Probabilidades y Estadística (C)

Clase Práctica 17

Matías Data

Departamento de Matemática, FCEyN, Universidad de Buenos Aires

20/10/2020

Introducción

- Hoy veremos la Ley de los Grandes Números (LGN) y el Teorema Central de Límite (TCL) en R.
- Recordemos los enunciados de estos teoremas:

Teorema

(Ley Débil de los Grandes Números) Sean variables aleatorias i.i.d. $X_1, X_2, X_3, \dots$ con $\mathbb{E}(X_1) = \mu$ y $\mathbb{V}(X) = \sigma^2$, entonces:

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

Teorema

(Teorema Central del Límite) Sean variables aleatorias i.i.d. $X_1, X_2, X_3, \dots$ con $\mathbb{E}(X_1) = \mu$ y $\mathbb{V}(X) = \sigma^2$, entonces:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} N(0, 1)$$

Ley de los Grandes Números

- La convergencia en probabilidad de la LGN nos dice que para todo $\varepsilon > 0$:

$$P(|\overline{X}_n - \mu| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$$

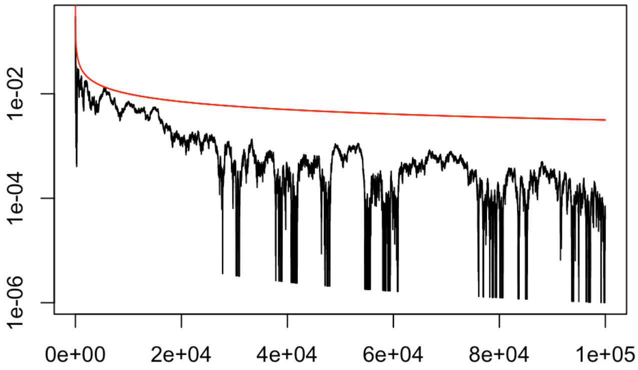
- Esto se deduce inmendiatamente de la Desigualdad de Chebyshev ya que:

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\mathbb{V}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$$

- Como la varianza del promedio $\overline{X}_n$ es σ^2/n, la convergencia a μ será del orden de $1/\sqrt{n}$, ya que el desvio estándar de $\overline{X}_n$ será $\sigma/\sqrt{n}$, y como es aproximadamente normal se mantendrá casi siempre a menos de tres desvios estándar de μ.

Ley de los Grandes Números

- Veamos como ejemplo la convergencia para $X_i \sim \mathrm{Be}(p)$ para $p = 0.3$.



Ley de los Grandes Números

- Una aplicación de la LGN es integración de Monte Carlo.
- Supongamos que queremos integrar una función $h : [a,b] \to \mathbb{R}$ (integrable).
- Sea $U \sim \mathcal{U}[a,b]$, entonces:

$$\mathbb{E}(h(U)) = \int_{-\infty}^{+\infty} h(u) \frac{1}{b-a} \mathbb{1}_{[a,b]}(u)\, du =$$
$$= \frac{1}{b-a} \left(\int_a^b h(u)\, du \right)$$

- Luego por la LGN tenemos que:

$$(b-a) \left(\frac{\sum_{i=1}^{n} h(U_i)}{n} \right) \xrightarrow{P} \int_a^b h(u)\, du$$

- Veamoslo en R.

Ley de los Grandes Números

- Veamos un no ejemplo de la LGN.
- De hecho la LGN tiene una formulación más fuerte, en realidad vale que la sucesión converge con probabilidad uno a μ, y además no es necesario pedir que tenga segundo momento finito (varianza finita), solo es necesario que tenga primer momento finito (media).
- Consideremos el siguiente ejemplo, llamado la Paradoja de San Petersburgo.
- Un apostador nos propone el siguiente juego, se tira una moneda sucesivas veces hasta que sale una ceca. Si salen k caras seguidas, entonces el juego paga 2^k dólares.
- Cuánto está dispuesto a pagar para jugar este juego?
- Si somos neutrales al riesgo, deberiamos estar dispuestos a pagar el valor esperado del juego.

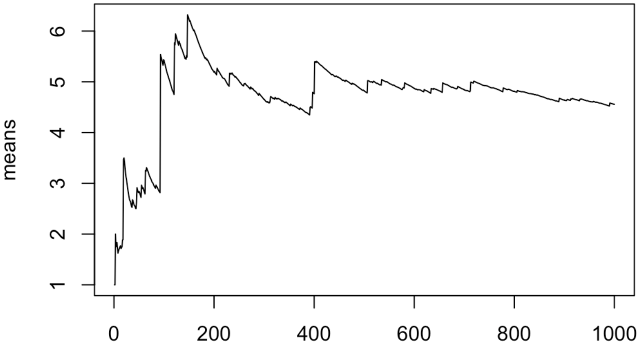
Ley de los Grandes Números

- Pero si X es el pago del juego, entonces $X = 2^k$ con $p = 1/2^k$, luego:

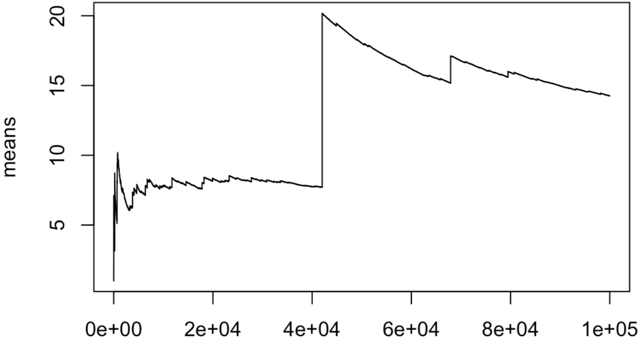
$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} 2^k \frac{1}{2^k} = 1 + 1 + 1 + \cdots = +\infty$$

- Luego a cualquier precio que nos propongan, deberiamos aceptar jugar.
- En estudios empiricos, la gente paga un promedio menor a \$5 dólares.
- Veamos con una simulación en R que si tomamos el promedio este no converge y de hecho crece en el tiempo.

Ley de los Grandes Números



Ley de los Grandes Números



Teorema Central del Límite

- Hemos visto en la clase 15 la aproximación de la binomial por una normal por el TCL, es decir cuando los experimentos son Bernoullis.
- Veremos ahora otros dos casos para ilustrar como funciona en la práctica el teorema, ver que tan buena es la aproximación.
- Consideramos primero el caso de sumas de uniformes indendientes. Más precisamente tomamos $U_i \sim \mathcal{U}[-1, 1]$, para $i = 1, \dots, n$.
- Como la distribución exacta es difícil de calcular, vamos a simular N sumas, para N grande (por ejemplo $N = 100000$), y en lugar de graficar la densidad de la suma $S_n = U_1 + \cdots + U_n$, vamos a graficar el histograma obtenido y compararlo con la densidad de la normal que la aproxima.

Teorema Central del Límite

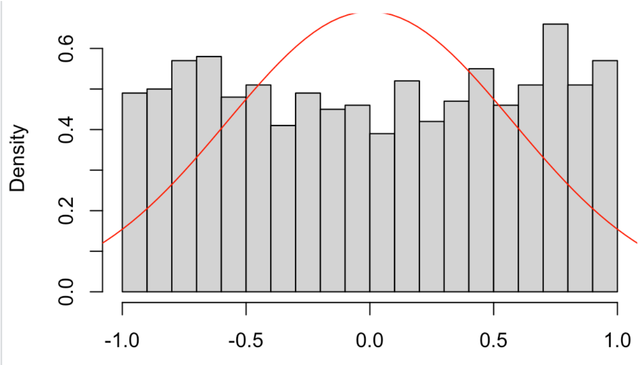


Fig.: $n = 1$

Teorema Central del Límite

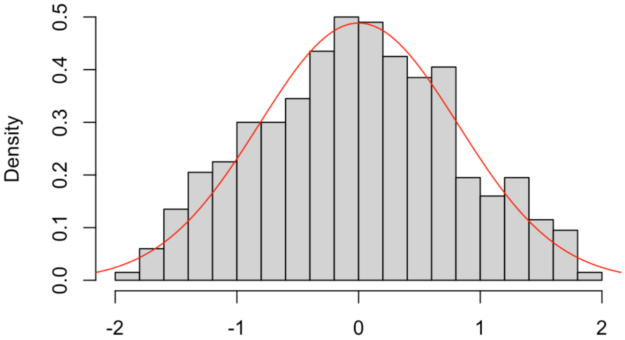


Fig.: $n = 2$

Teorema Central del Límite

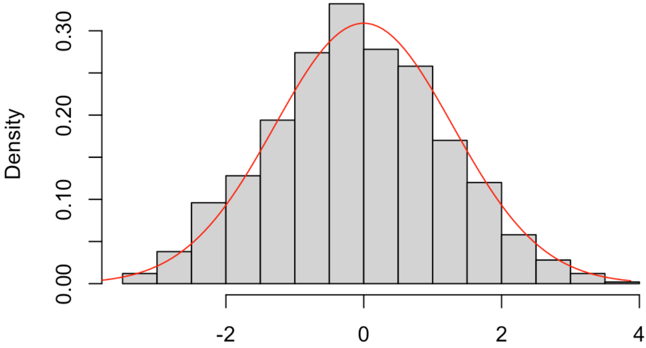


Fig.: $n = 5$

Teorema Central del Límite

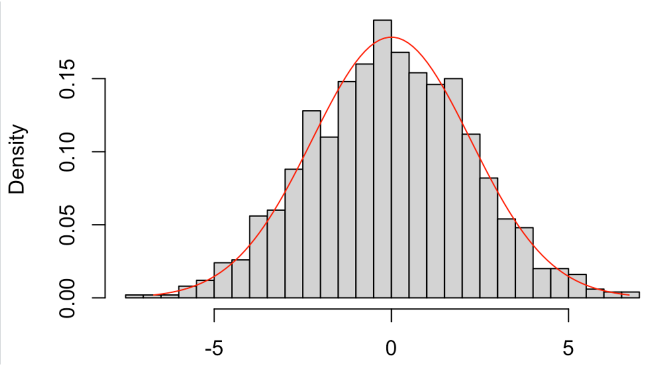


Fig.: $n = 15$

Teorema Central del Límite

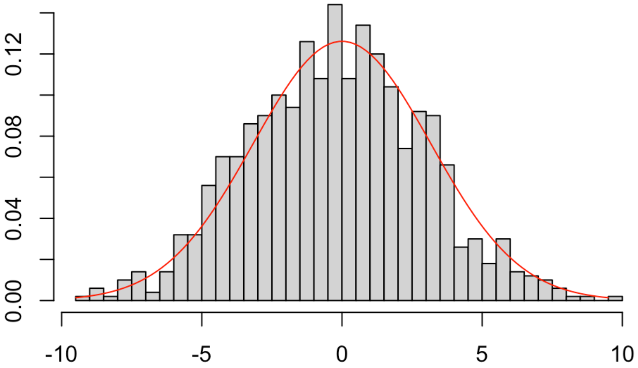


Fig.: $n = 30$

Teorema Central del Límite

- Para distribuciones simétricas la convergencia del TCL es muy rápida, como observamos en el caso uniforme.
- Veamos el caso de suma de exponenciales, es decir $X_i \sim \mathcal{E}(1)$, que son distribuciones muy asimétricas y por lo tanto debería ser más lenta la convergencia.
- Otra ventaja de este caso es que la suma de n exponenciales λ es $\Gamma(n, \lambda)$, con lo cual podemos plotear su densidad exacta.

Teorema Central del Límite

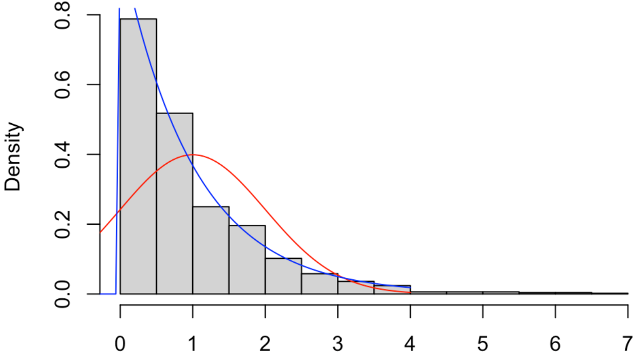


Fig.: $n = 1$

Teorema Central del Límite

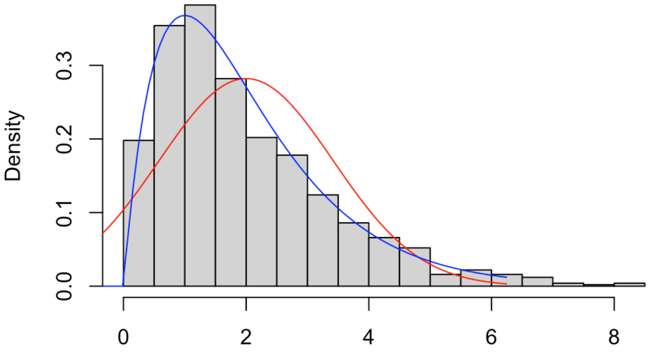


Fig.: $n = 2$

Teorema Central del Límite

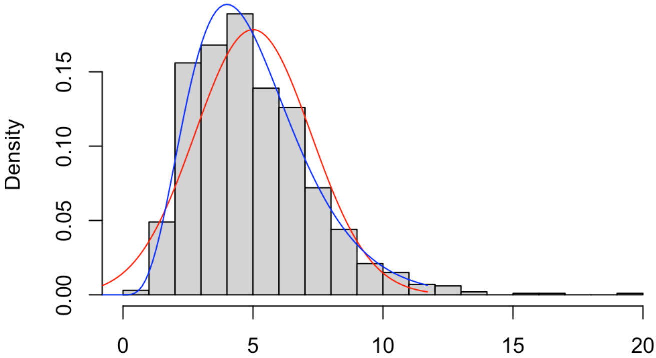


Fig.: $n = 5$

Teorema Central del Límite

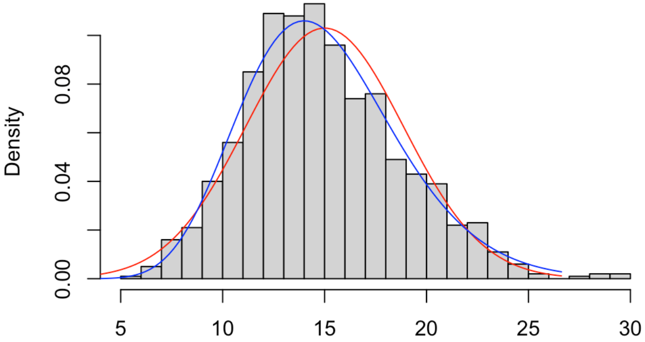


Fig.: $n = 15$

Teorema Central del Límite

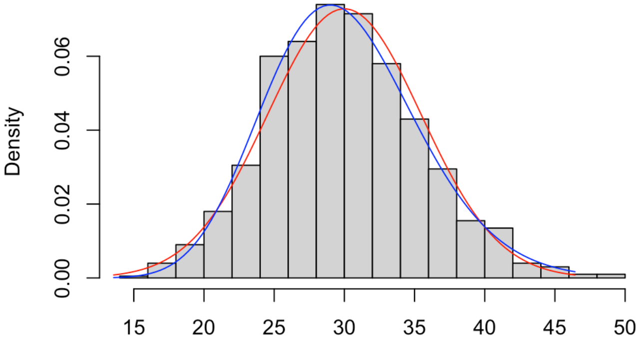


Fig.: $n = 30$