N(0, 1),$$

cuya función de distribución es Φ.

1. Disponemos de una m.a.s. de gastos de los turistas franceses, es decir, de una colección de variables i.i.d.,

$$F_1, F_2, \ldots, F_{300},$$

con $E(F_i) = 21$ y $V(F_i) = 36$, pero con distribución desconocida.

Tenemos que calcular una probabilidad sobre la variable aleatoria gastos **totales**,

$$T = F_1 + F_2 + \ldots + F_{300}.$$

Solución al Ejercicio 9 (continuación)

- Aunque la distribución de las F_i es desconocida, como disponemos de una **muestra grande**, el **Teorema del Límite Central** garantiza que la distribución de su suma es aproximadamente normal, con esperanza la suma de las medias, es decir,

$$E(T) = \sum_{i=1}^{300} E(F_i) = 300 \cdot 21 = 6300 \text{ euros},$$

y, por tratarse de variables independientes, con varianza,

$$V(T) = \sum_{i=1}^{300} V(F_i) = 300 \cdot 36 = 10800 \text{ euros}^2,$$

es decir,

$$T \approx N(\mu = 6300, \sigma^2 = 10800).$$

Solución al Ejercicio 9 (continuación)

- Por tanto, la probabilidad de que el total de los gastos de esta muestra de tamaño 300 supere los 7000 euros es, aproximadamente,

$$\begin{aligned}
P\left[T > 7000\right] &= P\left[\frac{T-6300}{\sqrt{10800}} > \frac{7000-6300}{\sqrt{10800}}\right] \\
&\simeq P\left[Z > 6.7357\right] \\
&= 1-\Phi(6.7357) \\
&\simeq 0.
\end{aligned}$$

Solución al Ejercicio 9 (continuación)

2 En este caso se nos pregunta una probabilidad que se refiere al **gasto medio** de franceses e ingleses, es decir, a los **promedios de los gastos de las muestras** $\overline{F}$ y $\overline{I}$, respectivamente, donde

$$I_1, I_2, \ldots, I_{200},$$

es la muestra de los gastos de los turistas ingleses para la cual se verifica

$$E(I_i) = 18 \text{ euros}$$

y

$$V(I_i) = 3^2 = 9 \text{ euros}^2.$$

De nuevo en virtud del **Teorema del Límite Central** se tiene que, por ser ambos tamaños grandes,

$$\overline{F} \simeq N\left(\mu = 21, \sigma^2 = \frac{36}{300}\right)$$

e

$$\overline{I} \simeq N\left(\mu = 18, \sigma^2 = \frac{9}{200}\right).$$

Solución al Ejercicio 9 (continuación)

- Tenemos que calcular la probabilidad de que la diferencia entre los **gastos medios** diarios en alimentación de los turistas franceses y los ingleses de las muestras sea inferior a 2.5 euros, es decir

$$P(\overline{F} - \overline{I} < 2.5).$$

La distribución aproximada de la diferencia de estos promedios es

$$\overline{F} - \overline{I} \approx N\left(\mu = 21 - 18, \sigma^2 = \frac{36}{300} + \frac{9}{200}\right) \equiv N(\mu = 3, \sigma^2 = 0.165).$$

Por tanto,

$$P(\overline{F} - \overline{I} < 2.5) = P\left(\frac{\overline{F} - \overline{I} - 3}{\sqrt{0.165}} < \frac{2.5 - 3}{\sqrt{0.165}}\right) \simeq P(Z < -0.121),$$

es decir,

$$P(\overline{F} - \overline{I} < 2.5) \simeq 1 - \Phi(0.121) = \mathbf{0.1092}$$

Ejemplo TCL: aproximación suma de exponenciales

- **Ejercicio 10:** Consideremos una colección $\{X_i\}_{i\in\mathbf{N}}$ de variables aleatorias e idénticamente distribuidas con distribución común

$$X_i \sim Exp(0.5).$$

Determinar la distribución aproximada de

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

y de

$$\overline{X}_n = \frac{\sum\limits_{i=1}^{n} X_i}{n},$$

para n grande.

resolución:.........pizarra

Aproximación de la binomial

por la distribución normal

Sonia Hernández Alonso
Probabilidad-Grado en Matemáticas (URJC)

Calcular probabilidades sobre $Bin(n, p)$ con n grande

- El cálculo de probabilidades sobre una binomial cualdo el número de repeticiones n es grande puede resultar largo y tedioso, salvo que se utilice algún programa como R.

- **Ejemplo 4:** La proporción de piezas defectuosas que fabrica cierta máquina es del 40 %.

  Si se seleccionan al azar 500 piezas, ¿cuál es la probabidad de que el número de defectuosas esté entre 197 y 205?

  resolución:.........pizarra *Solución:* 0.2821

- Como consecuencia inmediata del Teorema 1, podemos encontrar una **aproximación de la distribución binomial a la normal cuando n es grande**.

Teorema de De Moivre-Laplace

- **Teorema 2:** Si $H_n \sim Bin(n,p)$, entonces, asintóticamente se verifica que

$$H_n \approx N(\mu = np, \sigma^2 = npq).$$

demostración:.........pizarra

- Este resultado nos permite calcular probabilidades aproximadas sobre binomiales con n grande de manera más cómoda.

¿Cuándo puede aplicarse esta aproximación?

- En general, suele considerarse que la aproximación que proporciona la aproximación de De Moivre-Laplace, es buena si se verifica

$$n\,p \geq 5,$$

y

$$n\,q \geq 5.$$

Corrección por continuidad

- Nótese que la distribución binomial es discreta, mientras que la normal es una distribución continua.
- Cuando aproximamos una distribución binomial mediante una normal, estamos aproximando una variable discreta por una continua.
- Esto implica, entre otras cosas, que si aplicamos directamente la aproximación de De Moivre-Laplace, obtendríamos, por ejemplo, que, para cualquier $k \in \{0, 1, 2, \ldots, n\}$,

  $$P(H_n = k) \approx 0,$$

  lo que evidentemente no parece una buena aproximación.
- Por ello, cuando se aproximan probabilidades de una binomial por una normal, conviene aplicar la **corrección por continuidad de Yates**, que consiste en ampliar en el intervalo considerado en 0.5 unidades en cada extremo del intervalo considerado.

Ejemplo: aproximación binomial por normal

- **Ejercicio 11:** La proporción de piezas defectuosas que fabrica cierta máquina es del 40 %.

  Se seleccionan al azar 500 piezas. Calcular la probabidad de que el número de defectuosas esté entre 197 y 205.

  resolución:.........pizarra

- *Solución:*
  - Exacta: 0.2821
  - Sin corrección por continuidad: 0.2484
  - Con corrección por continuidad: 0.2825

Otras versiones del T.C.L.

URJC

¿En qué casos puede aplicarse el teorema 1?

- El **Teorema 1** establece que

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z,$$

o equivalentemente que

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{d} Z,$$

siempre y cuando las variables $X_1, X_2, \ldots,$ sean independientes e **idénticamente distribuidas**.

- Pero **¿qué ocurre cuando las variables no tienen todas la misma distribución?**

- **¿Puede aplicarse el TCL cuando las medias y/o varianzas de las variables no son todas iguales?**

- Existen distintas versiones el teorema del límite central que no requieren esta condición.

Condiciones para T.C,L.'s para variables no i.i.d.

- El problema de los teoremas del límite central para variables que no tienen todas la misma distribución es que requieren condiciones que, en este punto del grado, os resultarán difíciles de intepretar, ya que a primera vista no resultan intuitivas.
- Pero vamos a tratar de entender las conclusiones a las que llegan estos teoremas.
- Las condiciones que se imponen son condiciones sobre las varianzas de las X_i, que recordemos que en el caso general no tienen porque ser todas iguales.

¿Como son los T.C,L.'s para v.al. no i.d?

- Sea $X_1, X_2, \ldots,$ una sucesión de variables aleatorias independientes con $\mu_i = E(X_i)$ y $\sigma_i^2 = V(X_i)$.

  Definimos

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

  Entonces, bajo determinadas condiciones se verifica

$$\frac{S_n - \sum\limits_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \xrightarrow{d} Z.$$

¿Qué condiciones se requieren?

- Para aplicar el resultado anterior es necesario que se verifiquen ciertas restricciones que, como ya hemos comentado, no resultan muy intuitivas.

- Las siguientes condiciones (que se dejan para vuestra consulta), proporcionan **condiciones suficientes** para asegurar que que la aproximación de la página anterior es válida:

  - **Condiciones de Lindeberg**

    Mirar, por ejemplo, Yohai, V.J. *Notas de Probabilidades y Estadística*, Teorema 11.12

  - **Condiciones de Lyapunov**

    Mirar, por ejemplo, Yohai, V.J. *Notas de Probabilidades y Estadística*, Teorema 11.13

Bibliografía

- Yohai, V.J. *Notas de Probabilidades y Estadística*. 2008,

  Capítulos 10 y 11.

- Wackerly D., Mendenhall W. y Scheaffer R.L. *Estadística Matemática con Aplicaciones*. Thompson, 2008.

  Capítulo 7.

- Grimmet, G. y Welsh, D. *Probability: An Introduction*. Oxford University Press, 1996.

  Capítulo 8.