Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

Teoremas Límites

Alberto Vera Azócar, albvera@ing.uchile.cl

1. Nociones de convergencia

Digamos que tenemos una colección de v.a. $(X_n)_{n\in\mathbb{N}}$ y nos interesa su comportamiento si hacemos $n \to \infty$, por ejemplo nos preguntamos ¿qué pasa con el promedio de las v.a.?.

Consideremos el experimento donde hay una moneda que sale cara con probabilidad $\frac{3}{4}$. Lanzamos n veces la moneda y sea X_n la v.a. que vale 1 si salió cara y 0 si salió sello la n-ésima tirada.

Uno esperaría que si n es grande, la proporción de caras que salen se parezca mucho a $\frac{3}{4}$, más aún quisiéramos escribir lo siguiente:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{3}{4}$$

Lo anterior se interpreta como la proporción de caras que salieron en las n tiradas, o bien, el promedio empírico de las n variables.

Recordemos convergencia de funciones de $\mathbb{R}$ en $\mathbb{R}$: sea por ejemplo (f_n) una sucesión de funciones dada por $f_n(x) = x + \frac{1}{n}$, afirmamos que la sucesión (f_n) converge puntualmente a la función $g(x) = x$. En efecto, sea $x \in \mathbb{R}$ (cualquiera), tenemos que $\lim_{n\to\infty} f_n(x) = x + \lim_{n\to\infty} \frac{1}{n} = g(x)$, luego es cierto que $\forall\, x \in \mathbb{R}, f_n(x) \to g(x)$.

En nuestro experimento de lanzar infinitas monedas podemos definir Ω como tuplas infinitas del estilo $(c, c, s, c, \cdots)$ donde $c =$cara y $s =$sello. Sea $w' \in \Omega$ dado por $w' = (c, c, c, c, \cdots)$, es decir, el evento donde en cada lanzamiento siempre salió cara, tenemos que $\forall\, n \in \mathbb{N}, X_n(w') = 1$, luego

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k(w') = 1$$

Por lo tanto para w' no es cierto que el promedio de los X_n converge a $\frac{3}{4}$, pero todos estamos de acuerdo que w' tiene probabilidad muy baja de ocurrir (de hecho ocurre con probabilidad 0, pues que n veces seguidas salga sello tiene probabilidad $(\frac{1}{4})^n$).

Concluimos que la buena noción de convergencia no es la que usamos para funciones de $\mathbb{R}$ en $\mathbb{R}$, es decir, no queremos que sea cierto "el promedio de los $X_n(w)$ converge a $\frac{3}{4}$ para todo $w \in \Omega$", lo que en realidad queremos decir es "el promedio de los $X_n(w)$ converge a $\frac{3}{4}$ con probabilidad 1".

Definición 1.1. *Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. y sea X otra v.a., decimos que*

1. *$X_n \to X$ casi seguramente si $\mathbb{P}(X_n \to X) = 1$, denotado $X_n \xrightarrow{c.s.} X$.*

2. *$X_n \to X$ en probabilidad si $\forall\, \varepsilon > 0, \mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$, denotado $X_n \xrightarrow{\mathbb{P}} X$*

3. $X_n \to X$ *en distribución si* $\forall\, x \in \mathbb{R}$, $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$, *denotado* $X_n \xrightarrow{d} X$.

Interpretamos la convergencia casi segura como "con probabilidad 1 se observa que las v.a. convergen", es la noción más fuerte de convergencia que estudiaremos.

La convergencia en probabilidad se lee "la probabilidad de que X_n no se parezca a X converge a cero".

Por último, la convergencia en distribución nos dice que para todo $A \subseteq \mathbb{R}$, $\mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A)$.

Intuitivamente todas las nociones de convergencia nos dicen lo mismo: "cuando n es grande podemos aproximar X_n por X", pero matemáticamente son distintas y tienen propiedades diferentes.

El siguiente ejemplo es muy sencillo, pero ayuda a comprender las nociones que acabamos de definir.

Ejemplo 1.1. *Sea* $\Omega = \{a, b\}$ *y dotemos el conjunto con la probabilidad uniforme, i.e.* $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \frac{1}{2}$.

Sea $X : \Omega \to \mathbb{R}$ *dada por* $X(a) = X(b) = 1$, *es decir, la v.a. constante. Definamos la sucesión de v.a.* (X_n) *por*

$$X_n(w) = \begin{cases} 1 & \text{si } w = a \\ \frac{n-1}{n} & \text{si } w = b \end{cases}$$

Se pide estudiar la convergencia de la sucesión (X_n)

Solución: Estudiaremos las tres nociones, comenzando por convergencia c.s., debemos probar que $\mathbb{P}(X_n \to X) = 1$, *equivalentemente, si definimos* $A \subseteq \Omega$ *como* $A = \{w \in \Omega : X_n \to X\}$, *hay que probar que* $\mathbb{P}(A) = 1$.

Veamos entonces que $a, b \in A$: *como* $X_n(a) = 1$ *claramente* $X_n(a) \to X(a)$. *Para* b *tenemos que* $X_n(b) = \frac{n-1}{n} \to 1$, *es decir* $X_n(b) \to X(b)$. *Concluimos que* $A = \Omega$ *y por lo tanto* $\mathbb{P}(A) = 1$.

Para ver convergencia en probabilidad tomemos un $\varepsilon > 0$, *particionemos el conjunto dependiendo del valor que toma* X_n:

$$\begin{aligned} \mathbb{P}(|X_n - X| > \varepsilon) &= \mathbb{P}(|X_n - 1| > \varepsilon) \\ &= \mathbb{P}(|X_n - 1| > \varepsilon, X_n = 1) + \mathbb{P}(|X_n - 1| > \varepsilon, X_n \neq 1) \\ &= 0 + \mathbb{P}(|X_n - 1| > \varepsilon, X_n \neq 1) \\ &= \mathbb{P}\left(\left|\frac{n-1}{n} - 1\right| > \varepsilon\right) \\ &= \mathbb{P}\left(\frac{1}{n} > \varepsilon\right) \end{aligned}$$

Claramente para un n *grande la última probabilidad vale 0, por lo que* $X_n \xrightarrow{\mathbb{P}} X$.

Finalmente para ver convergencia en distribución basta ver que para $x < 1$ *tenemos que* $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$, *pues para* $x = 1$ *ambas probabilidades valen 1. Por un lado vemos que si* $x < 1$, *entonces* $\mathbb{P}(X \leq x) = 0$. *Para* (X_n) *se tiene que*

$$\mathbb{P}(X_n \leq x) = \mathbb{P}\left(\frac{n-1}{n} \leq x\right) \longrightarrow 0$$

Queda demostrado entonces que nuestra sucesión converge a X *según los tres criterios.*

El siguiente teorema (que aceptaremos sin demostración) nos dice que la noción más fuerte es la convergencia c.s. y la más débil es la convergencia en distribución (usando este teorema en el ejemplo anterior, basta probar convergencia c.s.).

Teorema 1.1. *Sea* (X_n) *sucesión de v.a. y* X *v.a.*

1. *Si* $X_n \xrightarrow{c.s.} X$*, entonces* $X_n \xrightarrow{\mathbb{P}} X$*.*
2. *Si* $X_n \xrightarrow{\mathbb{P}} X$*, entonces* $X_n \xrightarrow{d} X$*.*

El resultado que veremos ahora fue formulado como conjetura en 1853 por el matemático francés Bienaymé, pero fue el matemático ruso Chebyshev (pronúnciese Chebihshyov) quien lo demostró en 1867, por eso lleva su nombre.

Teorema 1.2 (Desigualdad de Chebyshev)**.** *Sea* X *una v.a. con varianza y esperanza finitas. Si definimos* $\mu := \mathbb{E}(X)$ *y* $\sigma^2 := \mathbb{V}ar(X)$*, entonces* $\forall\, k > 0$

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Mediante la desigualdad de Markov (pronúnciese Markof), quien fue alumno de Chebyshev, se puede probar fácilmente el teorema anterior.

Teorema 1.3 (Desigualdad de Markov)**.** *Sea* X *una v.a. positiva, i.e.* $\mathbb{P}(X \geq 0) = 1$*, entonces* $\forall a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Demostración. Tenemos que por ser X positiva $a\mathbb{1}_{\{X\geq a\}} \leq X$, luego si tomamos esperanza a ambos lados por monotonía $\mathbb{E}(a\mathbb{1}_{\{X\geq a\}}) \leq \mathbb{E}(X)$. Además si usamos linealidad y que $a > 0$ tenemos $\mathbb{E}(\mathbb{1}_{\{X\geq a\}}) \leq \frac{\mathbb{E}(X)}{a}$.

Finalmente $\mathbb{1}_{\{X\geq a\}}$ es una v.a. que toma sólo dos valores, 0 y 1, por lo que su esperanza es fácil de obtener: $\mathbb{E}(\mathbb{1}_{\{X\geq a\}}) = 1 \cdot \mathbb{P}(\mathbb{1}_{\{X\geq a\}} = 1) + 0 \cdot \mathbb{P}(\mathbb{1}_{\{X\geq a\}} = 0) = \mathbb{P}(X \geq a)$ □

2. Ley Fuerte de Grandes Números

Este resultado es probablemente el avance más importante que se ha hecho después del teorema fundamental del cálculo. La primera versión fue probada por Jacob Bernoulli en 1713 usando las v.a. de Bernoulli. Después fueron saliendo versiones cada vez más potentes hasta la que estudiaremos ahora.

Teorema 2.1 (Ley de Grandes Números)**.** *Sea* (X_n) *una sucesión i.i.d. de v.a., tales que* $\mu := \mathbb{E}(X_1)$ *es finita, entonces*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{c.s.} \mu$$

Lo que nos dice el teorema es que con probabilidad 1 el promedio empírico de las v.a. converge a la esperanza.

Una forma de verlo es el ejemplo con que comenzamos, si lanzamos 100 monedas que salen cara con probabilidad $\frac{3}{4}$, entonces uno esperaría observar aproximadamente 75 caras, o bien, que la proporción de caras sea 75 %, donde la proporción de caras es exactamente $\frac{1}{n}\sum_{k=1}^{n} X_n$ (pues X_n vale 1 si la n-ésima moneda fue cara y 0 sino).

La ley de grandes números nos dice que en efecto cuando $n \to \infty$ observaremos que la proporción de caras es $\frac{3}{4}$, pues la esperanza de cada variable es justamente esa.

Ejemplo 2.1. *Considere una persona que con probabilidad $\frac{1}{2}$ retrocede $1[m]$ en el instante n y con probabilidad $\frac{1}{2}$ avanza $1[m]$ (siempre elige de forma independiente).*

Muestre que si definimos Z_n como la distancia a la que se encuentra del punto de partida en el instante n , entonces con probabilidad 1 se tiene que $\frac{1}{n}Z_n \to 0$ cuando $n \to \infty$.

Solución: Debemos hacer aparecer una suma para usar ley de grandes números, sea entonces X_n la v.a. que vale 1 si en el instante n la persona avanzó 1 y que vale -1 si retrocedió 1.

Tenemos que $Z_n = \sum_{k=1}^{n} X_k$. Las v.a. (X_n) son independientes e idénticamente distribuidas por enunciado, más aún $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$, por lo tanto $\mathbb{E}(X_n) = 0 \, \forall \, n \in \mathbb{N}$. Por ley de grandes números sigue que

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{c.s.} 0$$

es decir, con probabilidad 1 se observa $\frac{1}{n}Z_n \to 0$, que es lo que queríamos probar.

3. Teorema Central del Límite

Este teorema se denominó "central" por la importancia que tiene en la teoría de las probabilidades. Los primeros resultados importantes se deben a Pierre-Simon, marquis de Laplace, aunque hay algunos trabajos previos.

En esencia el teorema nos dice que cualquier suma de n variables i.i.d. se parece a una normal cuando n es grande.

Teorema 3.1 (Teorema del Límite Central)**.** *Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. con esperanza y varianza finita. Sea $Z \sim \mathcal{N}(0,1)$, si denotamos $\mu := \mathbb{E}(X_1)$ y $\sigma^2 := \mathbb{V}ar(X_1)$, entonces*

$$\frac{\sum_{k=1}^{n}(X_k - \mu)}{\sqrt{n}\sigma} \xrightarrow{d} Z$$

Recordando la noción de convergencia en distribución, en particular tenemos $\forall \, z \in \mathbb{R}$ lo siguiente:

$$\mathbb{P}\left(\frac{\sum_{k=1}^{n}(X_k - \mu)}{\sqrt{n}\sigma} \leq z\right) \approx \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Por la observación anterior, la siguiente función nos será útil para los cálculos:

Definición 3.1. *Definimos la función $\Phi : \mathbb{R} \to [0,1]$ como la probabilidad acumulada de una normal standard, vale decir*

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Ejemplo 3.1. *Se lanza una moneda 100 veces que sale cara con probabilidad $\frac{1}{2}$, se pide aproximar mediante TCL la probabilidad de obtener 60 o más caras.*

Solución: Definamos X_n como una v.a. Bernoulli que dice si la n-ésima tirada sale cara con $n = 1, \cdots, 100$. Claramente los (X_n) son independientes y la probabilidad que nos piden es

$$\mathbb{P}\left(\sum_{n=1}^{100} X_n \geq 60\right)$$

Definamos $\mu := \mathbb{E}(X_1) = \frac{1}{2}$ *y* $\sigma^2 = \mathbb{V}ar(X_1) = \frac{1}{2}(1-\frac{1}{2})$ *(fórmulas de esperanza y varianza conocidas de una Bernoulli).*

$$\begin{aligned}
\mathbb{P}\left(\sum_{n=1}^{100} X_n \geq 60\right) &= \mathbb{P}\left(\sum_{n=1}^{100} X_n - 100\mu \geq 60 - 100\mu\right) \\
&= \mathbb{P}\left(\sum_{n=1}^{100} (X_n - \mu) \geq 60 - 100\mu\right) \\
&= \mathbb{P}\left(\frac{\sum_{n=1}^{100}(X_n-\mu)}{\sqrt{100}\sigma} \geq \frac{60-100\mu}{\sqrt{100}\sigma}\right) \\
\textit{(por TCL)} \quad &\approx \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{60-100\mu}{\sqrt{100}\sigma}\right) \\
&= \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{60-50}{5}\right) \\
&= \mathbb{P}\left(\mathcal{N}(0,1) \geq 2\right) \\
&= 1 - \Phi(2)
\end{aligned}$$

Usando cualquier software se puede obtener lo anterior, en particular usando matlab tenemos que la aproximación de la probabilidad es $0,0227$.

Ahora para "comprobar" el resultado, notamos que la cantidad de caras es una binomial de parámetros 100 y $\frac{1}{2}$*, por lo que se puede calcular exactamente la probabilidad*

$$\mathbb{P}\left(\sum_{n=1}^{100} X_n \geq 60\right) = \sum_{k=60}^{100} \binom{100}{k} (0,5)^k (0,5)^{100-k} = (0,5)^{100} \sum_{k=60}^{100} \binom{100}{k}$$

La última expresión es complicadísima de calcular, pero puede hacerse también mediante un software, por ejemplo usando maple nos da que la probabilidad (exacta) es 0,0284.

Ambos números (el exacto y el usando TCL) se parecen mucho y claramente la expresión $1-\Phi(2)$ *es más informativa, fácil de graficar y de interpretar.*