PROBABILIDAD 2

Licenciatura en Estadística — Licenciatura en Economía
Facultad de Ciencias Económicas y de Administración
Universidad de la República

1 de abril de 2024

Comentarios

Estas notas fueron realizadas por Diego Armentano para el curso *Probabilidad 2* de la Licenciatura en Estadística de la Facultad de Ciencias Económicas y de Administración, en el primer semestre de 2023.

Índice general

Capítulo 1

Abreboca

Comencemos estas notas con una motivación de los teoremas límites que veremos en este curso.[1]

Supongamos que una persona va al casino a jugar a la ruleta. La misma tiene una rueda con 37 “slots”, del 0 al 36, el 0 de color verde, 18 números de color rojo, y 18 números de color negro.

Figura 1.1: Imagen de wikipedia

Si nuestro jugador apuesta 1\$ dólar a color rojo, entonces entonces ganará un dólar si sale rojo, o perderá un dólar si sale negro o el 0.

Supongamos que nuestra persona se lleva una valija de dinero, y empieza a apostar de manera sucesiva. Sean $X_1, X_2 \dots$ el monto ganado $(+1)$, o perdido (-1), en cada apuesta de manera sucesiva.

Formalmente podemos pensar X_i como variables aleatorias *independientes*, que toman los valores 1 o -1, con las siguientes probabilidades

$$\mathbb{P}(X_1 = 1) = \frac{18}{37} \approx 0{,}486, \qquad \mathbb{P}(X_1 = -1) = \frac{19}{37} \approx 0{,}514.$$

Nuestro apostador se interesa por sus ganancias (o pérdidas) en tiempo n, esto es,

$$S_n = X_1 + \ldots + X_n, \qquad n = 0, 1, 2, \ldots,$$

y como no sabe mucho de probabilidades nos pide que lo asesoremos.

[1]Esta introducción está motivada y guionada por el libro Durret[RD], el cual usaremos de manera esporádica en este curso.

Lo primero que le comentamos es que en cada jugada su ganancia esperada es

$$\mu := \mathbb{E}(X_1) = 1 \cdot \mathbb{P}(X_1 = 1) + (-1) \cdot \mathbb{P}(X_1 = -1)$$

Pero como no le gustaban mucho los símbolos anteriores le pasamos en limpio la siguiente expresión de su ganancia esperada:

$$(1\$) \cdot \frac{18}{37} - (-1\$) \cdot \frac{19}{37} = -\frac{1}{37}\$ \approx -0{,}027\$.$$

Además le comentamos que si sigue jugando, en su apuesta número n tendría ganancias $\mathbb{E}(S_n) = n\mathbb{E}(X_1) = n\mu = -n/37\$$.

Nuesto apostador, con experiencia en el tema, nos dice que le parece raro que en su n-ésima apuesta pierda ese monto porque sus ganancias son números enteros, y $n/37$ no lo parece para n arbitrario.

♢ **Ejer 1.1.** De hecho su observación es correcta, y nosotros podemos preguntarnos, cuán probable es que la ganancia sea exactamente $n/37$. ¿Se animan a pensarlo? Definiendo $a_n = \mathbb{P}(S_n = -n/37)$, ¿qué podemos decir de a_n?

Una vez convencidos de que en realidad el suceso $\{S_n = n\mu\}$ tiende a cero, nuestro apostador nos tranquiliza y comenta que su experiencia es que la ganancia está muy cerca de ese número.

El siguiente resultado va en esa dirección

Ley (débil) de los grandes números:
Sean $X_1, X_2, \ldots$ variables aleatorias i.i.d. con esperanza $\mathbb{E}(X_1) = \mu$, entonces

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \to 0 \quad \text{cuando} \quad n \to 0.$$

En otras palabras, si n es grande, S_n/n está muy cerca de μ con probabilidad muy alta.

Por ejemplo, dado $\varepsilon = 1/100$, si n es suficientemente grande tenemos que el promedio de ganancia S_n/n está en el invervalo $(\mu - 1/100, \mu + 1/100)$ con probabilidad muy alta. Y casi con probabilidad 1 si tomamos n aún más grande.

`Dibujito`

Bien, pero supongamos que nuestro apostador lograra jugar infinitas veces. De la ley débil sabemos S_n/n está en el intervalo $(\mu-\varepsilon,\mu+\varepsilon)$ con probabilidad casi 1. Será que S_n/n tiende realmente a μ en la tirada infinita? La respuesta es sí, y la da el siguiente resultado.

Ley (fuerte) de los grandes números:
Sean $X_1,X_2,\dots$ variables aleatorias i.i.d. con esperanza $\mathbb{E}(X_1)=\mu$, entonces con probabilidad 1 se tiene

$$\frac{S_n}{n}\xrightarrow{n\to\infty}\mu.$$

Este resultado es de una naturaleza distinta al anterior. Como veremos en el caso que estamos viendo, la prueba de la ley débil es un tema de conteo, algo similar al Ejercicio 1.1, donde en vez de pedir la igualdad exacta de $S_n/n=\mu$ queremos que $|S_n/n-\mu|>\varepsilon$, para un ε prefijado. Sin embargo, la ley fuerte habla de una probabilidad en el espacio de tiradas infinitas $\{0,1,\dots,36\}^{\mathbb{N}}$. Eso es un tema más complejo que requiere de la teoría creada por Kolmorogov en los años $30'$, comunmente denomidada "teoría de la medida". Ya hablaremos de esto en el curso.

Volviendo a nuestro apostador, una consecuencia de la ley fuerte es que con probabilidad 1 se tiene $S_n\to-\infty$. Por lo que su valija con dinero se vaciará en algún momento.

Pero todo lo anterior parece muy asintótico, y nuestro apostador le gustaría que seamos más precisos respecto a n no tan grandes. Por ejemplo, para $n=100$ qué puede esperar de sus ganancias. O más específico, cuál es la probabilidad de que tenga ganancia positiva?

El siguiente resultado, conocido como *Teorema Central del Límite*, es uno de los resultados más importantes en matemática y en particular nos permitirá estimar las probabilidades anteriores.

Teorema Central del Límite (TCL):
Sean $X_1,X_2,\dots$ variables aleatorias i.i.d. con esperanza $\mathbb{E}(X_1)=\mu$, y varianza $\sigma^2=\mathbb{E}((X_1-\mu)^2)<\infty$, entonces

$$\mathbb{P}\left(\frac{S_n-n\mu}{\sigma\sqrt{n}}\leqslant u\right)\xrightarrow{n\to\infty}\Phi(u):=\frac{1}{2\pi}\int_{-\infty}^{u}e^{-x^2/2}\,dx,\quad u\in\mathbb{R} \tag{1.1}$$

Observar que esa normalización extraña resulta natural de observar que $\mathbb{E}(S_n)=n\mu$ y que $\operatorname{var}(S_n)=n\sigma^2$, y por lo tanto la v.a.

$$Z_n:=\frac{S_n-n\mu}{\sigma\sqrt{n}},$$

satisface $\mathbb{E}(Z_n)=0$ y $\operatorname{var}(Z_n)=1$. En este sentido Z_n es la correcta normalización para tener un resultado tipo el TCL.

Dicho de otro modo, el TCL dice que la variable aleatoria $\frac{S_n-n\mu}{\sigma\sqrt{n}}$ se aproxima a v.a. $\mathcal{N}$, siendo esta una v.a. con distribución normal estándar. Y por lo tanto podemos escribir

$$S_n\approx\sigma\sqrt{n}\mathcal{N}+n\mu.$$

Más allá de que describe el comportamiento aproximado de S_n, esta aproximación nos ayuda a dar información precisa de S_n utilizando las tablas con datos de Φ.

Por ejemplo podemos contestar a la pregunta si puede volver contento a su casa o no después de 100 apuestas, i.e. $\mathbb{P}(S_{100}>0)$. Para eso observar que

$$\mathbb{P}(S_n>0)\approx\mathbb{P}(\sigma\sqrt{n}\mathcal{N}+n\mu>0).$$

Para estimar lo anterior observemos que

$$\sigma^2=\mathbb{E}(X_1^2)-\mu^2=1-\mu^2=1-(1/37^2)\approx 0{,}999.$$

Luego tomando $\sigma=1$ para facilitar cuentas tenemos

$$S_{100}\approx 10\mathcal{N}-2{,}70.$$

Entonces

$$\begin{aligned}\mathbb{P}(S_{100}>0)&\approx\mathbb{P}(10\mathcal{N}-100/37>0)\\&=\mathbb{P}(\mathcal{N}>0{,}270..)\approx 0{,}4.\end{aligned}$$

Concluimos que nuestro apostador en 100 jugadas, deberia perder en promedio 2,7\$, y tiene probabilidad 0,4 de salir victorioso, algo que él corrobora con su experiencia.

Así no parace tan pesimista el juego de la ruleta, pero si pensamos que hay 100 apostadores jugando, y cada uno juega 100 veces tenemos que la la ganancia total de los apostadores es

$$S_{100^2}\approx 100\mathcal{N}-100^2/37\approx 100\mathcal{N}-270,$$

y por lo tanto la probabilidad de que el casino salga ganando es

$$\mathbb{P}(S_{100^2}<0)\approx\mathbb{P}(\mathcal{N}<2{,}70)\approx 0{,}999.$$

Sabiendo que $\mathbb{P}(\mathcal{N}<1{,}3)\approx 0{,}9$, tenemos que $\mathbb{P}(S_{100^2}<-140)\approx 0{,}9$, por lo que el casino va de manera lenta pero segura a quedarse con la plata de los apostadores.

Con esto concluimos que es mejor no ir apostar, o en su defecto, comprar un casino.

El postre:

Nuestro apostador que ya va entendiendo un poco nos comenta hay algo tramposo en lo anterior. Nos dice que el TCL es asintótico en n, por lo que el símbolo $\approx$ sería para n suficientemente grande al igual que antes, y por lo tanto es un poco engañoso tomar $n=100$.

Es correcto su comentario, pero como comentaremos en este curso, existen teoremas (cuantitativos) que nos dicen que la aproximación en (1.1) es muy rápida para nuestro caso y por lo tanto para n tipo 100 la aproximación tiene un error casi despreciable.

Parte I

Matemática en una cáscara de nuez

Capítulo 2

Introduccion al lenguaje matemático

Resumen: En esta clase haremos una introducción al pensamiento matemático. Para tal fin comenzaremos revisitando las definciones y operaciones básicas que ya conocen y usaremos en este curso. La presentación será rápida y el objetivo principal es tener una puesta en común de las notaciones, y en especial de la forma de pensar, que usaremos. Luego daremos un breve paseo por lo que entendemos es una *prueba matemática*, mostrando distintas técnicas y ejemplos.

(Utilizaremos a lo largo del curso la notación ◇ para indicar los ejercicios.)

Para más referencias se recomienda una lectura del recomendable libro "*A Discrete Transition to Advanced Mathematics*" [?RR].

2.1. Conjuntos y sus operaciones

Definición 2.1. Entendemos por un *conjunto* a una colección definida de objetos[1].

Denotaremos la los conjuntos por letras mayúsculas, por ejemplo A, X,..., etc. y a veces mayúsculas caligráficas $\mathcal{A}, \mathcal{X}$,...,etc.

Existen varias formas de definir conjuntos, via "una frase" bien definida, o de forma descriptiva utilizando "$\{\cdot\}$" para describir una lista de objetos.

Veamos algunos ejemplos.

Ejemplo 2.2. El conjunto X de números naturales entre 1 y 5, que lo podemos representar como $X = \{1,2,3,4,5\}$,

Ejemplo 2.3. El conjunto formado por los departamentos de Uruguay.

Ejemplo 2.4. Sea $\mathcal{C}$ el conjunto determinado por los departamentos de Uruguay que comienzan con "C", el cual podemos representar por $\mathcal{C} = \{$"Canelones" ,"Colonia"$\}$.

Ejemplo 2.5. Sea X el conjunto de soluciones de la ecuación $x^2 - x = 0$. El conjunto X se puede representar como $X = \{0,1\}$.

[1] Para no entrar en formalismos, se puede cambiar *objetos* por *cosas* y listo.

Ejemplo 2.6. Sea $p(x) = \pi x^3 - ex^2 + \sqrt{2}x + 1$. Consideremos el conjunto de soluciones de $p(x) = 0$.

Observar que en el último ejemplo el conjunto está bien definido, pero a veces no es tán sencillo dar una descripción explítica del conjunto, como sí lo hicimos en el Ejemplo 2.5.

◇ **Ejer 2.7.** Dar un argumento para probar que el conjunto definido en el Ejemplo 2.6 no es vacío. (Sug: recordar el Teorema de Bolzano.)

Veamos algunas definiciones y propiedades de conjuntos:

Definiciones 2.8.

- (Pertenencia) El símbolo "$\in$" se utiliza para indicar si un objeto *pertenece* a un conjunto (y "$\notin$" para cuando no pertenece). Por ejemplo, si $X = \{1,2,3,4,5\}$ entonces $1 \in X$, pero $7 \notin X$, o que "Montevideo" $\notin \mathcal{C}$, siendo $\mathcal{C}$ el Ejemplo 2.4.

- (Inclusión) El símbolo "$\subset$" se utiliza para indicar los *subconjuntos*, siendo estos una subcolección de objetos de un conjunto dado. Por ejemplo $\{1,2\} \subset \{1,2,3,4,5\}$.[2]

- (Partes) Denotaremos por $\mathcal{P}(X)$, a las *partes de* X, que es el conjunto formado por todos los subconjuntos de X: es decir, $\mathcal{P}(X) = \{A : A \subset X\}$. Por ejemplo, $\mathcal{P}(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$.

- (Cardinal) El símbolo # indica el número de elementos de un conjunto finito. Si X es un conjunto con infinitos elementos definimos $\#X = \infty$.[3] También utilizaremos la notación $|X|$ para el cardinal de un conjunto. Por ejemplo $\#\{0,1,2\} = 3$, $|\{1,2,\text{“}hola''\}| = 3$.

  ◇ **Ejer 2.9.** ¿Para cuales conjuntos A más abajo se satisface $|A| \in A$?

  1. $A = \{1,2,3,4\}$
  2. $A = \{5,6,7,8\}$
  3. $A = \{1,3,5\}$

[2] Se asume que el conjunto vacío $\varnothing$ y el conjunto total, son sobconjuntos también.

[3] Veremos más adelante que hay distintos tipos de ∞.

4. $A = \{1,3,5,7\}$

♢ **Ejer 2.10.** Sea A un conjunto finito. Encontrar $|A|$ si $A = \{2,3,|A|\}$.

♢ **Ejer 2.11.** Calcular $\#\mathcal{P}(\{1,2,3\})$, y $\#\mathcal{P}(\{1,2,3,4\})$. (Sugerencia: calcular de manera ordenada los conjuntos con 0-elementos, 1-elemento, 2-elementos, 3-elementos, 4-elementos.) Se animan a calcular el cardinal de las las partes de $\{1,\dots,n\}$.

- La *intersección* y *unión* de conjuntos, que se denota por "$\cap$" y "$\cup$" respectivamente, indica los elementos que están en ambos conjuntos, y los elementos que están en por los menos en uno de los conjuntos (pero podrían estar en los dos) respectivamente: es decir,

$$A \cap B = \{x : x \in A, \text{y } x \in B\},$$
$$A \cup B = \{x : x \in A, \text{y/o } x \in B\}$$

Por ejemplo $\{1,2\} \cup \{2,3\} = \{1,2,3\}$, y $\{1,2\} \cap \{2,3\} = \{2\}$.

Observar que siempre es válida la inclusión del subconjunto $A \cap B$ en $A \cup B$.

♢ **Ejer 2.12.** Sean A y B conjuntos finitos. Probar que $|A \cup B| = |A| + |B| - |A \cap B|$.

♢ **Ejer 2.13.** Sean A, B, y C conjuntos finitos. Hallar $|A \cup B \cup C|$. (Sugerencia: usar los diagrama de Venn.)

- (Complemento) Para hablar del complemento de un conjunto necesitamos tener claro cuál es el *universo* donde vive nuestro conjunto, i.e. formalmente se define el complemento de un subconjunto. Si A es un subconjunto de U, entonces el *complemento de A en U* se define por $A^c = U \setminus A = \{x \in U : x \notin A\}$.[4] Por ejemplo, el complemento de $\{0,1\}$ en $\mathbb{R}$, es el conjunto de todos los reales diferentes de 0 y 1, lo cual podemos escribir de manera pícara $\{x \in \mathbb{R} : x^2 - x \neq 0\}$. Pero por otro lado si nuestro universo son los enteros, entonces $\mathbb{Z} \setminus \{0,1\} = \{\dots,-2,-1,2,3,\dots\}$.

Con estas definiciones, ya estamos en condiciones de recordar algunos subconjuntos de los números reales $\mathbb{R}$, que son ya conocidos para ustedes.

Ejemplo 2.14.
- conjunto de números naturales $\mathbb{N} = \{1,2,\dots,\}$;
- conjunto de números enteros $\mathbb{Z} = \{\dots,-2,-1,0,1,2,\dots\}$;
- conjunto de números racionales $\mathbb{Q} = \{m/n : m,n \in \mathbb{Z}, n \neq 0\}$
- conjunto de números irracionales $\mathbb{R} \setminus \mathbb{Q} = \{x \in \mathbb{R} : x \notin \mathbb{Q}\}$
- conjunto de números positivos $\mathbb{R}^+ = \mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$.
- intervalos de $\mathbb{R}$, por ejemplo el conjunto de números reales x tal que $0 \leqslant x < 1$, se denota por $[0,1)$.

[4] La notación $\cdot^c$ sólo se usa cuando es claro el ambiente donde está el subconjunto. De lo contrario es una notación ambigua. Sólo la usaremos cuando el contexto sea claro, por ejemplo, estamos trabajando en subconjuntos de $\mathbb{R}$.

Definición 2.15. Más en general se puede definir la intersección y unión arbitraria de conjuntos. Sean $\{A_\alpha\}_{\alpha \in I}$ una colección de conjuntos indexados por un parámetro $\alpha \in I$. Se define

$$\bigcup_{\alpha \in I} A_\alpha = \{x : x \in A_\alpha, \text{ para al menos un } \alpha \in I\};$$
$$\bigcap_{\alpha \in I} A_\alpha = \{x : x \in A_\alpha, \text{ para todo } \alpha \in I\};$$

♢ **Ejer 2.16.** Considere la colección de intervalos $A_n = \left(0, 1 + \frac{1}{n}\right)$, $n \in \mathbb{N}$. ¿Hallar los conjuntos $\bigcap_{n \in \mathbb{N}} A_n$ y $\bigcup_{n \in \mathbb{N}} A_n$?

♢ **Ejer 2.17.** Hallar $\bigcup_{n \in \mathbb{Z}} [n, n+1)$, y el complemento del conjunto $\bigcup_{n \in \mathbb{Z}} (n, n+1)$.

2.1.1. Producto Cartesiano

Supongamos que tenemos un conjunto X. Dado $n \in \mathbb{N}$, podemos considerar una lista ordenada de *n-tuplas*, $(x_1,\dots,x_n)$, donde cada $x_i \in X$. El conjunto de todas las n-tuplas se denota por X^n y escribimos

$$X^n = \{(x_1,\dots,x_n) : x_i \in X, i = 1,\dots,n\}.$$

Dada la n-tupla $(x_1,\dots,x_n)$, las entradas x_i se llama la *i-ésima componente* o *coordenada*

Ejemplo 2.18. Un ejemplo muy conocido para ustedes es $\mathbb{R}^2$ formada por los duplas de números reales (x,y), $x \in \mathbb{R}$, $y \in \mathbb{R}$, y de al cual la podemos representar gráficamente en una hoja. Este es un espacio natural por ejemplo para graficar funciones de $\mathbb{R}$ en $\mathbb{R}$, o describir el conjunto de puntos (x,y) que verifican $x^2 + y^2 = 1$.

Ejemplo 2.19. Supongamos que queremos estudiar los precios de tres items en las góndolas de un supermercado, a saber, pan-leche-agua. Un espacio natural para considerar es $(\mathbb{R}^+)^3$, donde una terna (x,y,z) puede representar los precios, en pesos uruguayos, de 1kg de pan, 1 litro de leche y 1 litro de agua.

Observar que la definición anterior se puede generalizar a componentenes en distintos conjuntos. Por ejemplo podríamos considerar pares (x,y) donde x represente el nombre de alguien de esta clase, e y represente el mes de cumpleaños, donde por ejemplo $(Diego, Junio)$ podría ser dupla a considerar.

En este sentido, tenemos la definición siguiente.

Definición 2.20. Dados dos conjuntos X e Y, se define el *producto cartesiano de* $X \times Y$ como al conjunto

$$X \times Y = \{(x,y) : x \in X, y \in Y\}.$$

Esta definición se extiende de manera natural a producto de n-uplas.

◊ **Ejer 2.21.** Sean X e Y conjuntos no vacíos, y finitos. Entonces $|X \times Y| = |X| \times |Y|$. (Sugerencia: Como los conjuntos son finitos, podemos escribir $X = \{x_1, \dots, x_n\}$ e $Y = \{y_1, \dots, y_m\}$.)

◊ **Ejer 2.22.** Sea $n \in \mathbb{N}$. ¿Cuántos elementos tiene $\{0,1\}^n$? (Sugerencia: probar para $n = 2,3$, primero.)

Definición 2.23. De manera análoga, si X_i, para $i = 1, \dots, n$, son conjuntos, se define el producto cartesiano

$$\prod_{i=1}^{n} X_i = \{(x_1, \dots, x_n) : x_i \in X_i, i = 1, \dots, n\}.$$

(Aquí la notación $X_1 \times \cdots \times X_n = \prod_{i=1}^{n} X_i$.)

2.1.2. Funciones

Una *función* es una regla que le asigna a una entrada una única salida. Por ejemplo podemos considerar la función que le asigna a cada estudiante de esta clase, su mes de cumpleaños. (Dado que cada uno tiene un único nacimiento, la función queda definida).

Por ejemplo la regla que a un número real le asigna la raíz cuadrada, no es una función, dado que los números reales negativos no tienen definida una raíz cuadrada. Sin embargo la regla que a un número positivo le asigna la raíz cuadrada, está bien definida.

Un poco más formal, una *función* de un conjunto A en un conjunto B (que se escribe $f : A \to B$), es una regla que asigna a cada $a \in A$ un único elmento $b \in B$, que lo denotamos por $f(a)$.

Observar que una función $f : A \to B$ induce un subconjunto de $A \times B$, a saber,

$$\{(a, f(a)) : a \in A\} \subset A \times B,$$

El conjunto A se llama *dominio*, y B *codominio*.

Veamos algunos ejemplos.

Ejemplo 2.24.
1. La regla que asigna a un círculo su área es una función. Si a los círculos los identificamos por su radio $r > 0$, la función área puede representarse por el mapa $\mathrm{Ar} : \mathbb{R}^+ \to \mathbb{R}$, dada por $\mathrm{Ar}(r) = \pi r^2$.
2. El costo del envío de un paquete por encomienda es una función definida. En general las empresas calculan el costo como función del peso (y/o también del volumen; pero en cualquier caso la función está bien definida).
3. $f : \mathbb{R}^+ \to \mathbb{R}$ dada por $f(x) = 1/x$ es una función.
4. $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$.
5. Si tenemos un producto cartesiano $\prod_{i=1}^{n} X_i$, (ver Def. 2.23), la *proyección* en la j-ésima componente como a la función

$$\pi_j : \prod_{i=1}^{n} X_i \to X_j, \quad \text{tal que} \quad \pi_j(x_1, \dots, x_n) = x_j,$$

es una función bien definida en el producto cartesiano.
6. Dado un conjunto A, la función *identidad* $\mathrm{id}_A : A \to A$ está definida por $\mathrm{id}_A(a) = a$ para todo $a \in A$.

Recordemos algunas definiciones.

Definiciones 2.25.
- El *gráfico* de $f : A \to B$ se define por como el subconjunto $\mathrm{G}(f) \subset A \times B$, definido por

$$\mathrm{G}(f) = \{(a,b) \in A \times B : b = f(a)\}.$$

- La *imagen* de f se define como el subconjunto de B definido por

$$\mathrm{Im}(f) = \{b \in B : \text{existe } a \in A, \text{ con } f(a) = b\}.$$

- Si $B' \subset B$, definimos la *imagen inversa de B' por f* como el subconjunto de A definido por

$$f^{-1}(B') = \{a \in A : f(a) \in B'\}.$$

La función $f(x) = x^2$ tiene la particularidad de que la imagen de x y $-x$ es la misma. Por lo que todo $y > 0$ tiene dos preimágenes, que se denotan por $\sqrt{y}$ (que es la solución positiva de $x^2 = y$, y $-\sqrt{y}$. En particular esta función es lo que se denomina *dos a uno*, i.e., que por cada punto en la imagen hay dos preimagenes (salvo el cero que tiene como preimágen única el 0 pero con multiplicidad 2).

Una clase importante de funciones son las que son *uno a uno*.

Definición 2.26. Decimos que $f : A \to B$ es *inyectiva* si siempre que $a \neq a'$ en A, se tiene $f(a) \neq f(a')$. O equivalentemente, si para todo $b \in \mathrm{Im}(f)$, existe un único elemento $a \in A$ tal que $f(a) = b$;

Definición 2.27. Decimos que $f : A \to B$ es *sobreyectiva* si para todo $b \in B$, existe $a \in A$ tal que $f(a) = b$.

Definición 2.28. Decimos que $f : A \to B$ es *biyectiva* si es inyectiva y sobreyectiva.

Dadas $f : A \to B'$, y $g : B \to C$, donde $B' \subset B$, definimos la *composición* de f y g a la función $g \circ f : A \to C$ definida por $(g \circ f)(a) = g(f(a))$, $a \in A$.

Dada $f : A \to B$ biyectiva, entonces definimos la *inversa* de f como la función $f^{-1} : B \to A$, tal que $f^{-1} \circ f = \mathrm{id}_A$. (Observar que también vale $f \circ f^{-1} = \mathrm{Id}_B$).

2.1.3. Particiones

Muchas veces es de mucha utilidad poder "partir" un conjunto en subconjuntos disjuntos. Esto procedimiento nos ayudó a resolver los Ejercicios 2.11 y 2.13 para calcular el cardinal de conjunto finitos, simplemente, contabilizando los subconjuntos con k-elementos para luego sumarlos.

Se define una *partición* de un conjunto X como una colección de subconjuntos (no vacíos) disjuntos, que su unión es X. Es decir $\{X_\alpha\}_{\alpha\in I}$ es una partición de X , si satisfacen

- $X_\alpha \cap X_{\alpha'} = \varnothing$, siempre que $\alpha \neq \alpha'$;
- $\bigcup_{\alpha\in I} X_\alpha = X$.

Ejemplo 2.29. Una partición del conjunto $A = \{1,2,3,4,5\}$ es por ejemplo la colección $\{\{1\},\{2,5\},\{3,4\}\}$. También lo es $\{\{1,2,3,4\},\{5\}\}$

Como comentamos más arriba, las particiones nos permiten calcular cardinales de conjuntos. Supongamos que $\{X_1,\dots,X_n\}$ es una partición de un conjunto X finito. Entonces tenemos

$$|X| = |X_1| + \cdots + |X_n|.$$

Por ejemplo, si $X = \mathcal{P}(\{1,\dots,n\})$, entonces podemos tomar como X_i a todos los subconjuntos de i-elementos. (Observar que los "elementos" de nuestro X son subconjuntos de $\{1,\dots,n\}$.)

Luego, sabemos de otros cursos que el cardinal de X_i es exactamente las combinaciones de n tomados de i, y escribimos

$$\binom{n}{i} = \frac{n!}{(n-i)!i!}.$$

Luego podemos concluir que $|X| = \sum_{i=0}^n \binom{n}{i}$.

◇ **Ejer 2.30.** Concluir de la identidad del *binomio de Newton* $(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$, que las partes de un conjunto de n elementos tiene 2^n elementos.[5]

◇ **Ejer 2.31.** Sean $A = \{a_1,\dots,a_n\}$ y $B = \{b_1,\dots,b_m\}$ conjuntos finitos. Sea $\pi_1 : A\times B \to A$ la proyección en la primera coordenada.

1. Describir el conjunto $\pi_1^{-1}(a_1)$, y probar que la colección $\{\pi_1^{-1}(a_i) : i = 1,\dots,n\}$ es una partición de $A\times B$.
2. ¿Cuál es el cardinal de $\pi_1^{-1}(a_i)$, para $i = 1,\dots,n$?
3. Calcular $|A\times B|$. (Comparar con Ejercicio 2.21.)

2.1.4. Cuantificadores

Los dos cuantificadores más famosos son:

- (Existe) $\exists$ que indica que "*existe*"
- y en particular $\exists!$ que indica que "*existe un único*"
- (Para todo) $\forall$ que indica "*para todo*"

Veamos ejemplos.

Ejemplo 2.32. Consideremos la función $f(x) = x^2$, con $x \in \mathbb{R}$. Podemos decir que $\exists p \in \mathbb{R}$ tal que $f(p) = p$. Por ejemplo $p = 0$ es una posible solución, pero también $p = 1$. Sin embargo si $g(x) = x^2+1$, con $x \in \mathbb{R}$, entonces $\nexists p \in \mathbb{R}$ tal que $g(p) = p$.

◇ **Ejer 2.33.** Dar un argumento para la no existencia en el ejemplo 2.32.

Ya veremos que un punto p que satisface $f(p) = p$ es lo que se llama *punto fijo* y será un tema importante en este curso.

Ejemplo 2.34. Por ejemplo, decir que la ecuación $x^2 + 1 = 0$ no tiene soluciones reales, es equivalente a decir $\forall x \in \mathbb{R}$, $x^2 \neq -1$.

Con estas definiciones podemos reescribir la intersección y union arbitrarias de conjuntos de la siguiente manera: si $\{A_\alpha\}_{\alpha\in I}$ es una colección de conjuntos, entonces,

$$\bigcup_{\alpha\in I} A_\alpha = \{x : \exists \alpha \in I \text{ tal que } x \in A_\alpha\};$$
$$\bigcap_{\alpha\in I} A_\alpha = \{x : \forall\, \alpha \in I, x \in A_\alpha\}.$$

Un punto importante es entender las negaciones de $\exists$ y $\forall$. Veamos esto con un ejemplo.

Ejemplo 2.35. Estudiemos el siguiente enunciado.

"$\forall x \in \mathbb{R}$, $\exists n \in \mathbb{N}$ tal que $n > x$."

Lo que dice este enunciado es muy simple, no importa cuál es el número real x que consideremos, siempre existe un número natural que es mayor que él. Esta propiedad se conoce como *Propiedad Arquimedeana*.

¿Qué implicaría que este enunciado no sea cierto? Implicaría que $\exists z \in \mathbb{R}$ tal que $z > n$, $\forall n \in \mathbb{N}$. Esto implicaría que z sería una cota superior de los naturales, y esto sabemos que no es cierto.[6]

2.1.5. Implicaciones

Un enunciado de la forma A implica B, o A entonces B y se escribe $A \Rightarrow B$, indica que si A ocurre, entonces ocurre B. Y esa implicación, no implica nada más. Si B ocurre, nada dice sobre A. Veamos algunos ejemplos

[5] Hay otra forma de probar este hecho, y es probando que existe una función biyectiva de las parte de un conjunto de n elementos con $\{0,1\}^n$: a un cojunto $A \subset\subset \mathcal{P}(\{1,\dots,n\}$ se le asocia la n-upla de $0's$ y $1's$ que nos dicen si el i-'esimo elmento está o no en el conjuinto.

[6] La prueba formal de este hecho utiliza propiedades finas de los números reales, a saber, el Axioma de Completitud de los reales, el cual dice que todo subconjunto acotado superiormente en $\mathbb{R}$ tiene un supremo. Es un lindo desafío probar que $\mathbb{N}$ no puede ser acotado. La prueba puede entrar en el margen de esta hoja.

Ejemplo 2.36. "$x = 2$" $\Rightarrow$ "x es solución de la ecuación $x^2 = 4$".

Este ejemplo dice algo cierto, que es que si x toma el valor 2, entonces, 2 es solución de la ecuación $x^2 = 4$. Sin embargo es claro que no vale la implicación en el otro sentido (a lo que designamos como *recíproco*), i.e., es incorrecto decir que si "x es solución de la ecuación $x^2 = 4$" entonces "$x = 2$", dado que x puede ser también -2.

Sin embargo, un resultado importante se deduce cuando tenemos una implicación $A \Rightarrow B$, y es el hecho que si B no ocurre, entonces no puede ocurrir A (i.e. **no** $B \Rightarrow$ **no** A). A esto se le denomina *contrarecíproco*. En nuestro ejemplo sería la afirmación

$$\text{“}x \textbf{ no } \text{es solución de } x^2 = 4'' \Rightarrow x \neq 2.$$

Lo interesante es que ambos enunciados son equivalentes, es decir, que $A \Rightarrow B$ es verdad, sí y solo sí, **no** $B \Rightarrow$ **no** A es verdad.[7]

Esto sirve como método de prueba de muchas cosas, como ya veremos en la siguiente sección.

2.2. Demostraciones

Una *demostración* es una sucesión de consecuencias lógicas que, a partir de ciertas hipótesis, culminan en un resultado preciso.

En el proceso de resolver un problema se pueden identificar *tres etapas*, aunque a veces esta separación no es tan clara. En una primera etapa uno tiene que *entender el problema*. Es una parte muy importante del proceso de resolución de un problema. Muchas veces podemos tener a mano "la buena pregunta", y eso nos permite tener claro cuál es el problema a resolver. En esta etapa es de mucha utilidad buscar los elementos importantes, y despojarnos de los elementos que nos alejan del verdadero problema. De alguna manera limpiar y dejar en claro cuál es el problema.

La siguiente etapa es *tener una idea de la demostración*. Esta es la etapa más creativa del proceso. Para esta etapa sirve mucho de intentar resovler problemas particulares, o más sencillos antes de abordar la demostración general.

La tercer etapa, es *escribir la prueba de manera clara y correcta*.

En las siguientes secciones daremos algunos ejemplos y técnicas utilizadas en demostraciones.

2.2.1. Pruebas directas

Quizás esta es la técnica más clara de demostración donde a partir de ciertas hipótesis, y deducciones e implicaciones lógicas llegamos al resultado.

Veamos un ejemplo.

Proposición 2.37. *Probar que* la suma de dos números impares, es un número par.

Comencemos el proceso de resolver este problema.

Para entender el problema necesitamos la definición de lo que es un número *par* e *impar*. Recordar que n es un número par (por ejemplo $2, 8, -10, \dots$) si se puede escibir como el doble de un número entero, i.e., $n = 2k$ para algún $k \in \mathbb{Z}$. Análogamente, m es impar (tipo $1, 3, -7$) si es de la forma $m = 2j + 1$, para algún $j \in \mathbb{Z}$.

Con las definiciones anteriores, y leyendo nuevamente el problema, queda claro qué es lo que tenemos que probar.

Ahora la parte más creativa de tener una idea de la demostración. Para eso podemos particularizar el problema tomando un ejemplo.

Si tomamos dos números impares, por ejemplo, 3 y 9, vemos que su suma es 12, la cual es par. Ok, pero ¿dónde está la importancia de que 3 y 9 sean impares? Por ejemplo si cambiamos 9 por el número par 8 vemos que el resultado es 11 y no es más par. el resultado es 11 que no es par. (De hecho, haciendo lo anterior para otros ejemplos, podemos conjeturar el resultado de que si sumamos un par más un impar entonces el resultado es impar.)

Volviendo a nuestro ejemplo, quizás valga la pena escribir los impares en su forma 2·"entero" +1. En nuestro caso tenemos $3 + 9 = (2 \cdot 1 + 1) + (2 \cdot 4 + 1) = 2 \cdot (1 + 4) + 1 + 1$. Con esto en mente vemos $3 + 9$ se puede obtener como el doble de un número más 2, y por lo tanto podemos escribir $3 + 9 = 2 \cdot (1 + 4 + 1)$. El papel que juegan el primer 1 y el 4 es irrelevante, y el último 1 viene de juntar la suma $1 + 1$. Acá tenemos ya clara la prueba. Si sumo dos impares $2k + 1$ con $2\ell + 1$, el resultado es $2(k + \ell + 1)$ que es par. Ya estamos en condiciones de escribir prolija la solución.

Demostración. Sean m y n dos números impares. Entonces existe $k \in \mathbb{Z}$ y $\ell \in \mathbb{Z}$ tales que $m = 2k + 1$ y $n = 2\ell + 1$.

Realizando operaciones algebraicas sencillas, obtenemos

$$\begin{aligned} m + n &= (2k + 1) + (2\ell + 1) \\ &= 2 \cdot (k + \ell) + 1 + 1 = 2 \cdot (k + \ell + 1). \end{aligned}$$

Definiendo $(k + \ell + 1)$ como j, tenemos que $m + n = 2 \cdot j$, con $j \in \mathbb{Z}$. Luego concluimos que $m + n$ es par. $\square$

Observar que el enunciado del Proposición 2.37 puede ser reparafraseado como una implicación:

Si "m y n son números impares" $\Rightarrow$ "$m + n$ es par".

Veamos otro ejemplo.

[7] La prueba de que son equivalentes resulta de que el directo implica el contrarecíproco, y luego si intercambiamos los roles de A con **no** A, y de B con **no** B, obtenemos la equivalencia.

Proposición 2.38. *Si m y n son números impares, entonces $m \cdot n$ es impar.*

◊ **Ejer 2.39.** Probar que la Proposición 2.38.

Un *corolario* es en general una consecuencia directa de un resultado más fuerte. Un corolario del la Proposición 2.38 es aplicarlo para el mismo número impar.

Corolario 2.40. *Si m es un número impar, entonces m^2 es un número impar.*

De hecho, podemos convencernos que el resultado es un "*si y sólo si*", dado que si m^2 es un número impar entonces m^2 no puede ser divisible por 2, y en particular m tampoco.

2.2.2. Pruebas indirectas

Supongamos que queremos probar la implicación $P \Rightarrow Q$. Pruebas indirectas se refieren a que no comenzamos con que la hipótesis P es cierta, como el caso anterior (P en la Proposición 2.38 era "m y n son impares").

Comúnmente hay dos tipos de pruebas indirectas:

- el *contrarecíproco*: i.e. probar que "**no** $Q \Rightarrow$ **no** P"
- el *absurdo*: que consiste en suponer el resultado es falso y llegar a una contradicción, i.e. que no pueden convivir **no** Q y P. Esto muestra que el resultado no puede ser falso.

Lo que tienen en común ambas técnicas es que se comienza asumiendo que la tesis no se cumple, i.e., **no** Q.

Ejemplo 2.41. Sea $n \in \mathbb{Z}$. Si n^2 es par, entonces n es par.

(Prueba por Contrarecíproco). Hay que probar que si n es **no** par, entonces n^2 es **no** par, es decir, si n es impar, entonces n^2 es impar. Pero eso ya lo sabemos, y es lo que probamos en la sección anterior, por lo que no hay nada que probar. □

(Prueba por Absurdo). Una prueba por absurdo sería de la siguiente manera. Sea n^2 par, y supongamos por absurdo que n es impar. Entonces existe $k \in \mathbb{Z}$ tal que $n = 2k + 1$. Luego elevando al cuadrado tenemos

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1,$$

pero definiendo $j = 2k^2 + 2k$ tenemos que podemos escribir

$$n^2 = 2j + 1,$$

con $j \in \mathbb{Z}$, y por lo tanto n^2 es impar lo cual es un absurdo dado que teníamos n^2 par. □

Veamos un ejemplo un poco más interesante de prueba por absurdo.

Proposición 2.42. *La solución positiva de la ecuación $x^2 = 2$ es irracional.*

Este enunciado es equivalente a decir que $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$, pero de la manera enunciada queda más en evidencia que es un enunciado tipo $P \Rightarrow Q$. [8]

Nuestra definición de ser irracional es que es un número real pero que no es racional, i.e. que no se puede escribir de la manera m/n con $m, n \in \mathbb{Z}$. Este tipo de resultados es casi una tentación usar estos métodos indirectos. Lo mismo ocurre con resultados donde se pide probar que algo es infinito (no finito), etc., aunque no hay una regla general.

Hagamos una prueba por el absurdo de este hecho.

Demostración. Supongamos que $\sqrt{2}$ es racional. Esto es, existen $m, n \in \mathbb{Z}$, (con $n \neq 0$) tales que

$$\sqrt{2} = \frac{m}{n}.$$

De hecho, existen infinitos pares que sirven, pero podemos asumir que m y n no tengan divisores comunes, porque podemos cancelarlos (por ejemplo $\frac{3 \cdot 4}{5 \cdot 4} = \frac{3}{5}$).

Luego tenemos que $n = \sqrt{2}m$, y elevando al cuadrado tenemos:

$$n^2 = 2m^2.$$

Pero entonces n^2 es un número par, y por el Corolario 2.40 concluimos que n tiene que ser par, i.e. $n = 2 \cdot k$, con $k \in \mathbb{Z}$. Luego, sustituyendo n por $2k$ y simplficando, tenemos

$$(2 \cdot k)^2 = 2m^2 \quad \Rightarrow \quad 2k^2 = m^2.$$

Luego m^2 también es par, y nuevamente por el Corolario 2.40 tenemos que m tiene que ser par. Pero esto es un absurdo dado que habíamos elegido m y n sin divisores comunes. □

2.2.3. Existencia y unicidad

La *existencia*, y *unicidad* de objetos es un tema transversal en la matemática. Veremos varios ejemplos a lo largo de este curso, como por ejemplo la existencia de puntos fijos, o la existencia de raíces de un polinomio, y en el caso de existir podemos preguntarnos si son únicos o no.

De hecho cuando hablamos de funciones inversas típicamente estamos apelando a la existencia y unicidad. Por ejemplo ¿qué número es $\sqrt{2}$? Sabemos que no lo podemos representar como un número entero, ni de la forma m/n, con $m, n \in \mathbb{Z}$. Cuando hablamos de $\sqrt{2}$ es una etiqueta de un número que sabemos que existe y además que es único. Para evitar continuar con el suspenso, $\sqrt{2}$ es una etiqueta para representar la única solución positiva de la ecuación $x^2 = 2$.

[8] Este resultado se conocía desde la época de Pitágoras, donde la etiqueta $\sqrt{2}$ se sustituía por la etiqueta *longitud de la diagonal del cuadrado unidad*, y se sabía que esa longitud no era *inconmensurable* respecto a los lados del cuadrado.

Para la *existencia* se distinguen dos métodos, uno directo en el cual se muestra explícitamente el objeto que conjeturamos que existe (este sería el método más razonable y directo). Por ejemplo sabemos que existe una solución real de la ecuación $x^2 - 1 = 0$, porque $1^1 - 1 = 0$. Pero en la mayoría de las ocasiones interesantes e importantes no tenemos herramientas para darlos explícitamente, (como el caso de $\sqrt{2}$). A lo largo de la historia, y con un gran despliegue hoy día con el uso de las computadoras, se han construido métodos para aproximar cada vez mejor esos objetos que sabemos su existencia.

Para probar la *unicidad* de cierto objeto con cierta propiedad se utilizan generalmente métodos indirectos: por ejemplo de existir otro objeto que satisface la propiedad se llegaría a un absurdo.

Ejemplo 2.43. La ecuación $x^3 - x = 0$ tiene algún raíz entera. En este caso no hay que recurrir a nada muy sofisticado, el número 0, satisface $0^3 - 0 = 0$ y por lo tanto es una solución del problema. Hemos probado la existencia de una solución entera de manera explícita.

Ejemplo 2.44. Existe una solución real de la ecuación $\pi x^3 - ex^2 + \sqrt{2}x + 1 = 0$. La existencia resulta del Teorema de Bolzano observando que si b toma un valor muy muy grande, la función $f(x) = \pi x^3 - ex^2 + \sqrt{2}x + 1$ evaluada en b es positiva, y si a toma un valor muy muy negativo $f(a)$ es negativo, luego por la continuidad de f, sabemos que existe $x^* \in [a,b] \subset \mathbb{R}$, tal que $f(x^*) = 0$.

2.3. Inducción Matemática

Observar lo siguiente:

$$\begin{aligned} 1 &= \frac{1\cdot 2}{2} \\ 1+2 &= \frac{2\cdot 3}{2} \\ 1+2+3 &= \frac{3\cdot 4}{2} \\ 1+2+3+4 &= \frac{4\cdot 5}{2} \end{aligned}$$

Podemos verificar a mano que las 4 igualdades son correctas, y podríamos aventurarnos a que la ecuación

$$(S_n) \qquad 1+2+\cdots+n = \frac{n\cdot(n+1)}{2} \tag{2.1}$$

también es correcta para todo $n \in \mathbb{N}$. ¿Pero cómo podemos probarlo?[9]

El *principio de inducción matemática* nos dice cómo hacerlo, y básicamente funciona como el "efecto dominó". Supongamos que tenemos una linea infinita de fichas de dominó paradas como en la foto.

Sabemos que si están suficientemente juntas, resulta que si una ficha se cae, entonces empuja a la siguiente y esta también caerá. Por lo tanto, si empujamos la primera, sabremos que toda las fichas caerán.

Definición 2.45 (Principio de Inducción)**.** Supongamos que S_n es un enunciado que depende de un número natural n (como el dado en (2.1)).

1. Si el enunciado es cierto para $n = 1$, (i.e., S_1 es válida);

2. Si la implicación $S_k \Rightarrow S_{k+1}$, para todo $k \in \mathbb{N}$,

entonces, el enunciado S_n es cierto para todo $n \in \mathbb{N}$.

El paso 1. se denomina *caso base* o *hipótesis de inducción*. El paso 2. se denomina *paso inductivo*.

Probemos por el principio de inducción que (2.1) es verdad para todo $n \in \mathbb{N}$.

Para $n = 1$ es válido dado que $1 = 1 \cdot 2/2$.

Dado $k \in \mathbb{N}$, asumimos que el válido que

$$1+2+\cdots+k = \frac{k\cdot(k+1)}{2}.$$

Luego tenemos, de la igualdad anterior que

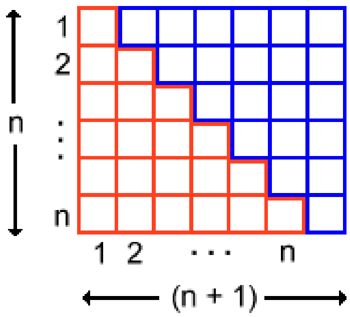
$$\begin{aligned} 1+2+\cdots+k+(k+1) &= \frac{k\cdot(k+1)}{2} + (k+1) \\ &= \left(\frac{k}{2}+1\right)\cdot(k+1) \\ &= \frac{(k+1)\cdot(k+2)}{2}, \end{aligned}$$

por lo que S_k implica S_{k+1} terminando la prueba.

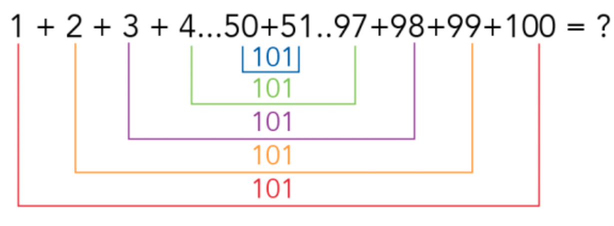
2.3.1. Una imágen puede más que mil palabras

En varias ocasiones podemos encontrar argumentos más directos que nos evidencie que el resultado es correcto. Acá va una imagen que evidencia que (2.1) es correcta para todo $n \in \mathbb{N}$.

[9] La igualdad anterior se hizo muy conocida por una historia asociada a Gauss en su edad escolar.



La leyenda dice que a Gauss le pidieron que sumara los primeros 10 números naturales, y lo hizo tan rápido que la maestra lo desafió a que sumara los primeros 100 números, a lo cual la respuesta de Gauss fue tan rápida como la primera. Se cree que la prueba que utilizó es la que se muestra en la imagen.



Pattern showing addition of pairs from one to ten (©2021 Let's Talk Science).

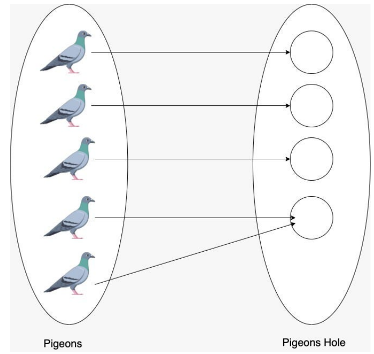
Dependiendo de la paridad o no de n la agrupación anterior me deje un elemento sin agrupar. Una prueba alternativa es la siguiente. Si $T(n)$ es la suma $1+2+\cdots+n$, entonces

$$\begin{array}{r} T(n) = 1+2+\cdots+(n-1)+n \\ T(n) = n+(n-1)+\cdots+2+1 \\ \hline 2T(n) = (n+1)\cdot n \end{array}$$

2.4. Principio el Palomar

El *Principio del Palomar* es un principio muy básico pero que tiene consecuencias inesperadas.

Básicamente el principio dice que si tenemos 6 palomas que vuelan para entrar en 5 casitas, entonces hay alguna casita que tiene al menos 2 palomas.



Principio del Palomar: *"Si más de m objetos son colocados en m categorías, entonces alguna categoría recibe más de un objeto."*

Este principio se utiliza de muchas maneras, pero una que es frecuentemente utlizada es como un *análisis del peor caso*[10]. Veamos un ejemplo.

Ejemplo 2.46. Supongamos que tenemos 5 casilleros, y queremos colocar manzanas en cada uno. ¿Cuántas manzanas necesito colocar para que en algún casillero pueda asegurar que al menos tiene dos manzanas?

La respuesta es 6, claramente.

Veamos otro ejemplo extraido de [?RR].

Ejemplo 2.47. Supongamos que que tenemos en nuestro cajón de medias 7 medias de color negro y 12 de color blanco. La luz está apagada y quiero sacar medias de manera que

1. dos medias sean del mismo color
2. dos sean negras
3. dos sean blancas.

¿Cuántas necesito sacar en cada caso para asegurarme lo solicitado?

Veamos la respuesta de 1, pero les sugiero pensar las respuestas antes de seguir.

Demostración. Necesitamos sacar por los menos dos medias, de lo contrario no tendremos un par de medias. Ahora, pensando en el principio del palomar, tenemos dos categorías, color blanco y color negro. Por lo que para asegurarme de tener dos del mismo color, basta con sacar tres medias. □

◊ **Ejer 2.48.** Responder a las preguntas 2. y 3. anteriores.

Este princpio tiene aplicaciones divertidas a problemas geométricos. Para terminar pongamos de ejercicios dos resultados geométricos interesantes.

[10] Este tipo de análisis es muy utilizado en distintas áreas de la matemática.

♢ **Ejer 2.49.** Dado 5 puntos en un triángulo equilátero de lado 1, hay al menos dos puntos que su distancia es menor o igual a $1/2$.

♢ **Ejer 2.50.** Dado 5 puntos en la esfera, hay 4 que están en un mismo hemisferio cerrado (i.e. con su ecuador incluido). (Sugerencia: tomar con algo de astucia un buen ecuador.)

Para terminar pongamos un ejemplo divertido de la utilización de este principio para poner a prueba nuestra intuición.[11]

Ejemplo 2.51. ¿Cuál es la probabilidad de que dos personas en Brasil tengan exactamente la misma cantidad de pelo? O más preciso, la misma cantidad de folículos pilosos activos?

Se estima que una persona tiene en el entorno de mil pelos por centímetro cuadrado. En Brasil hay en el entorno de 200 millones de personas. Por lo que si suponemos que todas las personas tienen una cantidad de pelos diferente entonces utilizando el principio tenemos que por los menos hay una persona con más de 200 millones de pelos. Esto significa que tendría una cabeza en el entorno de 200 mil cm^2, lo que equivale a 20 metros cuadrados. Una estimación grosera, la cabeza tiene un radio de 1 metro 20 centímetros. En conclusión, es seguro que hay personas con la misma cantidad de pelo.

2.5. Numerabilidad

El cardinal de un conjunto, concepto que vimos más arriba, está definido cuando el conjunto es finito, y este número nos dice cuántos elementos tiene el conjunto.

Una forma un poco más avanzada de pensar el cardinal es el siguiente. Decimos que dos conjuntos tienen el *mismo cardinal* si existe una función biyectiva entre ambos. Esta forma de definir el cardinal (que formalmente es una clase de equivalencia) vale también para conjuntos infinitos como veremos en seguida.

En este sentido el conjunto $X = \{a,b,c\}$, donde a,b,c son tres elementos diferentes, tiene el mismo cardinal que el conjunto $\{1,2,3\}$, dado que podemos definir una función biyectiva entre ambos conjuntos. Con esta definición el cardinal de cualquiera de estos conjuntos es 3 y es la forma de representar la clase. Por ejemplo, un conjunto de cardinal n, es un conjunto que está en correspondencia biyectiva con $\{1,2,\dots,n\}$.

Esta definición nos permite dar un salto al infinito. Podríamos comparar conjuntos infinitos y ver si tienen el mismo cardinal. Por ejemplo, ¿tendrá $\mathbb{N}$ el mismo cardinal que $\mathbb{R}$? ¿Tendrá $\mathbb{N}$ el mismo natural que $\mathbb{N}\cup\{0\}$? La *teoría de conjuntos* se ocupó de todas estas cuestiones a finales del siglo 19, siendo Cantor uno de los principales exponentes.

Hay muchas "paradojas", o argumento no intuitivos, asociadas a este problema. Uno de estos es el *hotel de Hilbert*.

[11] Creo recordar que el ejemplo sale el libro de geometría métrica de P. Adams, en su introducción.

En un hotel común, o al menos los que me he alojado, la cantidad de habitaciones es finita. Por lo que si están todas ocupadas, no queda otra que irme a otro lado.

Pero he aquí el famoso hotel de Hilbert. En hotel podemos numerar las habitaciones, $1,2,3,\dots$, donde para cada natural, tendremos un habitación (con su número en la puerta). Claramente hay infinitas habitaciones.

Supongamos que todas las habitaciones están ocupadas. ¿Podrá el hotel alojar a un nuevo huésped, sin tener que pedirle a alguien que abandonde el hotel? La respuesta es sí, basta pedirle a las personas de cada habitación que se mueva a la habitación contigua con el siguiente número de puerta. Es decir, quien esté alojado en la habitación n se le pide que se mueva a la $n+1$. Luego la primer habitación queda libre.

Repitiendo este argumento es posible alojar a una cantidad arbitraria (pero finita) de húespedes.

Ahora la pregunta, supongamos que viene una canidad infinita de húespedes. Podremos alojarlos a todos? Nuevamente la respuesta es sí. Basta pedirle a cada húesped que se mueva a la habitación con número el doble de la que tiene, es decir, que si está en la habitación n, se moverá a la $2n$. De esta manera todas las habitaciones impares estaran libres, y podremos alojar a nuestros infinitos huéspedes que han llegado, (al húesped n lo alojamos en habitación vacía $2n-1$).

Definición 2.52. Decimos que un conjunto es *numerable* si es finito, o tiene el mismo cardinal que $\mathbb{N}$.

Cuando el conjunto es numerable y no finito, se le puede decir *infinito numerable*.

Con estas definiciones podemos plantear la siguiente pregunta:

¿Es el conjunto de números reales, $\mathbb{R}$, numerable?

En otras palabras, ¿existe una función biyectiva entre $\mathbb{R}$ y $\mathbb{N}$? En las siguientes líneas mostraremos que **no**, utilizando la prueba debida a Cantor, usualmente denominada *proceso diagonal de Cantor*.

Primero una pequeña simplificación. Es muy fácil construir una función biyectiva entre $\mathbb{R}$ y $(0,1)$, por lo que basta ver si podemos contruir una función biyectiva entre $(0,1)$ y $\mathbb{N}$.

Por absurdo, supongamos que tenemos una función biyectiva de $(0,1)$ a $\mathbb{N}$, i.e., podemos contarlos.

Sabemos que podemos representar a todos los números de la forma entre 0 y 1 con una expansión de la forma

$$0.a_1a_2,\dots a_n\dots, \qquad a_i\in\{0,1,\dots 9\}.$$

Formalmente, estamos hablando de la expansión *decimal* de un número el número x asociado a la expansión anterior es simplemente el que satisface

$$x = \sum_{n=1}^{\infty} \frac{a_n}{10^n}.$$

Esta expansión tiene una pequeña desventaja, la representación de un número no es única. Por ejemplo el número $1/10 \in (0,1)$ lo puedo representar de dos formas distintas

$$0{,}100000000\ldots = 0{,}09999999\ldots.$$

◇ **Ejer 2.53.** Probar que ambas expansiones coinciden.

Luego, podemos decir que en un caso como este nos quedamos con la expansión finita, es decir la que termina con infinitos ceros. Se puede probar que tomando esta elección hay una correcta biyección entre x y su expansión decimal.

Volvamos a nuestra prueba.

Si tenemos una biyección entre los elemento entre $(0,1)$ y su expansión entonces podemos "numerar y ordenar" todos los números entre $(0,1)$:

$$\begin{array}{l} 0{,}a_{11}\,a_{12}\,a_{13}\,a_{14}\ldots \\ 0{,}a_{21}\,a_{22}\,a_{23}\,a_{24}\ldots \\ 0{,}a_{31}\,a_{32}\,a_{33}\,a_{34}\ldots \\ 0{,}a_{41}\,a_{42}\,a_{43}\,a_{44}\ldots \\ \vdots \end{array}$$

donde $a_{ij} \in \{0,1,\ldots,9\}$. (Aquí la primera expansión se refiere al asociado por la biyección al número 1, la segunda al número 2, etc.).

Luego hacemos lo siguiente. Construiremos una nueva expansión

$$0{,}b_1\,b_2\,b_3\,b_4\ldots$$

donde b_n son 0 o 1 de forma tal que $b_n \neq a_{nn}$ (por ejemplo, si $a_{nn} \in \{1,\ldots,9\}$ definimos $b_n = 0$, y si $a_{nn} = 0$, entonces definimos $b_n = 1$.

Con esta construcción logramos construir una nueva expansión décimal, o sea, un número entre $(0,1)$ que es distinta a las anteriores. Pero habíamos asumido que estaban todos los números entre $(0,1)$, lo que es un absurdo.

Este prueba, que se repite en varios lugares del análisis matemático se debe a Cantor, y se conoce como el *proceso diagonal de Cantor*.

2.6. Recopilación de ejercicios

Ejercicio 2.54. Dar un argumento para probar que el conjunto definido en el Ejemplo 2.6 no es vacío. (Sug: recordar el Teorema de Bolzano.)

Ejercicio 2.55. ¿Para cuales conjuntos A más abajo se satisface $|A| \in A$?

1. $A = \{1,2,3,4\}$
2. $A = \{5,6,7,8\}$
3. $A = \{1,3,5\}$
4. $A = \{1,3,5,7\}$

Ejercicio 2.56. Sea A un conjunto finito. Encontrar $|A|$ si $A = \{2,3,|A|\}$.

Ejercicio 2.57. Calcular $\#\mathcal{P}(\{1,2,3\})$, y $\#\mathcal{P}(\{1,2,3,4\})$. (Sugerencia: calcular de manera ordenada los conjuntos con 0-elementos, 1-elemento, 2-elementos, 3-elementos, 4-elementos.) Se animan a calcular el cardinal de las las partes de $\{1,\ldots,n\}$.

Ejercicio 2.58. Sean A y B conjuntos finitos. Probar que $|A\cup B| = |A| + |B| - |A\cap B|$.

Ejercicio 2.59. Sean A, B, y C conjuntos finitos. Hallar $|A\cup B\cup C|$. (Sugerencia: usar los diagrama de Venn.)

Ejercicio 2.60. Considere la colección de intervalos $A_n = \left(0, 1+\frac{1}{n}\right)$, $n \in \mathbb{N}$. ¿Hallar los conjuntos $\bigcap_{n\in\mathbb{N}} A_n$ y $\bigcup_{n\in\mathbb{N}} A_n$?

Ejercicio 2.61. Hallar $\bigcup_{n\in\mathbb{Z}}[n,n+1)$, y el complemento del conjunto $\bigcup_{n\in\mathbb{Z}}(n,n+1)$.

Ejercicio 2.62. Sean X e Y conjuntos no vacíos, y finitos. Entonces $|X\times Y| = |X|\times|Y|$. (Sugerencia: Como los conjuntos son finitos, podemos escribir $X = \{x_1,\ldots,x_n\}$ e $Y = \{y_1,\ldots,y_m\}$.)

Ejercicio 2.63. Sea $n \in \mathbb{N}$. ¿Cuántos elementos tiene $\{0,1\}^n$? (Sugerencia: probar para $n = 2,3$, primero.)

Ejercicio 2.64. Concluir de la identidad del *binomio de Newton* $(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^ib^{n-i}$, que las partes de un conjunto de n elementos tiene 2^n elementos.[12]

Ejercicio 2.65. Sean $A = \{a_1,\ldots,a_n\}$ y $B = \{b_1,\ldots,b_m\}$ conjuntos finitos. Sea $\pi_1 : A\times B \to A$ la proyección en la primera coordenada.

1. Describir el conjunto $\pi_1^{-1}(a_1)$, y probar que la colección $\{\pi_1^{-1}(a_i) : i = 1,\ldots,n\}$ es una partición de $A\times B$.
2. ¿Cuál es el cardinal de $\pi_1^{-1}(a_i)$, para $i = 1,\ldots,n$?
3. Calcular $|A\times B|$. (Comparar con Ejercicio 2.62.)

Ejercicio 2.66. Dar un argumento para la no existencia en el ejemplo 2.32.

Ejercicio 2.67. Probar que la Proposición 2.38.

Ejercicio 2.68. . Responder a las preguntas 2 y 3 del **??**.

Ejercicio 2.69. Dado 5 puntos en un triángulo equilátero de lado 1, hay al menos dos puntos que su distancia es menor o igual a $1/2$.

Ejercicio 2.70. Dado 5 puntos en la esfera, hay 4 que están en un mismo hemisferio cerrado (i.e. con su ecuador incluido). (Sugerencia: tomar con algo de astucia un buen ecuador.)

[12] Hay otra forma de probar este hecho, y es probando que existe una función biyectiva de las parte de un conjunto de n elementos con $\{0,1\}^n$: a un cojunto $A \subset\subset \mathcal{P}(\{1,\ldots,n\}$ se le asocia la n-upla de $0's$ y $1's$ que nos dicen si el i-'esimo elmento está o no en el conjuinto.

Capítulo 3

Preliminares de Análisis

(*Este capítulo está basado en unas notas elaboradas por Federico Molina en 2023.*)

Estas son unas notas de repaso de análisis real respecto a conceptos que se utilizan en el curso de Probabilidad 2.

3.1. Resumen análisis real

3.1.1. Sucesiones y convergencia

Definición 3.1 (Sucesión o Secuencia)**.** Una *sucesión* de números reales es un mapa de naturales a números reales, es decir, $a : \mathbb{N} \to \mathbb{R}$. La notación puede ser de la forma: $(a_1, a_2, \dots)$, $(a_n)_{n\in\mathbb{N}}$, $(a_n)_{n=1}^{\infty}$,,$\{a_n\}_{n=1}^{\infty}$ etc.

Por ejemplo: $(a_n)_{n\in\mathbb{N}} = (\frac{1}{n})_{n\in\mathbb{N}}$.

Definición 3.2 (Convergencia)**.** Decimos que la sucesión $(a_n)_{n\in\mathbb{N}}$ es *convergente* a un $b \in \mathbb{R}$ si para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$, tal que se cumple $|a_n - b| < \varepsilon$, siempre que $n \geqslant n_0$. En tal caso escribimos $\lim_n x_n = b$, o $x \overset{n}{\to} b$.

Si no existe un b que cumple lo anterior la sucesión *no converge*. (Se reserva el término *diverge* si la función "converge" a $+\infty$, o $-\infty$. Por ejemplo $\lim_n x_n = +\infty$ si dado $K > 0$, existe N tal que $x_n > K$, $\forall n \geqslant N$.)

Ejemplo 3.3. Es fácil probar que la sucesión $\{1/n\}$ converge a 0. Dado $\varepsilon > 0$, basta tomar $n_0 > 1/\varepsilon$.

♢ **Ejer 3.4.** Encontrar $n_0(\varepsilon)$ para que la sucesión $\{1/n^2\}$ satisfaga $1/n^2 < \varepsilon$ si $n \geqslant n_0$.

Es evidente que si una sucesión es convergente, el límite es único. Eso ocurre dado que no puede aproximarse a dos puntos diferentes al mismo tiempo.

♢ **Ejer 3.5.** Formalizar lo anterior. Suponer por absurdo que la sucesión converge a b_1 y b_2 distintos.

Definición 3.6 (Sucesión acotada)**.** Decimos que la sucesión a_n es *acotada* si $\exists K \in \mathbb{R}$ tal que $\forall n \in \mathbb{N} : |a_n| \leq K$. En otro caso decimos que a secuencia *no esta acotada*.

Observación 3.7. Es fácil ver si una sucesión es convergente, entonces está acotada. Esto sucede ya que dado $\varepsilon = 1$, existe un n_0 donde a partir de ese n_0 todos los términos de la sucesión están a distancia 1 de su límite. Y como quedan una una cantidad finita de términos, a saber, los primero n_0 términos, el resultado sigue.

Los límites cumplen siguientes propiedades:

- $\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$
- $\lim_{n\to\infty}(a_n b_n) = \lim_{n\to\infty} a_n \lim_{n\to\infty} b_n$
- $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$, siempre que $\lim_{n\to\infty} b_n \neq 0$

3.1.2. Extremos de conjuntos

Definición 3.8 (Cota superior, máximo)**.** Dado un conjunto $C \subset \mathbb{R}$ decimos que $K \in \mathbb{R}$ es una *cota superior* de C si $x \leqslant K$, para todo $x \in C$. En el caso que la cota K sea un elemento de C, decimos que K es el *máximo* del conjunto C.

Cuando existe una cota superior para un conjunto, decimos que el conjunto está *acotado superiormente*.

Sea C un conjunto acotado superiormente. Un concepto fundamental en análisis es la existencia de la "menor de las cotas superiores".

Definición 3.9 (Supremo)**.** Sea $C \subset \mathbb{R}$ un conjunto acotado superiormente. El número real $s \in \mathbb{R}$ se dice que es el *supremo* de C (lo denotamos $\sup C$) si:

- s es cota superior;
- $\forall\, \varepsilon\ > 0$, existe $c \in C$ tal que: $s - \varepsilon < c$, (i.e., $s - \varepsilon$ no es una cota superior de C).

(Si C no esta acotado superiormente escribimos $\sup C := \infty$.)

Análogamente podemos definir conjunto *acotado inferiormente*, *mínimo* e *ínfimo*.

♢ **Ejer 3.10.** Definir los conceptos anteriores.

Ejemplo 3.11. El intervalo $I = (0,1]$ es acotado, y su ínfimo es 0, y su supremo 1. Además como $1 \in I$, tenemos que es máximo.

3.1.3. Axioma de Completitud

Una pregunta fundamental es saber si un conjunto acotado superiormente siempre tiene un supremo. Es interesante que con las propiedades algebraicas que cumplen los números, más que es un conjunto ordenado (donde puedo decir que real es más grande que cuál), no alcanza para asegurar la existencia de un supremo. Y es por esta razón que se necesita un axioma, denominado, *axioma de completitud*, que dice que todo conjunto acotado superiormente tiene un supremo.

Una definición importante es la siguiente.

Definición 3.12 (Sucesión de Cauchy)**.** Sea $\{a_n\}$ una sucesión. Decimos que es de *Cauchy* si $\forall \varepsilon > 0\ \exists\ n_0 \in \mathbb{N}$ tal que $|a_n - a_m| < \varepsilon$ para cualesquiera $n, m \geqslant n_0$.

Una particularidad fundamental de estas sucesiones es lo siguiente:

Teorema 3.13 *Toda sucesión de Cauchy es convergente.*

Esto teorema es muy importante, porque podemos saber si una sucesión es convergente aún sin saber quién es su límite. Esta propiedad es de suma utilidad. Por ejemplo, consideremos la sucesión definida por recurrencia como:

$$x_0 = 1$$
$$x_{n+1} = \sin(x_n)$$

Esta sucesión se puede probar que es de Cauchy y por tanto convergente.

La noción de sucesión de Cauchy se extiende de manera análoga a $\mathbb{R}^n$ utilizando ...

Ejemplo 3.14 (Suma de Neumann)**.** Consideremos una matriz cuadrada $A = \begin{pmatrix} 0 & 1/2 & 1/3 \\ 1/15 & 0 & 1/\sqrt{2} \\ 0 & 1/10 & 0 \end{pmatrix}$ Supongamos que queremos estudiar la suma

$$S_n = \mathrm{Id} + A + A^2 + \cdots + A^n, \quad , n \in \mathbb{N}.$$

¿Podemos saber si esta suma convege a alguna matriz 3×3? Es un ejercicio probar que esta suma es de Cauchy, utilizando la norma de operador en el espacio de matrices.

Subsucesiones

Dada una sucesión, una *subsucesión* es una nueva sucesión que se construye a partir de la primera sacando términos de la sucesión (y reindexando). Por ejemplo, dada la sucesión $\{1/n\}_{n\in\mathbb{N}} = (1, 1/2, 1/3, \ldots)$, tenemos la subsucesión formada sólo por los denominadores impares: $\{1/(2n-1)\}_{n\in\mathbb{N}}$.

Formalmente, una subsucesión se define de la siguiente manera.

Definición 3.15 (Subsucesión)**.** Sea $\{a_n\}_{k\in\mathbb{N}}$ una sucesión. Consideremos una sucesión de naturales $b_k : \mathbb{N} \to \mathbb{N}$ monótona creciente estricta ($n_{k+1} > n_k$ para todo $k \in \mathbb{N}$). Luego $\{a_{n_k}\}_{k\in\mathbb{N}}$ es una *subsucesión* de $\{a_n\}_{n\in\mathbb{N}}$.

◇ **Ejer 3.16.** Considere T la transformación en el espacio e sucesiones, dada por $T\{a_n\}_{n\in\mathbb{N}} = \{a_{n+9}\}_{n\in\mathbb{N}}$. ¿Cómo describiría esta transformación?

Ejemplo 3.17. $(a_n)_{n\in\mathbb{N}}$ definida por $(a_n) = \frac{1}{n} = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots)$. Notar que podemos tomar una subsucesión de la forma $(a_{n_k}) = (a_{2^k}) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots,)$

Definición 3.18. Sea $C \subset \mathbb{R}$ un conjunto. Decimos que $a \in \mathbb{R}$, *punto de acumulación* de C si existe una sucesión $(c_n)_{n\in\mathbb{N}}$ de elementos de C, con $c_n \neq a$, tal que $\lim_{k\to\infty} c_{n_k} = a$.

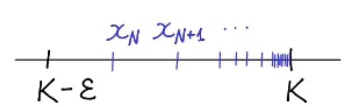
Definición 3.19. Diremos que la sucesión $\{x_n\}$ es *monótona creciente* cuando $x_n \leqslant x_{n+1}$, para todo $n \in \mathbb{N}$.
Análogamente, diremos que la sucesión $\{x_n\}$ es *monótona decreciente* cuando $x_n \geqslant x_{n+1}$, para todo $n \in \mathbb{N}$.
Cuando las desigualdades $\leqslant$ o $\geqslant$ son estrictas ($<$ o $>$) diremos que son *estrictamente creciente* o *estrictamente decreciente* respectivamente.

El siguiente teorema nos da una condición suficiente de convergencia.

Teorema 3.20 *Toda sucesión monótona y acotada es convergente.*

Idea de la prueba: Supongamos que $\{x_n\}$ es monótona creciente y acotada. Por ser acotada, tenemos que existe el supremo de $\{x_n\}$, que denotamos por K. Veamos que K tiene que ser el límite de la sucesión.

Como K es cota superior se tiene que $x_n \leqslant K$ para todo $n \in \mathbb{N}$. Como K es el menor de los números que verifica esa propiedad, se tiene que para cualquier $\varepsilon > 0$ el número $K - \varepsilon$ (que es menor a K) no puede ser cota superior. Es decir que existe por lo menos un término x_N suficientemente grande tal que $K - \varepsilon < x_N \leqslant K$. Ahora como al sucesión es monótona creciente tenemos que $K - \varepsilon < x_N \leqslant x_n \leqslant K$ para todo $n > N$. Es decir que $|K - x_n| < \varepsilon$ para todo $n > N$.



Como el ε es arbitrario concluimos que la sucesión tiene límite y es igual a K. $\square$

Teorema 3.21 (Bolzano-Weierstrass) *Toda sucesión acotada tiene una subsucesión convergente.*

La prueba es muy sencilla y se basa en el principio del palomar. Veamos la idea de la prueba.

Demostración. Sea $\{a_n\}_{n\in\mathbb{N}}$ la sucesión acotada. Para simplificar supongamos que está acotada entre $[0,1]$. Luego tomando los intervalos $I_1 = [0,1/2]$, e $I_2 = [1/2,1]$, tenemos que la sucesión $\{a_n\}_{n\in\mathbb{N}}$ tiene infinitos términos en al menos alguno de los intervalos, que denotamos por $I^{(1)}$. Prosiguiendo de la misma manera, relizando de manera consecutiva subdivisiones de la mitad del tamaño, obtenemos una sucesión de intervalos cerrados $I^{(n)}$ encajados ($I^{(n+1)} \subset I^{(n)}$) de tamaño $1/2^n$, tales que en cada uno hay infinitos términos de la sucesión. Luego definimos una subsucesión $\{a_{n_k}\}_{k\in\mathbb{N}}$, que satisfaga $a_{n_k} \in I^{(k)}$, $(k = 1,2,\ldots)$. Por construcción, los términos de la sucesión mayores a k_0, están en $I^{(k_0)}$, y por lo tanto la distancia entre ellos es menor a $1/2^{k_0}$. Esto implica que la subsucesión es de de Cauchy, y por lo tanto convege. □

Este Teorema se puede generalizar a sucesiones en $\mathbb{R}^n$. La prueba consiste en ver que las componentes tienen que converger, y por lo tanto la sucesión converge.

3.2. Sucesión de funciones

Para terminar este capítulo, vamos a introducir nociones de convergencias en espacios de funciones.

Vamos a trabajar con funciones definidas en un conjunto arbitrario X, a valores en $\mathbb{R}$. Podríamos considerar sucesión de funciones definidas en conjuntos arbitarios, pero para los fines de este curso (en particular estudiar variables aleatorieas) podemos reducirnos a funciones a valores reales.

Sin embargo, para algunos resultados particulares necesitaremos más estructura en X, como para poder integrar, y en tales casos tomaremos $X \subset \mathbb{R}$.

Una *sucesión de funciones* $\{f_n\}_{n\in\mathbb{N}}$, es simplemente una colección de funciones $f_n : X \to \mathbb{R}$, indexadas por $n \in \mathbb{N}$.

Veamos a continuación algunos ejemplos sencillos que servirán para más adelante. Se sugiere realizar bosquejos en cada caso.

Ejemplo 3.22. Consideremos la sucesión de funciones

$$f_n : [0,1] \to \mathbb{R}, \qquad f_n(x) = x^n.$$

Ejemplo 3.23.

$$f_n : \mathbb{R} \to \mathbb{R}, \qquad f_n(x) = \frac{\sin(nx)}{n}.$$

Ejemplo 3.24.

$$f_n : [0,1] \to \mathbb{R}, \qquad f_n(x) = \begin{cases} 2n^2 x & x \in [0, \frac{1}{2n}] \\ -2n^2(x - 1/n) & x \in [\frac{1}{2n}, \frac{1}{n}] \\ 0 & x \in [\frac{1}{n}, 1], \end{cases}$$

(No se asusten! es una función continua, lineal a trozos, y por lo tanto para bosquejarla sólo basta calcular su valor en los extremos de los intervalos.)

3.3. Convergencia puntual

Definición 3.25 (Convergencia puntual)**.** Consideremos una sucesión de funciones $\{f_n\}_{n\in\mathbb{N}}$. Supongamos que la sucesión de números reales $\{f_n(x)\}_{n\in\mathbb{N}}$, para cada $x \in X$, es convergente. Entonces existe una función

$$f : X \to \mathbb{R}, \quad \text{definida por} \quad f(x) := \lim_{n\to+\infty} f_n(x).$$

En este caso decimos decimos que la sucesión $\{f_n\}$ *converge puntualmente* a la función f, o que f es límite puntual de f_n, y escribiremos $f_n \to_n f$, o $\lim_n f = f$.

Para la definición anterior es imporante que el codominio sea $\mathbb{R}$, y no un espacio arbitrario. De hecho lo importante es que el codominio tenga una noción de distancia entre pares de puntos para que la definición se pueda extender con naturalidad. Este tipo de espacios se denominan, *espacios métricos*. En cambio, al conjunto X no se le pide nada.

Observación 3.26. En términos "$\varepsilon - \delta$", $f_n \to f$ si para todo $\varepsilon > 0$ y $x \in X$, se tiene que existe $n_0 = n_0(\varepsilon, x)$, tal que

$$|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geqslant n_0.$$

(Cuando escribimos $n_0 = n_0(\varepsilon, x)$, es para reforzar la idea de la dependencia de n_0 de ε y x.

Una vez definido el límite puntal una pregunta interesante surge.

¿Cuándo podemos asegurarnos que propiedades de la sucesión de funciones son heredadas a su límite?

Por ejemplo, si las funciones f_n son continuas, ¿su límite f lo es? Análogo para la diferenciabilidad: si las f_n son derivables, ¿f lo es? y en tal caso se tiene $f_n'(x) \to_n f'(x)$? Análogo para integrabilidad.

A continuación intentaremos analizar posibles respuestas a estas preguntas. Para fijar ideas consideremos X un subconjunto de los reales (la recta misma o un intervalo alcanzan para el análisis).

Continuidad

Recordar que si una función f es continua en $x_0 \in X$, entonces

$$\lim_{x\to x_0} f(x) = f(x_0).$$

Por lo tanto, en el caso que f sea límite puntual de $\{f_n\}$, i.e. $f = \text{lím}_n f$, lo que nos gustaría saber es si es verdad que

$$\overbrace{\underset{x\to x_0}{\text{lím}}\underbrace{\left(\underset{n}{\text{lím}}\, f_n(x)\right)}_{f(x)}}^{\text{lím}_{x\to x_0} f(x)} \overset{(?)}{=} \overbrace{\underset{n}{\text{lím}}\underbrace{\left(\underset{x\to x_0}{\text{lím}}\, f_n(x)\right)}_{f_n(x_0)}}^{f(x_0)}.$$

En este sentido nuestra interrogante se transforma en la siguiente pregunta

¿podemos intercambiar límites?

Para encontrar una respuesta basta recordar los ejemplos dados más arriba.

En el Ejemplo 3.22 es fácil ver que

$$f(x) = \underset{n}{\text{lím}}\, x^n = \begin{cases} 0 & x \in [0,1] \\ 1 & x = 1 \end{cases} \tag{3.1}$$

Pero entonces, tomando $x_0 = 1$ tenemos

$$0 = \underset{x\to 1^-}{\text{lím}} f(x) \neq f(1) = 1.$$

Con esto concluimos que límite puntual de funciones continuas, no tiene neceseriamente que ser una función continua.

Diferenciabilidad*[1]

Para este caso basta mirar el Ejemplo 3.23. Observar que $\text{lím}_n \sin(nx)/n = 0$ para todo $x \in \mathbb{R}$, y por lo tanto la función límite puntual f es la identicamente nula, i.e. $f \equiv 0$. ¿Se podrá asegurar que $f_n'(x) \to f'(x) = 0$?

Derivando obtenemos $f_n'(x) = \cos(nx)$, y en particular $f_n'(\pi) = \cos(n\pi) = (-1)^n$, y por lo tanto no existe su límite $f_n'(\pi)$. Es decir, no está definido el límite puntal de la sucesión de las derivadas.

Integración

Recordando nuestra consulta inicial, en el caso de integración nos gustaría saber si el límite de funciones integrables es integrable. Y si ese es el caso nos gustaría saber si es verdad que $\int f_n \to \int f$, i.e., si es correcto el intercambio de límites:

$$\underset{n}{\text{lím}} \int f_n(x)\,dx \overset{(?)}{=} \int \underset{n}{\text{lím}}\, f_n(x)\,dx.$$

Mirando el Ejemplo 3.24, es fácil ver que independiente de x, el límite es cero y por lo tanto $f = \text{lím}_n f \equiv 0$.

Además es claro que $\int_0^1 f_n(x)\,dx = 1/2$, y por lo tanto

$$\frac{1}{2} = \underset{n}{\text{lím}} \int_0^1 f_n(x)\,dx \neq \int_0^1 \underset{n}{\text{lím}}\, f_n(x)\,dx = 0.$$

[1] Puede omitirse esta sección en una primera lectura. No lo usaremos en lo que resta del curso.

Los párrafos anteriores muestran que hay que tener cuidado con intercambiar "límites" en las operaciones anteriores (de continuidad, diferenciación e integración) cuando tenemos convergencia puntual de una sucesión de funciones.

En la siguiente sección veremos una noción de límite más fuerte (más restrictiva) que nos permitirá trabajar con comodidad el intercambio de límites.

3.4. Convergencia Uniforme

Definición 3.27 (Convergencia uniforme)**.** Decimos que una sucesión de funciones $\{f_n\}$ *converge uniformemente* a f si dado $\varepsilon > 0$, existe $n_0 = n_0(\varepsilon) \in \mathbb{N}$ tal que

$$|f_n(x) - f(x)| < \varepsilon, \qquad \forall\, n \geqslant n_0,\ \forall\, x \in X. \tag{3.2}$$

En tal caso escribimos $f_n \rightrightarrows f$.

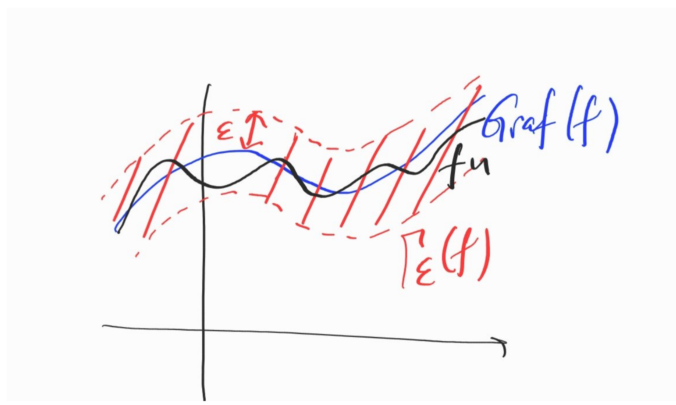


Figura 3.1: Convergencia uniforme.

Comentario 3.28. Una forma visual de entender la convergencia uniforme es via el gráfico $\text{Graf}(f_n)$ de las funciones de la sucesión.

Sea f el límite uniforme de $\{f_n\}$. Definimos el ε-*entorno tubular* del gráfico de f:

$$\Gamma_\varepsilon(f) := \{(x,y) \in X \times \mathbb{R} : |y - f(x)| < \varepsilon\}.$$

Entonces, la condición descrita en (3.2), es análogo a decir que $\text{Graf}(f_n) \subset \Gamma_\varepsilon(f)$, para todo $n \geqslant n_0$.

Observación 3.29. Una forma equivalente de escribir la fórmula (3.2) es la siguiente:

$$\sup_{x\in X} |f_n(x) - f(x)| \leqslant \varepsilon, \qquad \forall\, n \geqslant n_0,$$

(¿Por qué el menor igual?)

Observación 3.30. Dado que la condición (3.2) involucra a todos los $x \in X$, es claro que si $f_n \rightrightarrows f$, entonces $f_n \to f$ puntualmente. Sin embargo el recíproco está lejos de ser cierto. Lo que hace la diferencia (gran diferencia por cierto) entre la convergencia uniforme y puntual es que el caso de convergencia puntual el n_0 depende del ε y del $x \in X$ (ver Observación 3.26), y por lo tanto no podemos asegurarnos que en todos los puntos x se aproximan a su límite de manera simultanea.

Ejemplo 3.31. En el Ejemplo 3.22 y Ejemplo 3.24 no hay un n_0 universal.

♢ **Ejer 3.32.** En el Ejemplo 3.22, dado un $x \in [0,1)$, y $\varepsilon > 0$, hallar $n_0 = n_0(\varepsilon, x)$ tal que $x^n < \varepsilon$ para todo $n \geqslant n_0$. Observar que si $x \to 1$, entonces $n_0 \to +\infty$.

♢ **Ejer 3.33.** Si cambiamos el dominio $[0,1]$ en el Ejemplo 3.22 por el subconjunto $[0,1)$. ¿$f_n|_{[0,1)}$ converge uniformemente a 0? Puede describir algún subintervalo $[0,a]$, con $a < 1$ donde haya convergencia uniforme?

Nota: En la siguiente proposición establecemos un criterio para saber cuándo hay convergencia uniforme. Quizás exista un poco de redundancia en la prueba pero dado que recień comenzamos está bueno establecer el lenguaje que trabajaremos en el curso, además de ser una entrada en calor al mismo.

Proposición 3.34 (Criterio de convergencia)**.** *Sean $f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, y $f : X \to \mathbb{R}$. Definamos*

$$d_n := \sup_{x \in X} |f_n(x) - f(x)|. \tag{3.3}$$

Entonces $f_n \rightrightarrows f$ si y solo si $\lim_{n\to+\infty} d_n = 0$.

Demostración. ($\Leftarrow$) Ejercicio

($\Rightarrow$) Queremos probar que $d_n \to 0$ cuando $n \to \infty$. Dado un $\varepsilon > 0$, sabemos que existe $n_0 \in \mathbb{N}$ tal que $|f_n(x) - f(x)| < \varepsilon$, para todo $n \geqslant n_0$, y $x \in X$. Entonces, tomando supremo en $x \in X$ se tiene

$$d_n := \sup_{x \in X} |f_n(x) - f(x)| \leqslant \varepsilon, \qquad \forall n \geqslant n_0,$$

esto es exactamente la definición de convergencia a 0 de la sucesión $\{d_n\}$.

Con esto hemos probado que dado $\varepsilon > 0$, existe $n_0 = n_0(\varepsilon)$ tal que $0 \leqslant d_n = |d_n| \leqslant \varepsilon$ para todo $n \geqslant n_0$, lo que significa $\lim_n d_n = 0$. □

Comentario 3.35. Sea $\mathcal{B}(X)$ el espacio de funciones $f : X \to \mathbb{R}$ acotadas, i.e., $f : X \to \mathbb{R}$ tales que existe $k_f > 0$ tal que $|f(x)| \leqslant k_f$, para todo $x \in X$. Este conjunto tiene estructura de espacio vectorial (es decir sumar funciones acotadas, y multiplicar por número, siguen siendo funciones acotadas) y además tiene una *norma*, llamada *norma infinito en X* definida por

$$\|f\|_\infty := d(f,0) = \sup_{x \in X} |f(x)|.$$

De la observación anterior resulta que

$$f_n \rightrightarrows f \quad \text{si y solo si} \quad \|f_n - f\|_\infty \to 0.$$

es decir, que la convergencia uniforme dada en la definición 3.4, coincide con la convergencia en la norma infininto en $\mathcal{B}(X)$.

Ejemplo 3.36. Retomando el Ejemplo 3.22, vimos que la sucesión de funciones $f_n(x) = x^n$ converge puntualmente a la función f dada en (3.1). Observar que $d_n = d(f_n, f) = 1$, y por lo tanto no hay convergencia uniforme.

Ejemplo 3.37. En el Ejemplo 3.22, se tiene que la distancia uniforme de $f_n(x) = \sin(x)/n$ a la función 0 es $1/n$. Por lo tanto f_n converge uniformemente a la función 0.

Ejemplo 3.38. En el Ejemplo 3.24 es fácil ver que $d_n = n$ y por lo tanto f_n no converge uniformemente a su límite puntual.

Ejemplo 3.39. Sea $a \in (0,1)$. Consideremos la sucesión de funciones $f_n(x) = x^n$, $x \in [0,a]$. Entonces f_n converge puntualmente a la función 0 en el intervalo $[0,a]$. Pero además $d_n = a^n \to 0$, y por lo tanto la convergencia es uniforme.

3.4.1. Convergencia uniforme y continuidad

Es fácil convencerse con un dibujo (utilizando entornos tubulares) que si tenemos una discontinuidad de una función f en x_0 (pero con límites laterales definidos), difícilmente la podamos aproximar uniformemente por funciones continuas.

El siguiente resultado formalizaremos esa idea mostrando que la convergencia uniforme de funciones continuas es continua.

Teorema 3.40 *Sea $X \subset \mathbb{R}$, $f_n : X \to \mathbb{R}$ continua en $z_0 \in X$. Sea $f : X \to \mathbb{R}$ tal que $f_n \rightrightarrows f$. Entonces f es continua en z_0.*

Demostración. Queremos probar que dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$|f(z) - f(z_0)| < \varepsilon \quad \text{siempre que} \quad |z - z_0| < \delta.$$

La idea detrás de la prueba es que si n es suficientemente grande los valores de $f_n(z)$ y $f(z)$ son arbitrariamente cercanos para todo z en el dominio, luego triangulando de la siguiente manera

$$|f(z) - f(z_0)| \leqslant |f(z) - f_n(z)| + |f_n(z) - f_n(z_0)| + |f_n(z_0) - f(z_0)|, \tag{3.4}$$

tendríamos que los términos de los extremos son arbitrariamente pequeños sin importar el punto donde están evaluados, y el del medio es pequeño si z es próximo a z_0 utilizando la continuidad de las f_n.

Más precisamente, existe n_0 tal que $\|f - f_n\|_\infty < \varepsilon/3$ para todo $n \geqslant n_0$. Por otro lado, para n_0 fijo, tenemos de la continudad de f_{n_0} en z_0, que existe $\delta > 0$ tal que $|f_n(z) - f_n(z_0)| < \varepsilon/3$ siempre que $|z - z_0| < \delta$. Por lo tanto de la desigualdad (3.4), para $n = n_0$, concluimos que si $|z - z_0| < \delta$ entonces

$$\begin{aligned} |f(z) - f(z_0)| &\leqslant 2\|f - f_n\|_\infty + |f_n(z) - f_n(z_0)| \\ &< 2\varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

□

Corolario 3.41. *En las hipótesis del Teorema 3.40 resulta*

$$\lim_{n\to\infty}\lim_{z\to z_0} f_n(z) = \lim_{z\to z_0}\lim_{n\to\infty} f_n(z)$$

Demostración. De la continuidad de f_n en z_0, y del límite puntal se tiene que el miembro izquierdo coincide con $f(z_0)$. Por otro lado, el miembro derecho utilizando el límite puntual y luego de la continuada de f en z_0, se tiene que $\lim_{z\to z_0} f(z)$ es igual a $f(z_0)$ probando la igualdad requerida. □

3.4.2. Convergencia uniforme e integración

Teorema 3.42 *Sea $I\subset\mathbb{R}$, un intervalo acotado, y sean $f_n: I\to\mathbb{R}$ continuas. Supongamos que existe f tal que $f_n \rightrightarrows f$. Sea $x_0\in I$, y definimos*

$$F_n(x) := \int_{x_0}^{x} f_n(t)\,dt, \qquad F(x) := \int_{x_0}^{x} f(t)\,dt, \qquad x\in I.$$

Entonces resulta $F_n \rightrightarrows F$ en I.

La prueba de este teorema se puede ver fácilmente reflejada utilizando el Comentario 3.28 la diferencia $\|F-F_n\|_\infty$ está acotada por el área del entorno tubular, el cual tiende a cero cuando n crece. Veamos una prueba analítica.

Demostración. Observar que del Teorema 3.40 se tiene que f es continua, y por lo tanto F está bien definida.

Como $f_n \rightrightarrows f$, dado $\varepsilon>0$, existe n_0 tal que $\|f-f_n\|_\infty<\varepsilon$, para todo $n\geqslant n_0$.

Entonces

$$\begin{aligned}
|F(x)-F_n(x)| &= \left|\int_{x_0}^{x} f(t)\,dt - \int_{x_0}^{x} f_n(t)\,dt\right| \\
&= \left|\int_{x_0}^{x} (f(t)-f_n(t))\,dt\right| \\
&\leqslant \int_{x_0}^{x} |f(t)-f_n(t)|\,dt < \varepsilon|x-x_0|.
\end{aligned}$$

Como el intervalo I es acotado, tenemos que $|x-x_0|\leqslant \ell_I$, donde ℓ_I es la longitud del intervalo. Luego concluimos que

$$\|F-F_n\|\infty \leqslant \varepsilon\cdot\ell_I.$$

Luego “ajustando” ε obtenemos el resultado, i.e., si tomamos al comienzo n_0 tal que $\|f-f_n\|_\infty<\varepsilon/\ell_I$, entonces obtendremos que $\|F-F_n\|_\infty<\varepsilon$. □

♢ **Ejer 3.43.** Nuestro comentario previo a la prueba sugería una cota (quizás grosera) de que la diferencia $\|F-F_n\|_\infty$ estaba acotada por él área del ε-entorno tubular de f, que es $2\varepsilon\cdot\ell_I$. Sin embargo la prueba analítica fue más precisa. ¿Podrías perfeccionar nuestra estimativa gráfica para que diera la cota analítica?

Corolario 3.44. *En las hipótesis del Teorema 3.42 se tiene*

$$\lim_n \int_{x_0}^{x} f_n(t)\,dt = \int_{x_0}^{x} \lim_n f_n(t)\,dt, \qquad x\in I.$$

Parte II

¿Cara o número? Una introducción a teoremas límites

Capítulo 4

Modelo probabilístico

Comenzaremos estas notas estudiando el problema clásico de lanzar una moneda y observar el resultado. Para esta parte seguiremos de cerca el libro, de lectura altamente recomendable, Lesigne [EL].

La motivación de estudiar este problema en el arranque de este curso se debe a que 3 objetivos diferentes.

Primero servirá como un repaso de definiciones ya vistas en el curso anterior de Probabilidad, y por lo tanto un buen calentamiento para lo que viene.

Por otro lado, este modelo sencillo sirve para visualizar resultados fundamentales de la teoría como lo son la *ley de los grandes números* y el *teorema central del límite*.

Y por si fuera poco, nos dará pie para mostrar la importancia de la axiomatización de la probabilidad via teoría de la medida.

4.1. Experimento aleatorio

En esta primera parte consideramos modelos matemáticos de experimentos aleatorios que tienen una cantidad finita de resultados posibles.

Sea $\Omega = \{\omega_1, \ldots, \omega_k\}$ el espacio de posibles resultados. A cada $\omega_i \in \Omega$, para $i = 1, \ldots, k$, se le asigna una *probabilidad* p_i, donde cada $p_i \geqslant 0$, y además se satisface $\sum_{i=1}^{k} p_i = 1$.

Observar que el párrafo anterior es una formalización matemática de un experimento aleatorio como puede ser tirar un dado ($k = 6$) o lanzar una moneda ($k = 2$). En otras palabras, al igual que en muchas ramas de la ciencia, lo que estamos realizando es una *idealización* del experimento real.

Es importante mencionar que con la idealización anterior estamos dejando de lado discusiones sumamente interesantes sobre la asignación *a priori* de las probabilidades asignadas a cada resultado. ¿Está nuestro dado balanceado? Este tipo de preguntas, como ya saben, pertenecen al mundo de la *estadística*.

Un subconjunto de Ω se denomina *evento*, y la *probablidad* asociada a este se define por la suma de las probabilidades de los resultados en el conjunto. Esto es, si $A \subset \Omega$, se define

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i.$$

Observar que con esta definición tenemos que los eventos puntuales $\{\omega_i\}$ tienen probabilidad p_i, i.e., $\mathbb{P}(\{\omega_i\}) = p_i$.

Sea $\mathbb{1}_A$ o χ_A la *función indicatriz* del conjunto A, esto es, $\mathbb{1}_A : \Omega \to \{0,1\}$ toma el valor 1 en A y 0 fuera de este. Por lo tanto tenemos

$$\mathbb{P}(A) = \sum_{i=1}^{k} p_i \mathbb{1}_A(\omega_i).$$

Es fácil ver que con estas definiciones $(\Omega, \mathbb{P})$ consitituye un espacio de probabilidad discreto, como ya han visto en cursos anteriores. Es decir, Ω es un conjunto finito, y $\mathbb{P}$ es una función de subconjuntos de Ω que satisface:

- $\mathbb{P}(\Omega) = 1$,
- Si $A, B \subset \Omega$, son subconjuntos disjuntos, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

En el caso de que equiprobabilidad de sucesos elementales ω_i, decimos que el la función $\mathbb{P}$ es la *probabilidad uniforme*.

Ejemplo 4.1. El experimento de lanzar una moneda fiel puede modelarse con el espacio $\Omega = \{0,1\}$ y la probabilidad uniforme, i.e. $\mathbb{P}(0) = \mathbb{P}(1) = 1/2$. De manera similar, el experimento de tirar un dado se puede modelar considerarando la probabilidad uniforme en el espacio muestral $\{1,2,3,4,5,6\}$.

Ejemplo 4.2. Un espacio de probabilidad bastante conocido por ustedes es tomar $\Omega = \{0,1\}$ con probablidad de éxito $\mathbb{P}(1) = p$, y $\mathbb{P}(0) = 1 - p$, siendo $0 < p < 1$. Denotaremos este espacio $(1 - p, p)$.

4.2. Repeticiones del experimento

Consideramos el experimento básico de 1 y 0, (éxito y fracaso), siendo p la probabilidad éxito, y $q = 1 - p$ la de fracaso.

Si realizamos este experimento de forma repetida, n veces, podemos suponer que nuestro espacio muestral es $\Omega_n = \{0,1\}^n$, i.e., las tiras de números $(\omega_1,\dots,\omega_n)$ tales que $\omega_i \in \{0,1\}$. A continuación definiremos una probabilidad $\mathbb{P}_n$ en Ω_n.

Si suponemos que las experimentos se realizan de manera idéntica, podemos suponer

$$\mathbb{P}_n(\omega_i = 0) = q, \quad \text{y} \quad \mathbb{P}_n(\omega_i = 1) = p. \tag{4.1}$$

Acá entendemos por $\{\omega_i = 0\} \subset \Omega_n$, al evento formado por todas las tiradas tales que en el lugar i toma el valor 0. Por ejemplo $\{\omega_1 = 0\}$ es el formado por tiradas $(0,*,\dots,*)$ siendo $* \in \{0,1\}$.

Si suponemos que lo que ocurre en la tirada $(i+1)$, es independiente de lo que ocurre antes, podemos asumir

$$\begin{aligned}\mathbb{P}_n\left(\omega_{i+1} = 1 \text{ y } (\omega_1,\dots,\omega_i) = (a_1,\dots,a_i)\right)\\ = \mathbb{P}_n(\omega_{i+1} = 1)\cdot \mathbb{P}_n\left((\omega_1,\dots,\omega_i) = (e_1,\dots,e_i)\right),\end{aligned} \tag{4.2}$$

donde $a_\ell \in \{0,1\}$.

◇ **Ejer 4.3.** ¿Es suficiente la condición (4.1) para definir una probabilidad en Ω_2?

◇ **Ejer 4.4.** Pruebe por inducción, que las condiciones (4.1) y (4.2) definen una probabilidad en Ω_n.

◇ **Ejer 4.5.** Probar que si $S_n(\omega)$ cuenta la cantidad de éxitos en cada resultado $\omega \in \Omega_n$, entonces se tiene

$$\mathbb{P}_n(\omega) = p^{S_n(\omega)} \cdot q^{n-S_n(\omega)}.$$

Esta probablidad construida se denomina *probabilidad* producto, y la denotamos $\mathbb{P}_n = (q,p)^{\otimes n}$.

Un caso particular importante es cuando la probabilida de éxito es la misma que de fracaso en cada experimento realizado. En este caso $\mathbb{P}_n = (1/2,1/2)^{\otimes n}$, y resulta en probabilidad uniforme en Ω_n, siendo la probabilidad de un evento elemental igual a $1/2^n$, y la de un evento cualquier, su cardinal dividido 2^n.

4.3. Variables Aleatorias

Consideremos un espacio de probabilidad finito $(\Omega,\mathbb{P})$. A una función X definida en Ω con valores a $\mathbb{R}$, i.e. $X:\Omega\to\mathbb{R}$, la denominamos *variable aleatoria*.

La *distribución de probabilidad* está dada por las probabilidades correspondientes a los eventos $\{X = x_i\}$, $(i = 1,\dots,k)$, siendo $x_1,\dots,x_k$, valores de X. Observar que dichos conjuntos forman una partición (disjunta) de Ω.

Su *valor esperado* está dado por

$$\mathbb{E}(X) := \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i). \tag{4.3}$$

Básicamente, el $\mathbb{E}(X)$ es el promedio de los valores tomados por X, ponderados por la probabilidad de que X tome dicho valor.

Observar que podemos escribir X como la suma de indicatrices:

$$X = \sum_i x_i \mathbb{1}_{X=x_i}. \tag{4.4}$$

La esperanza cumple varias propiedadas conocidas.

Proposición 4.6.
1. *Si X es constante, enonces $\mathbb{E}(X)$ coincide con esa constante.*
2. *Si $X \geqslant 0$, entonces $\mathbb{E}(X) \geqslant 0$.*
3. *Si X e Y son v.a. discretas tales que $X \leqslant Y$, entonces $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.*
4. *$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$*
5. *La función $\mathbb{E}$ actúa como un operador lineal sobre el espacio de variables aleatorias discretas.*

◇ **Ejer 4.7.** Pruebe las propiedades anteriores. Para la última conviene obervar que la descomposición (4.4) no es única en el sentido de que no es la única combinación lineal de indicatrices que resulta en la variable aleatoria X. En particular se pueden particionar los eventos $\{X = x_i\}$ para tener una expresión tipo $X = \sum_k y_k \mathbb{1}_{A_k}$.

◇ **Ejer 4.8.** Sea f una función a valores reales, definida sobre la imagen de X. Considere la nueva variable aleatoria $f\circ X$. Pruebe que

$$\mathbb{E}(f(X)) = \sum_{i=1}^{k} f(x_i)\cdot \mathbb{P}(X = x_i). \tag{4.5}$$

A continuación veremos un resultado importante que usaremos para la ley de los grandes números

Proposición 4.9 (Chebyshev–Markov)**.** *Sea X una v.a. no negativa. Entonces, para cada $a > 0$, se tiene*

$$\mathbb{P}(X > a) \leqslant \frac{\mathbb{E}(X)}{a}.$$

□

◇ **Ejer 4.10.** Dar una prueba de este resultado. (Sugerencia $\mathbb{P}(X > a) = \sum_{i:x_i>a} \mathbb{P}(X = x_i)$.)

Como corolario de este resultado tenemos lo siguiente:

Corolario 4.11. *Sea X una v.a., y $a > 0$. Entonces*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > a\right) \leqslant \frac{\operatorname{var}(X)}{a^2},$$

donde

$$\operatorname{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2. \tag{4.6}$$

4.4. Independencia

Recordar que grosso modo la noción de *independencia* entre dos eventos indica que la ocurrencia de uno, no altera la ocurrencia del otro.

Si A y B son dos eventos en Ω, entonces decimos que son independientes si

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\cdot\mathbb{P}(B).$$

¿Pero que tiene que ver esta linda expresión con la frase anterior? Si asumimos que $\mathbb{P}(B)>0$ tenemos que

$$\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}=\frac{\mathbb{P}(A)}{\mathbb{P}(\Omega)}.$$

Recordar que lo anterior es equivalente a pedir que la probabilidad condicional $\mathbb{P}(A|B)=\mathbb{P}(A)$.[1]

Ejemplo 4.12. Consideremos el modelo de dos tiradas independientes de un experimento de éxito y fracaso con probabilidaes p y q.

Sabemos del capítulo anterior que $(\Omega_2,\mathbb{P}_2)$ se puede describir como:

$$\Omega_2=\{0,1\}^2=\{(0,0),(0,1),(1,0),(1,1)\}=\{(\omega_1,\omega_2):\omega_i=0,1\}$$
$$\mathbb{P}_2(0,0)=q^2,\mathbb{P}_2(0,1)=pq,\mathbb{P}_2(1,0)=pq,\mathbb{P}_2(1,1)=p^2,$$

Del Ejercicio 4.4 sabemos que $\mathbb{P}_2$ queda unívocamente definida por los valores $\mathbb{P}_2(\omega_i=1)$, para $i=1,2$, y la condición que $\{\omega_1=1\}$ es independiente de $\{\omega_2=1\}$.

Ejemplo 4.13. Un ejemplo un poquito más general al anterior, es considerar el resultado de dos experimentos (ambos con dos resultados posibles, de éxitos respectivos p_1 y p_2) distintos que son independientes. Por ejemplo, Ω_2 definir

$$\begin{aligned}\mathbb{P}(0,0)&=(1-p_1)(1-p_2),\\ \mathbb{P}_2(0,1)&=(1-p_1)p_2,\\ \mathbb{P}_2(1,0)&=p_1(1-p_2),\\ \mathbb{P}_2(1,1)&=p_1p_2\end{aligned}$$

Decimos que una familia de eventos $\{A_1,\dots,A_\ell\}$ son independientes si se tiene

$$\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})=\prod_{r=1}^{k}\mathbb{P}(A_{i_r}),$$

para cualesquiera $1\leqslant i_i<\cdots<i_k\leqslant\ell$.

Recordar del curso básico de probabilidad que no alcanza con pedir que sean dos a dos independientes, o que valga la anterior igualdad sólo para el producto de todos (es decir, $k=\ell$).

[1]Rercordar que si $\mathbb{P}(B)>0$, entonces la probabilidad condicional de A dado B, se define por $\mathbb{P}(A|B)=\mathbb{P}(A\cap B)/\mathbb{P}(B)$.

♢ **Ejer 4.14.** Probar que en el modelo de tirar la moneda fiel, los conjuntos $A_1=\{\omega_1=1\}$, $A_2=\{\omega_2=1\}$, $A_3=\{\omega_1=\omega_2\}$ son dos a dos independientes, pero la familia $\{A_1,A_2,A_3\}$ no es una familia independientes.

♢ **Ejer 4.15.** Probar que en el modelo definido en el Ejercicio 4.4 los eventos $A_i=\{\omega_i=1\}$ son independientes.

El concepto de independencia puede trasladarse a variables aleatorias. Decimos que las variables aletorias $\{X_1,\dots,X_n\}$ son independientes, si los eventos $\{X_1=x_1\},\dots\{X_n=x_n\}$ son independientes para cualesquiera reales $x_1,\dots,x_n$.

♢ **Ejer 4.16.** Probar que las variables aletorias $\{X_1,\dots,X_n\}$ son independientes, si y sólo si los eventos $\{X_1\in B_1\},\dots\{X_n\in B_n\}$ son independientes, para todo $B_1,\dots,B_n$ subconjuntos de $\mathbb{R}$.

♢ **Ejer 4.17.** Probar que una familia de eventos es independientes si lo es la familia de sus funciones características.

♢ **Ejer 4.18.** Probar que las variables aletorias $\{X_1,\dots,X_n\}$ son independientes, si y sólo los eventos $\{X_1=x_1\}\cap\cdots\cap\{X_{j-1}=x_{j-1}\}$ es independiente de $\{X_j=x_j\}$, para cualesquiera $x_1,\dots,x_j$ en $\mathbb{R}$.

Proposición 4.19. *Sean X e Y v.a. independientes. Entonces*

$$\begin{aligned}\mathbb{E}(XY)&=\mathbb{E}(X)\mathbb{E}(Y)\\ \mathrm{var}(X+Y)&=\mathrm{var}(X)+\mathrm{var}(Y).\quad\square\end{aligned}$$

♢ **Ejer 4.20.** Dar una prueba de la proposición anterior.

4.5. Ley de los grandes números

Ya tenemos los prerrequisitos necesarios para el estudio asintótico del experimento de lanzar una moneda n veces, i.e. el espacio $(\Omega_n,(p,1-p)^{\otimes n})$.

Antes de ir al resultado de la ley de grandes números repasaremos la distribución binomial.

4.5.1. Distribución binomial

Sea S_n la v.a. que cuenta la cantidad de éxitos, i.e. la v.a. $S_n:\Omega_n\to\mathbb{R}$, tal que

$$S_n(\omega)=\sum_{i=1}^{n}\omega_i,$$

Proposición 4.21. *La v.a. S_n toma valores $0,1,\dots,n$ con probabilidades*

$$\mathbb{P}(S_n=k)=\binom{n}{k}p^k(1-p)^{n-k},\qquad k\in\{0,\dots,n\}.$$

♢ **Ejer 4.22.** Probar este resultado.

Sabemos de cursos anteriores que la v.a. S_n es la realización de una v.a. con distribución *Binomial*, de parámetros n y p, Una v.a. X definida en $(\Omega,\mathbb{P})$ que toma el valor 1 con probabilidad p, y 0 con probabilidad $1-p$ se denomina v.a. *Bernoulli*. Observar que una v.a. Bernoulli tiene distribución binomial con parámetro $n=1$, p.

En el siguiente resultado se observa la distribución binomial puede ser obtenida como una suma de v.a. independientes Bernoulli.

Proposición 4.23. *Sean $X_1,\dots,X_n$ v.a. independientes definidas en en un espacio de probabilidad $(\Omega,\mathbb{P})$ con distribución Bernoulli de parámetro p. Entonces la v.a.*

$$S_{n,p}=X_1+\cdots+X_n,$$

tiene distribución binomial de parámetro n y p. En particular resulta

$$\mathbb{E}(S_{n,p})=np,\quad \mathrm{var}(S_{n,p})=np(1-p).\quad\square$$

◊ **Ejer 4.24.** Probar el resultado anterior. La fórmula de esperanza y varianza resultan aplicaciones sencillas de la Proposición 4.19. Se sugiere como un lindo ejercicio de manipulación de sumas, probar ambas identidades utilizando las definiciones (4.3) (4.6) y manipulando (derivando) la expresión del binomio de newton $(a+b)^n$. (Para la varianza recordar expresión (4.13).

4.5.2. El modelo y sus frutos

En la formalización de un experimiento aleatorio de éxito y fracaso, la asignación de una probabilidad específica responde a la frecuencia con que estos aparecen. Pero como comentamos al principio, este número asignado (la probabilidad p por ejemplo) es una definición. La ley de los grandes números es justamente un resultado que vincula ambas cosas. Básicamente nos dice que en nuestro modelo teórico, si repetimos el experimento muchas veces, entonces la frecuencia de éxitos tiende a la probabilidad designada a priori.

Teorema 4.25 (Ley (débil) de Grandes Números para la binomial) *Para todo $\varepsilon>0$ se tiene*

$$\mathbb{P}_n\left(\left|\frac{S_n}{n}-p\right|>\varepsilon\right)\xrightarrow{n\to+\infty}0,$$

Además la convergencia es uniforme en $p\in(0,1)$.

Demostración. Utilizando el Corolario 4.11 se tiene

$$\mathbb{P}_n(|S_n-np|>n\varepsilon)\leqslant\frac{\mathrm{var}\,S_n}{n^2\varepsilon^2}=\frac{p(1-p)}{n\varepsilon^2}\leqslant\frac{1}{4n\varepsilon^2}.$$

Por lo tanto en el límite en n tiende a cero, y además la convergencia se puede hacer independiente de p. $\square$

Una aplicación importante de la ley de grandes números es dar una prueba constructiva del siguiente resultado fundamental en análisis.

Teorema 4.26 (Aproximación de Weierstrass) *Sea $f:[0,1]\to\mathbb{R}$ continua. Entonces f es aproximada uniformemente por polinomios.*

Demostración. Dejemos la prueba como ejercicio guiado. $\square$

◊ **Ejer 4.27.** Sean $X_1,\dots,X_n$ una colección de v.a. i.i.d. con distribución Bernoulli de parámetro p. Sea S_n las sumas parciales.

- Probar que la función
$$B_n(p)=\mathbb{E}\left(f\left(\frac{S_n}{n}\right)\right),$$
es un polinomio en p de grado a lo sumo n.
- Usar la desigualdad de Chebyshev para probar que para todo $p\in[0,1]$, y todo $\varepsilon>0$, se tiene
$$\sum_{k\in K}\binom{n}{k}p^k(1-p)^{n-k}\leqslant\frac{1}{4n\varepsilon^2},$$
donde $K=\{k\in\{0,1,\dots,n\},\,|k/n-p|>\varepsilon\}$.
- Utilizando la continuidad uniforme de f probar que
$$\lim_{n\to\infty}\sup_{0\leqslant p\leqslant 1}|f(p)-B_n(p)|=0.$$

4.6. Grandes desvíos

De la ley de los grandes números, tenemos la cota explícita para el

$$\mathbb{P}_n\left(\left|\frac{S_n}{n}-p\right|>\varepsilon\right)\leqslant\frac{p(1-p)}{n\varepsilon^2}.\tag{4.7}$$

¿Cuán precisa es esa desigualdad? Sabemos que esa desigualdad resulta de la desigualdad de Chebyshev, por lo que podemos preguntarnos cuán precisa es esta.

Supongamos que X es una variable aleatoria positiva, con esperanza finita $\mathbb{E}(X)$. Observar que la desigualdad de Chebyshev nos dice que

$$\mathbb{P}(X\geqslant 2\mathbb{E}(X))\leqslant\frac{1}{2}.$$

Este resultado tiene mucho sentido y puede ser interpretado de la siguiente manera. Supongamos que X da la ganancia individual en una población (o cualquier número positivo en una población). Si la desigualdad fuera un “>”, entonces diría que más de la mitad de la población gana más que el doble del promedio. Y esto es un absurdo, dado que el promedio en la población sería mayor a $\mathbb{E}(X)$. De hecho, aún en el caso que el resto gane 0, el promedio sería superior a $\mathbb{E}(X)$.

De manera análoga se puede observar que

$$\mathbb{P}(X\geqslant a\mathbb{E}(X))\leqslant\frac{1}{a},\quad\forall a>0.$$

Esta discusión nos da una primera debilidad de la desigualdad de Chebyshev, que es que si X es constante igual a 1, entonces la desigualdad de Chebyshev no aporta nueva información sobre la probabilidad $\mathbb{P}(X \geqslant 2)$.

Un simple ejercicio dice que la desigualdad de Chebyshev no puede ser mejorada sin hipótesis adicionales.

◇ **Ejer 4.28.** Dar un ejemplo de v.a. donde $\mathbb{P}(X \geqslant a) = \mathbb{E}(X)/a$, para $X \geqslant 0$, y $a > 0$

Que la desigualdad sea justa (o "sharp" en inglés) no nos impide encontrar mejores cotas para ciertos casos especiales con hipótesis adicionales.

◇ **Ejer 4.29.** Realizar una simulación en R que aproxime numéricamente la probabilidad de que en 100 tiradas la cantidad de caras que salgan sean mayores a 60. Análogo, para 1000 tiradas, y que salgan más de 600 caras.

Los resultados anteriores nos motivan a buscar una desigualdad más fina para (4.7). El área que se dedica a este tipo de refinamientos de la desigualdad anterior, pero para distintos tipos de v.a., se denomina *grandes desvíos*.

Sea $0 < \varepsilon < \min(p, 1-p)$, definiendo la función

$$h_+(\varepsilon) := (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (1-p-\varepsilon)\log\left(\frac{1-p-\varepsilon}{1-p}\right), \quad (4.8)$$

tenemos el siguiente resultado

Proposición 4.30 (Bernstein-1924'). *Para todo $0 < \varepsilon < \min(p, 1-p)$, definiendo $h_-(\varepsilon) := h_+(-\varepsilon)$, se tiene $h_+(\varepsilon), h_-(\varepsilon)$ son positivas y*

$$\mathbb{P}_n\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant e^{-nh_+(\varepsilon)} + e^{-nh_-(\varepsilon)}, \quad (4.9)$$

lo cual se aproxima a 0 *de manera exponencial cuando $n \to \infty$.*

Demostración. Sea $t > 0$ (el cual optimizaremos al final). Observa que el evento

$$\left\{\frac{S_n}{n} \geqslant p + \varepsilon\right\} = \left\{e^{t(S_n - np - n\varepsilon)} \geqslant 1\right\}.$$

Luego, aplicando la desigualdad de Chebyshev obtenemos

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} \geqslant p + \varepsilon\right) &\leqslant \mathbb{E}\left(e^{t(S_n - np - n\varepsilon)}\right) \\
&= e^{-nt(p+\varepsilon)}\mathbb{E}\left(e^{tS_n}\right) \\
&= e^{-nt(p+\varepsilon)}\sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\
&= e^{-nt(p+\varepsilon)}(e^t p + (1-p))^n
\end{aligned}$$

Es decir

$$\mathbb{P}\left(\frac{S_n}{n} \geqslant p + \varepsilon\right) \leqslant e^{-n[t(p+\varepsilon) - \log(e^t p + (1-p))]}.$$

Como la anterior desigualdad es válida para todo $t > 0$, tenemos que

$$\mathbb{P}\left(\frac{S_n}{n} \geqslant p + \varepsilon\right) \leqslant e^{-n\overline{h}},$$

siendo $\overline{h} = \sup_{t>0} t(p+\varepsilon) - \log(e^t p + (1-p))$.

Es un ejercicio verificar que la función $h(t) = t(p+\varepsilon) - \log(e^t p + (1-p))$, satisface $h(0) = 0$, $h'(0) = \varepsilon$, por lo que $\overline{h} > 0$. Además h' se anula en $\bar{t} = \log\left(\frac{(p+\varepsilon)(1-p)}{p(1-p-\varepsilon)}\right)$ que es donde se toma el máximo y se tiene que $h(\bar{t}) = h_+(\varepsilon)$.

Para la otra desigualdad, i.e. $\mathbb{P}(S_n/n \leqslant p - \varepsilon)$ procedemos de la siguiente manera. Sea W_n la cantidad de fracasos en las n tiradas, i.e. $S_n + W_n = n$. Luego el evento

$$\{S_n/n \leqslant p - \varepsilon\} = \{W_n/n \geqslant (1-p) + \varepsilon\}.$$

Por lo tanto basta utilizar nuesto resultado anterior cambiando p por $1-p$, y observar que el valor máximo, al sustituir p por $1-p$ en (4.8), resulta ser igual a $h_+(-\varepsilon)$. □

4.7. Ley fuerte de los grandes números

La ley débil de los grandes números, en el caso de la repetición independiente de lanzar una moneda, nos dice que dado la distribución empírica se aproxima a la media con alta probabilidad. Más formal, la probabilidad que la media de éxitos S_n/n esté a una distancia fija de la media, tiende a 0 con n.

La ley fuerte de los grándes números involucra medir sobre el espacio de tiradas infinitas. Lo que dice la ley fuerte es que salvo un conjunto que tiene probabilidad cero en este espacio, el siguiente límite $\lim_n S_n/n$ existe y es igual a la media.

En los siguientes párrafos intentaremos formalizar esto.

4.7.1. Eventos de tipo finito, conjuntos de medida 0, independencia

La formalización de nuestro espacio muestral de repetir infinitas veces, de manera independiente, el experimento de tirar una moneda con probabilidad de éxito p es

$$\Omega := \{\omega = (\omega_n)_{n\geqslant 1} : \omega_n = 0 \text{ o } 1, n \geqslant 1\} = \{0,1\}^{\mathbb{N}}.$$

Sobre Ω podemos definir la función $S_n : \Omega \to \mathbb{N}$, dada por

$$S_n(\omega) = \omega_1 + \cdots + \omega_n,$$

es decir, la función que cuenta la cantidad de éxitos en los primeros n experimentos.

Decimos que un subconjunto A de Ω, es un *evento de tipo finito* si está definido por ciertas restricciones en una cantidad finita de coordenadas. Por ejemplo $\{(1,1,\omega_3,\dots) : \omega_n \in \{0,1\}, n \geqslant 3\}$.

Observar $\{S_n = k\}$ es un un evento de tipo finito para $k = 1,\dots,n$.

Formalmente podemos decir que $A \subset \Omega$ es de tipo finito si existe un natural $n := n(A) \geqslant 1$, y un subconjunto $A' \subset \Omega_n$ tal que

$$A = \{\omega \in \Omega : \omega^{(n)} \in A'\}, \quad \text{donde} \quad \omega^{(n)} = (\omega_1, \dots, \omega_n),$$

es decir $\omega^{(n)}$ es la proyección de $\Omega \to \Omega_n$ en las primeras n coordenadas.

En particular podemos definir la *probabilidad* de estos conjuntos como

$$\mathbb{P}(A) := \mathbb{P}_{n(A)}(A') = \sum_{\omega^{(n)} \in A'} p^{S_n(\omega^{(n)})}(1-p)^{n-S_n(\omega^{(n)})}. \quad (4.10)$$

Observar que en la definición de evento finito, el natural $n(A)$ no está unívocamente determinado. De hecho si para cierto n, el conjunto A queda definido, entonces lo vale para cualquier $n' \geqslant n$. Esto nos obliga a ver que la definición de la probabilidad anterior no depende del n elegido.

♢ **Ejer 4.31.** Probar que la definición anterior no depende del $n(A)$ elegido.

Consideremos la familia $\mathcal{E}$ de eventos de tipo finito. Observar que agregando $\varnothing$ y Ω, tenemos que esta familia es cerrada por tomar complementos, por intersecciones finitas, y por uniones finitas.

Definición 4.32. Una familia que es cerrada respecto a esas operacione se denomina *álgebra*.[2]

Con esto tenemos que la familia $\mathcal{E}$ de eventos tipo finito es un álgebra.

Observar que la función $\mathbb{P}$ definida anteriormente es una función bien definida de $\mathcal{E}$ en $[0,1]$, y satisface

- $\mathbb{P}(\Omega) = 1$;
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, siempre que $A, B \in \mathcal{E}$ sean disjuntos.

Definición 4.33. Decimos que un conjunto $N \subset \Omega$ tiene *medida nula* (*negligible* en inglés) si para todo $\varepsilon > 0$ existe una colección numerable $\{A_k\}$ de conjuntos en $\mathcal{E}$ tal que

$$N \subset \bigcup_k A_k, \qquad \sum_k \mathbb{P}(A_k) < \varepsilon.$$

Diremos que un subconjunto de Ω es un *evento casi seguro* si su complemento tiene medida cero.

Para probar nuestra ley de grandes números vamos a restringirnos a trabajar con eventos de tipo finito, de medida cero o eventos casi seguros. Por esta razón de ahora en adelante llamaremos evento a alguno de estos tres casos.

[2] De hecho para que sea un álgebra basta que $\varnothing \in \mathcal{E}$; que si $A \in \mathcal{E}$ entonces $A^c \in \mathcal{E}$ y que si $A, B \in \mathcal{E}$ entonces $A \cup B \in \mathcal{E}$.

Observar que nuestra probabilidad $\mathbb{P}$ correctamente definida sobre $\mathcal{E}$, satisface la hipótesis de *invarianza por traslaciones*. Esto es que para todo $j \in \mathbb{N}$, la probabilidad del evento

$$\{\omega \in \Omega : (\omega_j, \omega_{j+1}, \dots) \in A\}$$

tiene la misma probabilidad que el evento A.

Por dar un ejemplo, la probabilidad del evento que en las primeras 10 tiradas salió cara, es la misma que la probabilidad que en las tiradas 10 a 19 salió cara.

Proposición 4.34. *1. Todo subconjunto de un conjunto de medida cero, tiene medida cero.*

2. Unión numerable de conjuntos de medida cero, tiene medida cero.

3. Asumiendo $0 < p < 1$, todo conjunto numerable de Ω tiene medida cero. □

♢ **Ejer 4.35.** Probar la anterior proposición.

Decimos que una familia de eventos $\{A_i\} \subset \mathcal{E}$ es *independiente* si

$$\mathbb{P}(\bigcap_{k=1} A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}),$$

para cualquier subonjunto finito de la familia.

Proposición 4.36. *Probar que eventos definidos por índices disjuntos, son independientes.*

Proposición 4.37. *Probar que los complementos también son independientes.*

Lo que dice la anterior proposición es lo siguiente. Si tenemos una colección $\{A_i\}_{i \in I}$ de eventos parametrizados por $I \subset \mathbb{N}$, y tenemos que para cada $i \in I$ hay un conjunto finito $E_i \subset \mathbb{N}$, y un conjunto $A_i' \subset \{0,1\}_i^E$ tal que $E_i \cap E_j = \varnothing$, si $i \neq j$, y

$$A_i = \{\omega \in \Omega : (\omega_n)_{n \in E_i} \in A_i'\},$$

luego los eventos $\{A_i\}_{i \in I}$ son independientes.

(Convencerse de la anterior proposición. Una forma)

Veamos algunos ejemplos más elaborados de conjuntos de medida cero. Para esto $0 < p < 1$.

> **Ejemplo 4.38.** Consideremos una palabra w formada por ceros y unos, esto es un conjunto finito $w \in \{0,1\}^j$. El conjunto formados por sucesiones que no contienen la palabra w tienen medida cero.

Veamos esto. Sea $W \subset \Omega$ el conjunto formado por sucesiones que no contienen la palabra w.

Para fijar ideas, supongamos que la palabra es $w = (1,1)$. Considerar primero el evento A_1 formado por sucesiones $(e_1, e_2, *, *, \dots)$ tales que $(e_1, e_2) \neq (1,1)$. Es claro que $W \subset A_1$. Observar que $0 < \mathbb{P}(A_1) = 1 - p^2 < 1$.

Consideremos $A_2 = (*,*,e_3,e_4,*,*,\dots)$ tales que $(e_3,e_4) \neq (1,1)$. Observar que $W \subset A_2$, y por la invarianza por traslaciones tenemos que $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1-p^2 < 1$. Procediendo de igual manera obtenemos una colección de conjuntos $\{A_k\}_{k\in\mathbb{N}}$ independientes de igual probabilidad tales que $W \subset A_1 \cap \cdots \cap A_n$ para todo n. Además $\mathbb{P}(A_1 \cap \cdots \cap A_n) = (1-p^2)^n \to 0$ cuando $n \to +\infty$, por lo que W tiene medida cero.

◊ **Ejer 4.39.** Escribir la prueba formal para una palabra $w = (w_1,\dots,w_j) \in \{0,1\}^j$ dada.

◊ **Ejer 4.40.** Probar que casi seguramente $\omega \in \Omega$ contiene todas las palabras posibles. (De aquí la frase de que con probabilidad 1 un mono tipeando en una computadora tarde o temprano escribirá exactamente alguna obra famosa.)

◊ **Ejer 4.41.** Probar que el conjunto formado por sucesiones periódicas, a partir de cierto índice, forman un conjunto de medida cero.

Ya con la idea de lo que es un conjunto de medida cero, podemos enunciar y probar la *ley fuerte de los grandes números*, debido a los trabajos de Émile Borel, en la primer década de 1900, donde se construyen las bases de la teoría de la medida y la formalizción de la probabilidad.

Teorema 4.42 (Ley fuerte de los grandes números) *Sea S_n la distribución binomial de parámetros n y p. Casi seguramente se tiene*

$$\lim_{n\to+\infty} \frac{S_n(\omega)}{n} = p. \tag{4.11}$$

Es decir, la condición (4.11) es válida para todo $\omega \in \Omega$ a menos de un conjunto medida cero.

Demostración. Consideremos $R_n(\omega) = \frac{1}{n}S_n(\omega) - p$. La sucesión $\{R_n(\omega)\}_n$ no tiende a cero si existe $k \in \mathbb{N}$ tal que para infinitos n se tiene $R_n(\omega) \geqslant \frac{1}{k}$. O dicho de otro modo, si existe $k \in \mathbb{N}$ tal que para cualquier $n \in \mathbb{N}$, existe un $\ell > n$ tal que $R_\ell(\omega) \geqslant \frac{1}{k}$.

Esta condición de no convergencia se puede escribir con teoría de conjuntos de la siguiente manera. Los ω's donde no hay convergencia de la sucesión (4.11) pertenecen a

$$N := \bigcup_{k=1}^{+\infty}\bigcap_{n=1}^{+\infty}\bigcup_{\ell=n}^{+\infty}\left\{\omega\in\Omega : |R_\ell(\omega)| \geqslant \frac{1}{k}\right\}.$$

Queremos ver que N tiene medida cero. Por la Proposición 4.34 sabemos que basta probar que el conjunto

$$N_k := \bigcap_{n=1}^{+\infty}\bigcup_{\ell=n}^{+\infty}\left\{\omega\in\Omega : |R_\ell(\omega)| \geqslant \frac{1}{k}\right\},$$

tiene medida cero.

Consideremos el evento

$$A_{k,\ell} := \left\{\omega\in\Omega : |R_\ell(\omega)| \geqslant \frac{1}{k}\right\}.$$

(Observar que este sí podemos medirlo porque es un evento tipo finito).

Utilizando el resultado de Bernstein de grandes desvíos, la Proposición 4.30, tenemos que existe una constante $c > 0$ (que depende de k de la probabilidad p), tal que

$$\mathbb{P}(A_{k,\ell}) \leqslant e^{-c\ell}.$$

Luego como la serie $\sum_{\ell=1}^{+\infty} e^{-c\ell}$, es convergente, resulta que la cola de la serie tiende a cero. Y por lo tanto, dado $\varepsilon > 0$ existe un $n_0 \in \mathbb{N}$ tal que

$$\sum_{\ell=n_0}^{+\infty} \mathbb{P}(A_{k,\ell}) < \varepsilon.$$

Luego como $N_k \subset \bigcup_{\ell=n} A_{k,\ell}$ para todo $n \geqslant 1$, se tiene que N_k es de medida cero. □

Observar que podemos reescribir la ley fuerte de la siguiente manera. Definiendo los eventos $A_k = \{\omega\in\Omega : \omega_n = 1\}$, con $k \in \mathbb{N}$, tenemos que

$$\frac{S_n(\omega)}{n} = \frac{1}{n}\#\big\{k : 1 \leqslant k \leqslant n, \text{ tal que } \omega \in A_k\big\}.$$

Observar que los eventos A_k son independientes y con igual probabilidad, a saber, $\mathbb{P}(A_k) = p$.

De la ley de los grandes números se desprende el siguiente resultado.

Teorema 4.43 *Sean $\{A_k\}$ una sucesión de eventos independientes y equiprobables. Entonces casi seguramente se tiene*

$$\lim_n \frac{1}{n}\#\big\{k : 1 \leqslant k \leqslant n, \text{ tal que } \omega \in A_k\big\} = \mathbb{P}(A_1). \tag{4.12}$$

La prueba de este resultado es muy similar a la ley fuerte.

Demostración. A cada $\omega \in \Omega$, se le puede hacer identificar una sucesión de 0's y 1's de manera tal que la coordenada k nos diga si ω está o no A_k. Esto es para cada $k \in \mathbb{N}$ considerar el mapa $\rho_k : \Omega \to \{0,1\}$ tal que $\rho_k(\omega) = 1$ si y sólo si $\omega \in A_k$, (i.e. $\rho_k = \mathbb{1}_{A_k}$), y el mapa requerido es $\omega\in\Omega \mapsto (\rho_1(\omega),\rho_2(\omega),\dots) \in \{0,1\}^{\mathbb{N}}$.

Luego la condición (4.12) puede ser reescrita como

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} \rho_k(\omega) = \mathbb{P}(A_1), \qquad \text{casi seguramente.}$$

Observar que de las hipótesis resulta que si $(e_1,\dots,e_n) \in \{0,1\}^n$, entonces

$$\mathbb{P}(\rho_1(\omega) = e_1,\dots,\rho_n(\omega) = e_n) = \mathbb{P}(A)^s(1-\mathbb{P}(A))^{n-s},$$

siendo $s = e_1 + \cdots + e_n$.

Luego podemos repetir cada paso de la prueba Teorema 4.42 para obtener el resultado. Dejamos estos detalles a cargo del lector. □

4.7.2. ¿Existe $\mathbb{P}(A)$ para cualquier $A \subset \Omega$?

Una pregunta razonable es saber si la función $\mathbb{P}$, que está bien definida en elementos de $\mathcal{E}$, puede ser extendida a cualquier subconjunto de Ω.

Esta pregunta crucial tiene un análgo similar a preguntarnos si podemos extender la definición de "longitud" a cualquier subconjunto del intervalo $[0,1]$, asumiendo que sabemos medir longitudes de intervalos y al álgebra generada por estos.

Como veremos en breve en este curso, la existencia de conjuntos que no podemos calcular su probabilidad motiva la pregunta de a qué familia de subconjuntos de Ω podemos asignarle una probabilidad, más allá de la familia (álgebra) $\mathcal{E}$.

Antes de enunciar el teorema realicemos algunos comentarios sobre qué hipótesis debería satisfacer nuestra medida. Para simplificar utilicemos el caso de $p = 1/2$.

Teorema 4.44 *Consideremos $\Omega = \{0,1\}^{\mathbb{N}}$, entonces no existe una función definida en las partes $\mathcal{P}(\Omega)$ en $[0,1]$ que satisfaga las siguientes hipótesis.*

(i) $\mathbb{P}(\Omega) = 1$;

(ii) Si $A_1, A_n, \ldots \subset \Omega$ son dos a dos disjuntos entonces

$$\mathbb{P}(\bigcup_{n \geqslant 1} A_n) = \sum_{n \geqslant 1} \mathbb{P}(A_n)$$

(iii) Para cada $n \geqslant 1$, $\mathbb{P}(F_n(A)) = \mathbb{P}(A)$, donde $F_n : \Omega \to \Omega$ está definido por

$$F_n : \omega = (\omega_1, \omega_2, \ldots) \mapsto (\omega_1, \ldots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \ldots)$$

Comentario 4.45. La segunda hipótesis se denomina σ-aditividad. La tercera dice que $\mathbb{P}$ es invariante por F_n, para todo $n \geqslant 1$. Esta hipótesis está motivada por que estamos repitiendo experimentos idénticos en el caso $p = 1/2$.

El siguiente ejemplo, es motivado por el conjunto de *Vitali*, (ejemplo típico de conjunto no medible en el intervalo $[0,1]$).

dada en (4.10).

Demostración. Para comenzar definimos una clase de equivalencia $\sim$ sobre Ω de la siguiente manera: $\omega \sim \omega'$ si ambos tienen la misma cola, i.e. si $\omega_n = \omega'_n$ a partir de cierta coordenada.

Consideremos el conjunto $A \subset \Omega$ definido por un representante en cada clase de equivalencia.[3]

Veamos que este conjunto nos generará problemas.

Consideremos la familia de subconjuntos finitos de $\mathbb{N}$, este es, $\mathcal{F} = \{C \subset \mathbb{N} : \#C < \infty\}$. Es fácil ver que $\mathcal{F}$ es un conjunto numerable, dado que es una unión numerable de conjuntos numerables (por ejemplo, es la unión en n de todos los subconjuntos de tamaño n).

Dado $C = \{n_1, \ldots, n_k\} \in \mathcal{F}$, definimos el mapa $F_C = F_{n_k} \circ \cdots \circ F_{n_1}$ que cambia el resultado en los lugares $n_1, \ldots, n_k$.
Afirmación 0: F_C deja invariante la clase (i.e. $F_C(\omega) \sim \omega$, para todo $C \in \mathcal{F}$), y además $\{F_C(\omega)\}_{C \in \mathcal{F}}$ genera la clase sin repeticiones. (En particular las clases de equivalencia son numerables.)
Afirmación 1: $\Omega = \bigcup_{C \in \mathcal{F}} F_C(A)$.
Si $\omega \in \Omega$, entonces está en alguna clase de equivalencia. Llamemos $\omega' \in A$ tal que $\omega' \sim \omega$. Luego de la Afirmación 0 se tiene que existe algún $C \in \mathcal{F}$ tal que $F_C(\omega') = \omega$.
Afirmación 2: $\{F_C(A)\}_{C \in \mathcal{F}}$ son disjuntos:
Esto resulta de que si $F_C(A) = F_{C'}(A)$, entonces existen $\omega, \omega' \in A$ tal que $F_C(\omega) = F_{C'}(\omega')$, y por lo tanto $\omega \sim \omega'$. Pero de nuestra elección de A resulta que esto implica que $\omega = \omega'$, y además por la Afirmación 0 se tiene que $C = C'$.

Luego tenemos una partición numerable disjunta de Ω y por lo tanto

$$1 \overset{(i)}{=} \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{C \in \mathcal{F}} F_C(A)\right) \overset{(ii)}{=} \sum_{C \in \mathcal{F}} \mathbb{P}(F_C(A)) \overset{(iii)}{=} \sum_{C \in \mathcal{F}} \mathbb{P}(A).$$

Luego estamos frente a una inconsistencia, dado que el miembro derecho vale $+\infty$ si $\mathbb{P}(A) > 0$, y 0 si $\mathbb{P}(A) = 0$. □

4.8. Recopilación de ejercicios

Ejercicio 4.46. ¿Es suficiente la condición (4.1) para definir una probabilidad en Ω_2?

Ejercicio 4.47. Pruebe por inducción, que las condiciones (4.1) y (4.2) definen una probabilidad en Ω_n.

Ejercicio 4.48. Probar que si $S_n(\omega)$ cuenta la cantidad de éxitos en cada resultado $\omega \in \Omega_n$, entonces se tiene

$$\mathbb{P}_n(\omega) = p^{S_n(\omega)} \cdot q^{n - S_n(\omega)}.$$

Ejercicio 4.49. Pruebe las propiedades anteriores. Para la última conviene observar que la descomposición (4.4) no es única, y en particular se pueden particionar los eventos $\{X = x_i\}$ para tener una expresión tipo $X = \sum_k y_k \mathbb{1}_{A_k}$.

Ejercicio 4.50. Sea f una función a valores reales, definida sobre la imagen de X. Considere la nueva variable aleatoria $f \circ X$. Pruebe que

$$\mathbb{E}(f(X)) = \sum_{i=1}^{k} f(x_i) \cdot \mathbb{P}(X = x_i). \tag{4.13}$$

Ejercicio 4.51. Dar una prueba de este resultado.
(Sugerencia $\mathbb{P}(X > a) = \sum_{i: x_i > a} \mathbb{P}(X = x_i)$.)

Ejercicio 4.52. Probar que en el modelo de tirar la moneda fiel, los conjuntos $A_1 = \{\omega_1 = 1\}$, $A_2 = \{\omega_2 = 1\}$, $A_3 = \{\omega_1 = \omega_2\}$ son dos a dos independientes, pero la familia $\{A_1, A_2, A_3\}$ no es una familia independientes.

Ejercicio 4.53. Probar que en el modelo definido en el Ejercicio 4.47 los eventos $A_i = \{\omega_i = 1\}$ son independientes.

[3] Formalmente acá estamos aplicando el Axioma de elección.

Ejercicio 4.54. Probar que las variables aletorias $\{X_1,\ldots,X_n\}$ son independientes, si y sólo si los eventos $\{X_1 \in B_1\},\ldots\{X_n \in B_n\}$ son independientes, para todo $B_1,\ldots,B_n$ subconjuntos de $\mathbb{R}$.

Ejercicio 4.55. Probar que una familia de eventos es independientes si lo es la familia de sus funciones características.

Ejercicio 4.56. Probar que las variables aletorias $\{X_1,\ldots,X_n\}$ son independientes, si y sólo los eventos $\{X_1 = x_1\} \cap \cdots \cap \{X_{j-1} = x_{j-1}\}$ es independiente de $\{X_j = x_j\}$, para cualesquiera $x_1,\ldots,x_j$ en $\mathbb{R}$.

Ejercicio 4.57. Dar una prueba de la proposición anterior.

Ejercicio 4.58. Probar el resultado anterior. La fórmula de esperanza y varianza resultan aplicaciones sencillas de la Proposición 4.19. Se sugiere como un lindo ejer2cicio de manipulación de sumas, probar ambas identidades utilizando las definiciones (4.3) (4.6) y manipulando (derivando) la expresión del binomio de newton $(a+b)^n$. (Para la varianza recordar expresión (4.13).

Ejercicio 4.59. Sean $X_1,\ldots,X_n$ una colección de v.a. i.i.d. con distribución Bernoulli de parámetro p. Sea S_n las sumas parciales.

- Probar que la función
$$B_n(p) = \mathbb{E}\left(f\left(\frac{S_n}{n}\right)\right),$$
es un polinomio en p de grado a lo sumo n.
- Usar la desigualdad de Chebyshev para probar que para todo $p \in [0,1]$, y todo $\varepsilon > 0$, se tiene
$$\sum_{k\in K} \binom{n}{k} p^k (1-p)^{n-k} \leqslant \frac{1}{4n\varepsilon^2},$$
donde $K = \{k \in \{0,1,\ldots,n\}, |k/n - p| > \varepsilon\}$.
- Utilizando la continuidad uniforme de f probar que
$$\lim_{n\to\infty} \sup_{0\leqslant p\leqslant 1} |f(p) - B_n(p)| = 0.$$

Ejercicio 4.60. Dar un ejemplo de v.a. donde $\mathbb{P}(X > a) = \mathbb{E}(X)/a$, para $X \geqslant 0$, y $t > 0$

Ejercicio 4.61. Realizar una simulación en R que aproxime numéricamente la probabilidad de que en 100 tiradas la cantidad de caras que salgan sean mayores a 60. Análogo, para 1000 tiradas, y que salgan más de 600 caras.

Ejercicio 4.62. Probar que la definición anterior no depende del $n(A)$ elegido.

Ejercicio 4.63. Probar la anterior proposición.

Ejercicio 4.64. Escribir la prueba formal para una palabra $w = (w_1,\ldots,w_j) \in \{0,1\}^j$ dada.

Ejercicio 4.65. Probar que casi seguramente $\omega \in \Omega$ contiene todas las palabras posibles. (De aquí la frase de que con probabilidad 1 un mono tipeando en una computadora tarde o temprano escribirá exactamente alguna obra famosa.)

Ejercicio 4.66. Probar que el conjunto formado por sucesiones periódicas, a partir de cierto índice, forman un conjunto de medida cero.

Parte III

Probabilidad via Teoría de la Medida

Capítulo 5

Conceptos Básicos

En este capítulo revisitaremos definciones básicas en teoría de probabilidades pero desde un punto de vista de la teoría de la medida.

5.1. σ-álgebra

Dado un conjunto Ω, recordemos que una familia no vacía de conjuntos $\mathcal{A} \subset \Omega$ es una $\underline{\sigma\text{-}\textit{álgebra}}$ si es cerrada por complementos y uniones numerables, esto es:

(i) si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$;

(ii) si $A_i \in \mathcal{A}$, es una colección *numerable* indexada por i, entonces $\bigcup_i A_i \in \mathcal{A}$.

Observar que de (i) y (ii) se desprende que las intersecciones numerables de elementos de una σ-álgebra, también están en la σ-álgebra.

Existen σ-álgebras muy sencillas como las partes de Ω (la σ-álgebra más grande), o la familia con dos elementos $\{\varnothing, \Omega\}$ (la σ-álgebra más pequeña). Pero en general hay otras en el medio más complejas y útiles como veremos.

Ejemplo 5.1. Consideremos el espacio $\Omega = \{a,b,c,d\}$, luego la familia $\mathcal{A} = \{\varnothing, \{a,b\}, \{c,d\}, \{a,b,c,d\}\}$ es una σ-álgebra.

Ejemplo 5.2. Si tenemos una partición dijsunta numerable $\Omega^{(1)}, \Omega^{(2)}, \ldots$ del espacio Ω, entonces la familia formada por todas las uniones posibles de $\Omega^{(k)}$, más el conjunto vacío, es una σ-álgebra. (Observar que el ejemplo anterior es un caso particular.)

Ejemplo 5.3. Consideremos un conjunto Ω arbitrario. Considermos $\mathcal{A}$ la famlila de conjuntos $A \subset \Omega$ tales que A es numerable, o que su complemento es numerable. Entonces $\mathcal{A}$ es una σ-álgebra.

Una σ-álgebra muy importante es la denominada *σ-álgebra de Borel* que es la σ-álgebra *generada* por los intervalos de $\mathbb{R}$. Esta familia se construye a partir de intervalos, realizando las operaciónes dadas en la definición de σ-álgebra. Para ver esto necesitamos entender la idea de σ-álgebra generada por una familia de subconjuntos. Este concepto es muy importante en la teoría por lo que le ponemos en un sección apartada.

5.1.1. σ-álgebra generada por una familia

La *σ-álgebra generada por una familia* $\mathcal{F}$ es la menor σ-álgebra que contiene la familia.

◇ **Ejer 5.4.** Observar que si $\mathcal{A}_\alpha$ es una colección arbitraria de σ-álgebras en Ω, entonces $\bigcap_\alpha \mathcal{A}_\alpha$ es también una σ-álgebra.

Esto ejercicio nos permite definir formalmente la σ-álgebra generada por la familia $\mathcal{F}$, que denotaremos por $\sigma(\mathcal{F})$, como la intersección de todas las σ-álgebra que la contienen, i.e. $\sigma(\mathcal{F}) := \bigcap_\alpha \mathcal{A}_\alpha$, siendo α un índice que varía por todas las σ-álgebra $\mathcal{A}_\alpha$ tales que $\mathcal{F} \subset \mathcal{A}_\alpha$. Al menos sabemos que $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, por lo tanto esa intersección es no vacía, y por el ejercicio es de hecho una σ-álgebra. Por otro lado si $\mathcal{A}$ es una σ-álgebra tal que $\mathcal{F} \subset \mathcal{A}$, entonces $\mathcal{A}$ es un miembro de la intersección y por lo tanto $\sigma(\mathcal{F}) \subset \mathcal{A}$. Con esto podemos concluir que si $\mathcal{F}$ es una familia arbitraria de conjuntos de Ω resulta:

1. $\sigma(\mathcal{F})$ es una σ-álgebra;
2. $\mathcal{F} \subset \sigma(\mathcal{F})$;
3. si $\mathcal{F} \subset \mathcal{A}$, y $\mathcal{A}$ es una σ-álgebra, entonces $\sigma(\mathcal{F}) \subset \mathcal{A}$.
4. Si $\mathcal{F}$ es una σ-álgebra, entonces $\sigma(\mathcal{F}) = \mathcal{F}$;
5. Si $\mathcal{F}' \subseteq \mathcal{F}$, entonces $\sigma(\mathcal{F}') \subseteq \sigma(\mathcal{F})$.

◇ **Ejer 5.5.** Probar que si tomamos la familia de subconjuntos de Ω formados por un elemento, entonces la σ-álgebra coincide con la del Ejemplo 5.3.

Ejemplo 5.6 (σ-álgebra de Borel)**.** Se define la *σ-álgebra de Borel* como la intersección de todas las σ-álgebra que contiene la familia de intervalos (a,b) con $a<b$, i.e.,

$$\mathcal{B}(\mathbb{R}) = \bigcap_\alpha \mathcal{A}_\alpha, \quad \mathcal{A}_\alpha \quad \sigma\text{-álgebra tal que} \quad \mathcal{I} \subset \mathcal{A}_\alpha,$$

siendo $\mathcal{I} = \{(a,b) : a,b \in \mathbb{R},\ a<b\}$.

La σ-álgebra de Borel es una familia muy rica de conjuntos, en particular tiene a todos los conjuntos abiertos. Esto resulta del hecho que si A es un conjunto abierto, para cada $x \in A$, existe un intervalito (a_x, b_x) con extremos racionales tales que $x \in (a_x, b_x) \subset A$, y por lo tanto $A = \bigcup_{x\in A}(a_x, b_x)$. Pero como solo hay una cantidad numerable de estos intervalitos, la unión anterior es de hecho una unión numerable.

♢ **Ejer 5.7.** La σ-álgebra de Borel también puede ser obtenida si cambiamos la familia $\mathcal{I}$ por la familia de intervalos semi-abiertos $(a,b]$, o intervalos cerrados $[a,b]$, o intervalos más generales tipo $(-\infty, b)$, o $(\infty, b]$, variando $a, b \in \mathbb{R}$ con $a < b$. Basta probar que cada uno de estos nuevos intervalos puede ser formado por elementos de $\mathcal{B}(\mathbb{R})$.

A modo de ejemplo, en el ejercicio anterior $(0,1] = \bigcap_{n\geqslant 1}(0, 1+1/n)$, y por lo tanto $(0,1] \in \mathcal{B}(\mathbb{R})$. También están los conjuntos de un elemento $\{2\} \in \mathcal{B}(\mathbb{R})$.

5.2. Medidas

Un *espacio medible* es un par $(\Omega, \mathcal{A})$, donde Ω es un conjunto, y $\mathcal{A}$ es una σ-álgebra de subconjuntos de Ω.

Una *medida* es un espacio medible $(\Omega, \mathcal{A})$ es una función $\mu : \mathcal{A} \to \mathbb{R}$ que cumple

(i) $\mu(A) \geqslant \mu(\varnothing) = 0$, para todo $A \in \mathcal{A}$;

(ii) si A_i es una colección numerable y disjunta de miembros de $\mathcal{A}$, entonces $\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$.

A la propiedad (ii) se le conoce como propiedad σ-aditiva. (En algunos textos, se defina una medida "finitamente aditiva" como aquella donde la propiedad (ii) se modifica por uniones finitas.)

A la terna $(\Omega, \mathcal{A}, \mu)$ se le llama *espacio de medida*. En el caso de que $\mu(\Omega) = 1$, decimos que μ es una *medida de probabilidad*, y en tal caso la denotaremos por $\mathbb{P}$, y decimos que $(\Omega, \mathcal{A}, \mathbb{P})$ es un *espacio de probabilidad*.

Veamos algunas propiedades.

Proposición 5.8. *Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Entonces se satisfacen las siguientes propiedades:*

1. *Si $A \in \mathcal{A}$, entonces $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*
2. *Si $A \subseteq B$ son subconjuntos de $\mathcal{A}$, entonces $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A) \geqslant 0$.*
3. *Sean A_i una colleción numerable en $\mathcal{A}$, entonces $\mathbb{P}(\bigcup_i A_i) \leqslant \sum_i \mathbb{P}(A_i)$.*
4. *Sean A_i una colleción numerable en $\mathcal{A}$ creciente (es decir $A_i \subseteq A_{i+1}$), entonces $\mathbb{P}(\bigcup_i A_i) = \lim_{i\to+\infty} \mathbb{P}(A_i)$.*
5. *Sean A_i una colleción numerable en $\mathcal{A}$ decreciente ($A_i \supseteq A_{i+1}$), entonces $\mathbb{P}(\bigcap_i A_i) = \lim_{i\to+\infty} \mathbb{P}(A_i)$.*

♢ **Ejer 5.9.** Dar una prueba de la proposición anterior. Para la parte 4, considerar la sucesón $B_i = A_i - A_{i-1}$, y observar que la colección de B_i es disjunta y su unión coincide con la unión de los A_i.)

Veamos algunos ejemplos de espacios de medida conocidos. El primero de ellos ya lo trabajamos en la primera parte del curso.

Ejemplo 5.10 (Espacios de probabilida discretos)**.** Consideremos un conjunto Ω numerable, y $\mathcal{A} = \mathcal{P}(\Omega)$ las partes de Ω. Luego definiendo $\mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$, donde $p : \Omega \to [0,1]$ tal que $\sum_{\omega\in\Omega} p(\omega) = 1$, resulta una medida en dicho espacio. Como ya vimos en la parte uno, si Ω es finito, y $p(\omega)$ es constante, entonces decimos que la medida inducida es la medida uniforme en Ω.

Ejemplo 5.11 (Medida de Lebesgue)**.** En el conjunto de los números reales $\mathbb{R}$, considerando la σ-álgebra de Borel, podemos considerar la *medida de Lebesgue* λ, como la *única* medida que satisface $\lambda((a,b]) = b - a$, para todo $a < b$. Observar que $\lambda(\mathbb{R}) = +\infty$. Para obtener un espacio de probabilidad podemos considerar $\Omega = (0,1]$, y $\mathcal{A}$ y $\mathbb{P}$ como las restricciones al intervalo, esto es

$$\mathcal{A} := \mathcal{B}((0,1]) := \{B \cap (0,1] : B \in \mathcal{B}(\mathbb{R})\} \subset \mathcal{B}(\mathbb{R}),$$

y $\mathbb{P}(A) = \lambda(A)$ para todo $A \in \mathcal{A}$. restricción.

Para probar la existencia de la medida de Lebesgue, la cual es análogo a la existencia de una medida de probabilidad en $\Omega = \{0,1\}^{\mathbb{N}}$, es necesario un resultado profundo de la teoría de la medida, lo cual contaremos brevemente.

5.2.1. Construcción de medidas: *Teorema de Carathedory**

Un álgebra es una familia de conjuntos que es cerrada por complementos y uniones finitas. Una *medida en un álgebra* $\mathcal{A}$, es una función $\mu : \mathcal{A} \to \mathbb{R}$ que satisface:

- $\mu(A) \geqslant \mu(\varnothing) = 0$, para todo $A \in \mathcal{A}$;
- si A_i es una colección numerable y disjunta de miembros de $\mathcal{A}$, tal que $\bigcup_i A_i$ también está en $\mathcal{A}$, entonces se tiene $\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$.

Se puede probar una proposición similar a la Proposición 5.8, que omitiremos la prueba.

Proposición 5.12. *Sea μ una medida en un álgebra $\mathcal{A}$. Entonces se satisfacen las siguientes propiedades:*

1. *Si $A \subseteq B$ son subconjuntos de $\mathcal{A}$, entonces $\mu(B-A) = \mu(B) - \mu(A) \geqslant 0$.*
2. *Sean A_i una colleción numerable en $\mathcal{A}$, tales que $\bigcup_i A_i \in \mathcal{A}$, entonces $\mu(\bigcup_i A_i) \leqslant \sum_i \mu(A_i)$.*
3. *Sean A_i una colleción numerable en $\mathcal{A}$ creciente tales que $\bigcup_i A_i \in \mathcal{A}$, entonces $\mu(\bigcup_i A_i) = \lim_{i\to+\infty} \mu(A_i)$.*
4. *Sean A_i una colleción numerable en $\mathcal{A}$ decreciente, tal que $\bigcup_i A_i \in \mathcal{A}$, y además $\mu(A_1) < \infty$, entonces $\mu(\bigcap_i A_i) = \lim_{i\to+\infty} \mu(A_i)$.* □

Dada una medida μ en un álgebra $\mathcal{A}$ de subconjuntos de Ω, decimos que μ es *σ-finita*, si podemos partir el espacio Ω en una cantidad numerable de conjuntos disjuntos Ω_i, tales que $\mu(\Omega_i) < \infty$.

El siguiente resultado es una pieza fundamental en la construcción de medidas en σ-álgebras.

Teorema 5.13 (Carathedory) *Sea $\mathcal{A}$ un álgebra, y μ una medida σ-finita sobre $\mathcal{A}$. Entonces μ tiene una (única) extensión a la σ-álgebra generada por $\mathcal{A}$.* □

La prueba de este teorema se da en cursos de teoría de la medida, y acá brevemente comentamos como realizar la construcción de la *medida de Lebesgue* utilizando esste teorema.

Construcción de la medida de Lebesgue

Primero se define lo que es un *semi–anillo* que básicamente modela las propiedades de los rectángulos dentro del cuadrado unidad. Esto es, una familia que es cerrada por intersecciones de dos elementos, y que el complemento, es una unión finita disjunta de elementos de la familia.

Ejemplo 5.14. En $\Omega = (0,1]$, el conjunto $\mathcal{I}$ formado por intervalos $(a,b]$, con $0 \leqslant a < b \leqslant 1$, forman un semi–anillo.

Ejemplo 5.15. Más en general, la familia de subconjuntos del plano,

$$\mathcal{R} = \{(a_1,b_1] \times (a_2,b_2] : -\infty \leqslant a_1 < b_1 < +\infty, -\infty \leqslant a_2 < b_2 < +\infty 1\}.$$

forman un semi–anillo.

Se puede probar, sin mucha dificultad, que en los ejemplos anteriores la σ-álgebra generada por los semi–anillos coincide con la σ-álgebra de Borel en cada caso (basta utilizar las propiedad vistas de σ-álgebra generada).

El siguiente lema nos dice cómo es el álgebra generada por un semi-anillo.

Lema 5.16. *Sea $\mathcal{S}$ un semi-anillo. Entonces $\overline{\mathcal{S}} = \{$uniones finitas disjuntas de elementos de $\mathcal{S}\}$ es un álgebra, y se denomina* álgebra generada por $\mathcal{S}$.

Demostración. Para escribir poco, sean $A = \bigsqcup_i S_i$, y $B = \bigsqcup_j R_j$ elementos genéricos de $\overline{S}$ (uniones finitas disjuntas). Entonces $A \cap B = \bigsqcup_{i,j} S_i \cap S_j$, y por lo tanto también un elemento de $\overline{S}$. Además $A^c = \bigcap_i S_i^c$, y como S_i^c es una unión de elementos de $\mathcal{S}$, se tiene que $S_i^c \in \overline{S}$ y por la parte anterior se termina la prueba. □

Prosigamos con la construcción de la medida de Lebesgue en $(0,1]$. Sea $\mathcal{I}$ el semi–anillo de intevalos dado en el Ejemplo 5.14.

Observar que si definimos $\lambda : \mathcal{I} \to \mathbb{R}$ como la función de conjuntos que da la longitud de los intervalos, i.e.

$$\lambda((a,b]) := |(a,b]| = b - a, \qquad (0 \leqslant a < b \leqslant 1),$$

resulta muy simple de extenderla al álgebra generada $\overline{\mathcal{I}}$ como

$$\lambda(\bigsqcup_{i=1}^{n}(a_i,b_i]) = \sum_{i=1}^{n}(b_i - a_i). \tag{5.1}$$

Observar que la definición de $\lambda : \overline{\mathcal{I}} \to \mathbb{R}$ es consistente, en el sentido de que si un conjunto A se escribe de dos maneras diferentes como union finita disjunta de miembros de $\mathcal{I}$, entonces $\lambda(A)$ coincide en ambas representaciones. Esto es, si $\bigsqcup_{i=1}^n I_i = \bigsqcup_{j=1}^m J_j$, con $I_i, J_j \in \mathcal{I}$, entonces

$$\sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n}\sum_{j=1}^{m} |I_i \cap J_j| = \sum_{j=1}^{m} |J_j|$$

Faltaría ver que la función de conjuntos así definida, satisface la σ-aditividad de la definición de medida en un álgebra. Una vez probada esa propiedad podemos concluir del Teorema de Carathedory 5.13 la existencia y unicidad de la medida de Lebesgue λ definida en la σ-álgebra de Borel en $(0,1]$, que extiende a la medida en $\overline{\mathcal{I}}$ dada en (5.1)
.

Esto requiere de un lema técnico que incluimos para cerrar esta sección

Lema 5.17. *Consideremos $(0,1]$ y el semianillo $\mathcal{I}$. Todos los conjuntos I_k e I dados a continuación son miembros de $\mathcal{I}$, y utilizamos la notación $\bigsqcup$ para unión disjunta. Se tiene,*

1. *Si $\bigsqcup_k I_k \subset I$, entonces $\sum_k |I_k| \leqslant |I|$;*
2. *Si $I \subset \bigcup_k I_k$, entonces $|I| \leqslant \sum_k |I_k|$.*
3. *Si $I = \bigsqcup_k I_k$, entonces $|I| = \sum_k |I_k|$.*

Observar que la propiedad 3, que es la que estamos buscando, resulta inmediata de 1 y 2.

Demostración. Sea $I = (a,b]$.

1. (Caso finito:) Supongamos que $\bigsqcup_{k=1}^{n}(a_k,b_k] \subset (a,b]$, donde los intervalos los ordenamos de izquierda a derecha. Luego $a \leqslant a_1$, y $b_n \leqslant b$, y $b_k \leqslant a_{k+1}$, para $k = 1,\ldots,n-1$. Por lo tanto $\sum_{k=1}^n |(a_k,b_k]| = \sum_{k=1}^n (b_k - a_k) \leqslant b_n - a_1 \leqslant b - a = |(a,b]|$.

   (Caso infinito:) Dado que $\bigsqcup_{k=1}^{\infty} I_k \subset I$, también es valido para la unión truncada $\bigsqcup_{k=1}^{\infty} I_k \subset I$, y por la parte anterior se tiene $\sum_{k=1}^{n} |I_k| \leqslant |I|$, para cualquier $n \in \mathbb{N}$. Luego tomando límite a la izquierda, el resultado sigue.

2. (Caso finito:) Por inducción, observar que si $(a,b] \subset (a_1,b_1]$, el resultado es trivial. Supongamos que vale para n intervalos, i.e., si $(a,b] \subset \bigcup_{k=1}^n (a_k,b_k]$, entonces $b-a \leqslant \sum_{k=1}^n (b_k - a_k)$. Supongamos que $(a,b] \subset \bigcup_{k=1}^{n+1} (a_k,b_k]$. Ordenamos según $a_1 \leqslant \cdots \leqslant a_{n+1}$. Si $a_{n+1} \geqslant b$, entonces el intervalo $(a,b]$ puede ser cubierto con los primeros n intervalos y el resultado sigue de la hipótesis de inducción. Así que podemos suponer $a_{n+1} < b < b_{n+1}$. Luego tenemos que $(a,a_{n+1}] \subset \bigcup_{k=1}^n (a_k,b_k]$, y utilzando la hipótesis de inducción resulta $a_{n+1} - a \leqslant \sum_{k=1}^n (b_k - a_k)$. Concluimos entonces

$$b-a = (b-a_{n+1}) + (a_{n+1}-a) \leqslant (b_{n+1}-a_{n+1}) + (a_{n+1}-a) \leqslant \sum_{k=1}^{n+1} (b_k - a_k).$$

(Caso infinito:) Supongamos que $(a,b] \subset \bigcup_{k=1}^\infty (a_k,b_k]$. Acá usaremos un truco y un teorema importante sobre compactos. El primer truco es achicar el I para que quede compacto, y agrandar un poquito I_k para que queden abiertos. Por ejemplo, consideramos el intervalo compacto $[a+\varepsilon,b]$, con $0<\varepsilon<b-a$, y los intervalos abiertos $\{(a_k,b_k+\varepsilon/2^k)\}_{k=1,\dots}$, resulta

$$[a+\varepsilon,b] \subset \bigcup_{k=1}^\infty (a_k,b_k+\frac{\varepsilon}{2^k}).$$

Luego por la compacidad del primer intervalo, el Teorema de Heine-Borel nos dice dicho cubrimiento pose un subcubrimento finito i.e. $[a+\varepsilon,b] \subset \bigcup_{j=1}^\ell (a_j,b_j+\varepsilon/2^j)$, donde reordenamos los índices. Luego, para caer en el caso finito anterior, también tenemos $(a+\varepsilon,b] \subset \bigcup_{j=1}^\ell (a_j,b_j+\varepsilon/2^j]$, y por el caso finito tenemos

$$b-(a+\varepsilon) \leqslant \sum_{j=1}^\ell (b_j-a_j)+\varepsilon \leqslant \sum_{j=1}^\infty (b_j-a_j)+\varepsilon.$$

Es decir $|I| \leqslant \sum_{k=1}^\infty |I_k| + 2\varepsilon$. Como el ε es arbitrario, el resultado sigue.

□

Proposición 5.18. *La función $\lambda : \overline{\mathcal{I}} \to \mathbb{R}$ dada en (5.1), es σ-aditiva y por tanto es una medida en el álgebra.*

Demostración. Hay que probar que si tenemos una sucesión de conjuntos disjuntos $A_n \in \overline{\mathcal{I}}$, y además $\bigsqcup_n A_n \in \overline{\mathcal{I}}$, entonces se satisface

$$\lambda\left(\bigsqcup_n A_n\right) = \sum_{n\in\mathbb{N}} \lambda(A_n).$$

Como $A := \bigsqcup_n A_n \in \overline{\mathcal{I}}$, tenemos que $A = \bigsqcup_{j=1}^m J_j$, para $J_j \in \overline{\mathcal{I}}$. Por otro lado, como $A_n \in \overline{\mathcal{I}}$ se tiene $A_n = \bigsqcup_{i=1}^{m_n} I_i^{(n)}$, donde los J_j y $I_i^{(n)}$ son intervalos. Observar que como $J_j \subset A$, entonces

$$J_j = \bigsqcup_{n=1}^\infty J_j \cap A_n = \bigsqcup_{n=1}^\infty \bigsqcup_{i=1}^{m_n} J_j \cap I_i^{(n)}, \qquad (j=1,\dots,m). \tag{5.2}$$

Luego del Lema 5.17 tenemos que

$$\lambda(J_j) = \sum_{n=1}^\infty \sum_{i=1}^{m_n} \lambda(J_j \cap I_i^{(n)}) \qquad (j=1,\dots,m). \tag{5.3}$$

De (5.3), y manipulación de series, concluimos

$$\begin{aligned}\lambda\left(\bigsqcup_n A_n\right) &= \lambda\left(\bigsqcup_{j=1}^m J_j\right)\\ &= \sum_{j=1}^m \lambda(J_j)\\ &= \sum_{j=1}^m \sum_{n=1}^\infty \sum_{i=1}^{m_n} \lambda(J_j \cap I_i^{(n)})\\ &= \sum_{n=1}^\infty \sum_{j=1}^m \sum_{i=1}^{m_n} \lambda(J_j \cap I_i^{(n)})\\ &= \sum_{n=1}^\infty \sum_{j=1}^m \lambda(J_j \cap A_n) = \sum_{n=1}^\infty \lambda(A_n).\end{aligned}$$

□

Terminada esta digresión sobre la construcción de la medida a partir de una medida en un álgebra, continuamos con más ejemplos de espacio de medida.

El siguiente es una manera simle de fabricar nuevos ejemplos, y que será fundamental para estudiar la noción de *independencia* en la teoría de probabilidades.

Ejemplo 5.19 (Espacios productos)**.** Si consideremos los espacios de probabilidad $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $(i=1,\dots,n)$, podemos definir un nuevo espacio, denominado *espacio producto* dado por $(\Omega, \mathcal{A}, \mathbb{P})$, donde

- $\Omega = \Omega_1 \times \cdots \times \Omega_n$,
- $\mathcal{A}$ es el la σ-álgebra generada por el semianillo de "rectángulos" $A_1 \times \cdots \times A_n$, donde $A_i \in \mathcal{A}_i$
- $\mathbb{P}$ es la medida *producto* que es la única que satisface

$$\mathbb{P}(A_1 \times \cdots \times A_n) = \mathbb{P}_1(A_1) \cdots \mathbb{P}_n(A_n), \qquad \text{donde} \quad A_i \in \mathcal{A}_i.$$

A esta medida la denotamos $\mathbb{P} := \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$.

Observar que en el ejemplo anterior la existencia de esa medida utiliza nuevamente el Teorema de Carathedory 5.13 y sigue los pasos dados en la construcción de la medida de Lebesgue. Esto es, se prueba que la familia de "rectángulos"

$$A_1 \times \cdots \times A_n, \qquad A_i \in \mathcal{A}_i,$$

forma un semianillo. Luego se define la medida $\mathbb{P}$ sobre estos, y uniones finitas de estos elementos. Se prueba que esa medida es σ-aditiva, para luego concluir del Teorema de Carathedory 5.13 la existencia y unicidad de la medida producto.

Un ejemplo particular e importante es el de la medida producto en $\mathbb{R}^n$.

Ejemplo 5.20. Se define en $\mathbb{R}^n$, el espacio producto $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda_n)$, donde $\lambda_n = \otimes_{j=1}^n \lambda$ es la *medida de Lebesgue* en $\mathbb{R}^n$. En particular la medida de un rectángulo

$$\lambda_n\left(\prod_{j=1}^n [a_i,b_i]\right) = \prod_{j=1}^n (b_i - a_i).$$

Otra forma alternativa de obtener la medida de Lebesgue, es via la construcción realizada en la Sección 5.2.1. La razón de que son la misma medida es una consecuencia de la unicidad que prueba el Teorema de Carathedory.

Una vez construida la medida de Lebesgue en $\mathbb{R}^n$, podemos definir la medida uniforme sobre cualquier Boreliano $A \in \mathcal{B}(\mathbb{R}^n)$ que tenga medida finita, simplemente, considerando la σ-álgebra relativa a A, y definir la medida

$$\lambda_A = \frac{1}{\lambda(A)}\lambda.$$

En particular podemos considerar la medida uniforme en el disco unidad

$$\mathbb{D} = \{(x,y) : x^2+y^2 < 1\}, \qquad \lambda_{\mathbb{D}} = \frac{1}{\pi}\lambda_2.$$

5.3. Variables Aleatorias

La probabilidad se torna más interesante cuando se introduce el concepto de variable aleatoria.

5.3.1. Definiciones y ejemplos

Dada un espacio medible $(\Omega,\mathcal{A})$, una *variable aleatoria* es una función

$$X : \Omega \to \mathbb{R}, \quad \text{tal que} \quad X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A},$$

para cualquier $B \in \mathcal{B}(\mathbb{R})$, boreliano de $\mathbb{R}$.

La idea de esta definición es poder "medir", i.e. asignarle un valor numérico a que la variable aleatoria tome ciertos valores (a saber, valores en un boreliano).

Ejemplo 5.21. El ejemplo más sencillo de v.a. son las funciones indicatrices de conjuntos en la σ-álgebra. Dado $A \in \mathcal{A}$, se definine $\mathbb{1}_A : \Omega \to \mathbb{R}$ a aquella función que toma el valor 1 si $\omega \in A$, y 0 si $\omega \notin A$.

Definición 5.22. Una v.a. X en un espacio de probabilidad $(\Omega,\mathcal{A})$, induce una medida de probabilidad en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, denominada *distribución de* X de la siguiente manera

$$\mu_X(B) := \mathbb{P}(X \in B), \qquad B \in \mathcal{B}(\mathbb{R})$$

donde entendemos por $\{X \in B\} := X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$. (De manera análoga se puede definir la distribución de un vector aleatorio, siendo esta la medida de probabildad inducida en $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$.)indexdistribución!variable aleatoria

◊ **Ejer 5.23.** Probar que que la distribución de una v.a. es una medida en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

◊ **Ejer 5.24.** Hallar la distribución de una función indicatriz. Observar que la medida está concentrada en dos valores únicamente.

Una v.a. X induce una medida sobre $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, que denominamos distribución de la v.a. X. Otro concepto importante en la teoría el de *función de distribución* asociado a una variable aleatoria. Esto es,

$$F(x) = \mu((-\infty,x]) = \mathbb{P}(X \leqslant x), \qquad x \in \mathbb{R}.$$

Una función de distribución de una v.a. verifica ciertas propiedades que las detallamos a continuación.

Proposición 5.25. *Si F es una función de distribución de una v.a., entonces satisface:*

1. *F es creciente (no necesariamente estricta);*
2. $\lim_{x\to+\infty} F(x) = 1$, *y* $\lim_{x\to-\infty} F(x) = 0$*;*
3. *F es continua por derecha, i.e.,* $\lim_{y\downarrow x} F(y) = f(x)$.

◊ **Ejer 5.26.** Dar una prueba de este resultado.

El siguiente resultado nos dice que las propiedades anteriores alcanzan para determinar una funcion de distribución

Proposición 5.27. *Si F es una función que satisface las propiedades 1. 2. y 3. de la Proposición 5.25, entonces es la distribución de alguna variable aleatoria.* □

Omitimos la prueba de este resultado.

Corolario 5.28. *Si F es una función que satisface las propiedades 1. 2. y 3. de la Proposición 5.25, entonces existe una única medida μ en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ tal que $\mu((a,b]) = F(b)-F(a)$, para $a < b$.*

Demostración. De la Proposición 5.27 sabemos que existe una v.a. X con función de distribución F. Además X induce una distribución μ_X en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, que en particular satisface $\mu_X((a,b]) = F(b)-F(a)$, para cualesquiera $a < b$. Luego como lo intervalos $(a,b]$ generan la σ-álgebra, resulta de Carathedory que $\mu = \mu_X$. □

Hay una famlila importante de funciones de distribución $F(x) = \mathbb{P}(X \leqslant x)$ que satisfacen

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

En este caso decimos que la v.a. X tiene *densidad* f.

Observar que si f es una función "bien comportada", digamos continua, entonces

$$f(x) \approx \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt = \frac{F(x+\Delta x)-F(x)}{\Delta x} = \frac{1}{\Delta x}\mathbb{P}(X \in (x,x+\Delta x])r,$$

por lo que $f(x)$ puede ser pensada como la probabilidad de que $X \in (x,x+\Delta x)$ normalizado por Δx.

Recordemos algunos ejemplos del curso de Probabilidad.

Ejemplo 5.29 (Distribución uniforme en $(0,1)$)**.** Si consideramos la función $f(x) = 1$, para $x \in (0,1)$, y 0 de lo contrario. Luego la función de distribución

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

es la distribución de una v.a. uniforme en $(0,1)$.

Comentario 5.30. Cuando decimos que X es una variable aleatoria con distribución uniforme en $[0,1]$, lo que formalmente estamos diciendo es que tenemos una variable aleatoria $X : \Omega \to [0,1]$, definida en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, tal que $\mathbb{P}(X \in B) = \lambda(B)$, siendo λ la medida de Lebsegue en $[0,1]$, y B cualquier boreliano del interrvalo $[0,1]$. Muchas veces, y de manera un poco tautológica, especializamos este ejemplo en considerar $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$, y $X : [0,1] \to [0,1]$ como la función identidad, por lo que el conjunto $\{\omega \in [0,1] : X(\omega) \in (a,b]\} = (a,b]$. Dicho esto, lo que realmente interesa de una variable aleatoria es cómo se "distribuye" en $\mathbb{R}$, i.e., conocer su distribución.

"

Ejemplo 5.31 (Distribución uniforme en (a,b))**.** Sea

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ si } x \in [a,b] \\ 0 \text{ en otro caso} \end{cases}$$

de donde resulta $F = \begin{cases} 0 \text{ si } x < a \\ \frac{x-a}{b-a} \text{ si } a \leqslant x \leqslant b \\ 1 \text{ si } x > b \end{cases}$. En este caso diremos que una v.a. con esta función de distribución es uniforme $[a,b]$.

Ejemplo 5.32 (*Exponencial*)**.** Diremos que X es una v.a. exponencial de parámetro ("tasa") λ ($X \sim Exp(\lambda)$) si:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} \text{ si } x \geqslant 0 \\ 0 \text{ si } x < 0 \end{cases}$$

◇ **Ejer 5.33.** Probar que $F_X(x) = \begin{cases} 1 - e^{-\lambda x} \text{ si } x \geqslant 0 \\ 0 \text{ si } x < 0 \end{cases}$.

Otro ejemplo de suma importancia es la distribución gaussiana.

Ejemplo 5.34. La distribución gaussiana es la que tiene densidad

$$f(x) = \frac{1}{2\pi} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Hay otras distribuciones, que no tienen funciones densidad.

Ejemplo 5.35. La función $F(x) = 1$, si $x \geqslant 0$, y 0 si $x < 0$.

5.3.2. Funciones Medibles

El concepto de variable aleatoria es un caso de particular de función medible. Una función $f : (\Omega, \mathcal{A}) \to (S, \mathcal{S})$ definida entre dos espacios medibles se dice *función medible* si

$$\forall B \in \mathcal{S}, \text{ se tiene que } f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A},$$

i.e. si f trae conjuntos de la σ-álgebra del codominio, en conjuntos de la σ-álgebra del dominio.

Cuando el codominio es $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, una función medible se le dice *variable aleatoria*, y cuando codominio es $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, le decimos *vector aleatorio*.

¿Cómo saber si cierta función es medible? El siguiente resultado, de mucha utilidad nos dice que si tenemos un generador de la σ-álgebra del codominio, entonces basta verificar que la función f trae esos conjuntos en la σ-álgebra del dominio. En particular para el caso de variables aleatorias, basta probar que $f^{-1}((a,b])$ son conjuntos de la σ-álgebra $\mathcal{A}$.

Proposición 5.36. *Sea $f : (\Omega, \mathcal{A}) \to (S, \mathcal{S})$ una función. Supongamos que la σ-álgebra $\mathcal{S}$ es generada por $\mathcal{F}$, i.e. $\sigma(\mathcal{F}) = \mathcal{S}$. Si $f^{-1}(F) \in \mathcal{A}$, para todo $F \in \mathcal{F}$, entonces f es medible.*

Demostración. Observar que la inversa se comporta bien con complementos y uniones, i.e.

$$f^{-1}(B^c) = (f^{-1}(B))^c$$
$$f^{-1}\left(\bigcup_n B_n\right) = \bigcup_n f^{-1}(B_n).$$

Luego si consideremos la familia de conjuntos

$$\mathcal{S}_f := \{B : f^{-1}(B) \in \mathcal{A}\},$$

se tiene de las propiedades de la inversa que $\mathcal{S}_f$ es una σ-álgebra, y además contiene a $\mathcal{F}$. Por lo tanto contiene a $\sigma(\mathcal{F}) = \mathcal{S}$, i.e. todo preimagen de $\mathcal{S}$ está en $\mathcal{A}$. □

Como corolario, tenemos lo que ya adelantamos.

Ejemplo 5.37. Una función $X : (\Omega, \mathcal{A}) \to \mathbb{R}$ que verifica que $X^{-1}((a,b) \in \mathcal{A}$, para cualquier $a < b$, entonces es una variable aleatoria.

Ejemplo 5.38. Sea $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ una función continua. Como la preimagen de un abierto, es abierta, resulta que f es medible, (i.e. una variable aleatoria con dominio $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

◇ **Ejer 5.39.** Pruebe que composición de funciones medibles es medible. Esto es, si $X : (\Omega, \mathcal{A}) \to (S, \mathcal{S})$ y $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ son funciones medibles, entonces $f(X) : (\Omega, \mathcal{A}) \to (T, \mathcal{T})$, es medible.

Observar que del Ejercicio 5.39 y el Ejemplo 5.38 se concluye que si X es una v.a., entonces cX, $\sin(X)$, X^2 son todas variables aleatorias.

Y más en general se tiene.

Proposición 5.40. *Si $X_1,\dots,X_n$ son v.a. en $(\Omega,\mathcal{A})$, y $f : (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n)) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ es medible, entonces $f(X_1,\dots,X_n)$ es una variable aleatoria.*

Demostración. Basta ver que la función $(X_1,\dots,X_n) : (\Omega,\mathcal{A}) \to (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ es medible. Para eso, utilizando la Proposición 5.36, basta ver que la preimagen de rectangulos es medible. Esto resulta de que

$$(X_1,\dots,X_n)^{-1}(A_1\times\cdots\times A_n) = \{\omega\in\Omega : X_i(\omega)\in A_i\} = \bigcap_{i=1}^{n}\{X_i\in A_i\}$$

es una intersección de elementos de $\mathcal{A}$ y por tanto en $\mathcal{A}$. □

Corolario 5.41. *Si $X_1,\dots,X_n$ son variables aleatorias, entonces $X_1+\cdots+X_n$ es una variable aletaria.*

De esta manera podemos construir muchas de variables aleatorias de interés.

5.3.3. Convergencia casi segura y en probabilidad

Dado que este es un curso de teoremas límites nos interesaría estudiar el comportamiento del $\lim_n X_n$ de v.a.

Proposición 5.42. *Supongamos que tenemos una familia de variables aleatorias $X_1,X_2,\dots$. Entonces las funciones*

$$\inf_n X_n(\omega) \quad \sup_n X_n(\omega) \quad \liminf_n X_n(\omega) \quad \limsup_n X_n(\omega)$$

donde $\omega\in\Omega$, son variables aleatorias.[1]

Demostración. Dado $a\in\mathbb{R}$, observar que fijado $\omega\in\Omega$, si $\inf X_n(\omega) < a$ entonces algún $X_n(\omega) < a$. Por lo tanto

$$\{\inf_n X_n < a\} = \bigcup_n \{X_n < a\}.$$

Luego como cada $\{X_n < a\}\in\mathcal{A}$ se tiene $\{\inf_n X_n < a\}$. Se deja como prueba las otras partes. □

◇ **Ejer 5.43.** Finalizar prueba de la proposición anterior.

Dada una sucesión de variable aleatoria $X_1,X_2,\dots$, consideremos el subconjunto $\Omega_0\subset\Omega$ donde existe el límite de X_n, esto es

$$\Omega_0 := \{\omega\in\Omega : \lim_n X_n(\omega) \text{ existe}\} = \{\omega\in\Omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega)\}.$$

Luego Ω_0 es la preimagen de 0 de una variable aleatoria, a saber $\liminf_n X_n - \limsup_n X_n$, y por lo tanto $\Omega_0\in\mathcal{A}$.

[1] Se recuerda que, dada una sucesión $\{x_n\}_n$, el límite inferior y superior de x_n es el menor y el mayor de los puntos de acumulación. Alternativamente se puede definir: $\liminf x_n = \sup_n \inf_{m\geqslant n} x_m$, y $\limsup x_n = \inf_n \sup m \geqslant n x_m$.

Definición 5.44. Decimos que la sucesión de variables aleatorias $\{X_n\}$ *converge casi seguro* si $\mathbb{P}(\{\omega\in\Omega : \lim_n X_n(\omega) \text{ existe}\}) = 1$.

Análogamente, diremos que la sucesión de variables aleatorias $\{X_n\}$ converge casi seguro a una variable aleatoria Y, y escribimos

$$X_n \to Y \text{ c.s.},$$

cuando $\mathbb{P}(\lim_n X_n = Y) = 1$. En particular la función $Y = \lim_n X_n$ es una variable aleatoria.[2]

En un espacio de medida $(\Omega,\mathcal{F},\mu)$, este tipo de convergencia se dice "μ *en casi todo punto*", y se escribe "μ-ctp". Observar que es un concepto más débil que la convergencia puntual conocida en el espacio de funciones.

Ejemplo 5.45. Si consideramos el espacio $[0,1]$ con la medida de Lebesgue, tenemos que la sucesión de funciones x^n (que son funciones medibles por ser continuas) convergen a la función 0 casi seguro, pero no así puntualmente (en $x = 1$ la sucesión toma el valor 1). Pero este punto es insignificante para la medida de Lebesgue, y por tanto hay convergencia casi segura, pero no puntual.

Un ejemplo más débil, ya visitado anteriormente, es el de convergencia en medida o probabilidad.

Definición 5.46. Decimos que la sucesión de variables aleatorias $\{X_n\}$ *converge en probabilidad* a una variable aleatoria X, y escribimos $X_n \xrightarrow{\mathbb{P}} X$, si para todo $\varepsilon > 0$ se tiene

$$\lim_{n\to+\infty} \mathbb{P}(\{\omega\in\Omega : |X_n(\omega)-X(\omega)|\geqslant\varepsilon\}) = 0.$$

Si uno cambia las variables aleatorias por funciones medibles en un espacio de medida, y $\mathbb{P}$ por la medida en dicho espacio, entonces utilizamos el término *convegencia en medida*.

Ejemplo 5.47. Consideremos el espacio de probabilidad de la medida uniforme en $[0,1]$, y consideremos la sucesión de intervalos definidos como a continuación: $I_1 = [0,1/2], I_2 = [1/2,1], I_3 = [0,1/3], I_4 = [1/3,2/3], I_5 = [2/3,1],\dots$. Luego definiendo las variables aleatorias $X_n = \mathbb{1}_{I_n}$ en nuestro espacio, es fácil ver que tomando $\varepsilon\in(0,1)$, se tiene que $\mathbb{P}\{|X_n|\geqslant\varepsilon\} = \lambda(I_n) = 1/n \xrightarrow{n} 0$, y por lo tanto $X_n \xrightarrow{\mathbb{P}} 0$.

El siguiente ejercicio muestra que el límite de la convergencia en probabilidad único (a menos de conjuntos de probabilidad 1).

[2] Aquí hay un detalle, que formalmente la función límite $\lim_n X_n$ está definida en Ω_0 y no en Ω. Formalmente, como Ω_0 tiene medida cero, podemos extenderla a $\lim_n X_n$ en un conjunto de medida cero como querramos. En el fondo está que en teoría de la medida, los objetos están definidos a menos de un conjunto de medida cero.

◇ **Ejer 5.48.** Probar si $\{X_n\}$ es una sucesión de variables aleatorias, tales que $X_n \xrightarrow{\mathbb{P}} X$ y $X_n \xrightarrow{\mathbb{P}} Y$, entonces $\mathbb{P}(X = Y) = 1$.

5.3.4. Algunas equivalencias de convergencia

La convergencia en medida es más débil que la convergencia casi segura como lo muestra el siguiente ejercicio.

◇ **Ejer 5.49.** Probar que la sucesión de variables aleatorias definidas en el Ejemplo 5.47 no converge de manear casi segura.

Sin embargo la convegencia casi segura implica la convergencia en medida.

Proposición 5.50. *Si $X_n \to X$ c.s., entonces $X_n \to X$ en probabilidad.*

Demostración. Observar que dado $\varepsilon > 0$, el conjunto

$$\{|X_n - X| \geqslant \varepsilon\} \subset \bigcup_{m \geqslant n} \{|X_m - X| \geqslant \varepsilon\},$$

y por otro lado sabemos que $\bigcap_{n=1}^{\infty} \bigcup_{m \geqslant n} \{|X_m - X| \geqslant \varepsilon\}$ son puntos de no convergencia y por lo tanto

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geqslant n} \{|X_m - X| \geqslant \varepsilon\}\right) = 0.$$

Luego de la propiedad de la continuidad por intersecciones de la medida de probabilidad, resulta

$$\mathbb{P}(\{|X_n - X| \geqslant \varepsilon\}) \leqslant \mathbb{P}\left(\bigcup_{m \geqslant n} \{|X_m - X| \geqslant \varepsilon\}\right) \xrightarrow{n \to +\infty} 0.$$

□

Demostración. □

Proposición 5.51. *La sucesión de variables aleatorias $\{X_n\}$ satisface*

$$X_n \xrightarrow{n} X \text{ c.s.} \quad \text{si y sólo si} \quad \lim_N \mathbb{P}\left(\bigcap_{n \geq N} |X_n - X| < \varepsilon\right) = 1.$$

□

◇ **Ejer 5.52.** Dar una prueba de esta proposición.

5.4. Integral y Esperanza

En el curso básico de Probabilidad, y en la primera parte de este curso (ver definición (4.3)), hemos visto la noción de valor esperado o esperanza de variables aleatorias discretas o continuas. Esto es

$$\mathbb{E}(X) = \begin{cases} \sum_n x_n \mathbb{P}(X = x_n) & \text{si } X \text{ es discreta} \\ \int x f(x)\, dx & \text{si } X \text{ tiene densidad } f \end{cases}$$

Recordar que X es discreta si $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ es numerable.

En el caso de infinito numerable hay que pedir que $\sum_n |x_n| \mathbb{P}(X = x_n) < \infty$, para que la serie sea absolutamente convergente.

Estas nociones pueden ser unificadas con teoría de la medida, y es lo que haremos a continuación sin entrar en muchos detalles (los cuales pueden encontrar en cualquier libro de teoría de la medida).

Los siguientes pasos pueden ser realizados para cualquier espacio de medida $(\Omega, \mathcal{A}, \mu)$, pero por simplicidad expositiva nos restringimos al caso de espacios de medida finita.

Consideremos un espacio de medida $(\Omega, \mathcal{A}, \mu)$, con $\mu(\Omega) < \infty$. Los pasos para definir la integral de una función medible f, que denotaremos por $\int_\Omega f\, d\mu$ se dan a continuación.

(Paso 1) Comencemos definiendo la integral para funciones simples. Una *función simple* ξ es una combinación lineal de funciones indicatrices de conjuntos en $\mathcal{A}$.

Definición 5.53. Sea ξ una función simple dada por

$$\xi = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i},$$

donde $A_i \in \mathcal{A}$, y $a_i \in \mathbb{R}$. Se define la integral $\int \xi\, d\mu$ como

$$\int_\Omega \xi\, d\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

Recordar que si X es una variable aleatoria, en $(\Omega, \mathbb{P})$, que toma los valores $x_1, \ldots, x_n$, entonces podemos escribir $X = \sum_{i=1}^{n} x_i \mathbb{1}_{X=x_i}$, y por lo tanto $\int_\Omega X\, d\mathbb{P} = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)$ coincide con su valor esperado $\mathbb{E}(X)$. (Ver (4.3) y (4.4).)

◇ **Ejer 5.54.** Es un ejercicio ver que esta definición es independiente de la representación de ξ. Esto es, si tenemos otra representación $\xi = \sum_{j=1}^{m} b_m \mathbb{1}_{B_j}$, entonces hay que ver que $\sum_{j=1}^{m} b_j \mu(B_j) = \sum_{i=1}^{n} a_i \mu(A_i)$. (Cf. Ejercicio (4.7)).

Análogo a la Proposición 4.6, se tiene que la integral de funciones simples satisfacen todas la propiedades dadas en dicha proposición, las cuales reescribimos acá.

Proposición 5.55. a) *Si $\xi \geqslant 0$ es simple, entonces $\int_\Omega \xi d\mu \geqslant 0$.*

b) *Si ξ y η son simples tales que $\xi \leqslant \eta$, entonces $\int_\Omega \xi d\mu \leqslant \int_\Omega \eta\, d\mu$.*

c) $|\int_\Omega \xi d\mu| \leqslant \int_\Omega |\xi| d\mu$

d) *La función $\int_\Omega (\cdot)\, d\mu$ actúa como un operador lineal sobre el espacio de funciones simples, esto es:*

$$\int_\Omega (\alpha\xi_1 + \beta\xi_2)\, d\mu = \alpha \int_\Omega \xi_1\, d\mu + \beta \int_\Omega \xi_2\, d\mu,$$

siendo ξ_1, ξ_2 funciones simples, y $\alpha, \beta \in \mathbb{R}$.

(Paso 2) Este es el paso más relevante dado que definiremos como integral funciones medibles acotadas.

Consideremos una función medible $f:\Omega\to\mathbb{R}$, acotada por M, (i.e. $|f(\omega)|\leqslant M$, para todo $\omega\in\Omega$). Veamos que f puede ser aproximada puntualmente por funciones simples. Para eso basta considerar para cada $n\in\mathbb{N}$ la discretización

$$f_{(n)}(\omega):=\frac{Mk}{n},\quad \text{si } f(\omega)\in\left(M\frac{k}{n},M\frac{(k+1)}{n}\right],\qquad (-n\leqslant k\leqslant n). \tag{5.4}$$

Observar que $f_{(n)}$ es una función simple. Observar que constructivamente, la función $f_{(n)}$ resulta de partir el intervalo $[-M,M]$ en la imagen, en intervalos de longitud M/n, y luego definir $f_{(n)}$ como constante en cada preimagen de los intervalos.

Con esta construcción se tiene lo siguiente.

Lema 5.56. a) *Para todo $n\geqslant 1$, se tiene $f_{(n)}\leqslant f\leqslant f_{(n)}+M/n$.*

b) *La sucesión $\int_\Omega f_{(n)}\,d\mu$ es una sucesión de Cauchy, y por tanto convergente.*

Demostración. a) resulta por construcción de $f_{(n)}$.

Para la parte b) observar que $f_{(n)}\leqslant f\leqslant f_{(n)}+\frac{M}{n}$, y por lo tanto se satisface para cualesquiera $m,n\geqslant 1$,

$$f_{(n)}\leqslant f_{(m)}+\frac{M}{m}.$$

Luego integrando las funciones simples anteriores, e intercambiando roles entre m y n, resulta

$$\int_\Omega f_{(n)}\,d\mu\leqslant\int_\Omega f_{(m)}\,d\mu+\frac{M}{m}\mu(\Omega)$$
$$\int_\Omega f_{(m)}\,d\mu\leqslant\int_\Omega f_{(n)}\,d\mu+\frac{M}{n}\mu(\Omega),$$

de donde concluimos que

$$\left|\int_\Omega f_{(n)}\,d\mu-\int_\Omega f_{(m)}\,d\mu\right|\leqslant M\,\text{máx}\left\{\frac{1}{m},\frac{1}{n}\right\}\mu(\Omega),$$

probando la segunda afirmación. □

Definición 5.57. Dada f una función medible y acotada, se define la integral de f como

$$\int_\Omega f\,d\mu:=\text{lím}_n\int_\Omega f_{(n)},$$

siendo $f_{(n)}$ la sucesión de funciones simples , dadas en (5.4), que aproximan por debajo a f

Se puede probar, sin dificultad pero omitiremos la prueba que todas las propiedades enunciadas en (5.55) siguen siendo válidas cambiando las funciones simples por funciones medibles acotadas.

Una forma alternativa de definir la integral para funciones medibles acotadas es definirla como el supremo de integrales de funciones simples que aproximan por abajo, o el ínfimo de las integrales de funciones simples por arriba. Esto se puede ver fácilmente considerando la sucesión $g_{(n)}$ que aproxima por arriba a f dada por

$$g_{(n)}(\omega):=\frac{M(k+1)}{n},\quad \text{si } f(\omega)\in\left(M\frac{k}{n},M\frac{(k+1)}{n}\right],\qquad (-n\leqslant k\leqslant n). \tag{5.5}$$

Se puede probar de manera análoga que la sucesión $\int_\Omega g_{(n)}\,d\mu$ es de Cauchy y por tanto convergente. Pero como

$$\int_\Omega g_{(n)}\,d\mu-\int_\Omega f_{(n)}\,d\mu=\frac{M}{n}\mu(\Omega),$$

resulta que ambos límites coinciden. Tomando φ y ψ familia de funciones simples tales que $\varphi\leqslant f\leqslant\psi$ se tiene

$$\int_\Omega f_{(n)}\,d\mu\leqslant\sup_{\varphi\leqslant f}\int_\Omega\varphi\,d\mu\leqslant\text{ínf}_{\psi\geqslant f}\int_\Omega\varphi\,d\mu\leqslant\int_\Omega g_{(n)}\,d\mu,$$

y por lo tanto se puede dar como definición alternativa

$$\boxed{\int_\Omega f\,d\mu:=\sup_{\varphi\,\text{simple}\leqslant f}\int_\Omega\varphi\,d\mu=\text{ínf}_{\psi\,\text{simple}\geqslant f}\int_\Omega\varphi\,d\mu}$$

(Paso 3) Si $h\geqslant 0$ es una función medible, entonces se define la integral de h como

$$\int_\Omega h\,d\mu:=\sup\left\{\int_\Omega f\,d\mu;\ \text{donde } f \text{ es acotada y } f\leqslant h\right\}$$

Observar que con esta definición estamos permitiendo $\int_\Omega h\,d\mu$ tome el valor $+\infty$.

(Paso 4) Sea f una función medible arbitraria. Se define su *parte positiva* como $f^+=\text{máx}(f,0)$ y su *parte negativa* como $f^-=\text{máx}(-f,0)$. Como $\text{máx}(a,b)=(a+b)/2+|a-b|/2$, resulta que la parte positiva y negativa de f siguen siendo funciones medibles. Además se tiene

$$f=f^+-f^-.$$

Luego se define la integral de f respecto a μ como:

$$\int_\Omega f\,d\mu:=\int_\Omega f^+\,d\mu-\int_\Omega f^-\,d\mu.$$

siempre y cuando la diferencia tenga sentido, esto es si $\int_\Omega f^+\,d\mu<+\infty$ o $\int_\Omega f^-\,d\mu<+\infty$.

Al igual que el paso anterior, todas las propiedades enunciadas en la Proposición 5.55 siguen siendo válidas para integrales de funciones medibles a las cuales se pueda calcular su integral.

5.4.1. Definición de Esperanza

En nuestro curso estaremos mayormente interesados en espacios de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, por lo que si tenemos una variable aleatoria X definimos su *esperanza* o *valor medio*

$$\mathbb{E}(X) := \int_\Omega X \, d\mathbb{P},$$

De la construcción de la integral de una función medible es un espacio de medida, se concluye el siguiente teorema con las propiedades de la esperanza (heredadas de la definición de integral dada anterioremente).

Teorema 5.58 *Consideremos un espacio de probabilidad* $(\Omega, \mathcal{A}, \mathbb{P})$*. Sean X e Y variables aleatorias $\geqslant 0$, o tales que $\mathbb{E}(|X|)$ y $\mathbb{E}(|Y|)$ finitas. Entonces se tiene*

1. $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$*;*
2. $X \leqslant Y$*, entonces* $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$*.*
3. $|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$*.*

5.4.2. Integral de Riemann vs Integral de Lebesgue

Consideremos el espacio de probabilidad $([0,1], \mathcal{B}, \lambda)$ siendo λ la medida de Lebesgue.

Supongamos que tenemos una función $f : [0,1] \to \mathbb{R}$ continua.

> *¿Cuál es la relación entre la integral de Riemann $\int_0^1 f(x)\, dx$ y la integral respecto a la medida de Lebesgue $\int_{[0,1]} f \, d\lambda$?*

Veamos que ambas coinciden.

Para ver esto observar que utilizando las sumas de Riemann se tiene que

$$\int_0^1 f(x)\, dx = \lim_n \sum_{i=1}^n f(i/n) \frac{1}{n}. \tag{5.6}$$

Si definimos las funciones a trozos

$$f_n(x) = f(i/n), \quad x \in A_i = (i/n, (i+1)/n], \qquad (i = 0, \dots, n),$$

observar que f_n es la función simple

$$f_n := \sum_{i=0}^n f(i/n) \mathbb{1}_{A_i}.$$

Además la integral de Lebesgue de f_n resulta

$$\int_{[0,1]} f_n \, d\lambda = \sum_{i=0}^n f(i/n) \lambda(A_i) = \sum_{i=0}^n f(i/n) \frac{1}{n}, \tag{5.7}$$

que es exactamente las sumas de Riemann. De (5.6) sólo nos falta probar que $\text{lím}_n \int_{[0,1]} f_n \, d\lambda = \int_{[0,1]} f \, d\lambda$.

Eso resulta de que hay convergencia uniforme de f_n a f, y las propiedades de la integral de Lebesgue ya vistas. Veamos esto.

Utilizando la continudad uniforme de f en $[0,1]$, es fácil ver que

$$\|f - f_n\|_\infty := \sup_{x \in [0,1]} |f(x) - f_n(x)| \xrightarrow{n} 0,$$

y por lo tanto, utilizando las propiedades de la integral vistas en el Teorema 5.58), concluimos

$$\begin{aligned}
\left| \int_{[0,1]} f_n \, d\lambda - \int_{[0,1]} f \, d\lambda \right| &\leqslant \left| \int_{[0,1]} (f_n - f) \, d\lambda \right| \\
&\leqslant \int_{[0,1]} |f_n - f| \, d\lambda \\
&\leqslant \int_{[0,1]} \|f_n - f\|_\infty d\lambda \\
&\leqslant \|f_n - f\|_\infty \int_{[0,1]} 1 d\lambda = \|f_n - f\|_\infty \xrightarrow{n} 0.
\end{aligned}$$

Luego de (5.6) y (5.7) se sigue la identidad buscada $\int_0^1 f(x)\, dx = \int_{[0,1]} f_n \, d\lambda$. Lo que acabamos de probar es el siguiente resultado

Proposición 5.59. *Si consideremos el espacio* $([0,1], \mathcal{B}, \lambda)$*, siendo λ la medida de Lebesgue, y $f : [0,1] \to \mathbb{R}$ continua, entonces la integral de Lebesgue de f coincide con la integral de Riemann, i.e.*

$$\int_{[0,1]} f \, d\lambda = \int_0^1 f(x)\, dx.$$

En particular tenemos que en el espacio de probabilidad $([0,1], \mathcal{B}, \lambda)$, una función continua f (que por cierto es una variable aleatoria), su esperanza satisface

$$\mathbb{E}(f) = \int_0^1 f(x)\, dx,$$

siendo el miembro derecho la integral de Riemann de la función.

Observar que en realidad la continuidad de f no es algo relevante para la integral de Lebesgue. De hecho f puede ser totalmente discontinua y la integral de Lebesgue estar perfectamente definida, y no así la integral de Riemann.

Un ejemplo típico de esto es tomar el ejemplo de función no integrable Riemann dado por

$$f = \mathbb{1}_{\mathbb{Q} \cap [0,1]},$$

i.e. f toma el valor 1 en los racionales, y 0 en los irracionales, del intervalo $[0,1]$. Esta función es claramente no integrable Riemann dado que las sumas inferiores son trivialmente 0, y las superiores trivialmente 1. Sin embargo, se tiene trivialmente

$$\int_{[0,1]} f \, d\lambda = \mathbb{1}_{\mathbb{Q} \cap [0,1]} \int_{[0,1]} d\lambda = 1 \lambda(\mathbb{Q} \cap [0,1]) = 0.$$

Observar que los racionales miden cero dado que son numerables y $\lambda(\{x\}) = 0$ para todo $x \in [0,1]$.

5.4.3. Algunas desigualdades

Desigualdad de Jensen

Dados $a,b \in \mathbb{R}$, con $a < b$, el segmento delimitado por a e b está dado por todos los puntos de la forma

$$[a,b] := \{\lambda a + (1-\lambda)b : \lambda \in [0,1]\}.$$

(Basta aplicar algún teorema de continuidad a la función lineal $\lambda \mapsto \lambda(b-a)+a$.) Observar que la que el conjunto es el mismo si cambiamos λ por $(1-\lambda)$, por lo que podemos sacar la restricción de $a < b$.

Una *función convexa* es una función φ que satisface

$$\varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y),$$

para $x,y \in \mathbb{R}$ y $\lambda \in [0,1]$. En palabras, el valor entre cualquier punto entre x e y, es menor igual al promedio de valores en x y en y.

Teorema 5.60 (Desigualdad de Jensen) *Si φ es convexa, y X una variable aleatoria, entonces*

$$\varphi(\mathbb{E}(X)) \leqslant \mathbb{E}(\varphi(X)),$$

(siempre que ambas esperanzas existan.) □

(Omitimos la prueba de este resultado.)

Ejemplo 5.61. Dada que las funciones $x \mapsto |x|$, y $x \mapsto x^2$ son convexas tenemos

$$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|), \qquad \mathbb{E}(X) \leqslant \sqrt{\mathbb{E}(X^2)}$$

Siempre sirven de ayudas ejemplos sencillos para recordar el sentido de la desigualdad. Si consieramos el $[0,1]$ con la medida de Lebesgue, de la Proposición 5.59, para la función f la identidad, y φ elevar al cuadrado, se tiene

$$\frac{1}{4} = \left(\int_0^1 x\,dx\right)^2 \leqslant \int_0^1 x^2\,dx = \frac{1}{3}.$$

Otro ejemplo aún más básico es que si φ convexa, y X es una variable aleatoria discreta tal que $\mathbb{P}(X = x) = \lambda$, y $\mathbb{P}(X = y) = 1-\lambda$, entonces

$$\varphi(\mathbb{E}(X)) = \varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y) = \mathbb{E}(\varphi(X)).$$

Ejemplo 5.62. Considerando las funciones convexas $x \mapsto 1/x$, (si $x > 0$), y $x \mapsto e^{-x}$ tenemos que

$$\frac{1}{\mathbb{E}(|X|)} \leqslant \mathbb{E}\left(\frac{1}{|X|}\right), \qquad e^{-\mathbb{E}(X)} \leqslant \mathbb{E}(e^{-X}).$$

Desigualdad Chebyshev

Daremos la versión general ahora de esta desigualdad.

Teorema 5.63 (Desigualdad Chebyshev) *Sea $p > 0$, $\lambda > 0$, y X una variable aleatoria. Entonces*

$$\mathbb{P}(|X| \geqslant \lambda) \leqslant \frac{1}{\lambda^p}\mathbb{E}(|X|^p).$$

Demostración. Observar que

$$\mathbb{P}(|X| \geqslant \lambda) = \mathbb{P}(|X|^p \geqslant \lambda^p) = \mathbb{E}(\mathbb{1}_{\{|X|^p \geqslant \lambda^p\}}),$$

y como $\lambda^p \mathbb{1}_{\{|X|^p \geqslant \lambda^p\}} \leqslant |X|^p \mathbb{1}_{\{|X|^p \geqslant \lambda^p\}}$, resulta

$$\mathbb{P}(|X| \geqslant \lambda) \leqslant \frac{1}{\lambda^p}\mathbb{E}(|X|^p \mathbb{1}_{\{|X|^p \geqslant \lambda^p\}}) \leqslant \frac{1}{\lambda^p}\mathbb{E}(|X|^p).$$

□

Un corolario que podemos obtener de esta desigualdad es lo siguiente

Corolario 5.64. *Si X es una variable aleatoria tal que $\mathbb{E}(X^2) = 0$, entonces $X = 0$ casi seguro.*

Demostración. Observar que el evento $\{|X| > 0\}$ coincide con $\bigcup_n \{|X| > 1/n\}$, y por lo tanto $\mathbb{P}(|X| > 0) = \lim_n \mathbb{P}(|X| > 1/n)$. Luego utilizando la desigualdad de Chebyshev se tiene $\mathbb{P}(|X| > 1/n) \leqslant n^2\mathbb{E}(X^2) = 0$, para todo $n \in \mathbb{N}$. Luego$\mathbb{P}(|X| > 0) = 0$. □

Desigualdad de Cauchy-Schwartz

Una desigulada fundamental que han visto en otros cusos es la *desigualdad de Cauchy-Schwartz*:

Proposición 5.65.

$$\mathbb{E}(|XY|) \leqslant \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)},$$

(siempre que las esperanzas existan.) Además la igualdad vale sí y solo sí X o Y son cero casi seguro, o $X = \lambda Y$ casi seguro, para algún $\lambda \in \mathbb{R}$. □

La prueba es igual a la utilizada para probar que el producto escalar en $\mathbb{R}^n$ satisface $\langle v,w\rangle \leqslant \sqrt{\|v\|^2 \cdot \|w\|^2}$.

◇ **Ejer 5.66.** Dar una prueba de esta desigualdad. Observar que del Corlorio 5.64, se puede suponer $\mathbb{E}(X^2) > 0$ y $\mathbb{E}(Y^2) > 0$. Sugerencia, considerar la variable aleatoria $Z := \lambda X - Y$, con $\lambda \in \mathbb{R}$, y observar que

$$0 \leqslant \mathbb{E}(Z^2) = \lambda^2\mathbb{E}(X^2) - 2\lambda\mathbb{E}(XY) + \mathbb{E}(Y^2).$$

Luego como función en λ es cuadrática y positiva, y por lo tanto algo podemos decir de su discriminante.

5.4.4. Integral de Riemann-Stieljes*

5.4.5. Esperanza via la función de distribución

Hasta ahora hemos desarollado una linda teoría de integrales, pero la pregunta es cómo calcularlas en general.

Con ese fin es que la distribución juega un papel importante.

Supongamos que tenemos un variable aleatoria en el espacio $(\Omega, \mathcal{A}, \mathbb{P})$. Recordar que X induce un nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}, \mu_X)$, donde μ_X es su distribución, i.e. aquella que satisface

$$\mu_X((a,b]) = \mathbb{P}(X \in (a,b]).$$

(Ver página 45).

Por definición la esperanza de X es $\mathbb{E}(X) = \int_\Omega X\, d\mathbb{P}$. Nos gustaría podes escribir, o calcular esta integral mediante una integral en el nuevo espacio $(\mathbb{R}, \mathcal{B}, \mu_X)$.

Veamos primero el caso particular de que X es absolutamente continua, i.e. que su función distribución F satisface:

$$F(x) = \mu_X((-\infty, x]) = \int_{-\infty}^{x} f(t)\, dt, \qquad x \in \mathbb{R},$$

o equivalentemente se tiene

$$\mu_X((a,b]) = \int_a^b f(t)\, dt, \qquad a < b.$$

Comentario 5.67. Más en general, se dice que una medida μ en un espacio $(\mathbb{R}, \mathcal{B})$ es *absolutamente continua* si existe una función medible e integrable $f : \mathbb{R} \to \mathbb{R}$ tal que

$$\mu(A) = \int_A f\, d\lambda, \quad \text{para todo boreliano } A.$$

Decimos que la v.a. X es absolutamente continua, cuando lo es su distribución. Ver página 79 para más detalles.

Y para facilitar la exposición, asumimos X toma valores entre $[0,1]$.

Luego si consideramos la integral de la función

$$\begin{aligned}
\int_{\mathbb{R}} x f(x)\, dx &= \lim_n \sum_{k=0}^{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} x f(x)\, dx \\
&\geqslant \lim_n \sum_{k=0}^{n} \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)\, dx \\
&= \lim_n \sum_{k=0}^{n} \frac{k}{n} \mathbb{P}\left(X \in \left(\frac{k}{n}, \frac{k+1}{n}\right]\right) = \mathbb{E}(X)
\end{aligned}$$

donde la última igualdad surge de nuestra definición de integral. (Observar que es exactamente la integral de la función simple que aproxima a nuestra función por debajo, cf. (5.4).) De manera análoga, se prueba la desigualdad por arriba utilizando las funciones simples que aproximan por arriba (cf. (5.5)).

Con esto concluimos la siguiente proposición, que su prueba formal la daremos en otra sección.

Proposición 5.68. *Sea X v.a.con densidad f. Entonces, si X es integrable, se tiene*

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx.$$

□

Este es un resultado particular del teorema de cambio de variable que veremos en la Sección 5.4.7.

5.4.6. Tres teoremas límites de integrales

En esta sección enunciaremos, sin demostracion, tres teoremas importantes de la teoría para límites de variables aleatorias. Básicamente la idea es tener hipótesis que nos permitan asegurar que es correcto intercambiar límite con la esperanza.

El primero es un conocido y útil teorema, que tiene nombre de lema.

Teorema 5.69 (Lema de Fatou) *Si $X_n \geqslant 0$, entonces* $\mathbb{E}(\liminf_n X_n) \leqslant \liminf_n \mathbb{E}(X_n)$.

Para recordar el sentido de la desigualdad es útil tener el siguiente ejemplo en mente.

> **Ejemplo 5.70.** Consideremos, en $(0,1)$ con la medida de Lebesgue, las v.a. $X_n = n\mathbb{1}_{(0,1/n)}$, con $n \in \mathbb{N}$. Luego se tiene que $\liminf X_n = 0$, i.e., es la variable aleatoria constante igual cero. Pero por otro lado $\mathbb{E}(X_n) = 1$, para todo $n \in \mathbb{N}$.

Además, observar que en este ejemplo $\liminf X_n$ es de hecho el límite de X_n, y por lo tanto no es cierto que $1 = \lim_n \mathbb{E}(X_n) \neq \mathbb{E}(\lim_n X_n) = 0$.

Teorema 5.71 (Teorema de Convergencia Monótona) *Si* $0 \leqslant X_n \uparrow X$, *entonces* $\lim_n \mathbb{E}(X_n) = \mathbb{E}(X)$.

Lo que dice el teorema de convergencia monónona es que si $X_n \geqslant 0$, y $X_n \leqslant X_{n+1}$ c.s., (y por lo tanto existe su límite casi seguro posiblemente tomando el valor $+\infty$), entonces podemos intercambiar el símbolo de límite por el de integral:

$$\mathbb{E}(\lim X_n) = \lim \mathbb{E}(X_n).$$

Una aplicación de este resultado es los siguiente.

♢ **Ejer 5.72.** Sean X_n variables aleatorias positivas. Probar $\mathbb{E}\left(\sum_{n=1}^{+\infty} X_n\right) = \sum_{n=1}^{+\infty} \mathbb{E}(X_n)$.

Con el teorema de convergencia monónota tenemos un primer resultado que nos permite intercambiar la integral con un límite.

Observar que el ejemploEl siguiente ejemplo nos muestra que si

> **Ejemplo 5.73.** Consideramos, en $(0,1)$ con la medida de Lebsgue, las variables aleatorias $X_n = n\mathbb{1}_{(0,1/n)}$.

Teorema 5.74 (Teorema de Convergencia Dominada) *Si $X_n \to X$ c.s., y $|X_n| \leqslant Y$ con $\mathbb{E}(Y) < \infty$, entonces* $\text{lím}_n \mathbb{E}(X_n) = \mathbb{E}(X)$.

Este es otro teorema que nos da condiciones para intercambiar el símbolo de integral con el de límite. Si hay convergncia casi segura de la sucesión de X_n, y las variables aleatorias involucradas están acotadas por una integrable, entonces $\mathbb{E}(\text{lím} X_n) = \text{lím}\, \mathbb{E}(X_n)$.

5.4.7. Cálculo de integrales

Teorema 5.75 *Consideremos $X : \Omega \to \mathbb{R}^n$ un vector aleatorio, y sea μ_X su distribución en $\mathbb{R}^n$, i.e. $\mu_X(B) = \mathbb{P}(X \in B)$, para todo boreliano B en $\mathbb{R}^n$. Si $\varphi : \mathbb{R}^n \to \mathbb{R}$ medible, (positiva o $\mathbb{E}(|\varphi(X)|) < \infty$) entonces*

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^n} \varphi\, d\mu_X. \tag{5.8}$$

Demostración. Daremos la demostracion por pasos, al igual que la mayoría de las pruebas que vinculen integrales.

(Caso 1:) La función $\varphi = \mathbb{1}_B$ es una indicatriz, para B es un boreliano de $\mathbb{R}^n$. Observar que $\mathbb{1}_B(X) : \Omega \to \mathbb{R}$ es la indicatriz $\mathbb{1}_{X\in B}$, y por lo tanto se tiene

$$\mathbb{E}(\mathbb{1}_B(X)) = \mathbb{E}(\mathbb{1}_{\{X\in B\}}) = \mathbb{P}(X \in B) = \mu_X(B) = \int_{\mathbb{R}^n} \mathbb{1}_B\, d\mu_X.$$

(Caso 2:) $\varphi = \sum_{k=1}^n c_k \mathbb{1}_{B_k}$ es una función simple, con B_k borelianos en $\mathbb{R}^m$. Luego por la linealidad tenemos:

$$\begin{aligned}\mathbb{E}\Big(\sum_{k=1}^n c_k \mathbb{1}_{B_k}(X)\Big) &= \sum_{k=1}^n c_k \mathbb{E}(\mathbb{1}_{\{X\in B_k\}}) = \sum_{k=1}^n c_k \mathbb{P}(X \in B_k) \\ &= \sum_{k=1}^n c_k \mu_X(B_k) = \int_{\mathbb{R}^n} \left(\sum_{k=1}^n c_k \mathbb{1}_{B_k}\right) d\mu_X.\end{aligned}$$

(Caso 3:) Si $\varphi \geqslant 0$ podemos definir

$$\varphi_n(x) := \text{mín}\left(\frac{\lfloor 2^n \varphi(x) \rfloor}{2^n}, n\right),$$

que es una función simple que aproxima a f por abajo, y utilizar el Teorema de convergencia monótona 5.71 para concluir el resultado. Esto es

$$\mathbb{E}(\varphi(X)) = \text{lím}_n \mathbb{E}(\varphi_n(X)) = \text{lím}_n \int_{\mathbb{R}^n} \varphi_n\, d\mu_X = \int_{\mathbb{R}^n} \varphi\, d\mu_X.$$

□

Decimos que un vector aleatorio $X : \Omega \to \mathbb{R}^n$ es *absolutamente continuo*, i.e. su distribución μ_X en $\mathbb{R}^n$ satisface que

$$\mu_X(B) = \int_B \rho_X(x)\, dx,$$

para todo boreliano B en $\mathbb{R}^n$, y cierta función Borel medible $\rho_X : \mathbb{R}^n \to \mathbb{R}$.

La función ρ_X se llama *densidad* de X.

Como coralario tenemos el resultado que ya hicimos a mano antes, en el caso $n = 1$.

Corolario 5.76. *Consideremos $X : \Omega \to \mathbb{R}^n$ un vector aleatorio absolutamente continuo con densidad ρ_X. Si $\varphi : \mathbb{R}^n \to \mathbb{R}$ medible, (positiva o $\mathbb{E}(|\varphi(X)|) < \infty$) entonces*

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^n} \varphi(x) \rho_X(x)\, dx. \tag{5.9}$$

□

◇ **Ejer 5.77.** Realizar la prueba de este resultado imitando la prueba anterior.

Definición 5.78. Sea X una variable aleatoria tal que $\mathbb{E}(X^2) < \infty$. Se define la *varianza* de X como

$$\text{var}(X) := \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Observar que del Teorema 5.75 resulta

$$\text{var}(X) = \int_{\mathbb{R}} (\text{Id}_{\mathbb{R}} - \mu^2)\, d\mu_X = \int_R (x - \mu^2)\, d\mu_X(x);$$

siendo $\mu = \mathbb{E}(X)$.

Con este resultado en mano podemos calcular esperanzas de momentos de variables aleatorias, $\mathbb{E}(X^k)$, calcular varianzas de variables aleatorias conocidas donde se define Veamos algunos ejemplos conocidos.

Ejemplo 5.79 (Distribución Gaussiana)**.** Si ξ es un variable aleatoria gaussiana estándar, entonces

$$\mathbb{E}(\xi) = \int_{-\infty}^{+\infty} x \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = 0;$$
$$\text{var}(\xi) = \int_{-\infty}^{+\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = 1.$$

Luego de las propiedades de la esperanza tenemos que si $X = \sigma\xi + \mu$, tenemos $\mathbb{E}(X) = \mu$, y $\text{var}(X) = \sigma^2$.

Ejemplo 5.80 (Distribución exponencial)**.** De maneral similar, si X es una variable aleatoria con distribución exponencial, sus momentos, y varianza

$$\mathbb{E}(X^k) = \int_0^{+\infty} x e^{-x}\, dx = k!,$$
$$\text{var}(X) = \mathbb{E}(X^2) - 1^2 = 2 - 1 = 1.$$

Comentario 5.81. Observar que del Teorema 5.75, para $n = 1$, se tiene que la esperanza de una variable aleatoria es una función de su distribución, más precisamente, tomando $\varphi : \mathbb{R} \to \mathbb{R}$ la función identidad se tiene

$$\mathbb{E}(X) = \int_{\mathbb{R}} \text{id}_{\mathbb{R}}\, d\mu_X.$$

Esto sugiere que podemos definir la esperanza y varianza de una medida de probabilidad μ definida en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, como $\mathbb{E}(\mu) := \int_{\mathbb{R}} \mathrm{Id}_{\mathbb{R}}\, d\mu$, y su varianza como $\mathrm{var}(\mu) := \int_{\mathbb{R}}(\mathrm{Id}_{\mathbb{R}} - \mathbb{E}(\mu))^2\, d\mu$.

Por ejemplo la esperanza de la medida uniforme en $(0,1)$ es $1/2$ y la varianza es 1/3-1/4=1/12.

Ejercicios Complementarios

◇ **Ejer 5.82** (Principio de Inclusión–Exclusión)**.** El principio de Inclusión–Exclusión generaliza la conocida fórmula $\mathbb{P}(A\cap B) = \mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)$, y dice si tenemos n eventos $A_1,\dots,A_n$, entonces

$$\mathbb{P}(A_1\cup\cdots\cup A_n) = \sum_{1\leqslant j\leqslant n}\mathbb{P}(A_i) - \sum_{1\leqslant i<j\leqslant 1}\mathbb{P}(A_i\cap A_j)+\cdots(-1)^{n+1}\mathbb{P}(A_1\cap\cdots\cap A_n).$$

La idea de este ejercicio es dar una prueba de este resultado utilizando la esperanza de funciones simples.

Observando que $\bigcup_{i=1}^n A_i = \left(\bigcap A_i^c)\right)^c$, probar que

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = 1-\prod_{i=1}^n (1-\mathbb{1}_{A_i}),$$

y concluir el resultado.

◇ **Ejer 5.83.**

1. Sea X una variable aleatoria que toma valores en $\mathbb{Z}_+$. Probar que
$$\mathbb{E}(X) = \sum_{k=1}^{\infty}\mathbb{P}(X\geqslant k).$$
2. Utilizando aproximaciones por simples, probar que si $X\geqslant 0$ una variable aleatoria, entonces
$$\mathbb{E}(X) = \int_0^{\infty}\mathbb{P}(X\geqslant t)\,dt.$$

◇ **Ejer 5.84.** Utilizar la desigualdad de Jensen, para $\varphi(x) = e^x$, y $\mathbb{P}(X=\log y_m) = p(m)$, para concluir que si $\sum_{m=1}^n p(m) = 1$, y $p(m), y_m > 0$, entonces

$$\sum_{m=1}^n p(m)y_m \geqslant \prod_{m=1}^n y_m^{p(m)}.$$

Observar que si $p(m) = 1/n$, para todo $m = 1,\dots,n$, entonces este resultado prueba que la media aritmética es mayor o igual a la media geométrica.

◇ **Ejer 5.85.** Sea X integrable, y A_n conjuntos disjuntos tales que A es su unión. Entonces probar que $\sum_{n=0}^{\infty}\mathbb{E}(X\mathbb{1}_{A_n}) = \mathbb{E}(A)$.

5.5. Independencia y medida producto

Uno de los conceptos más importantes en probabilidad, y la cual la distingue de teoría de la medida, es la noción y rol que juega la *independencia*. De hecho, como dice Terence Tao que dice Durret[3] “*measure theory ends and probability theory begins with the definition of independence*.”)

La *independencia* ya fue vista en la Sección 4.4, donde se dice que dos eventos son independientes si la ocurrencia de uno, no altera la ocurrencia del otro.

Recordar que en un espacio de probabilidad $(\Omega,\mathcal{A},\mathbb{P})$, la *probabilidad condicional de un evento A dado B*, con $\mathbb{P}(B)>0$, se define como

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

Luego decimos que A y B son independientes si $\mathbb{P}(A|B) = \mathbb{P}(A)$. Observar que si $\mathbb{P}(B)>0$ entonces también es válido que B es independiente de A.

De hecho, una definición más cerrada, simétrica, y que no se necesita pedir la condición de positividad de la medida, es decir que A y B son independientes si

$$\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B).$$

De hecho esta forma de representarlo ya nos da una pista importante. ¿Cómo podemos bosquejar en una hoja, una región Ω, por ejemplo el cuadrado unidad, con dos subconjuntos A y B que se solapen, y las proporciones satisfagan la condición de independencia?

Más en general, decimos que una colección de eventos $\{A_\alpha\}_{\alpha\in I}$ en Ω son *independientes* si

$$\mathbb{P}(A_{\alpha_1}\cap\cdots\cap A_{\alpha_n}) = \prod_{i=1}^n \mathbb{P}(A_{\alpha_i}),$$

para cualquier collección finita de índices distintos $\alpha_1,\dots,\alpha_n$.

La definición de independencia puede ser también definida sobre variables aleatorias.

Definición 5.86. Decimos que una colección $\{X_\alpha\}_{\alpha\in I}$ de variables aleatorias en $(\Omega,\mathcal{A})$ son *independientes* si los eventos $\{X_\alpha\in B_\alpha\} = X_\alpha^{-1}(B_\alpha)$ son independientes para cualquier Boreliano $B_\alpha\subset\mathbb{R}$.

Esta definición puede ser expresada de una manera más clara de la siguiente manera. Definiendo $X_i = X_{\alpha_i}$, y $B_i = B_{\alpha_i}$ con $i=1,\dots,n$, entonces tenemos

$$\begin{aligned}\mathbb{P}\left(\{X_1\in B_1\}\cap\cdots\cap\{X_n\in B_n\}\right) &= \mathbb{P}\left({X_1}^{-1}(B_1)\cap\cdots\cap {X_n}^{-1}(B_n)\right)\\ &= \mathbb{P}\left((X_1,\dots,X_n)^{-1}(B_1\times\ldots\times B_n)\right)\\ &= \mu_{(X_1,\dots,X_n)}\left(B_1\times\ldots\times B_n\right),\end{aligned}$$

siendo $\mu_{(X_1,\dots,X_n)}$ la distribución inducida por el vector aleatorio $(X_1,\dots,X_n)$, i.e., la medida inducida por el vector aleatorio en $(\mathbb{R}^n,\mathcal{B}_n)$ (cf. Definición 5.22).

[3] No lo encontré es su libro :(

Por otro lado, tenemos

$$\prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) = \prod_{i=1}^{n} \mu_{X_i}(B_i) = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}(B_1 \times \cdots \times B_n),$$

siendo μ_{X_i} la distribución de la variable aleatoria X_i, y $\mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$ la medida producto (ver Ejemplo 5.19).

Luego, como la σ-álgebra de Borel en $\mathbb{R}^n$, $\mathcal{B}(\mathbb{R}^n)$, es la generada por productos $B_1 \times \cdots \times B_n$ (con B_i borelianos en $\mathbb{R}$, o simplemente intervalos en $\mathbb{R}$), podemos concluir el siguiente resultado.

Teorema 5.87 *La colección $\{X_\alpha\}_{\alpha \in I}$ de variables aleatorias son independientes si y sólo si la distribución del vector aleatorio, de cualquier colección finita de la familia, es el producto de las distribuciones individuales.*

En el caso particular que tengamos una colección finita de variables aleatorias independientes, podemos concluir lo siguiente.

Teorema 5.88 *Sean $X_1, \dots, X_n$ variables aleatorias definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. Entonces, $X_1, \dots, X_n$ son independientes si y sólo si $\mu_{X_1 \times \cdots \times X_n} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.*

Como corolario tenemos lo siguiente.

Corolario 5.89. *Para que $X_1, \dots, X_n$, variables aleatorias sobre el mismo espacio sean independientes alcanza con que se satisfaga, para todo $x_1, \dots, x_n \in (-\infty, +\infty)$,*

$$\mathbb{P}(X_1 \leqslant x_1, \dots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leqslant x_i).$$

Demostración. La demostración sigue de que los conjuntos $(-\infty, x_1] \times \cdots \times (-\infty, x_n]$, al moverse librementes con los x_i, generan $\mathcal{B}(\mathbb{R}^n)$, y por lo tanto las medidas $\mu_{(X_1,\dots,X_n)}$ y $\otimes_{i=1}^{n} \mu_{X_i}$ coinciden sobre un generador del álgebra, y por tanto en la σ-álgebra generada. □

Dejamos como ejericios condiciones suficientes para la independencia de variables aleatorias con densidad, y discretas.

◇ **Ejer 5.90.** Supongamos que el vector aleatorio $(X_1, \dots, X_n)$ tiene *densidad* $\rho : \mathbb{R}^n \to \mathbb{R}$, i.e.

$$\mathbb{P}((X_1, \dots, X_n) \in B) = \int \cdots \int_B \rho \, d\lambda_n, \qquad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

esto es en notación usual

$$\mathbb{P}((X_1, \dots, X_n) \in B) = \int \cdots \int_B \rho(x_1, \dots, x_n) \, dx_1 \dots dx_n, \qquad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

Si ρ puede escribirse como

$$\rho(x_1, \dots, x_n) = g(x_1) \cdots g(x_n),$$

para ciertas funciones medibles $g_i : \mathbb{R} \to [0, +\infty)$, entonces $X_1, \dots, X_n$ son independientes. (Observar que no es necesario pedir que las funciones g_i sean densidades.)

◇ **Ejer 5.91.** Supongamos que las variables aleatorias $X_1, \dots, X_n$, toman valores en conjuntos numerables $S_1, \dots, S_n$. Luego, para que sean independientes es suficiente que

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i),$$

para cualesquiera $x_i \in S_i$.

Teorema de Fubini-Tonelli

Observar que en el ejercicio 5.90 se utiliza la propiedad de cambio de orden en las integrales de varias variables, conocido como Fubini. De hecho, es un caso particular de un resultado más general debido a Fubini y Tonelli:

Teorema 5.92 (Teorema de Fubini-Tonelli) *Consideremos que el espacio $(\Omega, \mathcal{A}, \mu)$ es el espacio producto de espacios de medida σ-finitas $(\Omega_1, \mathcal{A}_1, \mu_1)$ y $(\Omega_2, \mathcal{A}_2, \mu_2)$, i.e. $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$, y $\mu = \mu_1 \otimes \mu_2$. Si $f \geqslant 0$, o $\int |f| d\mu < \infty$, se tiene*

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(x,y) \, d\mu_2(y) \right) d\mu_1(x) = \int_{\Omega} f \, d\mu = \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y) d\mu_1(x) \right) d\mu_2(y).$$

El teorema parece inofensivo, pero hay que tener cierto cuidado antes de intercambiar las integrales. Un ejemplo sencillo que muestra que no siempre es válido es el siguiente. Si consideramos $X = Y = \mathbb{N}$, con $\mathcal{A}_i$ las partes de $\mathbb{N}$, y μ_i la medida de conteo. Definamos la función $f(m,n)$ de la siguiente manera: $f(m,m) = 1$, $f(m+1,m) = -1$, y cero en el resto, (i.e., vale 1 en la diagonal, y -1 en la sub diagonal, y cero en el resto. Entonces se tiene

$$\sum_m \sum_n f(m,n) = 1, \qquad \sum_n \sum_m f(m,n) = 0.$$

$$\begin{pmatrix} 1 & 0 & \dots & \\ -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \end{pmatrix}$$

Comentario 5.93. La prueba del Teorema de Fubini requiere de varios pasos, y no la demostraremos en este curso, pero comentaremos algunos puntos de la prueba.

El primero es ver que las funciones que se están integrando son de hecho medibles. Esto es, si consideramos la primera igualdad en el Teorema de 5.92, tenemos que asegurarnos de los siguiente. Fijado $x \in \Omega_1$, tenemos que ver la función $y \mapsto f(x,y)$ es $\mathcal{A}_2$ medible, y luego ver que la función $x \mapsto \int_{\Omega_2} f(x,y) \, d\mu_2(y)$ es $\mathcal{A}_1$ medible, para poder ser integradas con las medidas respectivas.

Un caso particular e importante en la prueba es considerar $f(x,y) = \mathbb{1}_E(x,y)$, la indicatriz de cierto conjunto $E \in \mathcal{A}$. Observar que fijado $x \in \Omega_1$, el mapa $y \mapsto \mathbb{1}_E(x,y)$ es la indicatriz del conjunto $E_x := \{y \in \Omega_2 : (x,y) \in E\} \subset \Omega_2$. Este conjunto es una “sección” de E. Por lo tanto hay que probar que el conjunto E_x está en la σ-álgebra $\mathcal{A}_2$. En particular, si $E = A_1 \times A_2$, entonces las secciones, al variar $x \in \Omega_1$ se reducena al vacío o A_2, dependiendo si $x \in A_1$.

Aplicando el Teorema 5.92 a la indicatriz de E, con $E \in \mathcal{A}$, resulta

$$\mu(E) = \mu_1 \otimes \mu_2(E) = \int_{\Omega_1} \mu_2(E_x) \, d\mu_1(x) = \int_{\Omega_2} \mu_2(E_y) \, d\mu_2(y). \tag{5.10}$$

El siguiente resultado, bastante tedioso de escribir, pero fácil de convencerse, dice que si tenemos una familia de variables aleatorias independientes, y tomamos funciones medibles de subcolecciones finitas disjuntas, entonces las nuevas variables aleatorias son independientes.

Proposición 5.94. *Sean $\{X_{nj} : 1 \leqslant j \leqslant J(n),\ 1 \leqslant n \leqslant N\}$ son variables aleatorias independientes, y sean $f_n : \mathbb{R}^{J(n)} \to \mathbb{R}$ funciones Borel medibles para $1 \leqslant n \leqslant N$. Entonces las variables aleatorias $Y_n = f_n(X_{n1}, \dots, XnJ(n))$, para $1 \leqslant n \leqslant N$, son independientes.*

Demostración. Definamos los N vectores aleatorios $\mathbf{X}_n = (X_{n1}, \dots, X_{nJ(n)})$. Si $B_1, \dots, B_N$ son borelianos en $\mathbb{R}$, entonces tenemos

$$Y_n \in B_n = \mathbf{X}_n{}^{-1}\left(f_n{}^{-1}(B_n)\right)$$

(continuará) □

Proposición 5.95. *Sean X e Y variables aleatorias independientes con distribuciones μ_X y μ_Y respectivamente. Sea $h : \mathbb{R}^2 \to \mathbb{R}$ medible tal que $h > 0$ o $\mathbb{E}(|h(X,Y)|) < \infty$. Entonces*

$$\mathbb{E}(h(X,Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y)\, d\mu_X(x) \right) d\mu_Y(y).$$

En particular, si $h(x,y) = f(x)g(y)$ para $f,g : \mathbb{R} \to \mathbb{R}$ medibles, y se tiene $f,g \geqslant 0$, o se tiene $\mathbb{E}(|f(X)|) < \infty$ y $\mathbb{E}(|g(Y)|) < \infty$, entonces

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

Demostración. La prueba es una aplicación inmediata del Teorema de Fubini. Del Teorema 5.75 y del Teorema 5.88 tenemos

$$\mathbb{E}(h(X,Y)) = \int_{\mathbb{R}^2} h(x,y)\, d\mu_{X\times Y} = \int_{\mathbb{R}^2} h(x,y)\, d(\mu_X \otimes \mu_Y),$$

y luego aplicando el Teorema de Fubini concluimos

$$\mathbb{E}(h(X,Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y)\, d\mu_X(x) \right) d\mu_Y(y).$$

Para el caos particular, supongamos que $f,g \geqslant 0$. Luego por la parte anterior tenemos

$$\mathbb{E}(f(X)g(Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(y)\, d\mu_X(x) \right) d\mu_Y(y).$$

Como la función $x \mapsto f(x)g(y)$ es la función $x \mapsto f(g)$ multiplicada por la consante $g(y)$, resulta que

$$\mathbb{E}(f(X)g(Y)) = \int_{\mathbb{R}} g(y) \left(\int_{\mathbb{R}} f(x)\, d\mu_X(x) \right) d\mu_Y(y).$$

Luego, observando que la integral de adentro es $\mathbb{E}(f(X))$, tenemos

$$\begin{aligned}\mathbb{E}(f(X)g(Y)) &= \int_{\mathbb{R}} g(y) \left(\int_{\mathbb{R}} f(x)\, d\mu_X(x) \right) d\mu_Y(y) \\ &= \int_{\mathbb{R}} g(y)\mathbb{E}(f(X)) d\mu_Y(y) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).\end{aligned}$$

De esta manera si $f,g \geqslant 0$ obtenemos el resultado. Para el caso en que tengamos $\mathbb{E}(|f(X)|) < \infty$ y $\mathbb{E}(|g(Y)|) < \infty$, entonces observar que ya sabemos que $\mathbb{E}(|f(X)||g(Y)|) = \mathbb{E}(|f(X)|)\mathbb{E}(|g(Y)|)$, por lo que $\mathbb{E}(|f(X)g(Y)|) < \infty$ y podemos utilizar el argumento nuevamente para finalizar la prueba. □

Corolario 5.96. *Sean $X_1, \dots, X_n$ variables aleatorias independientes, tales que $\mathbb{E}(|X_i|) < \infty$ para cada $i = 1, \dots, n$. Entonces se tiene $\mathbb{E}(|\prod_{i=1}^n X_i|) < \infty$, y se tiene*

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Comentario 5.97. Es una observación no menor que si $\mathbb{E}(|X|) < \infty$, y $\mathbb{E}(|Y|) < \infty$ entonces no tiene que suceder que $\mathbb{E}(|XY|) < \infty$. Esto ocurre si hay independencia entre las variables aleatorias. Un ejemplo sencillo es considerar la medida uniforme en $[0,1]$, la función función $x \mapsto 1/\sqrt{x}$ es integrable, pero no su cuadrado.

Definamos el espacio L^p de variables aleatorias con p-momento integrable.

> **Definición 5.98.** Sea X una variable aleatoria en $(\Omega, \mathcal{A}, \mathbb{P})$. Decimos que $X \in L^p := L^p(\Omega, \mathcal{A}, \mathbb{P})$, para $p \geqslant 1$, si $\mathbb{E}(|X|^p) < \infty$.

Corolario 5.99. *Sean $X_1, \dots, X_n$, variables aleatorias independientes en L^2. Entonces*

$$\operatorname{var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \operatorname{var}(X_i).$$

◊ **Ejer 5.100.** Probar el resultado anterior.

La demostración sale por inducción y queda de ejercicio.

Algunos ejercicios complementarios

◊ **Ejer 5.101.** Sean X,Y,Z, variables aleatorias positivas, independientes, con distribución común μ. Sea F su distribución, i.e. $F(t) = \mu([0,t])$ con $t \geqslant 0$. Probar que la probabilidad de que el polinomio $t \mapsto Xt^2 + Yt + Z$ tenga raíces reales es igual a $\int_0^{+\infty}\int_0^{+\infty} F(t^2/4s)\, d\mu(t)\, d\mu(s)$.
(Sug: Recordar Comentario 5.93, y fórmula 5,10.

◊ **Ejer 5.102.** Sea X una variable aleatoria positiva. Probar utilizando el Teorema de Fubini que

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)\, dt.$$

(Sugerencia: recordar que para este caso tenemos $\mathbb{E}(X) = \int_0^\infty t\, d\mu_X(t)$, siendo μ_X la distribución. Escribir esta integral como una integral en la medida producto de la medida de Lebesgue y μ_X en $[0,\infty)^2 \subset \mathbb{R}^2$.)

5.5.1. Construcción de variables aleatorias independientes*

(en construcción)

En esta sección se sugiere la construcción de la medida producto en el producto infinito de espacios de probabilidad.

Un estudio completo de este problema se hace en el curso de *Procesos estocásticos*.

Construcción de infinitas variables aleatorias independientes

El objetivo de esta sección es construir una sucesión infinita de variables aleatorias $X_1, X_2, \dots$ que sean independientes con una distribuciones $F_1, F_2 \dots$ dadas.

Procedamos como en el caso finito.

Sean $\{\mathbb{P}_i\}_{i\in\mathbb{N}}$ medidas de probabilidad inducidas en $\mathbb{R}$ por las distribuciones $\{F_i\}_{i\in\mathbb{N}}$. Esto es, medidas de probabilidad $\mathbb{P}_i$ que satisfacen $\mathbb{P}_i((a,b]) = F_i(b) - F_i(a)$.

Consideremos el espacio producto

$$\Omega := \mathbb{R}^{\mathbb{N}} := \{\omega : \mathbb{N} \to \mathbb{R}\},$$

donde escribimos $\omega_i := \omega(i)$, $i \in \mathbb{N}$. (Recordar digresión sobre espacios productos en página **??**.)

Sea $\mathcal{F}$ la σ-álgebra producto generado por los rectángulos finito-dimensionales

$$\{\omega \in \Omega : \omega_i \in A_i, i \in I, \#I < \infty\},$$

para borelianos $A_i \subset \mathbb{R}$, $i \in I$. En otras palabras, un rectángulo es dar condiciones en finitas coordenadas de ω. (Recordar la Sección **??**, para la noción de σ-álgebra generada.)

Motivado por lo anterior, buscamos una medida $\mathbb{P}$ definida en $\mathcal{F}$ que satisfaga

$$\mathbb{P}(\{\omega \in \Omega : \omega_i \in A_i, i \in I, \#I < \infty\}) = \prod_{i\in I} \mathbb{P}_i(A_i), \tag{5.11}$$

para borelianos $A_i \subset \mathbb{R}$, $i \in \mathbb{N}$.

Para probar la existencia de una medida con esta propiedad se utiliza el siguiente resultado conocido como *teorema de extensión de Kolmogorov*

Teorema 5.103 *Sean μ_n medidas de probabilidad definidas en $\mathbb{R}^n$ para la σ-álgebra de Borel. Supongamos que las medidas verifican la siguiente condición de consistencia*

$$\mu_{n+1}((a_1,b_1] \times \cdots \times (a_n,b_n] \times \mathbb{R}) = \mu_n((a_1,b_1] \times \cdots \times (a_n,b_n]).$$

Entonces, existe una única medida de probabilidad $\mathbb{P}$ en $\mathbb{R}^{\mathbb{N}}$, con σ-álgebra generada por rectángulos finito-dimensionales, que verifica

$$\mathbb{P}(\{\omega \in \Omega : \omega_i \in (a_i,b_i], i \in I, \#I < \infty\}) = \prod_{i\in I} \mu_i((a_i,b_i]).$$

Por lo tanto, si definimos $X_i : \Omega \to \mathbb{R}$, con $X_i(\omega) := \omega_i$, $i \in \mathbb{N}$, y definimos medidas finito-dimensionales que satisfagan la igualdad (5.11), resulta del Teorema 5.103, que existe $\mathbb{P}$ medida de probabilidad en $\mathbb{R}^{\mathbb{N}}$, tal que las v.a. X_i están definidas sobre Ω, con medida de probabilidad inducidas $\mathbb{P}_i$, que son independientes.

Observar que nuevamente hemos construido una extensión $(\Omega, \mathcal{F}, \mathbb{P})$ de espacios de probabilidad $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}_i)$, para $i \in \mathbb{N}$.

5.6. Recopilación de ejercicios

Ejercicio 5.104. Observar que si $\mathcal{A}_\alpha$ es una colección arbitraria de σ-álgebras en Ω, entonces $\bigcap_\alpha \mathcal{A}_\alpha$ es también una σ-álgebra.

Ejercicio 5.105. Probar que si tomamos la familia de subconjuntos de Ω formados por un elemento, entonces la σ-álgebra coincide con la del Ejemplo 5.3.

Ejercicio 5.106. La σ-álgebra de Borel también puede ser obtenida si cambiamos la familia $\mathcal{I}$ por la familia de intervalos semi-abiertos $(a,b]$, o intervalos cerrados $[a,b]$, o intervalos más generales tipo $(-\infty,b)$, o $(\infty,b]$, variando $a,b \in \mathbb{R}$ con $a < b$. Basta probar que cada uno de estos nuevos intervalos puede ser formado por elementos de $\mathcal{B}(\mathbb{R})$.

Ejercicio 5.107. Dar una prueba de la proposición anterior. Para la parte **??**, considerar la sucesón $B_i = A_i - A_{i-1}$, y observar que la colección de B_i es disjunta y su unión coincide con la unión de los A_i.)

Ejercicio 5.108. Probar que que la distribución de una v.a. es una medida en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Ejercicio 5.109. Hallar la distribución de una función indicatriz. Observar que la medida está concentrada en dos valores únicamente.

Ejercicio 5.110. Dar una prueba de este resultado.

Ejercicio 5.111. Probar que $F_X(x) = \begin{cases} 1 - e^{-\lambda x} \text{ si } x \geqslant 0 \\ 0 \text{ si } x < 0 \end{cases}$.

Ejercicio 5.112. Pruebe que composición de funciones medibles es medible. Esto es, si $X : (\Omega, \mathcal{A}) \to (S, \mathcal{S})$ y $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ son funciones medibles, entonces $f(X) : (\Omega, \mathcal{A}) \to (T, \mathcal{T})$, es medible.

Ejercicio 5.113. Finalizar prueba de la proposición anterior.

Ejercicio 5.114. Probar si $\{X_n\}$ es una sucesión de variables aleatorias, tales que $X_n \xrightarrow{\mathbb{P}} X$ y $X_n \xrightarrow{\mathbb{P}} Y$, entonces $\mathbb{P}(X = Y) = 1$.

Ejercicio 5.115. Probar que la sucesión de variables aleatorias definidas en el Ejemplo 5.47 no converge de manear casi segura.

Ejercicio 5.116. Dar una prueba de esta proposición.

Ejercicio 5.117. Es un ejer2cicio ver que esta definición es independiente de la representación de ξ. Esto es, si tenemos otra representación $\xi = \sum_{j=1}^m b_m \mathbb{1}_{B_j}$, entonces hay que ver que $\sum_{j=1}^m b_j \mu(B_j) = \sum_{i=1}^n a_i \mu(A_i)$. (Cf. Ejercicio (4.49)).

Ejercicio 5.118. Dar una prueba de esta desigualdad. Observar que del Corlorio 5.64, se puede suponer $\mathbb{E}(X^2) > 0$ y $\mathbb{E}(Y^2) > 0$. Sugerencia, considerar la variable aleatoria $Z := \lambda X - Y$, con $\lambda \in \mathbb{R}$, y observar que

$$0 \leqslant \mathbb{E}(Z^2) = \lambda^2 \mathbb{E}(X^2) - 2\lambda \mathbb{E}(XY) + \mathbb{E}(Y^2).$$

Luego como función en λ es cuadrática y positiva, y por lo tanto algo podemos decir de su discriminante.

Ejercicio 5.119. Sean X_n variables aleatorias positivas. Probar $\mathbb{E}\left(\sum_{n=1}^{+\infty} X_n\right) = \sum_{n=1}^{+\infty} \mathbb{E}(X_n)$.

Ejercicio 5.120. Realizar la prueba de este resultado imitando la prueba anterior.

Ejercicio 5.121 (Principio de Inclusión–Exclusión)**.** El principio de Inclusión–Exclusión generaliza la conocida fórmula $\mathbb{P}(A\cap B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)$, y dice si tenemos n eventos $A_1,\ldots,A_n$, entonces

$$\mathbb{P}(A_1\cup\cdots\cup A_n)=\sum_{1\leqslant j\leqslant n}\mathbb{P}(A_i)-\sum_{1\leqslant i<j\leqslant 1}\mathbb{P}(A_i\cap A_j)+\cdots(-1)^{n+1}\mathbb{P}(A_1\cap\cdots\cap A_n).$$

La idea de este ejer2cicio es dar una prueba de este resultado utilizando la esperanza de funciones simples.

Observando que $\bigcup_{i=1}^n A_i=(\bigcap A_i^c))^c$, probar que

$$\mathbb{1}_{\bigcup_{i=1}^n A_i}=1-\prod_{i=1}^n(1-\mathbb{1}_{A_i}),$$

y concluir el resultado.

Ejercicio 5.122.

1. Sea X una variable aleatoria que toma valores en $\mathbb{Z}_+$. Probar que
$$\mathbb{E}(X)=\sum_{k=1}^{\infty}\mathbb{P}(X\geqslant k).$$
2. Utilizando aproximaciones por simples, probar que si $X\geqslant 0$ una variable aleatoria, entonces
$$\mathbb{E}(X)=\int_0^{\infty}\mathbb{P}(X\geqslant t)\,dt.$$

Ejercicio 5.123. Utilizar la desigualdad de Jensen, para $\varphi(x)=e^x$, y $\mathbb{P}(X=\log y_m)=p(m)$, para concluir que si $\sum_{m=1}^n p(m)=1$, y $p(m), y_m>0$, entonces

$$\sum_{m=1}^n p(m)y_m\geqslant\prod_{m=1}^n y_m^{p(m)}.$$

Observar que si $p(m)=1/n$, para todo $m=1,\ldots,n$, entonces este resultado prueba que la media aritmética es mayor o igual a la media geométrica.

Ejercicio 5.124. Sea X integrable, y A_n conjuntos disjuntos tales que A es su unión. Entonces probar que $\sum_{n=0}^{\infty}\mathbb{E}(X\mathbb{1}_{A_n})=\mathbb{E}(A)$.

Ejercicio 5.125. Supongamos que el vector aleatorio $(X_1,\ldots,X_n)$ tiene *densidad* $\rho:\mathbb{R}^n\to\mathbb{R}$, i.e.

$$\mathbb{P}((X_1,\ldots,X_n)\in B)=\int\cdots\int_B\rho\,d\lambda_n,\qquad\forall B\in\mathcal{B}(\mathbb{R}^n),$$

esto es en notación usual

$$\mathbb{P}((X_1,\ldots,X_n)\in B)=\int\cdots\int_B\rho(x_1,\ldots,x_n)\,dx_1\ldots dx_n,\qquad\forall B\in\mathcal{B}(\mathbb{R}^n),$$

Si ρ puede escribirse como

$$\rho(x_1,\ldots,x_n)=g(x_1)\cdots g(x_n),$$

para ciertas funciones medibles $g_i:\mathbb{R}\to[0,+\infty)$, entonces $X_1,\ldots,X_n$ son independientes. (Observar que no es necesario pedir que las funciones g_i sean densidades.)

Ejercicio 5.126. Supongamos que las variables aleatorias $X_1,\ldots,X_n$, toman valores en conjuntos numerables $S_1,\ldots,S_n$. Luego, para que sean independientes es suficiente que

$$\mathbb{P}(X_1=x_1,\ldots,X_n=x_n)=\prod_{i=1}^n\mathbb{P}(X_i=x_i),$$

para cualesquiera $x_i\in S_i$.

Ejercicio 5.127. Probar el resultado anterior.

Ejercicio 5.128. Sean X,Y,Z, variables aleatorias positivas, independientes, con distribución común μ. Sea F su distribución, i.e. $F(t)=\mu([0,t])$ con $t\geqslant 0$. Probar que la probabilidad de que el polinomio $t\mapsto Xt^2+Yt+Z$ tenga raíces reales es igual a $\int_0^{+\infty}\int_0^{+\infty}F(t^2/4s)\,d\mu(t)\,d\mu(s)$.
(Sug: Recordar Comentario 5.93, y fórmula 5,10.

Ejercicio 5.129. Sea X una variable aleatoria positiva. Probar utilizando el Teorema de Fubini que

$$\mathbb{E}(X)=\int_0^{\infty}\mathbb{P}(X>t)\,dt.$$

(Sugerencia: recordar que para este caso tenemos $\mathbb{E}(X)=\int_0^{\infty}t\,d\mu_X(t)$, siendo μ_X la distribución. Escribir esta integral como una integral en la medida producto de la medida de Lebesgue y μ_X en $[0,\infty)^2\subset\mathbb{R}^2$.)

Parte IV

Teoremas límites en probabilidad

Capítulo 6

Ley de los grandes números

[Duración: 1 Semanas]

6.1. Introducción

La *Ley de los grandes números (LGN)* y el *teorema central del límite (TCL)* son dos de los teoremas límites más conocidos en probabilidad. En esta sección estudiaremos la LGN, el cuál ya fue abordado en un caso particular en la Sección 4.5. En dicha sección estudiamos la relación entre la versión frecuentista de la probabilidad, y su contrapartida axiomática, para el caso de tiradas (independientes) de una moneda.

La LGN dice que si promediamos los valores de tomar muestras independientes de v.a. con igual distribución, (con esperanza μ y varianza finita) entonces la v.a. resultante "tiende" a μ. En particular, si las v.a. son la indicatriz de cierto evento A, (i.e. toma valor 1 si A ocurre, y 0 si no), entonces la LGN dice que la frecuencia relativa tiende a la probabilidad de A.

6.2. Ley débil de los grandes números

Ya estamos en condiciones de enunciar y probar la ley (débil) de los grandes números en general.

Teorema 6.1 (Ley débil de los grandes números) *Sean $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes con igual esperanza μ e igual varianza σ^2. Sea*

$$S_n = X_1 + \cdots + X_n, \quad n \in \mathbb{N}.$$

Entonces, $S_n/n \xrightarrow{\text{P}} \mu$. i.e. para todo $\varepsilon > 0$, se tiene

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \to 0, \quad \textit{cuando} \quad n \to +\infty.$$

◊ **Ejer 6.2.** Probar el teorema utilizando la desigualdad de Chebyshev 5.63. □

◊ **Ejer 6.3.** Mostrar con un ejemplo que si las variables son dependientes el resultado no siempre es cierto.

Una de las aplicaciones más importantes de la ley débil de los grandes números es en el caso de que las v.a. X_i son independientes con igual distribución, lo que comunmente se escribe como *sucesión de v.a. i.i.d.*. Si $\{X_i\}$ son v.a. i.i.d con esperanza μ y varianza finita σ^2, entonces S_n/n tiende a μ en el sentido anterior.

6.2.1. Convergencia en L^2

Una noción más fuerte que la convergencia en probabilidad es la convergencia en L^2.

Definición 6.4. Decimos que una sucesión de v.a. $\{X_i\}$ en L^2, converge a X en L^2, si

$$\mathbb{E}((X_n - X)^2) \to 0, \quad \text{cuando} \quad n \to \infty.$$

En tal caso escribimos $X_n \xrightarrow{L^2} X$.

◊ **Ejer 6.5.** Probar que si $X_n \xrightarrow{L^2} X$, entonces $X_n \xrightarrow{\text{P}} X$.

Comentario 6.6. Si miramos la prueba realizada de la ley débil de los grandes números, en realidad se prueba un resultado más fuerte: bajo las mismas hipótesis resulta $S_n/n \xrightarrow{L^2} \mu$. Veamos esto en un caso un poquito más general.

Proposición 6.7. *Sean $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes en L^2 tales que $\mathbb{E}(X_i) = \mu$ y $\mathrm{var}(X_i) \leqslant C < \infty$. Si $S_n = X_1 + \cdots + X_n$, entonces cuando $n \to +\infty$ tenemos $\frac{S_n}{n} \xrightarrow{L^2} \mu$.*

Demostración. Observar que $\mathbb{E}(S_n/n) = \mu$, por lo tanto

$$\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^2\right) = \mathrm{var}\left(\frac{S_n}{n}\right) = \frac{\mathrm{var}(S_n)}{n^2} = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) \leqslant \frac{C}{n} \to 0,$$

cuando $n \to +\infty$. □

Luego utilizando el Ejercicio 6.5 se concluye otra prueba de la ley débil de los grandes números.

Comentario 6.8. Es posible generalizar aún más este resultado, de distintas maneras. Manteniendo una prueba análoga, se puede ver que basta utilizar que las variables aleatorias son no correlacionadas, i.e. $\mathbb{E}(X_iX_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ siempre que $i \neq j$. Esto resulta del hecho en que también es válida la igualdad $\mathrm{var}(S_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)$. También se puede cambiar la condición sobre las varianzas uniformemente acotadas por algo que verifique $\frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(X_i) \to 0$, cuando $n \to +\infty$.

Comentario 6.9. Observar que para tener convergencia en probabilidad no es necesario pedirle a las v.a. que tengan varianza finita (i.e. que estén en L^2). Sin embargo, lo anterior sí es un requisito en el caso de convergencia en L^2. Es por esta razón, y por razones históricas, que la LGN se enuncia en estas dos modalidades por separado.

6.3. Aplicaciones

En el siguiente ejemplo veremos cómo nuestra intuición geométrica empieza a fallar a gran escala.

Ejemplo 6.10. Sean $\{X_i\}$ v.a. i.i.d con distribución uniforme en $(-1,1)$. Sean $Y_i = X_i^2$.

◊ **Ejer 6.11.** Probar que

$$\frac{X_1^2 + \cdots + X_n^2}{n} \xrightarrow{\mathrm{P}} \frac{1}{3} \quad \text{cuando} \quad n \to \infty.$$

◊ **Ejer 6.12.** Si $A_{n,\varepsilon} = \{x \in \mathbb{R}^n : (1-\varepsilon)\sqrt{n/3} \leqslant \|x\| \leqslant (1+\varepsilon)\sqrt{n/3}\}$, entonces, para todo $\varepsilon > 0$ se tiene

$$\lim_n \frac{\mathrm{vol}(A_{n,\varepsilon} \cap (-1,1)^n)}{2^n} = 1.$$

Esto último significa que la mayor parte del volumen del hipercubo $(-1,1)^n$ está concentrado en el borde de la bola de radio $\sqrt{n/3}$.

6.4. Ley fuerte de los grandes números

La idea de esta sección es probar una *ley fuerte de los grandes números*. (Ejercicio tomado del segundo parcial).

Teorema 6.13 (Ley fuerte de los grandes números) *Si $\{X_i\}$ es una sucesión de v.a. i.i.d. con $\mathbb{E}(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2 < +\infty$, y cuarto momento finito, i.e. $\mathbb{E}(X_1^4) < \infty$, entonces*
si $S_n = X_1 + \cdots + X_n$, se tiene

$$\mathbb{P}\left(\lim_n \left|\frac{S_n}{n} - \mu\right| = 0\right) = 1. \tag{6.1}$$

Es decir $S_n/n \overset{\text{c.s.}}{\to} \mu$.

Lema de Borel-Cantelli

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, y una sucesión de eventos $A_1, A_2, \ldots$ en $\mathcal{A}$. Se define el *límite superior* de la sucesión $\{A_n\}$ como

$$\limsup_n A_n := \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k.$$

Proposición 6.14. *(a) Se tiene que $\omega \in \limsup_n A_n$ si y sólo si $\omega \in A_n$ para infinitos $n \in \mathbb{N}$.*

(b) $\limsup_n A_n \in \mathcal{A}$.

Demostración. (a) ($\Rightarrow$ si $\omega \in \limsup A_n$ entonces $\omega \in A_n$ para infinitos $n \in \mathbb{N}$)

Si $\omega \in \limsup A_n$, entonces para cualquier $n \in \mathbb{N}$ ocurre que $\omega \in \bigcup_{k \geqslant n} A_k$, y por lo tanto existe algún $k \geqslant n$ que satisface $\omega \in A_k$. Es fácil ver que esto implica que hay infinitos A_k que contienen a ω. Una forma de verlo es por inducción. Para $n = 1$ sea n_1 tal que $\omega \in A_{n_1}$. Como $\omega \in \bigcup_{k > n_1} A_k$, existe $n_2 > n_1$ tal que $\omega \in A_{n_2}$. Procediendo de esta manera construimos una sucesión $n_1, n_2, \ldots$ tal que $\omega \in A_{n_j}$ para todo $j \geqslant 1$.

($\Leftarrow$) Recíproramente, si $\omega \in A_{n_k}$ para una subsucesión $n_1, n_2, \ldots$, entonces dado cualquier $n \in \mathbb{N}$ se tiene que $\omega \in \bigcup_{k \geqslant n} A_k$ dado que n_j es mayor a cualquier n para j suficientemente grande.

(b) Notar que $A_k \in \mathcal{A}$, para todo $k \in \mathbb{N}$, y por lo tanto $B_n := \bigcup_{k \geqslant n} A_k \in \mathcal{A}$ por ser $\mathcal{A}$ una σ-álgebra. Luego como $\limsup_n A_n = \bigcap_n B_n$, se tiene que $\limsup_n A_n$ es una intersección numerable de elementos de $\mathcal{A}$, y por tanto en $\mathcal{A}$.

□

Teorema 6.15 *(Lema de Borel - Cantelli)*
Sean $A_1, A_2 \ldots$ eventos en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$.

1. *Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, entonces $\mathbb{P}(\limsup_n A_n) = 0$.*
2. *Si los sucesos $\{A_n\}_n$ son independientes y $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, entonces $\mathbb{P}(\limsup_n A_n) = 1$.*

Demostración. 1. Como $\limsup_n A_n \subset \bigcup_{k=n}^{\infty} A_k$ para todo n,

$$\mathbb{P}(\limsup_n A_n) \leqslant \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leqslant \sum_{k=n}^{\infty} \mathbb{P}(A_k) \quad \forall n.$$

Entonces:

$$\mathbb{P}(\limsup_n A_n) \leqslant \lim_{n \to \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0,$$

por ser la cola de una serie convergente.

2. Recordar que $\limsup_n A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$. Consideremos entonces los sucesos:
$$B_n = \bigcup_{k=n}^{\infty} A_k\,.$$
Observar que $B_1 \supseteq B_2 \supseteq \ldots \supseteq B_n \supseteq \ldots$ y que $\limsup_n A_n = \bigcap_{n=1}^{\infty} B_n$. Luego, utilizando la propiedad de *continuidad de la probabilidad*, tenemos que:
$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = 1 - \lim_{n\to\infty}\mathbb{P}(B_n^c).$$
Probemos $\mathbb{P}(B_n^c) = 0$ para todo n. Fijemos n.
$$\mathbb{P}(B_n^c) = \mathbb{P}\left[\left(\bigcup_{k=n}^{\infty} A_k\right)^c\right] = \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) \leqslant \mathbb{P}\left(\bigcap_{k=n}^{N} A_k^c\right),$$
para todo $N \geqslant n$. Luego tenemos, para todo $N \geqslant n$,
$$\mathbb{P}(B_n^c) \leqslant \prod_{k=n}^{N}\mathbb{P}(A_k^c) = \prod_{k=n}^{N}(1-\mathbb{P}(A_k)) \leqslant \prod_{k=n}^{N} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{N}\mathbb{P}(A_k)}.$$
Luego, como al desigualdad anterior es válida para todo N, y $\sum_{k=n}^{\infty}\mathbb{P}(A_k) = \infty$, resulta y lo tanto cuando $\mathbb{P}(B_n^c) = 0$.

□

6.4.1. Demostración ley fuerte de los grandes números

A continuación guiaremos la prueba. Primero observar que podemos asumir $\mu = 0$.

1. Sea X una v.a. positiva, con cuarto momento finito. Probar que se tiene $\mathbb{P}(X > \varepsilon) \leqslant \frac{\mathbb{E}(X^4)}{\varepsilon^4}$.
2. Probar que existe una constante $C > 0$ tal que $\mathbb{E}(S_n^4) \leqslant Cn^2$.
3. Fijemos $\varepsilon > 0$. Sea $A_n^{\varepsilon} = \{\left|\frac{S_n}{n}\right| > \varepsilon\}$. Probar que $\mathbb{P}(A_n^{\varepsilon}) \leqslant \frac{C'}{n^2}$, para cierta constante $C' > 0$.
4. Probar que para todo $\varepsilon > 0$ se tiene $\mathbb{P}(\limsup_n A_n^{\varepsilon}) = 0$. Concluir la prueba de la ley fuerte de los grandes números.

6.5. Teorema de Glivenko-Cantelli

Una situación muy común en estadística es tener un fenómeno aleatorio donde la ley probabilística (i.e. su distribución P_X) es desconocida, y mediante el uso de métodos estadísticos podemos obtener información de la misma. La idea general en estadística es que tomando muchas muestras independientes del fenómeno aleatorio, a saber, $X_1, X_2, \ldots, X_n$, con $n \to +\infty$, podemos recuperar propiedades de la distribución desconicida $\mathbb{P}_X$.[1] Un caso particular de esto es la ley fuerte de los grandes números donde, bajo ciertas hipótesis, podemos estimar (o recuperar si tuvieramos infintitas muestras) la esperanza de la medida $\mathbb{P}_X$:
$$\lim_{n\to\infty}\frac{1}{n}(X_1+\cdots+X_n) = \int_{\mathbb{R}} x\,d\mathbb{P}_X(x)(=\mathbb{E}(X)), \quad \text{c.s.},$$

[1] Formalmente, estamos haciendo un abuso de notación en la descripción. Los datos recabados pueden ser pensados como realizaciones aleatorias de variables aleatorias independientes e igualmente distribuidas, en un cierto espacio $(\Omega, \mathcal{A}, \mathbb{P})$. En este sentido, nuestros datos observados $\{x_1, x_2, \ldots, x_n\}$ se piensan como realizaciones $\{X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\}$ para cierto $\omega \in \Omega$.

El *Teorema de Glivenko Cantelli*, también conocido como *Teorema Fundamental de la Estadística Matemática*, dice que estimar la distribución $\mathbb{P}_X$ es posible.

Se define la *distribución empírica* de la muestra como
$$\widehat{F_n}(\cdot) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{[X_i,+\infty)}(\cdot),$$
donde para cada $x \in \mathbb{R}$ se tiene
$$\widehat{F_n}(x) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leqslant x\}} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,x]}(X_i), \quad x \in \mathbb{R}.$$
Esto es, $\widehat{F_n}(x)$ es la frecuencia del número de realizaciones menores o iguales a x. (Tener cuidado que en las identidades anteriores hay cierta ambiguedad en los dominios de las indicatrices, pero una vez prefijados el $\omega \in \Omega$ y $x \in \mathbb{R}$, todo queda bien definido.

Observación 6.16. Observar que la función $\widehat{F_n}$ es una "función de distribución simple aleatoria". Esto es, para cada $\omega \in \Omega$, la función $(\widehat{F_n})_\omega(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{[X_i(\omega),+\infty)}(\cdot)$ es una función simple. Formalmente, $\widehat{F_n}$ es un proceso estocástico, i.e. una función bi-medible $\widehat{F_n}: \Omega\times\mathbb{R}\to\mathbb{R}$, denominado *proceso empírico*.

Teorema 6.17 (Teorema de Glivenko Cantelli) *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con función de distribución F en $\mathbb{R}$. Entonces*
$$\|\widehat{F_n} - F\|_\infty := \sup_{x\in\mathbb{R}}|\widehat{F_n}(x) - F(x)| \xrightarrow{c.s.} 0.$$

Observación 6.18. Antes de dar la prueba, observar que para x fijo tenemos que $\mathbb{1}_{\{X_i\leqslant x\}}$ son variables aleatorias i.i.d. Bernoulli de parámetro $F(x)$. Por lo tanto, utilizando la ley fuerte de los grandes números (Teorema 6.13 para variables aleatorias i.i.d. Bernoulli), resulta
$$\widehat{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leqslant x\}} \xrightarrow{c.s.} p = F(x).$$
Es decir que tenemos la convergencia puntual en un conjunto de probabilidad 1. Lo que dice el Teorema de Glivenko Cantelli es que la anterior convergencia es uniforme en x.

Demostración. Realizaremos la prueba en el caso de que F sea continua. La prueba general se basa en esta prueba pero las cuentas son más engorrosas. Dado $m \in \mathbb{N}$, consideremos una partición $-\infty = x_0 < x_1 < \cdots < x_{m-1} < x_m = +\infty$, de los reales, tales que
$$F(x_j) - F(x_{j-1}) = \frac{1}{m}, \qquad (j = 1, \ldots, m).$$

Luego, sabiendo que F y $\widehat{F}_n$ son no-decrecientes, tenemos que para todo $x \in [x_{j-1}, x_j]$ se tiene

$$\widehat{F}_n(x) - F(x) \leqslant \widehat{F}_n(x_j) - F(x_{j-1}) = \widehat{F}_n(x_j) - F(x_j) + \frac{1}{m},$$
$$\widehat{F}_n(x) - F(x) \geqslant \widehat{F}_n(x_{j-1}) - F(x_j) = \widehat{F}_n(x_{j-1}) - F(x_{j-1}) - \frac{1}{m},$$

y esto implica que

$$|\widehat{F}_n(x) - F(x)| \leqslant \max_{j=0,\dots,m} |\widehat{F}_n(x_j) - F(x_j)| + \frac{1}{m}, \quad \text{para todo } x \in \mathbb{R},$$

y por lo tanto

$$\|\widehat{F}_n - F\|_\infty \leqslant Y_n + \frac{1}{m},$$

donde $Y_n \geqslant 0$ es la variable aleatoria $Y_n := \max_{j=0,\dots,m} |\widehat{F}_n(x_j) - F(x_j)|$. Luego, por la Observación 6.18, tenemos que Y_n converge a 0 casi seguro, y por lo tanto $\limsup_n \|\widehat{F}_n - F\|_\infty \leqslant \frac{1}{m}$ en un conjunto de probabilidad 1. Como el m es arbitrario el resultado sigue. □

Observación 6.19. Un tema que pasamos por alto es la “medibilidad” de la cantidad $\|\widehat{F}_n - F\|_\infty = \sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$, i.e. es $\|\widehat{F}_n - F\|_\infty : \Omega \to \mathbb{R}$ una variable aleatoria? La respuesta es sí, y para convencerse miremos el caso continuo que es más fácil. Basta ver que el supremo es un máximo en un conjunto finito (aleatorio) a saber, los datos observados. Esto es, para cada $\omega \in \Omega$,

$$\sup_{x\in\mathbb{R}} |(\widehat{F}_n)_\omega(x) - F(x)| =$$
$$\max\left\{ \max_{i=1,\dots,n} |(\widehat{F}_n)_\omega(X_i(\omega)) - F(X_i(\omega))|, \max_{i=1,\dots,n} |(\widehat{F}_n)_\omega(X_i(\omega)^-) - F(X_i(\omega)^-)| \right\}.$$

Luego, $\|\widehat{F}_n - F\|_\infty$ es un máximo de variables aleatorias. Para el caso de F no continua, hay que incluir los límites laterales en cada dato, pero eso sigue siendo un máximo de variables aleatorias.

6.5.1. Aplicaciones y extensiones

¿Por qué Glivenko–Cantelli es tan importante?

La distribución P_X tiene toda la información de interés sobre nuestro fenómeno aleatorio. Por ejemplo, si conocemos P_X, sabemos el valor esperado del fenómeno, sabemos cuál es la probabilidad de observar un cierto evento. Pero no tenemos acceso a eso. Ante eso, podemos distinguir básicamente dos enfoques en la estadística. Un enfoque es suponer un modelo de fondo, para lo que necesitamos hacer algunos supuestos sobre el comportamiento de los datos (y deberíamos hacer el ejercicio de verificar que los datos no se apartan mucho de esos supuestos requeridos). Otro camino, es “confiar”que la información va a surgir de los datos. Glivenko-Cantelli nos dice que la distribución empírica aproximará bien a la teórica, por tanto, un cálculo que hagamos con la distribución empírica (el cálculo de cualquier probabilidad, o un promedio, por ejemplo), es razonable que se parezca a los cálculos exactos que podríamos hacer si conociéramos la distribución teórica. Esta afirmación que parece informal e intuitiva, se sustenta muy rigurosamente por este teorema.

También es importante tener un control sobre la tasa de convergencia del Teorema de Glivenko-Cantelli. Una versión un poco más sofisticada, que se la recomandamos a los interesados, es la siguiente.

Teorema 6.20 (Glivenko-Cantelli cuantitativo) *Sean $X_1, \dots, X_n$ variables aleatorias i.i.d. con función de distribución $F(x) = \mathbb{P}(Z \leqslant x)$, y sea $\widehat{F}_n(x)$ la función de distribución empírica. Entonces*

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| > \varepsilon \right) \leqslant 8(n+1)e^{-n\varepsilon^2/32}.$$

♢ **Ejer 6.21.** Utilizando el lema de Borel–Cantelli, probar que el Teorema 6.20 implica la $\|\widehat{F}_n - F\|_\infty \xrightarrow{c.s.} 0$.

Veamos un par de notaciones útiles antes de seguir.

Delta de Dirac y medida discretas

Dado $x \in \mathbb{R}$, la *medida de Dirac* δ_x se define sobre las partes de $\mathbb{R}$ como

$$\delta_x(A) := \begin{cases} 0, & x \notin A; \\ 1, & x \in A, \end{cases}$$

siendo $A \subset \mathbb{R}$. La medida de Dirac es una medida de probabilidad en $\mathbb{R}$.

Es fácil convencerse que si $f : \mathbb{R} \to \mathbb{R}$ es una función arbitraria, entonces integrar respecto a δ_x es la evaluación en x de la funci´on, i.e.

$$\int_{\mathbb{R}} f \, d\delta_x = f(x).$$

Con esta definición, podemos reescribir variables aleatorias discretas, como combinación lineal de deltas de Dirac. Vemos esto

Dada una variable aleatoria discreta Y tomando valores $y_1, \dots, y_k$, donde $p_i = \mathbb{P}(Y = y_i)$, la distribución de esta Y es una medida discreta que tiene “átomos” en los puntos y_i, y con peso p_i. En este sentido, podemos escribir la distribución μ_Y como

$$\mu_Y := \sum_{i=1}^{n} p_i \delta_{y_i},$$

donde δ_x es la Delta de Dirac en $x \in \mathbb{R}$. (En particular, tomando iguales pesos (i.e. $p_i = 1/k$), resulta que la medida $\frac{1}{k}\sum_{i=1}^{n} \delta_{y_i}$, es la medida uniforme en el conjunto $\{y_1, \dots, y_k\} \subset \mathbb{R}$.)

Observar que si $f : \mathbb{R} \to \mathbb{R}$ es cualquier función, entonces

$$\mathbb{E}(f(Y)) = \sum_{i=1}^{n} f(y_i) p_i = \int_{\mathbb{R}} f \, d\mu_Y.$$

Utilicemos la siguiente notación: si $f:\mathbb{R}\to\mathbb{R}$ es una función Borel medible, y ν es una medida de borel en $\mathbb{R}$ entonces denotamos

$$\nu(f) := \int_{\mathbb{R}} f\, d\nu.$$

Medida empírica

Una forma alternativa de ver lo el Teorema de Glivenko–Cantelli es considerar la *medida empírica* generada por los datos observados $X_1,\ldots,X_n$:

$$\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}.$$

Formalmente, lo que tenemos es que $\widehat{\mu}_n$ es una medida *aleatoria* soportada en los datos. (Ver nota al pie **??**.) Por lo tanto, fijado $\omega\in\Omega$ tenemos la sucesión de datos $X_1(\omega), X_2(\omega),\ldots$, y en este caso tenemos que $(\widehat{\mu}_n)_\omega$ es la medida uniforme soportada en los primeros n datos.

Observar que si $f:\mathbb{R}\to\mathbb{R}$ entonces

$$\widehat{\mu}_n(f) = \int f\, d\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n f(X_i),$$

es el promedio de una suma de variables aleatorias independientes con igual distribución. En particular, si $\mu=\mathbb{P}_X$ la distribución de la variable aleatoria X, y $\mathbb{E}(|f(X)|^4)<\infty$, por la ley fuerte de los grandes números, $\widehat{\mu}_n(f)$ converge casi seguro a $\mu(f)=\mathbb{E}(f(X))$.

Observación 6.22 (Método de Monte Carlo)**.** Un caso de particular de importancia es el *Método de Monte Carlo* que nos permite aproximar integrales. El método nos dice que si tomamos variables aleatorias $U_1,\ldots,U_n,\ldots$ con distribución uniforme independientes, y construimos nuestra medida empírica $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n \delta_{U_i}$, luego para toda $f:[0,1]\to\mathbb{R}$ función continua, resulta

$$\widehat{\mu}_n(f) = \frac{1}{n}\left(f(U_1)+\cdots+f(U_n)\right) \xrightarrow{c.s.} \mu(f) = \int_0^1 f(x)\,dx.$$

La convegencia también es en probabilidad (por ser más débil que la casi segura), pero utilizando la desigualdad de Chebyshev podemos acotar la probabilidad de que $|\widehat{\mu}_n(f)-\int_0^1 f(x)\,dx|>\varepsilon$.

♢ **Ejer 6.23.** Probar que

$$\mathbb{P}\left(\left|\widehat{\mu}_n(f)-\int_0^1 f(x)\,dx\right|>\frac{1}{\sqrt{n}}\right)\leqslant\left(\frac{1}{n^2}\int_0^1 |f(x)|^2\,dx-\left(\int_0^1 f(x)\,dx\right)^2\right).$$

Como $\widehat{F}_n(x)=\widehat{\mu}_n((-\infty,x])$, y $F(x)=\mu((-\infty,x])$, podemos reescribir la tesis del Teorema de Glivenko–Cantelli como

$$\sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} \mathbb{1}_{(-\infty,x]}\,d\widehat{\mu}_n-\int_{\mathbb{R}}\mathbb{1}_{(-\infty,x]}\,d\mu\right|\xrightarrow{c.s.}0.$$

Teorema 6.24 (Glivenko–Cantelli) *Sea* $\mathcal{I}=\{(-\infty,x]: x\in\mathbb{R}\}$. *Entonces*

$$\sup_{S\in\mathcal{I}}|\widehat{\mu}_n(\mathbb{1}_S)-\mu(\mathbb{1}_S)|\xrightarrow{c.s.}0.$$

Una pregunta importante es saber hasta dónde se puede extender la familia de intervalos $\mathcal{I}$ para que el la tesis del teorema siga valiendo. Este problema fue estudiado por Vapnik y Chervonenkis y tiene implicancias en teoría de aprendizaje y otras yerbas. Para referncias ver el libro de Devroye-Györfi-Lugosi [DGL].

6.6. Recopilación de ejercicios

Ejercicio 6.25. Probar el teorema utilizando la desigualdad de Chebyshev 5.63. □

Ejercicio 6.26. Mostrar con un ejemplo que si las variables son dependientes el resultado no siempre es cierto.

Ejercicio 6.27. Probar que si $X_n\xrightarrow{L^2}X$, entonces $X_n\xrightarrow{\mathrm{P}}X$.

Ejercicio 6.28. Probar que

$$\frac{X_1^2+\cdots+X_n^2}{n}\xrightarrow{\mathrm{P}}\frac{1}{3}\quad\text{cuando}\quad n\to\infty.$$

Ejercicio 6.29. Si $A_{n,\varepsilon}=\{x\in\mathbb{R}^n: (1-\varepsilon)\sqrt{n/3}\leqslant\|x\|\leqslant(1+\varepsilon)\sqrt{n/3}\}$, entonces, para todo $\varepsilon>0$ se tiene

$$\lim_n \frac{\mathrm{vol}(A_{n,\varepsilon}\cap(-1,1)^n)}{2^n}=1.$$

Ejercicio 6.30. Utilizando el lema de Borel–Cantelli, probar que el Teorema 6.20 implica la $\|\widehat{F}_n-F\|_\infty\xrightarrow{c.s.}0$.

Ejercicio 6.31. Probar que

$$\mathbb{P}\left(\left|\widehat{\mu}_n(f)-\int_0^1 f(x)\,dx\right|>\frac{1}{\sqrt{n}}\right)\leqslant\left(\frac{1}{n^2}\int_0^1 |f(x)|^2\,dx-\left(\int_0^1 f(x)\,dx\right)^2\right).$$

Capítulo 7

Teorema central de límite

Teorema central del límite, funciones características, aplicaciones.

En esta sección veremos uno de los teoremas más importantes de matemática, a saber, *el teorema central del límite* (TCL).

Este teorema trata sobre la convergencia de variables aleatorias y por esa razón comenzamos introduciendo el tipo de convergencia utilizado en la tesis de este teorema.

7.1. Convergencia en distribución

Hasta ahora hemos visto distintos tipos de convergencia, como la convergencia en probabilidad, la convergencia en L^p, la convergencia casi segura. Todas estas son importantes en probabilidad y tienen que ver con la proximidad de los valores tomados por la sucesión de variables aleatorias y su límite. Sin embargo, existe otro tipo de convergencia muy importante en probabilidad que es la *convergencia en ley* o *convergencia en distribución* que refiere a la convergencia "puntual" de las funciones de distribución. Más precisamente se tiene la siguiente definición.

Definición 7.1. Decimos que una sucesión de funciones de distribución F_n *converge en distribución* a una función de distribución F, y escribimos $F_n \Rightarrow F$, o $F_n \overset{\mathcal{D}}{\to} F$, si

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \text{en } x \text{ punto de continuidad de } F. \quad (7.1)$$

Ahora podemos considerar convergencia en distribución de variables aleatorias.

Definición 7.2. Decimos que una sucesión de variables aleatorias $X_1, X_2, \ldots$, *converge en distribución* a una variable aleatoria X, si sus funciones de distribución convergen en distribución.

Observar que lo que importa no es los valores en sí que toman X_n y X, sino las probabilidades de que tomen esos valores.

Observación 7.3. ¿Por qué enfatizar las dos definiciones y no solo definir la segunda utilizando la fórmula 7.1? La razón es que la primer definición refiere a convergencia de funciones de distribución en $\mathbb{R}$ y por lo tanto podemos no es necesario asumir que las variables aleatorias viven en un mismo espacio $(\Omega, \mathcal{A}, \mathbb{P})$. De hecho ocurre que una sucesión $\{X_n\}_n$ de variables aleatorias converge en distribución a una variable aleatoria X que puede no existir donde están definidas las X_n. Ya veremos esto en con el Teorema Central de Límite.

Ejemplo 7.4. El Teorema de Glivenko–Cantelli nos da un ejemplo de convergencia en distribución, si fijamos el azar. (Recordar el proceso empírico asociado, ver Observación 6.16.) Esto es, si $X_1, X_2, \ldots$, son i.i.d con función de distribución F, entonces fijado $\omega \in \Omega$, tenemos una sucesión numérica $\{X_n(\omega)\}$ y una función de distribución

$$(\widehat{F}_n)_\omega(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i(\omega)) \overset{n}{\to} F(x),$$

para todo $x \in \mathbb{R}$.

Ejemplo 7.5 (Máximo de uniformes)**.** Sean $U_1, U_2, \ldots$ i.i.d. uniformes en $(0,1)$, y sea $M_n := \max_{i=1,\ldots,n} U_i$. Veamos si M_n converge en distribución. Lo razonable es que M_n se acerque al valor 1 (en algún sentido) a medida que n crece. Tenemos que para todo $0 < \varepsilon < 1$,

$$\mathbb{P}(|M_n - 1| \geqslant \varepsilon) = \mathbb{P}(M_n < 1-\varepsilon) = (1-\varepsilon)^n.$$

Luego como lo anterior tiende a cero con n, concluimos que M_n tiende en probabilidad a 1. De hecho, mirando entre líneas tenemos la prueba también de que hay convergencia en distribución a la función de distribución $\mathbb{1}_{[1,+\infty)}$, (que no es otra cosa que la función de distribución de la variable aleatoria constante igual a 1). Esto resulta de que si $\mathbb{P}(M_n < 1-\varepsilon) \overset{n}{\to} 0$, y si $\mathbb{P}(M_n < 1+\delta) = 1$ para cualquier $\delta > 0$, para concluir $M_n \overset{\mathcal{D}}{\to} \mathbb{1}_{[1,\infty)}$.

Observar además que, tomando $\varepsilon = t/n$ tenemos que

$$\mathbb{P}(M_n \leqslant 1 - t/n) = (1-t/n)^n \overset{n}{\to} e^{-t},$$

y por lo tanto

$$\mathbb{P}(n(1-M_n) \leqslant t) \overset{n}{\to} 1 - e^{-1}.$$

Es decir la variable aleatoria $n(1-M_n)$ tiende en distribución a la distribución exponencial de parámetro 1.

En los ejemplos anteriores la convergencia se daba en todos los puntos de $\mathbb{R}$, y no solo en los de contiuidad. Por ejemplo, en el caso anterior, si tomamos $x=1$, que es el punto de discontinuidad de la función distribución de la variable aleatoria constante igual 1, también hay convergencia ya que $\mathbb{P}(M_m \leqslant 1) = 1$.

◇ **Ejer 7.6.** Probar que el conjunto de discontinuidades de las funciones de distribución es a lo sumo numerable. Idea: probar que si x es de discontinuidad, entonces $\mathbb{P}(X=x)>0$. ¿Puede haber no-numerables saltos? Cuántos saltos de tamaño mayor a $1/m$ puede haber?).

Luego la convergencia en distribución implica que hay convergencia en "casi" en todo $\mathbb{R}$. De hecho el "casi" es es que hay convegencia salvo en un conjunto de medida nula para la medida de Lebesgue λ_1 en $\mathbb{R}$.

O sea la convergencia se da en todos los puntos salvo en un conjunto de medida de Lebesgue nula en $\mathbb{R}$, pero ¿por qué razón los excluímos?

Ejemplo 7.7. Sea X una variable aleatoria con función de distribución F. Luego la variable aleatoria $X+1/n$ tiene función de distribución $F_n(x) = \mathbb{P}(X+1/n \leqslant x) = \mathbb{P}(X \leqslant x-1/n) = F(x-1/n)$. Luego $F_n(x) \xrightarrow{n} F(x^-)$, y por lo tanto hay convergencia sólo en los puntos de continuidad de F. Así que con la definición que tenemos, podemos afirmar que $X+1/n \xrightarrow{\mathcal{D}} X$.

A continuación vemos otro ejemplo similar

◇ **Ejer 7.8.** Sea X_n la variable aleatoria constante igual a 2^{-n}. Probar que X_n tiende en distribución a la función distribución F de la variable aleatoria constante igual cero. Observar que en el punto de discontinuidad de la distribución F, no hay convergencia.

Si la sucesión $\{X_n\}$ y X viven en un mismo espacio de probabilidad, entonces podemos preguntarnos cuál es la relación entre la convergencia en distribución y los otros tipo de convergencia. Veamos que es la más débil de las convergencias que hemos visto.

◇ **Ejer 7.9.** La idea de este ejercicio es probar la convergencia en distribución es más débil que la de probabilidad (cuando estamos en un mismo espacio). (Ver Observación 7.3.) Sean $\{X_n\}$ una sucesión de variables aleatorias , tal que $X_n \xrightarrow{\mathbb{P}} X$ para cierta variable aleatoria X. Entonces $X_n \xrightarrow{\mathcal{D}} X$.

1. Dados t punto de coninuidad de $F(x) = \mathbb{P}(X \leqslant x)$, y $\varepsilon > 0$ mostrar que $\mathbb{P}(X \leqslant t-\varepsilon) \leqslant \mathbb{P}(X_n < t) + \mathbb{P}(|X-X_n| \geqslant \varepsilon)$.
2. De manera similar, encontrar una cota superior para $\mathbb{P}(X_n < t)$, para concluir el resultado.

Convergencia débil de medidas

La convergencia en distribución también se denomina *convergencia débil de medidas* por la siguiente cuestión. Se puede probar que la convergencia $F_n \Rightarrow F$ es equivalente a

$$\mu_n(f) \xrightarrow{n} \mu(f), \quad \text{i.e.} \quad \int_{\mathbb{R}} f\, d\mu_n \xrightarrow{n} \int_{\mathbb{R}} f\, d\mu,$$

para $f:\mathbb{R}\to\mathbb{R}$ continua y acotada, siendo μ_n y μ las medidas respectivas inducidas por F_n y F.

7.2. Introducción al Teorema Central del Límite

Consideremos una sucesión $\{X_n\}$ de v.a. i.i.d. con esperanza μ y varianza σ^2. La LGN dice que las sumas parciales $S_n = X_1 + \cdots + X_n$ se aproximan $n\mu$ cuando n crece. Dado que la varianza de S_n es $n\sigma^2$, podemos decir que

$$S_n = n\mu + O(\sqrt{n}\sigma),$$

donde la igualdad anterior debe interpretarse que los valores de S_n son cercanos a $n\mu$ pero con una dispersión del orden de $\sqrt{n}$. Otra forma de convencerse es reparafrasear la LGN donde sabemos que S_n/n converge en L^2 a μ, por lo que $\mathbb{E}((S_n/n-\mu)^2) \overset{n}{=} C/n$, y por tanto su dispersión al readedor de la media μ es del orden de $1/\sqrt{n}$ y podemos escribir

$$\frac{S_n}{n} = \mu + O\left(\frac{1}{\sqrt{n}}\right)$$

Para estudiar el comportamiento asintótico de S_n es razonable normalizarlo, para que esté centrado en cero y su varianza no dependa de n. En este sentido se considera la v.a.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}, \quad n \in \mathbb{N},$$

donde Z_n es la normalización de S_n que satisface

$$\mathbb{E}(Z_n) = 0, \qquad \mathrm{Var}(Z_n) = 1.$$

Observar que la normalización Z_n es lineal en S_n, i.e., $Z_n = aS_n + b$, para ciertos $a, b \in \mathbb{R}$.

Ahora podemos preguntarnos,

> *¿Es esperable que las v.a. Z_n tengan una distribución límite conocida? En tal caso, ¿será la misma que la de X_1?*

La respuestas a estas preguntas son: SÍ y NO, (en ese orden). Esto es, Z_n tiene una distribución límite, pero no tiene por qué ser la misma que la distribución de X_1. Lo sorprendente es que, sin importar la distribución de X_1, la distribución límite es siempre la misma: es la distribución normal estándar $\mathcal{N}(0,1)$!

Este resultado se conoce como el *teorema central del límite* y es uno de los teoremas más importantes y bellos de la matemática.

Teorema 7.10 (Teorema central del límite) *Sean $\{X_n\}$ una sucesión de v.a. i.i.d. en L^2, con esperanza μ y varianza σ^2. Si $S_n = X_1 + \cdots + X_n$, con $n \in \mathbb{N}$, entonces*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

siendo $\mathcal{N}(0,1)$ una v.a. gaussiana estándar.

En otras palabras el TCL dice que la normalización $Z_n := \dfrac{S_n - n\mu}{\sqrt{n}\sigma}$ satisface

$$\lim_n \mathbb{P}(Z_n \leqslant x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \quad \forall\, x \in \mathbb{R}.$$

Este teorema fue probado, para el caso particular de variables aleatorias Bernoulli, por de Moivre y Laplace a mediados del siglo XVIII. La idea detrás de la prueba es simple, pero requiere "fuerza bruta" de cálculo, donde fórmula de Stirling mediante se prueba

$$\mathbb{P}(S_n \leqslant \sqrt{np(1-p)}x + np) = \sum_{j=1}^{k(n)} \binom{n}{j} p^j (1-p)^j \xrightarrow{n} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$$

siendo $k(n)$ la parte entera de $\sqrt{np(1-p)}x + np$.

La versión general del teorema la probó Lyapunov en 1901.

Hay muchas versiones y pruebas distintas de este teorema. En este curso seguiremos técnicas basadas en *análisis de Fourier*.

7.3. Función Característica

En esta sección daremos un pantallazo sobre *funciones características* de variables aleatorias, las cuales nos dan una versión alternativa para estudiar la distribución de las mismas. Veremos cómo determinan la distribución de una v.a., para luego ver cómo la función característica nos permite estudiar límites en distribución de sucesiones de variables aleatorias.

Definición 7.11. La *función característica* de una v.a. X, con función de distribución μ_X, se define como la función $\varphi_X : \mathbb{R} \to \mathbb{C}$ definida por

$$\varphi_X(t) := \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx}\, d\mu_X(x), \quad t \in \mathbb{R},$$

Aquí "i" es la unidad imaginaria de los números complejos, y extendemos naturalmente la definición de esperanza de variables aleatorias que toman valores complejos. Esto es, si $Z = X + iY$, donde X e Y son v.a., entonces definimos $\mathbb{E}(Z) := \mathbb{E}(X) + i\mathbb{E}(Y)$. Luego tenemos

$$\varphi_X(t) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)).$$

Observación 7.12. La función caracterísitca puede entenderse como una función que mapea una medida de Borel de probabilidad μ_X en $\mathbb{R}$, en una función $\varphi_X : \mathbb{R} \to \mathbb{C}$.

Como la función e^{itX} está acotada (por 1), entonces $\varphi_X(t)$ está bien definida en todo $t \in \mathbb{R}$ (la integral está definida).

7.3.1. Función generatriz

Un caso de particular importancia es considerar la función característica de variables aleatorias discretas que toman valores sobre $0,1,2,\ldots$. Si X es una variable aleatoria con distribución $\sum_k p_k \delta_k$, i.e., X toma el valor del natural k con probabilidad p_k, entonces resulta

$$\varphi_X(t) = \sum_{k=0}^{\infty} p_k e^{itk}, \qquad t \in \mathbb{R}.$$

Definiendo $G_X(s) = \sum_{k=0}^{\infty} p_k s^k$, tenemos $\varphi_X(t) = G_X(e^{it})$, para todo $t \in \mathbb{R}$. La función G_X es la función generatriz de la variable aleatoria X. Veamos esto.

Una forma de codificar una sucesión de números $\{a_n\}_{n\in\mathbb{N}}$ es considerar la *función generatriz* de la sucesión, a saber,

$$a_0 + a_1 s^1 + a_2 s^2 \cdots$$

Definición 7.13. Sea X una v.a. que toma valores naturales $0,1,2,\ldots$, y sea p_X la función de probabilidad puntual asociada. La *función generatriz de probabilidad de X*, se define por

$$G_X(s) := \sum_{k=0}^{\infty} p_X(k) s^k, \quad s \in \mathbb{C}$$

para s donde la serie converja absolutamente.

Observar que si $|s| \leqslant 1$, se tiene $\sum_{k=0}^{\infty} p_X(k)|s|^k \leqslant \sum_{k=0}^{\infty} p_X(k) = 1$, y por lo tanto G_X converge al menos en el disco unidad $\{z \in \mathbb{C} : |z| \leqslant 1\}$.

♢ **Ejer 7.14.** Observar que $G_X(s) = \mathbb{E}(s^X)$, y que además $G_X(0) = p_X(0)$ y $G_X(1) = 1$.

Si G_X es una función generatriz de probabilidad, entonces podemos obtener la función de probabilidad puntual tomando sucesivas derivadas en 0. Esto es,

$$p_X(k) = G^{(k)}(0)/k!, \quad k = 0,1,2,\ldots$$

De esta manera, las funciones generatrices de probabilidad determinan unívocamente la distribución de la variable aleatoria. Es decir, si $G_X(s) = G_Y(s)$ en un entorno de cero, entonces las distribuciones de X e Y son las mismas.

Veamos algunos ejemplos de funciones generatrices de probabilidad. (Recordar que dado $p \in (0,1)$, denotamos $q = 1-p$.)

Ejemplo 7.15 (Distribución geométrica)**.** Si X es una v.a. con distribución Geo(p),

$$G_X(s) = \frac{ps}{1-ps}.$$

Ejemplo 7.16 (Distribución Bernoulli)**.** Si X es una v.a. con distribución $\mathcal{B}(p)$,

$$G_X(s) = q + ps.$$

Ejemplo 7.17 (Distribución Binomial)**.** Si X es una v.a. con distribución $\mathcal{B}(n,p)$,

$$G_X(s) = (q+ps)^n.$$

Ejemplo 7.18 (Distribución Poisson)**.** Si X es una v.a. con distribución Geo(p),

$$G_X(s) = e^{\lambda(s-1)}.$$

Una de las virtudes de las funciones generatrices de probabilidad es que permiten obtener propiedades de la distribución de una v.a. de manera más elegante y sencilla.

Una aplicación importante en este sentido es conocer los *momentos* de una variable aleatoria definida en un espacio de probabilidad discreto.

Definición 7.19. Sea X una v.a. sobre un espacio de probabilidad discreto. Dado $k = 1,2,\ldots$, definimos el *momento de orden k* de X a la cantidad $\mathbb{E}(X^k)$.

Por ejemplo, para conocer la varianza de X es suficiente conocer los dos primeros momentos.

◇ **Ejer 7.20.** Si X es una v.a. que toma valores naturales, con función de generatriz de probabilidad G_X. Entonces

$$G_X^{(k)}(1) = \mathbb{E}(X(X-1)\cdots(X-k+1)), \quad k = 1,2,\ldots.$$

(Se puede utilizar el siguiente resultado de Abel: si $A(s) = \sum_{k=0}^{\infty} a_k s^k$ es convergente en $|s|<1$, entonces $\sum_k a_k = \lim_{s\to 1} A(s)$ (pudiendo ser ambos infinitos).

◇ **Ejer 7.21.** Si X es un v.a. que toma valores naturales, dar una expresión de la varianza de X en función de las derivadas de G_X.

◇ **Ejer 7.22.** Rehacer cálculos de esperanza y varianza de las distribuciones geométricas y Poisson mediante este método.

Como ya hemos visto a lo largo del curso, la teoría de Probabilidad tiene especial interés en entender *suma* de variables aleatorias independientes. En lo que queda de esta sección estudiaremos estas sumas de v.a. independientes definidas sobre un mismo espacio de probabilidad discreto, y utilizaremos la poderosa herramienta de funciones generatrices para el caso de v.a. que toman valores naturales.

Si X e Y son v.a. sobre un mismo espacio, ¿cómo se calcula la función de probaiblidad puntual de la v.a. $Z = X+Y$. Dado que Z toma el valor z si $X = x$ y $Y = z-x$, se tiene que

$$p_Z(z) = \mathbb{P}\left(\bigcup_{x\in\Omega_X} \{X=x\}\cap\{Y=z-x\}\right) = \sum_{x\in\Omega_X} \mathbb{P}(X=x, Y=z-x),$$

siendo Ω_X el recorrido, o imagen, de la variable aleatoria X. En el caso de ser independientes se obtiene la siguiente fórmula

$$p_Z(z) = \sum_{x\in\Omega_X} p_X(x)p_Y(z-x), \quad z\in\mathbb{R}. \tag{7.2}$$

Esta fórmula se conoce como la *convolución* de las funciones p_X y p_Y. Sin embargo, cuando tenemos más variables aleatorias, la fórmula anterior para encontrar p_Z resulta engorrosa dado que debemos realizar $n-1$ convoluciones consecutivas.

Para el caso de v.a. que toman valores naturales, podemos utilizar la poderosa herramienta de funciones generatrices de probabilidad.

Teorema 7.23 *Sean $X_1,\ldots,X_n$, v.a independientes tomando valores en los naturales, entonces*

$$G_{X_1+\cdots+X_n}(s) = G_{X_1}(s)\cdots G_{X_n}(s). \tag{7.3}$$

Demostración. Se tiene

$$\begin{aligned} G_{X_1+\cdots+X_n}(s) &= \mathbb{E}(s^{X_1+\cdots+X_n}) \\ &= \mathbb{E}(s^{X_1}\cdots s^{X_n}) \\ &= \mathbb{E}(s^{X_1})\cdots\mathbb{E}(s^{X_n}) = G_{X_1}(s)\cdots G_{X_n}(s) \end{aligned}$$

donde en la penúltima igualdad usamos la independencia de las v.a. s^{X_i}. □

◇ **Ejer 7.24.** Hallar distribución de la suma de v.a. indpendientes con distribución $\mathcal{B}(n,p)$ y $\mathcal{B}(m,p)$ respectivamente. Concluir que la distribución de la suma de n variables aleatorias de Bernoulli de parámetro p tiene distribución Binomial de parámetro n y p.

◇ **Ejer 7.25.** Mostrar que la suma de v.a. independientes con distribución $\mathcal{P}(\lambda)$ y $\mathcal{P}(\mu)$ respectivamente, tiene distribución $\mathcal{P}(\lambda+\mu)$.

7.3.2. Caso general

Observar que en este sentido la función característica permite extender de manera natural la noción de función generatriz (la cual tenía propiedades interesantes sobre suma de variables aleatorias).

En el caso continuo también tenemos una fórmula sencilla. Si X es una v.a. continua con densidad f, resulta

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx, \tag{7.4}$$

donde entendemos la integral anterior como el número complejo con parte real $\int_{-\infty}^{+\infty}\cos(tx)f(x)\,dx$ y parte imaginaria $\int_{-\infty}^{+\infty}\sin(tx)f(x)\,dx$.

La fórmula (7.4) se conoce como la *transformada de Fourier* de la función f.

Veamos algunas propiedades sencillas de la función característica.

◇ **Ejer 7.26.** $\varphi_X(0)=1$

◇ **Ejer 7.27.** $|\varphi_X(t)|\leqslant 1$.

◇ **Ejer 7.28.** $\varphi_X(-t)=\overline{\varphi_X(t)}$, donde $\bar{z}$ es el conjugado de $z\in\mathbb{C}$.

◇ **Ejer 7.29.** $\varphi_{aX+b}(t)=\varphi_X(at)e^{itb}$.

También se puede probar, usando la regla de Leibnitz de intercambio entre derivada y signo de integral, que las sucesivas derivadas satisfacen

$$\varphi_X'(0)=\mathbb{E}(iX),\quad \varphi_X''(0)=-\mathbb{E}(X^2)\quad,\ldots,\quad \varphi_X^{(n)}(0)=i^n\mathbb{E}(X^n),\tag{7.5}$$

siempre que los momentos $\mathbb{E}(X^k)$ existen.

◇ **Ejer 7.30.** Probar el resultado anterior utilizando que el límite del cociente incremental se puede intercambiar el símbolo de esperanza (la integral) observando que la función en cuestión está uniformemente acotada por una integrable (de hecho está acotada por 1).

Al igual que ocurrió con las funciones generatrices de probabilidad, las funciones características determinan unívocamente las distribuciones de probabilidad. (Lo asumimos sin demostración.)

Lema 7.31. *Si X e Y tienen iguales funciones características, entonces X e Y tienen la misma distribución.*

◇ **Ejer 7.32.** Probar el Lema anterior.

La prueba del lema anterior se basa en una “fórmula de inversión” donde se puede recuperar la distribución a través de la función característica. (Otra forma posible es similar al caso de funciones generatrices, donde tomando derivadas en cero podemos recuperar los momentos. Esto motiva una problema interesante, a saber, cuándo los momentos determinan la distribución de un v.a.)

El siguiente resultado es crucial, y es el que sugiere la utilización de funciones características para estudiar sumas de variables aleatorias.

Lema 7.33. *Sean X e Y v.a. independientes. Entonces $\varphi_{X+Y}(t)=\varphi_X(t)\varphi_Y(t)$, para todo $t\in\mathbb{R}$.*

Veamos ahora la propiedad de estabilidad de la distribución gaussiana, a saber, la función caracacterística de una v.a. gaussiana es una función de densidad gaussiana.

Teorema 7.34 *Si X tiene distribución $\mathcal{N}(0,1)$, entonces*

$$\varphi_X(t)=e^{-t^2/2},\quad t\in\mathbb{R}.$$

Demostración. Completando cuadrados resulta

$$\begin{aligned}\varphi_X(t)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{itx}e^{-x^2/2}\,dx\\&=\frac{e^{-t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-(x-it)^2/2}\,dx\end{aligned}$$

donde la última integral resulta que toma el valor uno por ser la densidad de una normal centrada en it.[1] □

Comentario 7.35. En otras palabras, el resultado anterior dice que la densidad gaussiana es un punto fijo (a menos de una constante) de la transformada de Fourier (7.4).

◇ **Ejer 7.36.** Utilizando las propiedades de la función característica se concluye lo siguiente.

$$X\sim\mathcal{N}(\mu,\sigma^2),\quad \text{entonces}\quad \varphi_X(t)=e^{it\mu-\sigma^2t^2/2}.$$

Esto nos permite concluir un resultado interesante.

Teorema 7.37 *Si $X_1,\ldots,X_n$, son v.a. gaussianas independientes, entonces $X_1+\cdots+X_n$ es nuevamente una v.a. gaussiana.*

◇ **Ejer 7.38.** Probar el teorema anterior. ¿Cuál sería la esperanza y varianza?

Comentario 7.39. Una prueba alternativa, y con menos prerequisitos, es hacerla por convolución. Pero es muy complicada. El resultado sigue siendo cierto para el caso de v.a. gaussianas dependientes. Sin embargo la varianza y la esperanza no serán lineales en sus miembros. (Cf. distribución gaussiana multivariada.)

A continuación enunciaremos un resultado crucial sobre la relación entre convergencia en distribución y convergencia de las funciones características. (Lo asumimos sin demostración.)

Teorema 7.40 (Teorema de continuidad) *Sea Z una v.a. y $\{Z_n\}$ una sucesión de v.a. tales que φ_Z es continua en 0, y se satisface*

$$\lim_n \varphi_{Z_n}(t)=\varphi_Z(t),$$

para todo t punto de continuidad de φ_Z.

Entonces, $Z_n\xrightarrow{\mathcal{D}}Z$, i.e.

$$\lim_n \mathbb{P}(Z_n\leqslant x)=\mathbb{P}(Z\leqslant x),$$

para todo punto x de continuidad de la distribución $x\mapsto\mathbb{P}(Z\leqslant x)$. □

[1] Para evitar justificar correctamente los pasos anteriores se da una prueba formal. Derivando bajo el signo de integral (regla de Leibnitz), se prueba que $\varphi_X'(t)=-t\varphi_X(t)$. Luego resulta $\frac{d}{dt}(\varphi_X(t)\exp^{t^2/2})=0$. Luego se concluye el resultado usando el hecho que $\varphi_X(0)=1$.

Corolario 7.41. *Si $Z_1, Z_2, \ldots$, son v.a. tales que sus funciones características $\varphi_{Z_1}, \varphi_{Z_1}, \ldots$, satisfacen*

$$\lim_n \varphi_{Z_n}(t) = e^{-t^2/2},$$

para todo $t \in \mathbb{R}$, entonces,

$$\lim_n \mathbb{P}(Z_n \leqslant x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du, \quad x \in \mathbb{R}.$$

En otras palabras, si las funciones características de unas sucesión de v.a. tienden puntualmente a la función característica de una v.a. $\mathcal{N}(0,1)$, entonces, la sucesión tiende en distribución a una $\mathcal{N}(0,1)$.

Demostración del Teorema 7.10 . Primero, por simplicidad, observemos que podemos escribir $Z_n = (Y_1 + \cdots + Y_n)/\sqrt{n}$, donde cada una de estas nuevas v.a. Y_k son i.i.d, centradas y con varianza 1. Entonces usando Lema 7.33 se tiene

$$\varphi_{Z_n}(t) = \varphi_{Y_1/\sqrt{n}}(t) \cdots \varphi_{Y_n/\sqrt{n}}(t) = \left(\varphi_{Y_1}(t/\sqrt{n})\right)^n, \quad t \in \mathbb{R}. \tag{7.6}$$

En las hipótesis del teorema resulta que φ_{Y_1} tiene primera y segunda derivada. Haciendo el desarrollo de Taylor de φ_{Y_1}, en un entorno de cero, se tiene

$$\varphi_{Y_1}(u) = 1 + \varphi'_{Y_1}(0)u + \varphi''_{Y_1}(0)u^2/2 + o(u^2).$$

Usando (7.5), resulta

$$\varphi_{Y_1}(u) = 1 - u^2/2 + o(u^2), \tag{7.7}$$

para u en un entorno de 0.

Fijemos $t \in \mathbb{R}$. De (7.6) y (7.7), (para n grande tal que t^2/n esté en el entono de cero del desarrollo (7.7)), se tiene

$$\varphi_{Z_n}(t) = \left(1 - t^2/(2n) + o(t^2/n)\right)^n,$$

donde resulta

$$\lim_n \varphi_{Z_n}(t) = e^{-t^2/2}.$$

□

Una forma de justificar el límite anterior es la siguiente. Para t fijo, si escribimos $a_n = 1 - t^2/(2n)$ y $b_n = 1 - t^2/(2n) + o(1/n)$, entonces

$$\begin{aligned} |a_n{}^n - b_n{}^n| &= |(a_n - b_n)(a_n{}^{n-1} + a_n{}^{n-2} b_n + \cdots + a_n{}^1 b_n{}^{n-2} + b_n{}^{n-1})| \\ &\leqslant |(a_n - b_n)| n = o(1/n) n, \end{aligned}$$

y por lo tanto ambas sucesiones tienen igual límite cuando $n \to +\infty$. Luego se observa que

$$\lim_n a_n{}^n = e^{-t^2/2}.$$

◊ **Ejer 7.42.** Asumiendo que la función característica de una v.a con distribución de Cauchy es $\varphi_X(t) = e^{-|t|}$, $t \in \mathbb{R}$, probar que el TCL no es cierto para suma de v.a. i.i.d. con distribución de Cauchy. ¿Qué distribución tiene la suma normalizada?

Proposición 7.43. *Observar que se puede recuperar la ley débil de los grandes números usando el TCL.*

7.4. Aplicaciones del TCL*

En esta última sección de estas notas veremos algunas aplicaciones del TCL. La idea básica en cada ejemplo es aproximar las sumas normalizadas Z_n por una v.a. con distribución normal. Es posible tomar esta aproximación con todo rigor, ya que se puede conocer la velocidad de convergencia y por lo tanto cuantificar el error en esta aproximación. Existen varios teoremas en este sentido, como el de *Berry-Essen*, que dice que si el tercer momento de las v.a. es finito, entonces la aproximación tiene orden $1/\sqrt{n}$, (donde la constante depende del segundo y tercer momento). Este tipo de resultados se pueden ver en cursos de *Procesos Estocásticos*.

7.4.1. Aproximación normal de la binomial

Si S_n tiene distribución binomial de parámetros n y p (i.e. $S_n \sim \mathcal{B}(n,p)$), entonces cuando n es grande, y p se mantiene constante, se tiene del TCL que

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant x\right) \approx \Phi(x), \quad x \in \mathbb{R},$$

Donde Φ es distribución de una $\mathcal{N}(0,1)$.

Este resultado se conoce como la *aproximación normal a la binomial*. Es posible probar este resultado mediante cálculos elementales (ver Feller [?Feller] por ejemplo.)

Ejemplo 7.44. Supongamos que en un estadio quieren hacer un concierto para 10 mil personas, y deciden utilizar la cancha para colocar sillas. Como hay dos entradas independientes, y por razones de seguridad, deciden hacer dos zonas divididas por una valla. Se distribuyen las sillas en igual cantidad a ambos lados. Si asumimos (la no tan real idea de) que las personas elijen al azar con igual probabilidad cada entrada (a la entrada tiran una moneda fiel y dependiendo del resultado elijen en qué puerta entrar). Si quieren estar seguros en un 99 % de que que nadie quede sin silla. ¿Cuántas sillas deben poner en cada lado?

Sea $n = 10000$. Con las hipótesis dadas, si S es el número de personas que entra a la primer entrada, podemos asumir que S es una v.a. con distribución $\mathcal{B}(n, 1/2)$.

Sea s la cantidad de sillas necesarias a cada lado. Lo que queremos es que s satisfaga $\mathbb{P}(S \leqslant s) = 0{,}99$. Entonces usando la aproximación a la normal se tiene

$$\begin{aligned} 0{,}01 = \mathbb{P}(S > s) &= \mathbb{P}\left(\frac{S - n/2}{\sqrt{n/4}} > \frac{s - n/2}{\sqrt{n/4}}\right) \\ &\approx 1 - \Phi\left(\frac{2s - n}{\sqrt{n}}\right). \end{aligned}$$

Se tiene $\Phi(2{,}34) \approx 0{,}99$, por lo cual debemos resolver $\frac{2s-n}{\sqrt{n}} = 2{,}34$, y de de donde resulta

$$s = \frac{n}{2} + 1{,}17\sqrt{n}.$$

Esto es, para $n = 10000$, $s = 5117$. Esto es, se necesitan comprar 234 sillas extras. Sorprendente, no?

Lo anterior resulta más interesante cuando n es más grande aún. Observar que $s \approx n/2 + \sqrt{n}$, y por lo tanto el cociente $s/(n/2)$ es aproximadamente $2/\sqrt{n}$, y por lo tanto tiende a cero a cuando n crece. Esto significa que cuando n crece la proporción de sillas extras tiende a cero. Este fenómeno de cancelación se conoce como *concentración de la medida*.

7.5. Muestra estadística*

Hay un porcentaje p de electores que votan en blanco en las elecciones. Queremos estimar p, haciendo una encuesta, con un error menor a 0.005. ¿qué tan grande tiene que ser la muestra?

Sea n el tamaño de la muestra (variable a determinar). Sea S_n la v.a. suma de votantes que votan en blanco. Podemos asumir que $S_n \sim \mathcal{B}(n, p)$. Por una convención utilizada en *estadística*, denotamos por $\widehat{p}$ al promedio anterior

Queremos encontrar n tal que $|\widehat{p} - p| \leqslant 0{,}005$. Es claro que esto no puede ocurrir siempre, dado que la v.a. $|S_n/n - p|$ puede tomar valores mayores(por ejemplo si todos los de la encuesta votaban en blanco). La forma correcta de plantear el problema es dar una probabilidad para que ocurra ese error. Por ejemplo, hallar n para que el error del $|\widehat{p} - p| \leqslant 0{,}005$ ocurra con probabilidad mayor al 95 %. Esto es,

$$\mathbb{P}(|\widehat{p} - p| \leqslant 0{,}005) \geqslant 0{,}95.$$

Usando la aproximación dada por el TCL obtenemos

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant 0{,}005\right) &= \mathbb{P}\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leqslant \frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&\approx \Phi\left(\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2\Phi\left(\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \\
&\geqslant 2\Phi\left(0{,}005\sqrt{4n}\right) - 1,
\end{aligned}$$

(donde la última desigualdad es consecuencia de que $p(1-p) \leqslant 1/4$. Resulta que si $0{,}005\sqrt{4n} \geqslant 1{,}96$, entonces resulta $2\Phi\left(0{,}005\sqrt{4n}\right) - 1 \geqslant 0{,}95$. Esto es $n > 40000$.

Sin embargo, si sólo demandamos un error del 3 %, se prueba que $n \approx 1000$ funciona. Este número es el que generalmente se utiliza en las encuestas que vemos a diario.

7.6. Aguja de Buffon reconsiderada*

En esta sección mostraremos cómo estimar π con el problema visto de la Aguja de Buffon (ver Ejercicio **??**).

Supongamos que tiramos un aguja de longitud ℓ sobre la hoja de un cuaderno a rayas. Si r es la distancia entre lineas, y $\ell \leqslant r$, resulta que la probabilidad p que la aguja tirada al azar intersecte alguna de las lineas es igual a

$$p = \frac{2\ell}{\pi r}.$$

Veamos una prueba distinta a la que realizaron los estudiantes en clase. A diferencia de la prueba vista, que se basaba en el cálculo de una integral, esta prueba utiliza argumentos probabilísticos y geométricos. La principal herramienta a usar se basa en la linealidad de la esperanza.

Existen varias referencias sobre la versión de la prueba que daremos. Recomendamos la biblia sobre geometría integral de Santaló [?Santalo], y en particular la introducción del libro de Klain y Rota [?Rota].

Supongamos por el momento que tiramos al azar una aguja de longitud ℓ_1 (sin la restricción $\ell_1 \leqslant r$). Entonces si definimos la v.a. X_1 que cuenta la cantidad de cruces con las lineas, resulta

$$\mathbb{E}(X_1) = 0p_0 + p_1 + 2p_2 + \cdots,$$

donde p_i es la probabilidad de que la aguja corte las lineas en exactamente i lugares. Por lo tanto si $\ell_1 \leqslant r$ resulta

$$\mathbb{E}(X_1) = p_1,$$

que es la cantidad buscada. Por lo tanto reduciremos nuestro problema a calcular $\mathbb{E}(X_1)$.

Consideremos otra aguja de longitud ℓ_2 que tiramos también al azar. Sea X_2 la v.a. que cuenta los puntos de intersección de esta nueva aguja. Supongamos que la trasladamos para luego unirla a un extremo de la otra aguja. De esta manera podemos pensar en una“visagra” aleatoria. El número de puntos de intersección de la visagra con las lineas es $X_1 + X_2$. Observar que ambas v.a. son dependientes, pero sin embargo, por la linealidad de la esperanza se tiene

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2.$$

(Observar que es natural asumir que la ley de X_2 no tiene ninguna preferencia, y por tanto puede considerarse que tiene el mismo azar que la aguja original. También se puede trazar un argumento más convincente de la siguiente manera. Si asumimos que un extremo de la aguja tiene distribución en altura uniforme, y el ángulo de la aguja con distribución uniforme e independientes, entonces, al unir la nueva aguja tirada con el mismo azar por el extremo que conocemos, obtenemos una nueva aguja que individualmente tiene el azar mencionado.)

Este razonamiento se puede extender a una “cadena” de k agujas que son unidas por uno de sus extremos a otra.

Es interesante, y llamativo, observar que si las agujas se unen de manera que quede un segmento de recta, entonces la esperanza del números de cortes total sigue siendo igual a la esperanza del número de cortes de la cadena. Es decir, la forma no importa (sólo que sean segmentos de rectas unidos).

Es claro que $\mathbb{E}(X_1)$ es una función que sólo depende de la longitud ℓ_1. Sea f la función dada por $f(\ell_1) = \mathbb{E}(X_1)$.

Afirmación 1: Resulta que f satisface

$$f(\ell_1 + \ell_2) = f(\ell_1) + f(\ell_2).$$

Esto resulta de que una aguja de longitud $\ell = \ell_1 + \ell_2$ puede ser pensada como la unión de dos agujas unidas por los extremos de longitudes ℓ_1 y ℓ_2 respectivamente, y entonces se tiene

$$f(\ell_1 + \ell_2) = \mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 = f(\ell_1) + f(\ell_2)$$

Afirmación 2: f es lineal sobre $\mathbb{Q}$:

Basta ver que la Afirmación 1 implica el resultado tomando segmentos unitarios. Se tiene $f(k\ell) = kf(\ell)$, y además $f(1/n)n = f(1)$. Entonces se tiene $f(k/n) = f(1)k/n$.

Afirmación 3: existe $k \in \mathbb{R}$, tal que $f(\ell) = k\ell$, para todo $\ell \in \mathbb{R}$.

La prueba resulta de ser lineal en $\mathbb{Q}$ y de ser monótona creciente.

Afirmación 4: Sea $\mathcal{C}$ es un alambre rígido de longitud ℓ (una curva rectificable de longitud ℓ)[2] que se tira al azar. Entonces si X es la v.a. que cuenta la cantidad de intersecciones, resulta

$$\mathbb{E}X = k\ell$$

La prueba sigue de aproximar $\mathcal{C}$ por poligonales. Esto es, si $\mathcal{C}_n$ es una poligonal fija de longitud ℓ_n que aproxima $\mathcal{C}$, entonces, el número de cortes de la poligonal (al azar) $\mathcal{C}_n$ tiende al número de cortes de $\mathcal{C}$ tirados al azar. Dado que la esperanza de número de cortes de esta poligonal es $f(\ell_n)$, resulta tomando límite en n que $\mathbb{E}X = f(\ell) = k\ell$.

Afirmación 5: Se tiene $k = 2/(\pi r)$.

Basta tomar una curva donde podamos computar la esperanza de cortes. Si consideramos $\mathcal{C}$ el círculo de diámetro r, resulta que el número de cortes, que denotamos por X, es una v.a. con distribución constante igual a 2. Entonces de la Afirmación 5 se tiene

$$2 = \mathbb{E}(X) = k\mathrm{long}(\mathcal{C}) = k\pi r, \quad \text{entonces} \quad k = \frac{2}{\pi r}.$$

Hemos concluido que la probabilidad p de intersección de una aguja de longitud ℓ tirada al azar es

$$p = \frac{2\ell}{\pi r}.$$

[2] Que su longitud se pueda aproximar por poligonales.

Comentario 7.45. Otro hecho interesante es que como no importa la forma del alambre, podemos tirar al azar un “fideo” de longitud ℓ y sobre la grilla para tener el mismo resultado. Claro está que para esto es un poco más difícil entender qué quiere decir tirar al azar.

Comentario 7.46. La técnica usada en esta prueba está basada en la *geometría integral*, la cual tiene como fin estudiar propiedades geométricas de objetos a través de integrar ciertas cantidades numéricas. A modo de ejemplo, se puede probar de manera similar a lo anterior, que la longitud de una curva encajada en la esfera S^2 se puede recuperar integrando sobre todos los “ecuadores” la cantidad de puntos de intersección del ecuador con la curva. (Una forma de considerar esta integral, es observar que podemos parametrizar los “ecuadores al azar” con el ecuador ortogonal a un vector con distribución uniforme sobre la esfera.

7.6.1. Estimación de π

Supongamos que lanzamos la aguja n veces. Queremos encontrar n tal que el error en la aproximación de π sea menor 0,001, con probabilidad mayor a 95%.

Sea N el número de veces que la aguja intersecta las lineas. Podemos suponer que N tiene distribución $\mathcal{B}(n,p)$.

Del TCL tenemos que

$$\frac{N - np}{\sqrt{np(1-p)}} \approx \gamma,$$

donde $\gamma \sim \mathcal{N}(0,1)$. (Acá estamos abusando de la notación, donde $Z_n \approx \gamma$ se lee, Z_n tiende en distribución a γ.) Entonces

$$N \approx np + \sqrt{np(1-p)}\gamma$$

i.e. N tiende en distribución a una v.a. con distribución $\mathcal{N}(np, np(1-p))$.

La probabildiad “empírica” es $\widehat{p} = N/n$, y sea $\widehat{\pi}$ la v.a. que aproxima a π. Esto es

$$\widehat{\pi} = \frac{2\ell}{r\widehat{p}}a = \pi\frac{p}{\widehat{p}} \quad \text{lo cual implica} \quad \widehat{\pi} - \pi = \pi\frac{p - \widehat{p}}{\widehat{p}},$$

Acá hacemos una aproximación nuevamente para poder simplificar los cálculos. Esto es considerar

$$\widehat{\pi} - \pi \approx \pi\frac{p - \widehat{p}}{p}.$$

Resulta entonces del TCL que

$$\widehat{\pi} - \pi \approx \mathcal{N}\left(0, \frac{\pi^2}{n}\frac{(1-p)}{p}\right).$$

Al igual que en la sección anterior, queremos independizarnos de p, por lo cual podemos tomar una aguja de longitud ℓ que minimice la varianza anterior. El mínimo ocurre cuando p es maximizado, y esto es si tomamos $\ell = r$. En tal caso se tiene $2/\pi$, y

$$\widehat{\pi} - \pi \approx \mathcal{N}\left(0, \frac{\pi^2}{n}\left(\frac{\pi}{2} - 1\right)\right).$$

Ahora estamos en condiciones de resolver nuestro problema.

Con las aproximaciones en cuestión se tiene que

$$\mathbb{P}(|\widehat{\pi} - \pi| \leqslant 0{,}001) \geqslant 0{,}95 \quad \text{si y sólo si} \quad \frac{0{,}001}{\frac{\pi}{\sqrt{n}}\sqrt{\left(\frac{\pi}{2} - 1\right)}} \geqslant 1{,}96,$$

donde utilizamos la aproximación de la sección anterior. Basta tomar

$$n > \left(\frac{1{,}96}{0{,}001}\right)^2 \pi^2 \left(\frac{\pi}{2} - 1\right)$$

Basta tomar n del orden de 21 millones. Es decir, mejor busquemos otro método para estimarlo :)

7.7. Estimación de integrales mediante el método de Monte Carlo*

Esta sección es la segunda entrega del curso.

El objetivo de esta tarea es dar una estimación de una integral $\int_a^b f(x)\mathrm{d}x$ que no seamos capaces de calcular utilizando las herramientas que aprendimos en el curso de Cálculo I, a partir de la *Ley (débil) de los Grandes Números*.

1. Probar que si la sucesión de v.a. $\{X_n\}_n$ y X son tales que $X_n \to_{\mathbb{P}} X$ y $g : \mathbb{R} \to \mathbb{R}$ es una función continua, entonces:

$$g(X_n) \to_{\mathbb{P}} g(X)$$

Sugerencia: suponga primero que g es uniformemente continua

2. Demostrar que si $X_1, X_2, \ldots, X_n$ son v.a. i.i.d. con $\mathbb{E}(X_1) = \mu$ y $\mathrm{var}(X_1) = \sigma^2$ finitos, entonces la v.a.

$$\sigma_n^2 := \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2$$

converge en probabilidad a σ^2.

3. Demostrar que si $U_1, U_2, \ldots, U_n$ son v.a. i.i.d. con distribución $U(a,b)$ y $f : [a,b] \to \mathbb{R}$ es integrable, entonces:

$$\frac{1}{n}\sum_{i=1}^n f(U_i) \to_{\mathbb{P}} \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x.$$

4. Utilizar el resultado anterior para dar una estimación de $\int_2^7 e^{-x^2}\mathrm{d}x$. Comparar con el resultado que se obtiene en *Wolfram Alpha* (por ejemplo) a medida que aumenta n.

5. Considere el *error (aleatorio)* que cometemos por aproximar la integral $\int_a^b f(x)\mathrm{d}x$ por $(b-a) \times \frac{1}{n}\sum_{i=1}^n f(U_i)$, llamémosle:

$$E_n := \frac{1}{n}\sum_{i=1}^n f(U_i) - \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x.$$

   a) Acotar la probabilidad de que E_n sea más grande que $\varepsilon > 0$ utilizando la *desigualdad de Chebyshev*. Indique a partir de esta cota cuán grande tiene que ser n si quiero estar seguro de que con probabilidad 0,98 el verdadero valor de la integral dista de mi estimación menos de 0,01.

   b) Suponiendo que:

$$\sigma^2 := \frac{1}{b-a}\int_a^b (f(x))^2\mathrm{d}x - \left(\frac{1}{b-a}\int_a^b f(x)\mathrm{d}x\right)^2$$

   fuera conocido, probar que la función de distribucoón de E_n puede ser aproximada por $\Phi\left(\frac{\sqrt{n}x}{\sigma}\right)$.

   Observar que si $f(x) = e^{-x^2}$, como antes, entonces σ^2 sigue siendo desconocida. No se preocupe, ya veremos cómo arreglamos esto.

   c) ¿Cómo estimaría σ?

6. *a*) Utilizando Φ y esa estimación de σ, calcule cuán grande tiene que ser n si quiero asegurar que el error que cometo al aproximar $\int_2^7 e^{-x^2}dx$ por $\frac{5}{n}\sum_{i=1}^n e^{-u_i^2}$ es menor que 0.01 con probabilidad (aproximada) 0.98.

   b) Para ese valor de n, realice varias simulaciones. ¿Puede dar un intervalo tal que pueda asegurar que con probabilidad (aproximada) 0.98 contiene al verdadero valor de la integral? Para esto podría resultarle conveniente graficar en el eje de las y el valor obtenido para cada simulación y en el eje de las x numerar las repeticiones de la simulación. Observe si para alguna simulación la estimación de la integral se escapa de dicho intervalo. [3]

7.7.1. Solución

A continuación agregamos una solución a la entrega realizada por el estudiante Facundo Almeida.

1. Lo que queremos demostrar es que dado $\varepsilon_0 > 0$ se cumple que para todo $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que

$$\mathbb{P}(|g(X_n) - g(X)| > \varepsilon_0) < \varepsilon \qquad n \geqslant n_0.$$

   Comencemos tomando $\varepsilon_0, \varepsilon > 0$ arbitrarios. Considérese la función de distribución $F_X : \mathbb{R} \to \mathbb{R}$. Como sabemos, se

[3] simulamos $u_1, u_2, \ldots, u_n$ M.A.S. provinientes de $U_1, U_2, \ldots, U_n$ i.i.d. $U(a,b)$ con la función **runif()** de **R**.

cumple que $\lim_{x\to-\infty} F_X(x) = 0$ y $\lim_{x\to+\infty} F_X(x) = 1$. Por lo tanto, podemos encontrar números reales α y β, con $\alpha < \beta$, tales que $F_X(\alpha) < \varepsilon/4$ y $F_X(\beta) > 1-\varepsilon/4$.

Sea $A = [\alpha, \beta]$. Entonces

$$\begin{aligned}\mathbb{P}(X \notin A) &= 1 - \mathbb{P}(X \in A)\\ &= 1-(F_X(\beta)-F_X(\alpha))\\ &< 1-(1-\varepsilon/4-\varepsilon/4)\\ &= \varepsilon/2. \end{aligned} \tag{7.8}$$

A continuación observemos que A es un subconjunto compacto de $\mathbb{R}$. Como g es una función continua, es uniformemente continua en A. Luego existe $\delta > 0$ tal que $|g(x)-g(y)| < \varepsilon_0 \ \ \forall x, y \in A : |x-y| \leqslant \delta$. Además por hipótesis $X_n \xrightarrow{\mathbb{P}} X$, así que podemos tomar $n_0 \in \mathbb{N}$ tal que $\mathbb{P}(|X_n - X| > \delta) < \varepsilon/2$ para todo $n \geqslant n_0$. Veamos que esta elección de n_0 es adecuada.

Tomemos $n \geqslant n_0$ cualquiera. Si $\omega \in \{|X_n - X| \leqslant \delta, X \in A\}$, entonces $|X_n(\omega) - X(\omega)| \leqslant \delta$ y $X(\omega) \in A$, así que $|g(X_n(\omega)) - g(X(\omega))| \leqslant \varepsilon_0$, i.e. $\omega \in \{|g(X_n)-g(X)| \leqslant \varepsilon_0\}$. Por lo tanto,

$$\begin{aligned}&\{|X_n - X| \leqslant \delta, X \in A\} \subseteq \{|g(X_n)-g(X)| \leqslant \varepsilon_0, X\in A\}\\ \Longrightarrow\ &\mathbb{P}(|X_n - X| \leqslant \delta, X \in A) \leqslant \mathbb{P}(|g(X_n)-g(X)| \leqslant \varepsilon_0, X\in A)\end{aligned} \tag{7.9}$$

Por otro lado,

$$\begin{aligned}\mathbb{P}(|X_n - X| \leqslant \delta) &= \mathbb{P}(|X_n - X| \leqslant \delta, X\in A) + \mathbb{P}(|X_n - X| \leqslant \delta, X\notin A)\\ &\leqslant \mathbb{P}(|X_n - X| \leqslant \delta, X\in A) + \mathbb{P}(X\notin A)\\ \Longrightarrow\ \mathbb{P}(|X_n - X| \leqslant \delta, X\in A) &\geqslant \mathbb{P}(|X_n - X| \leqslant \delta) - \mathbb{P}(X\notin A).\end{aligned} \tag{7.10}$$

Entonces

$$\begin{aligned}\mathbb{P}(|g(X_n)-g(X)| \leqslant \varepsilon_0) &\geqslant \mathbb{P}(|g(X_n)-g(X)| \leqslant \varepsilon_0, X\in A)\\ &\geqslant \mathbb{P}(|X_n - X| \leqslant \delta, X\in A) \qquad \text{(por 2)}\\ &\geqslant \mathbb{P}(|X_n - X| \leqslant \delta) - \mathbb{P}(X\notin A) \qquad \text{(por 3)}\\ &> \mathbb{P}(|X_n - X| \leqslant \delta) - \varepsilon/2 \qquad \text{(por 1)}\\ \Longrightarrow\ \mathbb{P}(|g(X_n)-g(X)| > \varepsilon_0) &< 1 - \mathbb{P}(|X_n - X| \leqslant \delta) + \varepsilon/2\\ &< \mathbb{P}(|X_n - X| > \delta) + \varepsilon/2\\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

2. Para esta parte será conveniente contar con un resultado que nos hable de lo que pasa con el "límite en probabilidad" de la suma de dos sucesiones de variables aleatorias.

Lema 7.47. *Si $X_n \xrightarrow{\mathbb{P}} X$ e $Y_n \xrightarrow{\mathbb{P}} Y$, entonces $X_n + Y_n \xrightarrow{\mathbb{P}} X+Y$.*

Demostración. Sea $\varepsilon > 0$. Entonces

$$\begin{aligned}\mathbb{P}(|X_n + Y_n - X - Y| > \varepsilon) &\leqslant \mathbb{P}(|X_n - X| + |Y_n - Y| > \varepsilon) \qquad \text{(desigualdad triangular)}\\ &= 1 - \mathbb{P}(|X_n - X| + |Y_n - Y| \leqslant \varepsilon)\\ &\leqslant 1 - \mathbb{P}(|X_n - X| \leqslant \varepsilon/2,\ \ |Y_n - Y| \leqslant \varepsilon/2)\\ &= \mathbb{P}(|X_n - X| > \varepsilon/2 \text{ o } |Y_n - Y| > \varepsilon/2)\\ &\leqslant \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|Y_n - Y| > \varepsilon/2).\end{aligned}$$

Por hipótesis $\lim_{n\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon/2) = \mathbb{P}(|Y_n - Y| > \varepsilon/2) = 0$, así que $\lim_{n\to+\infty} \mathbb{P}(|X_n + Y_n - X - Y| > \varepsilon) = 0$, y el lema queda demostrado. □

Supongamos ahora que $X_1, \dots, X_n$ son como en la letra. Como $X_1, \dots, X_n$ tienen idéntica distribución, es claro de la definición de función distribución que $X_1^2, \dots, X_n^2$ tienen idéntica distribución. Además vimos en el teórico que la independencia de $X_1, \dots, X_n$ implica la independencia de $X_1^2, \dots, X_n^2$ (ya que para cada i se cumple $X_i^2 = g(X_i)$, donde $g : \mathbb{R} \to \mathbb{R}$ está dada por $g(x) = x^2$). Entonces, por la ley débil de los grandes números, $\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)_n$ converge en probabilidad a $\mathbb{E}(X_1^2)$. Por otro lado, la LGN también implica que $\left(\frac{1}{n}\sum_{i=1}^n X_i\right)_n$ converge en probabilidad a $\mathbb{E}(X_1)$; por la parte 1, $\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right)_n$ converge en probabilidad a $\mathbb{E}(X_1)^2$. Entonces, por el lema 1, σ_n converge en probabilidad a $\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \sigma^2$.

3. Para cada i sea $X_i = f(U_i)$. Argumentando como en la parte anterior, deducimos que las variables aleatorias $X_1, \dots, X_n$ son i. i. d.. Entonces, por la LGN, la sucesión $(\frac{1}{n}\sum_{i=1}^n X_i)_n$ converge en probabilidad a

$$\begin{aligned}\mathbb{E}(X_1) &= \mathbb{E}(f(U_1))\\ &= \int_{-\infty}^{\infty} f(x) f_{U_1}(x)\, dx\\ &= \frac{1}{b-a}\int_a^b f(x)dx \qquad (U_1 \text{ tiene distribución uniforme en } [a,b\end{aligned}$$

4. El resultado anterior nos dice que muestras de la variable aleatoria $(7-2)\frac{1}{n}\sum_{i=1}^n e^{-U_i^2} = \frac{5}{n}\sum_{i=1}^n e^{-U_i^2}$ sirven como aproximación para $\int_2^5 e^{-x^2}\, dx$, donde tomamos $f : [2,7] \to \mathbb{R}$ dada por $f(x) = e^{-x^2}$ y consideramos a las variables aleatorias U_i con distribución $\mathcal{U}(2,7)$.

Para la estimación utilizamos la función `runif` de R (que simula muestras de una variable aleatoria con distribución uniforme) como sigue: para un valor fijo de $n \geqslant 1$ tomamos n muestras aleatorias de una v.a. con distribución $\mathcal{U}(2,7)$ con ayuda de la función `runif`; evaluamos la función f en el valor obtienido de cada muestra; sumamos todos los resultados; multiplicamos todo por $\frac{5}{n}$. En resumen, ejecutamos la línea

```
5/n * sum(exp(-runif(n, 2, 7)^2))
```

(`exp` denota la función exponencial). En la siguiente tabla están registrados los resultados para distintos valores de n junto con el error de la aproximación con respecto al valor proporcionado por Wolfram Alpha: $\int_2^7 e^{-x^2}\, dx \approx$ 0,0041553.

n	`5/n * sum(exp(-runif(n, 2, 7)^2))`	error
5	0.0107149	0.0065596
10	0.0034393	0.0007160
100	0.0037871	0.0003682
500	0.0038945	0.0002608
1000	0.0040949	0.0000604
10000	0.0043993	0.0002440
100000	0.0041864	0.0000311

5. a) Notar que $\mathbb{E}(\frac{1}{n}\sum_{i=1}^n f(U_i)) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(f(U_i)) = \mathbb{E}(f(U_1)) = \frac{1}{b-a}\int_a^b f(x)\ dx$ (a esto último ya lo probamos en la parte 3). Si ponemos $Y_n := \frac{1}{n}\sum_{i=1}^n f(U_i)$, entonces $E_n = Y_n - \mathbb{E}(Y_n)$, y

$$\begin{aligned}
\mathbb{P}(|E_n| > \varepsilon) &\leqslant \frac{\mathbb{E}(E_n^2)}{\varepsilon^2} \qquad \text{(desigualdad de Chebyshev)}\\
&= \frac{\mathbb{E}((Y_n - \mathbb{E}(Y_n))^2)}{\varepsilon^2}\\
&= \frac{\operatorname{var}(Y_n)}{\varepsilon^2}\\
&= \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^n \operatorname{var}(f(U_i)) \qquad (*)\\
&= \frac{1}{\varepsilon^2 n}\operatorname{var}(f(U_1))\\
&= \frac{1}{\varepsilon^2 n}(\mathbb{E}(f(U_1)^2) - \mathbb{E}(f(U_1))^2)\\
&= \frac{1}{\varepsilon^2 n}\left(\frac{1}{b-a}\int_a^b f(x)^2\ dx - \left(\int_a^b f(x)\ dx\right)^2\right) \qquad (**)
\end{aligned}$$

(en $(*)$ usamos la independencia de las v.a. U_i para desarrollar la suma; en $(**)$ usamos la parte 3).

Lo que se pide es encontrar un valor de n para el cual $\mathbb{P}(|E_n| > 0{,}01) < 0{,}02$. Tomando $\sigma^2 = \frac{1}{b-a}\int_a^b f(x)^2\ dx - \left(\int_a^b f(x)\ dx\right)^2$, la inecuación a resolver es $\frac{1}{0{,}01^2 n}\sigma^2 < 0{,}02$. Bastará entonces con tomar $n > \frac{\sigma^2}{0{,}01^2 \cdot 0{,}02} = 5 \cdot 10^5\sigma^2$.

b) Para cada n sea $S_n = \sum_{i=1}^n f(U_i)$, y sea $\mu = \frac{1}{b-a}\int_a^b f(x)\ dx$. Notar que $E_n = \frac{1}{n}(S_n - n\mu)$ (parte 3) y que σ^2 es la varianza de $f(U_1)$ (parte 5-a)). Por el teorema central del límite, la sucesión $\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right)_n = \left(\frac{\sqrt{n}}{\sigma}E_n\right)_n$ converge en distribución a $\mathcal{N}(0,1)$, esto es,

$$\lim_{n\to+\infty}\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}E_n \leqslant x\right) = \Phi(x) \qquad \forall x \in \mathbb{R}$$
$$\Longrightarrow \lim_{n\to+\infty}\mathbb{P}(E_n \leqslant x) = \Phi\left(\frac{\sqrt{n}}{\sigma}x\right) \qquad \forall x \in \mathbb{R}.$$

En otras palabras, la función distribución de E_n se puede aproximar por $\Phi\left(\frac{\sqrt{n}}{\sigma}x\right)$.

c) Lo que nos interesa es obtener una cota superior razonable para σ, ya que vamos a usar el valor de σ para acotar por debajo el valor de $\Phi(\frac{\sqrt{n}}{\sigma}x)$, y la función Φ es creciente. Para eso podemos buscar una cota superior para $\int_a^b (f(x))^2\ dx = \int_2^7 e^{-2x^2}\ dx$ y una cota inferior para $\int_a^b f(x)\ dx = \int_2^7 e^{-x^2}\ dx$. Bastará con acotar los integrandos; esto es sencillo porque ambos son funciones decrecientes. Tenemos

$$\int_2^7 e^{-2x^2}\ dx \leqslant (7-2)e^{-2\cdot 2^2} = 5e^{-8},$$
$$\int_2^7 e^{-x^2}\ dx \geqslant (7-2)e^{-7^2} = 5e^{-49}.$$

Así, $\sigma^2 = \frac{1}{5}\int_2^7 e^{-2x^2}\ dx - \left(\frac{1}{5}\int_2^7 e^{-x^2}\ dx\right)^2 \leqslant e^{-8} - (e^{-49})^2$. Luego $\sigma \leqslant \sqrt{e^{-8} - (e^{-49})^2} \approx 0{,}01832$.

6. a) Queremos encontrar n lo suficientemente grande para que $\mathbb{P}(|E_n| \leqslant 0{,}01) \geqslant 0{,}98$. Sustituyendo la función distribución de E_n por $\Phi(\frac{\sqrt{n}}{\sigma}x)$, la inecuación queda $\Phi(\frac{\sqrt{n}}{\sigma}0{,}01) - \Phi(-\frac{\sqrt{n}}{\sigma}0{,}01) \geqslant 0{,}98$, esto es, $\Phi(\frac{\sqrt{n}}{\sigma}0{,}01) \geqslant 0{,}99$. Ahora reemplazamos σ por la aproximación que calculamos en el punto anterior[4] para obtener $\Phi(0{,}5459\sqrt{n}) \geqslant 0{,}99$. Observando la tabla del capítulo 8 de las notas, vemos que será suficiente con que $0{,}5459\sqrt{n} \geqslant 2{,}33$. Despejando resulta que nos servirá n tal que $n \geqslant 18{,}2$, o sea, $n \geqslant 19$.

b) Reiterando las simulaciones ahora con $n = 19$, obtenemos las estimaciones:

$0{,}001864, 0{,}003707, 0{,}004052, 0{,}000938, 0{,}00357, 0{,}002575, 0{,}0030$

$0{,}001786, 0{,}003576, 0{,}003170, 0{,}005159, 0{,}003053, 0{,}004337, 0{,}00$

Para dar con un intervalo como el que se pide, consideremos una muestra cualquiera m de las obtenidas con $n = 19$. Sabemos que la distancia entre m y $\int_2^7 e^{-x^2}\ dx$ es, con probabilidad aproximada 0.98, menor a 0.01. Esto es lo mismo que decir que $\int_2^7 e^{-x^2}\ dx \in (m - 0{,}01, m + 0{,}01)$ con probabilidad 0.98. Tomando por ejemplo $m = 0{,}003576$ obtenemos el intervalo $(-0{,}006423, 0{,}013576)$. A continuación graficamos varias muestras y nos fijamos si caen en dicho intervalo:

En la imagen se aprecia que efectivamente todas las muestras tomadas en este caso caen dentro del intervalo elegido.

[4] Acá queda claro por qué acotamos σ *inferiormente*: para la resolución de la inecuación, la aproximación elegida σ' de σ debe ser una que haga que $\frac{1}{\sigma'}0{,}01$ no sea menor que el valor real, ya que de lo contrario el valor que tomemos para n podría no ser lo suficientemente grande.

7.8. Recopilación de ejercicios

Ejercicio 7.48. Probar que el conjunto de discontinuidades de las funciones de distribución es a lo sumo numerable. Idea: probar que si x es de discontinuidad, entonces $\mathbb{P}(X=x)>0$. ¿Puede haber no-numerables saltos? Cuántos saltos de tamaño mayor a $1/m$ puede haber?).

Ejercicio 7.49. Sea X_n la variable aleatoria constante igual a 2^{-n}. Probar que X_n tiende en distribución a la función distribución F de la variable aleatoria constante igual cero. Observar que en el punto de discontinuidad de la distribución F, no hay convergencia.

Ejercicio 7.50. La idea de este ejer2cicio es probar la convergencia en distribución es más débil que la de probabilidad (cuando estamos en un mismo espacio). (Ver Observación 7.3.) Sean $\{X_n\}$ una sucesión de variables aleatorias , tal que $X_n \overset{\mathbb{P}}{\to} X$ para cierta variable aleatoria X. Entonces $X_n \overset{\mathcal{D}}{\to} X$.

1. Dados t punto de coninuidad de $F(x)=\mathbb{P}(X\leqslant x)$, y $\varepsilon>0$ mostrar que $\mathbb{P}(X\leqslant t-\varepsilon)\leqslant \mathbb{P}(X_n<t)+\mathbb{P}(|X-X_n|\geqslant\varepsilon)$.
2. De manera similar, encontrar una cota superior para $\mathbb{P}(X_n<t)$, para concluir el resultado.

Ejercicio 7.51. Observar que $G_X(s)=\mathbb{E}(s^X)$, y que además $G_X(0)=p_X(0)$ y $G_X(1)=1$.

Ejercicio 7.52. Si X es una v.a. que toma valores naturales, con función de generatriz de probabilidad G_X. Entonces

$$G_X^{(k)}(1)=\mathbb{E}(X(X-1)\cdots(X-k+1)),\quad k=1,2,\ldots.$$

(Se puede utilizar el siguiente resultado de Abel: si $A(s)=\sum_{k=0}^{\infty}a_k s^k$ es convergente en $|s|<1$, entonces $\sum_k a_k=\lim_{s\to 1}A(s)$ (pudiendo ser ambos infinitos).

Ejercicio 7.53. Si X es un v.a. que toma valores naturales, dar una expresión de la varianza de X en función de las derivadas de G_X.

Ejercicio 7.54. Rehacer cálculos de esperanza y varianza de las distribuciones geométricas y Poisson mediante este método.

Ejercicio 7.55. Hallar distribución de la suma de v.a. indpendientes con distribución $\mathcal{B}(n,p)$ y $\mathcal{B}(m,p)$ respectivamente. Concluir que la distribución de la suma de n variables aleatorias de Bernoulli de parámetro p tiene distribución Binomial de parámetro n y p.

Ejercicio 7.56. Mostrar que la suma de v.a. independientes con distribución $\mathcal{P}(\lambda)$ y $\mathcal{P}(\mu)$ respectivamente, tiene distribución $\mathcal{P}(\lambda+\mu)$.

Ejercicio 7.57. $\varphi_X(0)=1$

Ejercicio 7.58. $|\varphi_X(t)|\leqslant 1$.

Ejercicio 7.59. $\varphi_X(-t)=\overline{\varphi_X(t)}$, donde $\bar{z}$ es el conjugado de $z\in\mathbb{C}$.

Ejercicio 7.60. $\varphi_{aX+b}(t)=\varphi_X(at)e^{itb}$.

Ejercicio 7.61. Probar el resultado anterior utilizando que el límite del cociente incremental se puede intercambiar el símbolo de esperanza (la integral) observando que la función en cuestión está uniformemente acotada por una integrable (de hecho está acotada por 1).

Ejercicio 7.62. Probar el Lema anterior.

Ejercicio 7.63. Utilizando las propiedades de la función característica se concluye lo siguiente.

$$X\sim\mathcal{N}(\mu,\sigma^2),\quad \text{entonces}\quad \varphi_X(t)=e^{it\mu-\sigma^2t^2/2}.$$

Ejercicio 7.64. Probar el teorema anterior. ¿Cuál sería la esperanza y varianza?

Ejercicio 7.65. Asumiendo que la función característica de una v.a con distribución de Cauchy es $\varphi_X(t)=e^{-|t|}$, $t\in\mathbb{R}$, probar que el TCL no es cierto para suma de v.a. i.i.d. con distribución de Cauchy. ¿Qué distribución tiene la suma normalizada?

Capítulo 8

Miscelánea

En este capítulo se estudiarán diversos temas relacionados a los vistos en el curso. Comenzaremos con una introducción a la noción de *esperanza condicional* (ya visto en otros cursos) pero dando su definición formal, para luego dar una introducción a *martingalas* y otras aplicaciones.

8.1. Esperanza Condicional

Sea $(\Omega, \overline{\mathcal{F}}, \mathbb{P})$ un espacio de probabilidad. Sea X una variable aleatoria que es $\overline{\mathcal{F}}$–medible, tal que $X \in L^1$ (i.e. $\mathbb{E}(|X|) < \infty$).

Consideremos una sub σ-álgebra $\mathcal{F} \subset \overline{\mathcal{F}}$. Observar que en general no tiene que suceder que X sea una variable aleatoria en el nuevo espacio $(\Omega, \mathcal{F}, \mathbb{P})$, esto es, que $X^{-1}(B) \in \mathcal{F}$ para todo Boreliano B.

Definición 8.1. Se define una esperanza condicional de X dado la σ-álgebra $\mathcal{F}$, y se denota por $\mathbb{E}(X|\mathcal{F})$, a una variable aleatoria Y que satisface:

- Y es $\mathcal{F}$–medible,
- para todo $A \in \mathcal{F}$ se tiene $\int_Z Y \, d\mathbb{P} = \int_X d\mathbb{P}$.

Observación 8.2. Si Y es una esperanza condicional de X dado $\mathcal{F}$, entonces $Y \in L^1$. En efecto, sea $A = \{Y > 0\} \in \mathcal{F}$, entonces $\int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P} \leqslant \int_A |X| \, d\mathbb{P} < \infty$. Por otro lado $0 \leqslant \int_{A^c} -Y \, d\mathbb{P} = \int_{A^c} -X \, d\mathbb{P} \leqslant \int_{A^c} |X| \, d\mathbb{P} < \infty$.

Como $|Y| = Y \mathbb{1}_A + (-Y)\mathbb{1}_{A^c}$, resulta $\mathbb{E}(|Y|) \leqslant \mathbb{E}(|X|) < \infty$.

Lema 8.3. *Si existe la esperanza condicional, es única.*

Demostración. Sea Y' otra versión de la esperanza condicional. Dado $\varepsilon > 0$, sea $A_\varepsilon = \{Y - Y' \geqslant \varepsilon\} \in \mathcal{F}$. Luego resulta

$$\int_{A_\varepsilon} Y \, d\mathbb{P} = \int_{A_\varepsilon} X \, d\mathbb{P} = \int_{A_\varepsilon} Y' \, d\mathbb{P},$$

y por lo tanto $0 = \int_{A_\varepsilon} (Y - Y') \, d\mathbb{P} \geqslant \varepsilon \mathbb{P}(A_\varepsilon)$. Es decir $\mathbb{P}(A_\varepsilon) = 0$, para todo ε. Por lo que $Y \leqslant Y'$ casi seguro. Cambiando los roles de Y e Y', sigue el resultado. $\square$

Todavía no hemos visto si existe la esperanza condicional de una variable aleatoria respecto a una σ-álgebra.

Definición 8.4. Consideremos un espacio de medida $(\Omega, \mathcal{F})$, y dos medidas μ y ν sobre este espacio. Decimos que una medida ν es *absolutamente continua* respecto a una medida μ si para todo $A \in \mathcal{F}$ tal que $\mu(A) = 0$, entonces se tiene $\nu(A) = 0$. Se escribe $\nu \ll \mu$.

Teorema 8.5 (Teorema de Radon-Nykodim) *Dadas dos medidas μ y ν σ–finitas en $(\Omega, \mathcal{F})$. Si $\nu \ll \mu$, entonces existe una única función f que es $\mathcal{F}$–medible tal que*

$$\nu(A) = \int_A f \, d\mu, \qquad \forall A \in \mathcal{F}.$$

Esta función f, que se denota $\frac{d\nu}{d\mu}$, se llama derivada de Radon-Nykodim.

Recordar Comentario 5.67.

8.2. Martingalas

Bibliografía

[RD] Rick Durret, *Probability: Theory and Examples.*

[EL] Emannuel Lesigne, *Heads or Tails. An introduction to Limit Theorems in Probability.*

[DGL] Luc Devroye, László Györfi, and Gábor Lugosi, *A probabilistic theory of pattern recognition*, Applications of Mathematics (New York), vol. 31, Springer-Verlag, New York, 1996.

Índice alfabético