PROBABILIDAD 2

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

25 de abril de 2023

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.

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

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,

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

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

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.

◇ **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 $X_1, X_2, \dots$ 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 $\mathrm{var}(S_n) = n\sigma^2$, y por lo tanto la v.a.

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

satisface $\mathbb{E}(Z_n) = 0$ y $\mathrm{var}(Z_n) = 1$. 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 prisa 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

¿Cara o número? Una introducción a teoremas límites

Capítulo 2

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.

2.1. 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

$i = 1, \dots, k$, 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 2.1. 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 considerando la probabilidad uniforme en el espacio muestral $\{1,2,3,4,5,6\}$.

Ejemplo 2.2. 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 $(1-p, p)$.

2.2. 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

$$\mathbb{P}_n(\omega_i = 0) = q, \quad \text{y} \quad \mathbb{P}_n(\omega_i = 1) = p. \tag{2.1}$$

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

$$\mathbb{P}_n\left(\omega_{i+1} = 1 \text{ y } (\omega_1, \dots, \omega_i) = (e_1, \dots, e_i)\right) = \tag{2.2}$$

$$\mathbb{P}_n(\omega_{i+1} = 1) \cdot \mathbb{P}_n\left((\omega_1, \dots, \omega_i) = (e_1, \dots, e_i)\right), \tag{2.3}$$

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

$\diamondsuit$ **2.3.** ¿Es suficiente la condición (2.1) para definir una probabilidad en Ω_2?

$\diamondsuit$ **2.4.** Pruebe por inducción, que las condiciones (2.1) y (2.2) definen una probabilidad en Ω_n.

♢ **2.5.** 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.

2.3. 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

$$\mathbb{E}(X) := \sum_{i=1}^{k} x_i \cdot \mathbb{P}(X = x_i). \tag{2.4}$$

Básicamente, el $\mathbb{E}(X)$ es el promedio de los valores tomados por x_i, ponderados por la probabilidad del evento $\{X = x_i\}$.

Observar que podemos escribir X como la suma de indicatrices:

$$X = \sum_i x_i \mathbb{1}_{X = x_i}. \tag{2.5}$$

La esperanza cumple varias propiedadas conocidas.

Proposición 2.6. *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.*

◊ **2.7.** Pruebe las propiedades anteriores. Para la última conviene obervar que la descomposición (2.5) no es única, y en particular se pueden particionar los eventos $\{X = x_i\}$ para tener una expresión tipo $X = \sum_k y_k \mathbb{1}_{A_k}$.

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

$$\mathbb{E}(f(X)) = \sum_{i=1}^{k} f(x_i) \cdot \mathbb{P}(X = x_i). \tag{2.6}$$

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

Proposición 2.9 (Chebyshev–Markov)**.** *Sea* X *una v.a. no negativa. Entonces, para cada* $a > 0$*, se tiene*

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

□

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

Como coralario de este resultado tenemos lo siguiente:

Corolario 2.11. *Sea* X *una v.a., y* $a > 0$*. Entonces*

$$\mathbb{P}\left(|X - \mathbb{E}(X)|^2 > 0\right) \leqslant \frac{\mathrm{var}(X)}{a^2},$$

donde

$$\mathrm{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2. \tag{2.7}$$

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

Ejemplo 2.12. 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 2.4 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 2.13. 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

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

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, $k = \ell$).

♢ **2.14.** 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 $\{A_1, A_2, A_3\}$ son independientes.

♢ **2.15.** Probar que en el modelo definido en el Ejercicio 2.4 los eventos $A_i = \{\omega_i = 1\}$ son independientes.

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

♢ **2.16.** 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 $\mathbb{R}$.

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

♢ **2.18.** 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 $\mathbb{R}$.

Proposición 2.19. *Sean X e Y v.a. independientes. Entonces*

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

♢ **2.20.** Dar una prueba de la proposición anterior.

2.5. Ley de los grandes números

Ya tenemos los prerequisitos 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.

2.5.1. 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 2.21. *La v.a. S_n toma valores $0, 1, \ldots, n$ con probabilidades*

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

◊ **2.22.** 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 2.23. *Sean $X_1, \ldots, X_n$ 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

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

◊ **2.24.** Probar el resultado anterior. La fórmula de esperanza y varianza resultan aplicaciones sencillas de la Proposición 2.19. Se sugiere como un lindo ejercicio de manipulación de sumas, probar ambas identidades utilizando las definiciones (2.4) (2.7) y manipulando (derivando) la expresión del binomio de newton $(a+b)^n$. (Para la varianza recordar expresión (2.6).

2.5.2. El modelo y sus frutos

En la formalización de un experimento 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 2.25 (Ley (débil) de Grandes Números para la binomial)**.** *Para todo $\varepsilon > 0$ se tiene*

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

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

Demostración. Utilizando el Corolario 2.11 se tiene

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

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

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

Teorema 2.26 (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. □

◊ **2.27.** Sean $X_1, \dots, X_n$ una colección de v.a. i.i.d. con distribución binomial 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

$$\lim_{n\to\infty} \sup_{0\leqslant p\leqslant 1} |f(p) - B_n(p)| = 0.$$

2.6. Grandes desvíos

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

$$\mathbb{P}_n\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2}. \tag{2.8}$$

¿Cuán precisa es esa desigualdad? Sabemos que esa desigualdad resulta de la desigualdad de Chebyshev, por lo que podemos preguntarnos cuán precisa es esta. Un simple ejercicio dice que la desigualdad de Chebyshev no puede ser mejorada sin hipótesis adicionales.

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

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

♢ **2.29.** 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 (2.8). 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 $0 < \varepsilon < \text{mín}(p, 1-p)$, definiendo la función

$$h_+(\varepsilon) := (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (1-p-\varepsilon)\log\left(\frac{1-p-\varepsilon}{1-p}\right), \tag{2.9}$$

tenemos el siguiente resultado

Proposición 2.30 (Bernstein-1924'). *Para todo* $0 < \varepsilon < \text{mín}(p, 1-p)$*, definiendo* $h_-(\varepsilon) := h_+(-\varepsilon)$*, se tiene* $h_+(\varepsilon), h_-(\varepsilon)$ *son positivas y*

$$\mathbb{P}_n\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant e^{-nh_+(\varepsilon)} + e^{-nh_-(\varepsilon)}, \tag{2.10}$$

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. acotar $\mathbb{P}(S_n/n \leqslant p-\varepsilon)$ 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 sustituire p por $1-p$ en (2.9), resulta ser igual a $h_+(-\varepsilon)$. □

2.7. Ley fuerte de los grandes números

La ley débil de los grandes números, en el caso de la repetición indpendiente 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 sigiuente límite $\lim_n S_n/n$ existe y es igual a la media.

En los siguientes párrafos intentaremos formalizar esto.

2.7.1. 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 monoda 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

$$\mathbb{P}(A):=\mathbb{P}_{n(A)}(A')=\sum_{\omega^{(n)}\in A'} p^{S_n(\omega^{(n)})}(1-p)^{n-S_n(\omega^{(n)})}. \tag{2.11}$$

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.

$\diamondsuit$ **2.31.** 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 2.32. Una familia que es cerrada respecto a esas operacione se denomina *álgebra*.[1]

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

[1]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 $A\cup B\in\mathcal{E}$.

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

Definición 2.33. Decimos que un conjunto $N \subset \Omega$ tiene *medida nula* (*negligible* en inglés) si para todo $\varepsilon > 0$ existen 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 a trabajar con eventos de tipo finito, de medida cero o evento casi seguro. Por esta razón de ahora en adelante llamaremos evento a alguna de estos tres casos.

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

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

tiene la misma probabilidad que el evento A.

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

Proposición 2.34.

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.* □

◊ **2.35.** Probar la anterior proposición.

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

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

para cualquier subonjunto finito de la familia.

Proposición 2.36. *Probar que eventos definidos por índices disjuntos, son indpendientes.*

Proposición 2.37. *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 2.38. 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 $\hat{W} \subset \Omega$ 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 $(e_1, e_2, *, *, \ldots)$ 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 $A_2 = (*, *, e_3, e_4, *, *, \ldots)$ 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.

$\diamondsuit$ **2.39.** Escribir la prueba formal para una palabra $w = (w_1, \ldots, w_j) \in \{0,1\}^j$ dada.

$\diamondsuit$ **2.40.** 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.)

♢ **2.41.** 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 2.42 (Ley fuerte de los grandes números)**.** *Sea S_n la distribución binomial de parámetros n y p. Casi seguramente se tiene*

$$\lim_{n\to+\infty} \frac{S_n(\omega)}{n} = p. \tag{2.12}$$

Es decir, la condición (2.12) 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 exste $k \in \mathbb{N}$ 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 (2.12) 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 2.34 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 2.30, 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. $\square$

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

$$\lim_n \frac{1}{n}\#\Big\{k : 1 \leqslant k \leqslant n, \textit{ tal que } \omega \in A_k\Big\} = \mathbb{P}(A_1). \tag{2.13}$$

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 $\omega \in \Omega \mapsto (\rho_1(\omega), \rho_2(\omega), \ldots) \in \{0,1\}^{\mathbb{N}}$.

Luego la condición (2.13) 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 2.42 para obtener el resultado. Dejamos estos detalles a cargo del lector. □

2.7.2. ¿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 2.44. *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

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

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

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

Comentario 2.45. 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 (2.11).

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.[2]

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 $C = \{n_1, \dots, n_k\} \in \mathcal{F}$, definimos el mapa $F_C = F_{n_k} \circ \dots \circ F_{n_1}$ que cambia el resultado en los lugares $n_1, \dots, n_k$.
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 $F_C(\omega) =$

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

$F_{C'}(\omega')$, 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 $\mathbb{P}(A) = 0$. $\square$

Parte II

Probabilidad via Teoría de la Medida

Capítulo 3

Definiciones Básicas

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

3.1. σ-álgebra

Dado un conjunto Ω, recordemos que una familia no vacía de conjuntos $\mathcal{A} \subset \Omega$ es una *σ-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 3.1. 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 3.2. Si tenemos una partición dijsunta numerable $\Omega^{(1)}, \Omega^{(2)}, \dots$ 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 3.3. 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.

3.1.1. σ-álgebra generada por una familia

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

$\diamond$ **3.4.** 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 $\sigma(\mathcal{F}') \subseteq \sigma(\mathcal{F})$.

$\diamondsuit$ **3.5.** Probar que si tomamos la familia de subconjuntos de Ω formados por un elemento, entonces la σ-álgebra coincide con la del Ejemplo 3.3.

> **Ejemplo 3.6** (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.

$\diamondsuit$ **3.7.** 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 $(0,1] = \bigcup_{n\geqslant 1}(0, 1+1/n)$, y por lo tanto $(0,1] \in \mathcal{B}(\mathbb{R})$. También están los conjuntos de un elemento $\{2\} \in \mathcal{B}(\mathbb{R})$.

3.2. 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 3.8. *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 $\mathbb{P}(\bigcup_i A_i) = \lim_{i \to +\infty} \mathbb{P}(A_i)$.*

$\diamondsuit$ **3.9.** Dar una prueba de la proposición anterior. Para la parte 3, 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 3.10 (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 3.11 (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.

3.2.1. 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 3.8, que omitiremos la prueba.

Proposición 3.12. *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 3.13 (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 3.14. 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 3.15. 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 3.16. *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 $A^c = \bigsqcap_i S_i^c$, 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 3.14.

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

$$\lambda(\bigsqcup_{i=1}^{n}(a_i,b_i]) = \sum_{i=1}^{n}(b_i - a_i). \tag{3.1}$$

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 3.13 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 (3.1) .

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

Lema 3.17. *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 $k = 1,\dots,n-1$. 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 $\{(a_k, b_k+\varepsilon/2^k)\}_{k=1,\ldots}$, 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 3.18. *La función $\lambda : \overline{\mathcal{I}} \to \mathbb{R}$ dada en (3.1), 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

$$J_j = \bigsqcup_{n=1}^{\infty} J_j \cap A_n = \bigsqcup_{n=1}^{\infty}\bigsqcup_{i=1}^{m_n} J_j \cap I_i^{(n)}, \qquad (j=1,\ldots,m). \tag{3.2}$$

Luego del Lema 3.17 tenemos que

$$\lambda(J_j) = \sum_{n=1}^{\infty}\sum_{i=1}^{m_n} \lambda(J_j \cap I_i^{(n)}) \qquad (j=1,\ldots,m). \tag{3.3}$$

De (3.3), 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 3.19 (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 3.13 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 3.13 la existencia y unicidad de la medida producto.

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

Ejemplo 3.20. 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 3.2.1. 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

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

3.3. Variables Aleatorias

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

3.3.1. 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 3.21. 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 $\omega \notin A$.

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

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

donde entendemos por $\{X \in B\} := X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$.

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

♢ **3.23.** 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 3.24. *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.,* $\lim_{y \downarrow x} F(y) = f(x)$.

♢ **3.25.** Dar una prueba de este resultado.

El siguiente resultado nos dice que las propiedades anteriores alcanzan para determinar una funcion de distribución

Proposición 3.26. *Si F es una función que satisface las propiedades 1. 2. y 3. de la Proposición 3.24, entonces es la distribución de alguna variable aleatoria.* □

Omitimos la prueba de este resultado.

Corolario 3.27. *Si F es una función que satisface las propiedades 1. 2. y 3. de la Proposición 3.24, 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 3.26 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 3.28 (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 3.29. 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 3.30 (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 3.31 (*Exponencial*)**.** Diremos que X es una v.a. exponencial de parámetro ("tasa") λ ($X\sim Exp(\lambda)$) si:

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x} \text{ si } x\geqslant 0\\ 0 \text{ si } x<0\end{cases}$$

◊ **3.32.** 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 3.33. 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 3.34. La función $F(x)=1$, si $x\geqslant 0$, y 0 si $x<0$.

3.3.2. 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 3.35. *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

$$\mathcal{S}_f:=\left\{B: f^{-1}(B)\in\mathcal{A}\right\},$$

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 3.36. 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 3.37. 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

♢ **3.38.** 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 3.38 y el Ejemplo 3.37 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 3.39. *Si $X_1, \ldots, X_n$ 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 $f(X_1, \ldots, X_n)$ es una variable aleatoria.*

Demostración. Basta ver que la función $(X_1, \ldots, X_n) : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ es medible. Para eso, utilizando la Proposición 3.35, basta ver que la preimagen de rectangulos es medible. Esto resulta de que

$$(X_1, \ldots, X_n)^{-1}(A_1 \times \cdots \times A_n) = \{\omega \in \Omega : X_i(\omega) \in A_i\} = \bigcap_{i=1}^{n} \{X_i \in A_i\}$$

es una intersección de elementos de $\mathcal{A}$ y por tanto en $\mathcal{A}$. □

Corolario 3.40. *Si $X_1, \ldots, X_n$ son variables aleatorias, entonces $X_1 + \cdots + X_n$ es una variable aletaria.*

De esta manera podemos construir muchas de variables aleatorias de interés.

3.3.3. 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 3.41. *Supongamos que tenemos una familia de variables aleatorias $X_1, X_2, \ldots$. 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. □

◊ **3.42.** Finalizar prueba de la proposición anterior.

Dada una sucesión de variable aleatoria $X_1, X_2, \ldots$, 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 $\Omega_0 \in \mathcal{A}$.

Definición 3.43. 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

$$X_n \to Y \text{ c.s.},$$

[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 $\limsup x_n = \inf_n \sup m \geqslant n x_m$.

cuando $\mathbb{P}(\text{lím}_n X_n = Y) = 1$. En particular la función $Y = \text{lím}_n X_n$ 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 3.44. 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 3.45. 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

$$\lim_{n\to+\infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geqslant \varepsilon\}).$$

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 3.46. Consideremos el espacio de probabilidad de la medida uniforme en $[0,1]$, y consideremos la sucesión de intervalos definidos como a continuación: $I_1 = [0,1/2]$, $I_2 = [1/2,1]$, $I_3 = [0,1/3]$, $I_4 = [1/3,2/3]$, $I_5 = [2/3,1], \ldots$. 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 $X_n \xrightarrow{\mathbb{P}} 0$.

[2] Aquí hay un detalle, que formalmente la función límite $\text{lím}_n X_n$ está definida en Ω_0 y no en Ω. Formalmente, como Ω_0 tiene medida cero, podemos extenderla a $\text{lím}_n X_n$ 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.

El siguiente ejercicio muestra que el límite de la convergencia en probabilidad único (a menos de conjuntos de probabilidad 1).

◊ **3.47.** 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 $\mathbb{P}(X = Y) = 1$.

3.3.4. Algunas equivalencias de convergencia

La convergencia en medida es más débil que la convergencia casi segura como lo muestra el siguiente ejercicio.

◊ **3.48.** Probar que la sucesión de variables aleatorias definidas en el Ejemplo 3.46 no converge de manear casi segura.

Sin embargo la convegencia casi segura implica la convergencia en medida.

Proposición 3.49. *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 3.50. *La sucesión de variables aleatorias $\{X_n\}$ satisface*

$$X_n \xrightarrow{n} X \text{ c.s.} \quad \text{si y sólo si} \quad \lim_N \mathbb{P}\left(\bigcap_{n\geqslant N} |X_n - X| < \varepsilon\right) = 1.$$

□

◇ **3.51.** Dar una prueba de esta proposición.

3.4. Integrales y Esperanza

En el curso básico de Probabilidad, y en la primera parte de este curso (ver definición (2.4)), 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 3.52. 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 $x_1, \dots, x_n$, 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 (2.4) y (2.5).)

$\diamondsuit$ **3.53.** 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 (2.7)).

Análogo a la Proposición 2.6, se tiene que la integral de funciones simples satisfacen todas la propiedades dadas en dicha proposición, las cuales reescribimos acá.

Proposición 3.54. 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 α, $\beta \in \mathbb{R}$.

(Paso 2) Este es el paso más relevante dado que definiremos como integraf 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

$$f_{(n)}(\omega) := \frac{Mk}{n}, \quad \text{si } f(\omega) \in \left(M\frac{k}{n}, M\frac{(k+1)}{n}\right], \qquad (-n \leqslant k \leqslant n). \tag{3.4}$$

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 3.55. 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

$$\left|\int_\Omega f_{(n)}\, d\mu - \int_\Omega f_{(m)}\, d\mu\right| \leqslant M \max\left\{\frac{1}{m}, \frac{1}{n}\right\}\mu(\Omega),$$

probando la segunda afirmación. □

Definición 3.56. Dada f una función medible y acotada, se define la integral de f como

$$\int_\Omega f\, d\mu := \lim_n \int_\Omega f_{(n)},$$

siendo $f_{(n)}$ la sucesión de funciones simples , dadas en (3.4), que aproximan por debajo a f

Se puede probar, sin dificultad pero omitiremos la prueba que todas las propiedades enunciadas en (3.54) 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

$$g_{(n)}(\omega) := \frac{M(k+1)}{n}, \quad \text{si } f(\omega) \in \left(M\frac{k}{n}, M\frac{(k+1)}{n}\right], \qquad (-n \leqslant k \leqslant n). \tag{3.5}$$

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

$$\int_\Omega f_{(n)}\, d\mu \leqslant \sup_{\varphi \leqslant f} \int_\Omega \varphi\, d\mu \leqslant \inf_{\psi \geqslant f} \int_\Omega \varphi\, d\mu \leqslant \int_\Omega g_{(n)}\, d\mu,$$

y por lo tanto se puede dar como definición alternativa

$$\int_\Omega f\, d\mu := \sup_{\varphi\, \text{simple} \leqslant f} \int_\Omega \varphi\, d\mu = \inf_{\psi\, \text{simple} \geqslant f} \int_\Omega \varphi\, d\mu$$

(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 3.54 siguen siendo válidas para integrales de funciones medibles a las cuales se pueda calcular su integral.

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*

$$\mathbb{E}(X) := \int_\Omega X\,d\mathbb{P},$$

A continuación incluiremos, sin demostración, el teorema con las propiedades de la esperanza (heredadas de la definición de integral dada anterioremente).

Teorema 3.57. *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. $|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$.

3.4.1. Integral de Riemann vs Integral de Lebesgue

Consideremos el espacio $([0,1],\mathcal{B},\lambda)$ 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

$$\int_0^1 f(x)\,dx = \lim_n \sum_{i=1}^n f(i/n)\frac{1}{n}. \tag{3.6}$$

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

$$\int_{[0,1]} f_n\,d\lambda = \sum_{i=0}^n f(i/n)\lambda(A_i) = \sum_{i=0}^n f(i/n)\frac{1}{n}, \tag{3.7}$$

que es exactamente las sumas de Riemann. De (3.6) sólo nos falta probar que $\lim_n \int_{[0,1]} f_n\,d\lambda = \int_{[0,1]} f\,d\lambda$.

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 3.57), 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 (3.6) y (3.7) 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 3.58. *Sea $f : [0,1] \to \mathbb{R}$ continua. Entonces la integral de Lebesgue de f coincide con la integral de Riemann, i.e.*

$$\int_{[0,1]} f\,d\lambda = \int_0^1 f(x)\,dx.$$

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 $x \in [0,1]$.

3.4.2. Integral de Riemann-Stieljes*

3.4.3. Tres teoremas límites de integrales

En esta sección enunciaremos, sin demostracion, tres teoremas importantes de la teoría para límites de variables aleatorias.

El primero es un conocido y útil teorema, que tiene nombre de lema.

Teorema 3.59 (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 3.60.** 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 $\liminf X_n$ es la v.a. constante igual cero. Pero por otro lado $\mathbb{E}(X_n) = 1$, para todo $n \in \mathbb{N}$.

Teorema 3.61 (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:

$$\mathbb{E}(\lim X_n) = \lim \mathbb{E}(X_n).$$

Teorema 3.62 (Teorema de Convergencia Dominada)**.** *Si* $X_n \to X$ *c.s., y* $|X_n| \leqslant Y$ *con* $\mathbb{E}(Y) < \infty$*, entonces* $\lim_n \mathbb{E}(X_n) = \mathbb{E}(X)$.

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 $\mathbb{E}(\lim X_n) = \lim \mathbb{E}(X_n)$.

3.4.4. 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 46).

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 3.63. 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 $g : \mathbb{R} \to \mathbb{R}$ 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.

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. (3.4).) De manera análoga, se prueba la desigulddad por arriba utilizando las funciones simples que aproximan por arriba (cf. (3.5)).

Con esto concluimos la siguiente proposición, que su prueba formal la daremos en otra sección.

Proposición 3.64. *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 siguiente sección.

3.4.5. Cálculo de integrales

Teorema 3.65. *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, entonces*

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^n} \varphi\, d\mu_X. \tag{3.8}$$

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 3.61 para concluir el resultado. Esto es

$$\mathbb{E}(\varphi(X)) = \lim_n \mathbb{E}(\varphi_n(X)) = \lim_n \int_{\mathbb{R}^n} \varphi_n \, d\mu_X = \int_{\mathbb{R}^n} \varphi \, d\mu_X.$$

□

Bibliografía

[RD] Rick Durret, *Probability: Theory and Examples.*

[EL] Emannuel Lesigne, *Heads or Tails. An introduction to Limit Theorems in Probability.*

Índice alfabético