PROBABILIDAD

Notas de Curso

Universidad de la República

2018

Cómo utilizar estas notas.

Estas notas fueron realizadas por Diego Armentano y Valeria Goicochea para el curso *Probabilidad* que se dictó en el primer semestre de 2017, como curso de la Licenciatura en Matemática de la Facultad de Ciencias, Universidad de la República. La duración del curso es de 15 semanas.

Las referencias básicas de estas notas son los libros de Feller [F1], de Grimmet-Walsh [GW], y también las notas de curso de McMullen [M] y Weber [W].

Las notas siguen una modalidad de *"teórico-práctico"*. En esta modalidad los estudiantes trabajaran en grupos, que se mantendrán durante el semestre. Durante las clases se seguirán de cerca las notas discutiendo en grupo y con el equipo docente. Los docentes recurrirán al pizarrón en pocos casos, por ejemplo cuando la introducción o discusión de ciertos contenidos lo ameriten. Se sugiere al estudiante acompañar estas notas con un cuaderno y un lapiz para completar todas las pruebas solicitadas.

Cada capítulo tiene al inicio palabras claves de los temas a tratar y la duración estimada en semanas para ese capítulo.

En el texto aparecerán símbolos $\diamond$ que indican que son problemas para trabajar en equipo, donde se consultará al docente en caso de ser necesario.

Como complemento de estas notas se utilizará el software *R* para el cual se han diseñado tutoriales para su manipulación. Además se dispondrá de una sala de máquinas para su utilización durante el curso.

Tanto el tutorial de *R* como las prácticas han sido escritas usando *R Markdown* (otra de las facilidades del uso de *RStudio*). El uso de R Markdown es ideal para cuando uno quiere ir presentando líneas de código en R y gráficas a la vez que se hacen comentarios. Por ejemplo, si en su monografía, tesis, informe, etc. tiene que agregar un capítulo con simulaciones/gráficas, puede escribir ese capítulo usando R Markdown, exportarlo como .pdf e incluirlo en su monografía (por ejemplo).

Puede agregar código LaTeX, así que de esta manera podrá evitar tener que estar exportando muchas imágenes creadas en R para ser incluídas en un archivo .tex (mandará todo el capítulo de una). Los archivos .Rmd estarán a su disposición para que les sirvan de ejemplo. Los mismos se econtrarán en la página del curso en EVA.

Con el fin de mejorar estas notas, cualquier sugerencia será más que bienvenida.

Índice general

Parte I

Espacios de probabilidad discretos

Capítulo 1

Introducción

Motivación, espacio muestral, eventos, probabilidad.

[Duración: 2 Semanas]

En el mundo en que vivimos la incertidumbre es casi una regla. Conocer con precisión ciertos acontecimientos es casi impensado. Sin embargo, en muchos casos es posible dar una medida de certeza de que algo ocurra , como por ejemplo, "mañana es probable que llueva", " la probabilidad de que la moneda salga cara es 50 %", "la probabilidad de ganar jugando al "rojo" o "negro" en la ruleta es de $18/37 = 0{,}486\ldots$".

La noción de probabilidad es un concepto muy antiguo, que formó y forma parte de todas las culturas. En algunos casos (como el de la moneda y la lluvia) está fuertemente nutrida por nuestra experiencia.

Sin embargo, la formalización de dicho concepto y su transformación en una teoría es reciente en la historia de la humanidad.

1.1. ¿Qué es la Probabilidad?

La *teoría de Probabilidad*, o simplemente Probabilidad, se origina en el estudio de los juegos de azar. La teoría comenzó a desarrollarse a mediados del siglo XVII por intercambios entre los matemáticos Pascal y Fermat que estaban interesados en entender estos juegos.

La Probabilidad es un área de la matemática que se ha convertido en una herramienta poderosa para entender aquellos fenómenos que no pueden describirse con leyes deterministas. Un ejemplo importante es la búsqueda de patrones en fenómenos "aleatorios" (aquí entendemos por *fenómeno aleatorio* aquel experimento o procedimiento, en el cual el resultado no puede predecirse con precisión). A modo de ejemplo, cuando lanzamos una moneda al aire no sabemos predecir si el resultado es cara o número, sin embargo tenemos la certeza que la frecuencia relativa de caras (si es una moneda "fiel") será de 50%. (En las Sección 1.5, analizaremos con cuidado esta frase.)

Para la Probabilidad es irrelevante la naturaleza de la aleatoriedad. De hecho, el lanzamiento de una moneda puede considerarse como un fenómeno determinista: conociendo con precisión la posición y velocidad inicial (y todos los otros factores que involucran el experimento: material donde aterriza, grosor de la moneda, etc.), el movimiento de la moneda está totalmente determinado. Sin embargo, en el mundo real es muy difícil conocer con precisión las condiciones iniciales del lanzamiento.[1] Es por este motivo que el resultado es incierto y por lo tanto lo consideramos como un experimento aleatorio. (Se recomienda la lectura del artículo Keller [] donde se analiza este experimento.)

La teoría de Probabilidad es una rama de la matemática como cualquier otra, donde tiene su conjuntos de axiomas y métodos como la *Geometría*, el *Álgebra*, o el *Análisis*. Sin embargo, para poder entender la Probabilidad, y apreciar su belleza, es necesario entender ejemplos de la vida real, o ejemplos motivados por otras áreas de la matemática y otras ciencias.

La axiomática de la Probabilidad fue desarrollada en 1930 por A. N. Kolmogorov, donde los fundamentos yacen sobre la teoría abstracta de la medida, pero la relación con esta teoría es similar a la relación del cálculo diferencial e integral con la teoría de ecuaciones diferenciales. En este sentido, para poder apreciar y disfrutar la teoría de Probabilidad como disciplina es importante no perder los ejemplos de vista.

Hoy día la Probabilidad se aplica en la mayoría de las disciplinas científicas y tecnológicas, además de otras ramas del conocimiento humano como la sociología, o la psicología. Además, como toda área de la matemática, la Probabilidad está presente en otras áreas de la matemática. Algunas de las aplicaciones interesantes en este sentido, que veremos en este curso, son el teorema

[1] Observar que sin mencionarlo estamos utilizando un modelo que simplifica la realidad basado en la mecánica Newtoniana, que no toma en cuenta otros factores que pueden jugar un papel importante.

de aproximación de Weierstrass de funciones continuas por polinomios, y la geometría integral.

1.2. Algunos problemas motivadores

Veamos algunos ejemplos concretos.

Problema 1 (*Monty Hall*)**.** Supongamos que estamos en un juego donde hay tres puertas y donde hay un premio para ganare el juego detrás de una de ellas. Se elige una puerta al azar. Luego un presentador abre una de las puertas que no se eligió, y donde no está el premio. Ahora el presentador le pregunta a la persona: ¿quiere cambiar de puerta, o se mantiene con su primer opción?

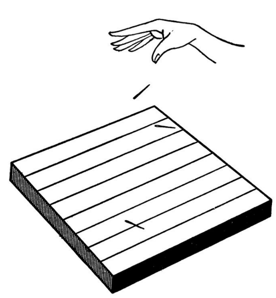
El problema a resolver es saber si hay una estrategia ganadora.

Problema 2 (*Cumpleaños*)**.** ¿Cuál es la probabilidad de que dos personas en una misma clase de facultad cumplan el mismo día?

Obviamente eso depende de la cantidad de personas. Si hay más de 366 personas, entonces la respuesta sería 1. Pero supongamos que hay 25 personas.

Problema 3 (*Problema de la ruina*)**.** Un jugador tira una moneda: si resulta número gana \$1, si resulta cara pierde \$1. Si la cantidad inicial de dinero era \$$x$ y el jugador pretende jugar hasta ganar el monto \$$m$, o quedarse sin dinero (en la ruina), ¿cuál es la probabilidad de arruinarse?

Problema 4 (*Aguja de Buffón*)**.** Considere una aguja de largo L (cm, por ejemplo) lanzada al azar sobre una superficie en la que hay dibujadas líneas paralelas separadas a una distancia t. ¿Cuál es la probabilidad de que la aguja cruce alguna línea?



1.3. Formalización de la aleatoriedad

Para comenzar, al igual que cualquier teoría matemática, es necesario *idealizar* el objeto a estudiar. En nuestro caso vamos a construir un modelo matemático de un experimento aleatorio que nos permita analizarlo.

Para definir nuestro modelo introduciremos tres conceptos que podemos pensar como conceptos primitivos: *espacio muestral*, *evento*, y *probabilidad*.

Brevemente, el espacio muestral es el conjunto de todos los posibles resultados de nuestro experimento; eventos son los subconjuntos del espacio muestral; a cada evento le asociamos una medida de cuán frecuente, o "probable" es.

Estos conceptos pueden considerarse como los conceptos primitivos análogos a *punto* y *recta* en la *geometría euclidiana*.

1.3.1. Espacio Muestral

Para comenzar a construir el modelo es necesario especificar los *resultados posibles de un experimento*. Para tal fin es necesario realizar una idealización que nos permita simplificar la teoría.

Por ejemplo, si nuestro experimento es lanzar un dado, podemos convenir que los resultados posibles son cada una de las posibles caras.

Sin embargo, observar que aquí estamos haciendo nuestra primera idealización. ¿A quién no le pasó que el dado quedó inclinado por estar sobre una superficie no plana? Es posible incluir el caso fallido como resultado posible, pero es más natural y menos tortuoso excluirlo de los resultados posibles para desarrollar una teoría más útil. En el caso del dado fallido, para evitar discusiones entre jugadores, se tira de vuelta el dado.

Definición. Cada resultado diferente del experimento se llama *evento elemental* o *suceso elemental* , y lo denotaremos por ω. Al conjunto de todos los sucesos elementales ω de un experimento lo llamaremos *espacio muestral*, y se denota por Ω.

Veamos algunos ejemplos de espacios muestrales de expermientos idealizados.

Ejemplo 1.3.1. (*Lanzamiento simple de una moneda*) $\Omega = \{C, N\}$, donde C y N indican cara o número.

Ejemplo 1.3.2. (*n-lanzamientos de una moneda*) $\Omega = \{C, N\}^n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{C, N\}\}$. Observar que Ω son las sucesiones de caras o números de longitud n. En particular

Ω tiene 2^n elementos (i.e. $\#\Omega = 2^n$).

Ejemplo 1.3.3. (*Lanzamiento de dos dados*) $\Omega = \{1,2,3,4,5,6\}^2 = \{(i,j) \in \mathbb{Z}\times\mathbb{Z} : 1 \leqslant i, j \leqslant 6\}$, i.e. $\#\Omega = 36$. Observar que aquí estamos identificando cada cara del dado con el número respectivo.

Ejemplo 1.3.4. (*Día de cumpleaños*) Para modelar el día del año de cumpleaños de n personas en una muestra de individuos podemos considerar $\Omega = \{(i_1,\dots,i_n) \in \mathbb{N}^n : 1 \leqslant i_1,\dots,i_n \leqslant 365\}$ (si asumimos que el año tiene 365 días).

Ejemplo 1.3.5. (*Barajar 52 naipes*) ¿De cuántas formas distintas podemos barajar un maso de 52 cartas? Para este experimento podemos tomar $\Omega = \{(i_1,\dots,i_{52}) \in \mathbb{N}^{52} : 1 \leqslant i_1,\dots,i_{52} \leqslant 52,\ i_k \neq i_\ell,\ \text{si } k \neq \ell\}$. Veremos que $\#\Omega = 52! = 8\times 10^{67}$.
(Para tener una noción de la magnitud de este número se recomienda la lectura de http://czep.net/weblog/52cards.html)

Ejemplo 1.3.6. (*Hora de llegada a la facultad*) Supongamos que queremos modelar la hora de llegada a la facultad de un estudiante en determinado día. Podríamos tomar Ω como el intervalo $[0,24)$, o una versión discreta determinando la hora de llegada ($\#\Omega = 24$), o minuto de llegada ($\#\Omega = 1440$), o segundos de llegada ($\#\Omega = 86400$).

Ejemplo 1.3.7. (*Tiro al blanco*) Supongamos que queremos modelar el tiro con arco y flecha a un blanco. Para este caso podemos modelar el espacio muestral de este experimento con el disco unidad $\Omega = \{z \in \mathbb{C} : |z| < 1\}$. Observar que este es un modelo continuo donde hay una cantidad no-numerable de resultados posibles.

Otro modelo posible, y más sencillo, es considerar como Ω el conjunto finito de posibles resultados del blanco (recordar que el blanco está compartimentado en anillos concéntricos que dan puntos según su distancia al centro). En este caso podemos considerar $\Omega = \{A_1,\dots,A_5\}$ donde A_i indica uno de los anillos.

Ejemplo 1.3.8. (*n-tiros al blanco*) El experimento de tirar n veces una flecha al blanco puede ser modelado por $\Omega = \{A_1,\dots,A_5\}^n$, (cf. Ejemplo 1.3.7). Sin embargo, si estamos solamente interesados en conocer el puntaje asociado en las n-tiradas, podemos considerar el espacio muestral $\Omega = \{(\alpha_1,\dots,\alpha_5) \in \mathbb{N}^5 : 0 \leqslant \alpha_i \leqslant n;\ \alpha_1+\cdots+\alpha_5 = n\}$ tal que α_i es la cantidad de veces (dentro de las n-veces) que se le embocó al anillo A_i.

Ejemplo 1.3.9. (*Ascensor*) Supongamos que hay n personas que se suben a un ascensor en un edificio de k pisos. Para modelar la forma en que las personas se retiran del ascensor podemos considerar $\Omega = \{0,1,\dots,k\}^n$.

Ejemplo 1.3.10. (*Distribución de k-bolas en r-celdas*) Supongamos que realizamos el "experimento" de colocar k-bolas de colores diferentes en r-celdas. Para fijar ideas supongamos que $k = 3$ y $r = 2$. Si a, b, c son las bolas, podemos describir el espacio muestral de la siguiente manera: $\{(abc,-),(ab,c),(bc,a),(ac,b),(a,bc),(b,ac),(c,ab),(-,abc)\}$. En general, podemos tomar como espacio muestral $\Omega=\{(\alpha_1,\dots,\alpha_r)\in\mathbb{N}^r : 0\leqslant\alpha_i\leqslant n;\ \alpha_1+\dots+\alpha_r=k\}$, siendo α_i la cantidad de bolas que caen en la celda i.

El ejemplo anterior es de una naturaleza distinta a los anteriores dado que no queda claro dónde está la aleatoriedad en el experimento.

Sin embargo, como ya comentamos, para la Probabilidad es irrelevante la naturaleza de la aleatoriedad, y lo que solo nos preocupa (por el momento) es la elección del espacio muestral.

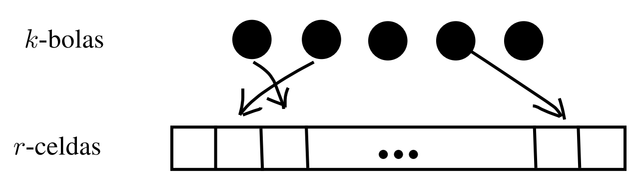
Si se mira con cuidado varios de los ejemplos anteriores (con espacio muestral discreto) pueden modelarse con el Ejemplo 1.3.10. Para ver esto comencemos con una digresión sobre productos cartesianos.

Digresión [Productos Cartesianos] Dado un conjunto de índices I arbitrario, y un conjunto $\mathcal{R}$, se define el *producto cartesiano* $\mathcal{R}^I$ como el conjunto de todas las funciones con dominio I y codomino $\mathcal{R}$, i.e.,

$$\mathcal{R}^I=\{\omega : I\to\mathcal{R} : \omega \text{ función}\}$$

Por ejemplo, si $I=\{1,\dots,n\}$, entonces $\mathcal{R}^I$ es el conjunto de funciones $\omega:\{1,\dots,n\}\to\mathcal{R}$. Definiendo $\omega_i:=\omega(i)$, $i\in\{1,\dots,n\}$, se tiene que $\mathcal{R}^I$ no es otra cosa que el conjunto de las n-uplas ordenadas $(\omega_1,\dots,\omega_n)$, o sea, es lo que habitualmente denotamos $\mathcal{R}^n$.

De la digresión anterior podemos reformular el Ejemplo 1.3.10 de la siguiente manera. Sea $\mathcal{R}$ el conjunto formado por las r-celdas, entonces a cada configuración distinta le podemos asociar un elemento de $\mathcal{R}^k$, y por lo tanto identificamos $\Omega=\mathcal{R}^k$.



- En el Ejemplo 1.3.1, los resultados posibles de lanzar una moneda corresponde a colocar 1-bolas, en 2 celdas;
- En el Ejemplo 1.3.2 el lanzamiento de una moneda n veces corresponde a colocar n bolas en 2 celdas.

- En el Ejemplo 1.3.4 los resultados posibles de los días de cumpleaños corresponde a colocar n bolas en 365 celdas.
- El Ejemplo 1.3.5 de barajar naipes es de una nueva clase. Cada naipe le corresponde un único lugar: es decir no está permitido poner dos naipes (2 bolas) en un mismo lugar (misma celda). Para modelar este experimento basta considerar las funciones inyectivas de $\{1,\dots,52\}^{52}$ (y por lo tanto biyectivas).

1.3.2. Eventos

A partir de ahora, siempre que hablemos de un experimento aleatorio sobreentenderemos que tenemos un espacio muestral Ω asociado.

En general, estamos interesados en conocer la probabilidad no sólo de los resultados del experimento sino de un conjunto de sucesos elementales. Por ejemplo, en el Ejemplo 1.3.3 nos gustaría saber si las siguientes condiciones ocurren:

- la suma de los dados es 11;
- la suma es par;
- el primer dado muestra un número mayor al segundo.
- los resultados son iguales.

Nuestra idea es formalizar esta noción intuitiva de evento.

Decimos que un evento A *está asociado* a un experimento aleatorio siempre que podemos decidir cuándo un evento elemental $\omega \in \Omega$ del experimento hace que el evento A ocurra, o no.
(Por ejemplo, es claro que en el Ejemplo 1.3.3: si conocemos los resultados individuales de cada dado, podemos determinar si ocurren algunas de las condiciones anteriores).

En este sentido tenemos una correspondencia natural entre evento y subconjuntos de Ω: un evento A se corresponde con el conjunto de sucesos elementales $\omega \in \Omega$ que hacen que A ocurra.

Esto motiva la siguiente definición.

Definición. Un *evento* de un experimento aleatorio es un subconjunto del espacio muestral Ω.

Con la descripción del espacio muestral mostrada en el Ejemplo 1.3.3, podemos dar una descripción de alguno de los eventos listados más arriba.

$$\text{“la suma es } 11'' \longleftrightarrow \{\,(5,6),\,(6,5)\,\} \subset \Omega$$

$$\text{“resultados iguales} \longleftrightarrow \{\,(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\,\} \subset \Omega$$

De ahora en adelante, siempre que hablemos de evento estaremos haciendo referencia a un subconjunto de Ω.

Relación entre eventos

Dado un espacio muestral Ω, a los eventos los indicaremos con letras mayúsculas del alfabeto.

Las combinaciones lógicas de eventos se corresponden de manera natural con las operaciones de la teoría de conjuntos. Esto es,

$$\text{“}A \text{ o } B \text{ ocurren”} \equiv A \cup B, \qquad \text{“}A \text{ y } B \text{ ocurren”} \equiv A \cap B, \qquad \text{“}A \text{ no ocurre”} \equiv A^c.$$

Veamos algunas definiciones.

Definiciones.

- Si A y B son dos eventos tales que A ocurre si y sólo si B ocurre entonces decimos que A y B son *iguales*, y escribimos $A = B$.
  (Por ejemplo, el evento en el experimento del los dados de que la suma es par coincide con el evento ambos dados salen pares, o ambos salen impares.)
- El evento Ω que siempre ocurre dado cualquier resultado del experimento se denomina *evento seguro*.
- El evento $\varnothing$ el aquel que nunca ocurre dado cualquier resultado del experimento, y lo denominamos *evento imposible*.
- Decimos que los eventos A y B son *incompatibles* si la ocurrencia de uno implica la no ocurrencia del otro, i.e. $A \cap B = \varnothing$.

Comentario 1. Como ya vimos, no todo espacio muestral Ω es un conjunto *discreto* (finito o infinito numerable). Cuando Ω es continuo (no discreto) se necesita ser más cuidadoso con la noción de eventos. Más adelante discutiremos este caso.

El espacio muestral y los eventos son las piezas básicas de la axiomática de la probabilidad. Pueden considerarse como conceptos primitivos de la teoría, y son análogos al concepto de punto y recta de la geometría euclidiana.

1.4. Probabilidades en espacios muestrales discretos

Cuando consideramos el experimento de tirar una moneda al aire (Ejemplo 1.3.1), no es descabellado afirmar que la "probabilidad de que el resultado sea cara" y la "probabilidad de que el resultado sea número" son iguales, es decir, $1/2$ cada uno. Estas afirmaciones están basadas en nuestra experiencia de que ambos eventos son *equiprobables*, o al menos parecen serlo.

En general, dado un experimento aleatorio, la *probabilidad de un evento A* es un valor numérico $\mathbb{P}(A) \in [0,1]$, que se asocia al evento y que brinda información sobre la frecuencia con que dicho evento sucede.

Ejemplo 1.4.1. Si consideramos el experimento de lanzar un dado, podemos pensar que cada resultado diferente en $\Omega = \{1,2,3,4,5,6\}$ es equiprobable, y por lo tanto podemos asignarle probabilidad $1/6$ a cada número entre 1 y 6. Consideramos el evento A correspondiente a que el resultado es 1 o 2. ¿Cuál es la probabilidad que le podemos asignar al evento A? Dado que cada suceso elemental es equiprobable, una respuesta razonable sería que la probabilidad de ese evento debería ser la misma que la de los eventos $B = \{3,6\}$, y $C = \{5,4\}$. En este sentido, sería natural definir $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/3$.

Ejemplo 1.4.2. Generalizando el ejemplo anterior, supongamos que tenemos un experimento aleatorio donde $\Omega = \{\omega_1, \ldots, \omega_n\}$. Si consideramos que los sucesos elementales son equiprobables se tiene que la probabilidad de cada suceso es $1/n$. Dado $k \leqslant n$, sea $A = \{\omega_1, \ldots, \omega_k\}$. ¿Cuál sería la probabilidad de A? Dado que cada ω_i es equiprobable, la respuesta tiene que depender solo del cardinal de A, i.e. k. Procediendo de manera similar al ejemplo anterior podemos asumir que $\mathbb{P}(A)$ es lineal en k. De esta manera concluimos que si los sucesos elementales son equiprobables entonces la probabilidad de A es

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{k}{n}. \tag{1.1}$$

Ejemplo 1.4.3 (La paradoja de De Méré)**.** Consideremos el experimento aleatorio de observar la suma del resultado de tirar tres dados. Para este experimento podemos considerar como espacio muestral $\Omega = \{3,4,\ldots,18\}$. ¿Sería razonable decir que cada suceso elemental sea equiprobable? De Méré era un apostador francés que observó -de realizar varias veces el mismo experimento- que la frecuencia con que aparecía el número 11 era mayor que la frecuencia del número 12.

◇ **1.1.** Realizar el *Ejercicio 1* de la *Práctica 1* en *R*.

1.4.1. Distintas nociones de probabilidad

A lo largo de la historia se utilizaron distintas formas de asignar probabilidades a eventos de un experimento aleatorio. Existen tres interpretaciones distintas de la probabilidad, a saber, *clásica*, *frecuentista*, y *subjetiva*.

1. La *probabilidad clásica* se aplica en experimentos aleatorios con una cantidad finita de resultados equiprobables. En este caso la probabilidad de un evento A es la fracción de sucesos elementales en el cual A ocurre. Es decir,

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

(Cf. con Ejemplo 1.4.2.)

2. La *interpretación frecuentista* de la probabilidad se utiliza cuando la repetición de un experimento (en mismas condiciones) muestra un comportamiento regular destacable (como el experimento de lanzar una moneda.) Supongamos que realizamos el mismo experimento n veces, y sea $n(A)$ el número de veces que el evento A ocurre. Entonces la razón

$$\mathrm{F}_n(A) := \frac{n(A)}{n}$$

se denomina *frecuencia relativa* del evento A, en n repeticiones del experimento. En ciertos experimentos se observa que la frecuencia relativa para diferentes repeticiones se aproxima a un valor fijo cuando n crece. En este caso se define la probabilidad del evento A como

$$\mathbb{P}(A) := \lim_n \frac{n(a)}{n}.$$

(Acá no pretendemos ser formales con la noción de límite. Sin embargo en la Sección 1.5 veremos un ejemplo particular donde este límite existe. (Cf. Ley de los grandes números.)

♢ **1.2.** Realizar el *Ejercicio 2* de la *Práctica 1* en *R*.

3. La interpretación *subjetiva* de la probabilidad representa una medida del grado de creencia o confianza que se tiene respecto a un evento de un experimento. Afirmaciones tales como la probabilidad de que mañana llueva es de 30 %, o la probabilidad de que la humanidad se extinga a fines de este siglo es de 50 %, pertenecen a esta interpretación.

1.4.2. La probabilidad como concepto primitvo

A continuación se dará una definición formal del concepto de probabilidad, que extiende a cada uno de las interpretaciones anteriores, siendo esta la versión moderna de la teoría.

En la axiomática de la teoría de la Probabilidad el concepto de probabilidad se asume como un concepto primitivo, y puede considerarse como una noción análoga a la de *masa* en Mecánica, o *longitud* en Geometría.

Definiciones.

- Sea Ω un espacio muestral discreto. Una *función de probabilidad* es una función $p : \Omega \to [0,1]$ que satisface $\sum_{\omega\in\Omega} p(\omega) = 1$.
- Una función de probabilidad en Ω induce una función $\mathbb{P}$ sobre los eventos, a la que llamaremos *medida de probabilidad* o *probabilidad*, de la siguiente manera

$$\mathbb{P}(A) := \sum_{\omega\in A} p(\omega), \qquad A \subset \Omega.$$

En particular, $\mathbb{P}(\{\omega\}) = p(\omega)$, $\forall\, \omega \in \Omega$.

De esta manera, un espacio muestral (discreto) Ω y una medida de probabilidad $\mathbb{P}$ forman un *modelo matemático de un espacio de probabilidad*, al que denotaremos $(\Omega, \mathbb{P})$. (Observar que la probabilidad $\mathbb{P}$ es una función definida en las partes de Ω, i.e. los subconjuntos de Ω).

En muchas situaciones, con el fin de ser lo más general posible, la medida de probabilidad $\mathbb{P}$ se utiliza como parámetro del modelo sin especificar algún valor concreto. Por ejemplo, nos gustaría analizar el experimento de lanzar una moneda donde la probabilidad de cara es p y la probabilidad de número es $q = 1 - p$.

Algunas consecuencias

◇ **1.3.** Sea Ω un espacio muestral. Dar una medida de probabilidad sobre Ω es equivalente a dar una función P definida en las partes de Ω en $[0,1]$ tal que

- $P(\Omega) = 1$;
- si A y B son suconjuntos disjuntos de Ω, entonces $P(A \cup B) = P(A) + P(B)$.

◊ **1.4.** Dado $(\Omega,\mathbb{P})$ un espacio de probabilidad discreto es fácil ver que se cumplen los siguientes resultados.

- $\mathbb{P}(\varnothing) = 0$;
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
- si $A, B \subset \Omega$, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

1.4.3. Ejemplos de espacios de probabilidad

Un ejemplo de particular importancia es cuando los resultados de un experimento se asumen equiprobables.

Ejemplo 1.4.4. Sea $\Omega = \{\omega_1, \ldots, \omega_n\}$ y sea $\mathbb{P}$ la medida de probabilidad inducida por la función constante p sobre Ω, es decir, $p(\omega_i) = 1/n$, $(i = 1, \ldots, n)$. En este caso decimos que los sucesos elementales son *equiprobables*, y la probabilidad $\mathbb{P}$ se denomina *probabilidad uniforme*. Observar que en este caso se tiene que para todo evento A resulta $\mathbb{P}(A) = \#A/n$. (Cf. con definición de probabilidad clásica en página 22 y Ejemplo 1.4.2.)

Varios de los espacios muestrales de los ejemplos tratados en la Sección 1.3.1 pueden ser modelados por una medida de probabilidad uniforme. En algunos casos se pueden aproximar más o menos a la realidad del experimento. ¿Es razonable asumir que el día de nacimiento es equiprobable?

Ejemplo 1.4.5. Supongamos que repetimos 10 veces el experimento de lanzar una moneda, y consideramos la probabilidad uniforme sobre el espacio muestral asociado ($\#\Omega = 2^{10}$). Es claro que la probabilidad de obtener 10 caras es $1/1024$. La probabilidad de obtener 5 caras es de $\binom{10}{5}/1024$, donde $\binom{10}{5} = 252$ es la cantidad de formas distintas de elegir 5 elementos de un conjunto de 10.

Ejemplo 1.4.6. Consideremos el experimento de lanzar 2 dados y estudiar la suma de los resultados. Dos posibles espacios muestrales a considerar son:

$$\Omega_1 = \{(1,1),(1,2),...,(1,2),...,(6,1),...,(6,6)\}$$

ó

$$\Omega_2 = \{2,3,\ldots,12\}$$

En Ω_1, una medida de probabilidad razonable a considerar es la uniforme. Sin embargo en Ω_2 no parace razonable un modelo de espacio de probabilidad equiprobable (recordar la paradoja de De Méré).

$\diamondsuit$ **1.5.** Partiendo del espacio de probabilidad $(\Omega_1, \mathbb{P})$ con $\mathbb{P}$ la probabilidad uniforme, hallar las probabilidades de los sucesos elementales de Ω_2 como eventos del espacio muestral Ω_1.

Ejemplo 1.4.7. Supongamos que elegimos r dígitos, entre 0 y 9, al azar de manera uniforme. Podemos considerar $\Omega = \{(a_1, \ldots, a_r) : 0 \leqslant a_i \leqslant 9\}$, como subconjunto de $\mathbb{Z}^k$, con la probabilidad uniforme.

$\diamondsuit$ **1.6.** La probabilidad de que ningún dígito excede k $(0 \leqslant k \leqslant 9)$ es $(k+1)^r/10^r$; la probabilidad que k sea el dígito más grande es de $((k+1)^r - k^r)/10^r$.

Veamos algunos ejemplos con espacios muestrales discretos infinitos.

Ejemplo 1.4.8. (*Uber*) Consideremos el experimento de contar el número de autos que trabajan para Uber que pasan frente a la facultad. Podemos considerar el espacio muestral $\Omega = \{0, 1, \ldots\}$. Hay algunos resultados que sugieren que una probabilidad razonable, para el caso de k autos, es que sea proporcional a $\lambda^k/k!$. En general, la "tasa"λ depende del caso a modelar.

$\diamondsuit$ **1.7.** ¿Cuál es la constante de proporcionalidad?

Ejemplo 1.4.9. Considerar el expermiento de contar el número de lanzamientos de una moneda hasta que salga cara. Un modelo posible es considerar

$$\Omega = \{[C], [N,C], [N,N,C], \ldots, [N, \ldots, N, C], \ldots\},$$

y la función de probabilidad

$$p([C]) = 2^{-1}, p([N,C]) = 2^{-2}, \ldots, p([\underbrace{N, \ldots, N}_{n-1}, C]) = 2^{-n}, \ldots$$

$\diamondsuit$ **1.8.** ¿Define esto una función de probabilidad?

En el ejemplo anterior, si queremos conocer el número promedio de lanzamientos realizados, podemos considerar el siguiente modelo: $\Omega = \mathbb{N} \cup \{\infty\}$ y $p(1) = 2^{-1}, p(2) = 2^{-2}, \ldots, p(n) = 2^{-n}, \ldots$. El número promedio de lanzamientos tirados es $\sum_{n=1}^{\infty} np(n)$. Observar que si $f(x) = \sum_{n=1}^{\infty} x^n/2^n = (x/2)/(1-(x/2))$, entonces el promedio resulta $f'(1) = 2$.

$\diamondsuit$ **1.9.** Consideremos el experimento que consiste en observar la suma del resultado de tirar un dado equilibrado tres veces. Hallar la probabilidad de los eventos A_k, $3 \leqslant k \leqslant 18$, donde A_k es el conjunto de $\omega \in \Omega$ tales que la suma es k.

Comentario 2. Dar una buena definición de la noción de probabilidad llevó mucho tiempo. Es importante destacar que la medida de probabilidad $\mathbb{P}$ es dada *a priori*, como concepto primitivo del modelo. La discusión de si un evento tiene realmente asignado de manera natural una probabilidad no pertenece a la teoría de probabilidad misma (comparar con el concepto de masa o longitud.) Además, la determinación de $\mathbb{P}$ a través de observaciones (como la del experimento de la moneda) es un tema que concierne a la *estadística*.

1.5. Probabilidad axiomática vs frecuentista

Una vez definido el espacio de probabilidad surge una pregunta natural. Suponiendo que repetimos un experimento varias veces.

> *¿Cuál es la relación entre la frecuencia relativa de éxitos de un evento y la probabilidad teórica del mismo?*

La respuesta a esta pregunta es uno de los resultados más importantes de la teoría de probabilidad, a saber, *la ley de los grandes números* (o *ley de promedios*): la frecuencia relativa asociada a un evento tiende a la probabilidad teórica del mismo. Este importante resultado es una primera justificación del modelo de probabilidad.

1.5.1. Repetición de tiradas de una moneda

Para motivar este resultado haremos un análisis cuidadoso del experimento de lanzar una moneda n-veces. Lo que dice la ley de los grandes números en este caso es que si consideramos el modelo de lanzar una moneda muchas veces, la proporción de veces que resulta cara es "aproximadamente" $1/2$.

Dado que nuestra moneda idealizada es "fiel", asumimos que nuestro espacio de probabilidad $(\Omega_n, \mathbb{P})$ es el siguiente:

- Ω_n está formado por las sucesiones $\omega = (\omega_1, \dots, \omega_n)$, donde cada ω_i representa "C" (cara) o "N" (número). ($\#\Omega_n = 2^n$.)

- $\mathbb{P}$ es la probabilidad uniforme: cada suceso elemental $\omega \in \Omega$ tiene probabilidad 2^{-n}.

(Más adelante formalizaremos la idea de “repetir varias veces” un experimento que nos llevará a justificar el espacio de probabilidad elegido.)

Nuestro objetivo es analizar cuántos $\omega \in \Omega_n$, de las 2^n posibilidades, tiene exactamente k caras. Denotemos por $A_k^{(n)}$ este evento. Calcular la probabilidad de $A_k^{(n)}$ es un problema puramente calculatorio: en n celdas, ¿de cuántas maneras podemos elegir k lugares?

♢ **1.10.** Resulta que el cardinal de $\#A_k^{(n)}$ es igual a

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

Realizando el gráfico de $\mathbb{P}(A_k^{(n)})$ en función de k, se observa que el evento más probable se toma cuando $k = n/2$ si n es par, y $k = (n-1)/2$ y $k = (n+1)/2$ si n es impar.

♢ **1.11.** Utilizando la *fórmula de Stirling*

$$n! = e^{-n}n^n\sqrt{2\pi n}(1+\delta_n), \qquad (\delta_n \to 0 \text{ cuando } n \to \infty)$$

probar que

$$\mathbb{P}(A_n^{(2n)}) \approx \frac{1}{\sqrt{\pi n}}.$$

Esto significa que en $2n$ tiradas, la probabilidad de que salgan exactamente la mitad de caras tiende a cero. Pero esto no es exactamente lo que queríamos poner a prueba con la afirmación que queremos analizar (recordar el “aproximadamente”).

Dado $\varepsilon > 0$, sea $C(\varepsilon, n)$ el número de n-tiradas tales que la cantidad de caras relativas diste menos de ε de $1/2$.

Lo que vamos a probar es el siguiente resultado.

Teorema 1.1 (Ley de los grandes números)**.**

$$\lim_n \frac{C(\varepsilon, n)}{2^n} = 1.$$

Este resultado dice que cuando n es grande la proporción de caras en tirar n-veces una moneda es la mitad, independientemente si comenzamos con una racha muy mala.

Este resultado es una aplicación de la *ley de los grandes números* aplicada al experimento de lanzar sucesivamente una moneda. Esta “ley” es uno de los resultados más importante de la teoría de la Probabilidad, y a la cual le dedicaremos un buen tiempo en este curso.

Del ejercicio 1.10, para probar este teorema basta probar el siguiente resultado

$$\lim_{n} \sum_{k:|k/n-1/2|>\varepsilon} \binom{n}{k} \frac{1}{2^n} = 0.$$

Históricamente la prueba se basaba en un análisis combinatorio de la fórmula anterior. A continuación daremos una prueba distinta del teorema que utiliza conceptos más modernos y que motivará varios conceptos que trabajaremos en este curso.

Una forma de cuantificar la cantidad de caras es introducir el concepto de variable aleatoria.

Definición. Una *variable aleatoria* en un espacio de probabilidad discreto Ω le hace corresponder a cada suceso elemental $\omega \in \Omega$ un número real.

En otras palabras, una variable aleatoria X en Ω es simplemente una función $X : \Omega \to \mathcal{R}$. [2]

Para $i = 1, \dots, n$, sean las variables aleatorias $X_i : \Omega_n \to \mathbb{R}$ definidas por

$$X_i(\omega) = \begin{cases} 1 & \omega_i = C \\ 0 & \omega_i = N \end{cases} \tag{1.2}$$

donde $\omega = (\omega_1, \dots, \omega_n) \in \Omega_n$. Es decir, $X_i(\omega)$ toma el valor 1 cuando la i-ésima coordenada de ω es cara. De esta manera podemos definir la nueva función $S_n = X_1 + \cdots + X_n : \Omega_n \to \mathbb{R}$ que cuenta la cantidad de caras en cada tirada.

Definición. Definimos la *esperanza* de una variable aleatoria X, definida en un espacio de probabilidad $(\Omega, \mathbb{P})$, como

$$\mathbb{E}X := \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \tag{1.3}$$

La esperanza de X es el promedio de sus valores sobre los sucesos elementales, ponderado por las probabilidades de los mismos. En particular, en el caso equiprobable resulta $\mathbb{E}X = \frac{1}{\#\Omega} \sum_{\omega\in\Omega} X(\omega)$.

Dado que el promedio (como función de funciones de $\Omega \to \mathbb{R}$), es una función lineal, resulta que si X e Y son dos variables aleatorias definidas en Ω, entonces

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y, \qquad \alpha, \beta \in \mathbb{R}.$$

[2]Si $(\Omega, \mathbb{P})$ es un espacio de probabilidad discreto, entonces una variable aleatoria $X : \Omega \to \mathbb{R}$ define un nuevo espacio de probabildad discreto, a saber, $(\mathbb{R}, \mathbb{P}_X)$, donde: $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$. Más adelante retomaremos esta discusión.

Reordenando la suma, observar que si X toma los valores $x_1,\dots,x_\ell$, entonces la esperanza coincide con

$$\mathbb{E}X = \sum_{i=1}^{\ell} x_i \mathbb{P}(X = x_i),$$

donde abusando notación $\mathbb{P}(X = x_i)$ denota la probabilidad del conjunto $\{\omega \in \Omega_n : X(\omega) = x_i\}$.

◊ **1.12.** Hallar $\mathbb{E}X_1$, y $\mathbb{E}S_n$.

Observar que

$$\frac{C(\varepsilon,n)}{2^n} = \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \leqslant \varepsilon\right),$$

y como $\mathbb{P}(\Omega_n) = 1$, basta probar que

$$\lim_n \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) = 0.$$

Para tal fin necesitamos la *desigualdad de Chebyshev*.

Lema 1.2.

$$\mathbb{P}(|X| > \varepsilon) \leqslant \frac{\mathbb{E}X^2}{\varepsilon^2}$$

Demostración. Observar que

$$\begin{aligned}\mathbb{P}(\{\omega : |X(\omega)| > \varepsilon\}) &= \sum_{\omega : |X(\omega)| > \varepsilon} \frac{1}{2^n} \leqslant \sum_{\omega : |X(\omega)| > \varepsilon} \frac{X(\omega)^2}{\varepsilon^2} \frac{1}{2^n} \\ &\leqslant \sum_{\omega \in \Omega_n} \frac{X(\omega)^2}{\varepsilon^2} \frac{1}{2^n} = \frac{\mathbb{E}X^2}{\varepsilon^2}.\end{aligned}$$

□

Comentario 3. Dado A un subconjunto de un espacio E, la función *indicatriz de A* es la función $\mathbb{1}_A$ definida por

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in E \setminus A. \end{cases}$$

Como el promedio es una función monótona y

$$\varepsilon^2 \mathbb{1}_{\{\omega \in \Omega_n : |X(\omega)| > \varepsilon\}} \leqslant X^2$$

entonces una demostración similar a la desiguladad de Chebyshev resulta de

$$\mathbb{P}(|X| > \varepsilon) = \mathbb{E}\left(\mathbb{1}_{\{\omega \in \Omega_n : |X(\omega)| > \varepsilon\}}\right) \leqslant \frac{\mathbb{E}X^2}{\varepsilon^2}$$

♢ **1.13.** Partiendo de la definición de X_i probar $\mathbb{E}(X_i - \frac{1}{2})(X_j - \frac{1}{2}) = \frac{\delta_{ij}}{4}$.

Demostración Teorema 1.1. Utilizando la desigualdad de Chebyshev resulta

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) &\leqslant \frac{1}{\varepsilon^2}\mathbb{E}\left(\frac{S_n}{n} - \frac{1}{2}\right)^2 \\
&= \frac{1}{\varepsilon^2}\mathbb{E}\sum_{i,j}(X_i - \frac{1}{2})(X_j - \frac{1}{2}) \\
&= \frac{1}{4n\varepsilon^2},
\end{aligned}$$

es decir,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) \leqslant \frac{1}{4n\varepsilon^2}.$$

Tomando límite en n concluimos

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega_n : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \varepsilon\right\}\right) = 0.$$

□

Ejercicios para hacer en R

Nota: Para estas primeras prácticas, le sugeriremos algunas líneas de código en **R**. Usted puede usarlas o programar a su *"gusto"*.

Ejercicio 1: Paradoja de De Méré

Consideremos el experimento que consiste en observar la suma del resultado de tirar tres dados.

a) ¿Cuáles son los resultados posibles de este experimento?

b) Utilizando la función *sample()*, simular *una* observación de este experimento:

```
dado = 1:6
tirar_tres_dados = sample (dado, 3, replace = TRUE)
observacion = sum (tirar_tres_dados)
observacion
```

c) Queremos ver qué ocurre (simulando lo observado por De Méré) al repetir muchas veces (N veces) el experimento anterior.

1. Repita el experimento de la parte **b)** N veces. Para ello podría usar la función *for* y crear un vector en el que se registre la suma de los tres dados correspondientes a cada experimento:

Nota: tenga en cuenta que es más eficiente primero crear un vector "vacío" y luego rellenarlo, que ir concatenando vectores.

```
N = 100

observaciones = rep(0, N) # creamos un vector de 0's, que lo rellenaremos con la suma
                          #correspondiente a cada repetición del experimento.



for (i in 1:N)
{
  tirar_tres_dados = sample (dado, 3, replace = TRUE)
  observaciones[i] = sum (tirar_tres_dados)
}

observaciones # es ahora un vector de N coordenadas con las sumas correspondientes a cada
              # experimento
```

¿Qué números aparecen con mayor frecuencia en el vector *observaciones*?

2. Podemos ver gráficamente lo ocurrido al contar la cantidad de veces en que la suma dio 3,4,...,18. Una manera es usar la función *'hist'*

```
help("hist")

hist(observaciones, breaks = 3:18, main= "Frecuencia con que la suma dio 3,4,...,18",
     sub=paste('Para N=', N), xlab= "Suma de los tres dados", ylab= "Frecuencia",
     col= "green")
```

d) Ahora consideraremos las *frecuencias relativas*, es decir la proporción de veces que ocurre cada uno de los resultados posibles:

$$\mathrm{F}_N(i) = \frac{\text{cant. de veces que aparece } i \text{ en el vector 'observaciones'}}{N}$$

para $i = 3, 4, ..., 18$.

Para eso, usamos nuevamente la función *hist* y para que divida las cantidades entre N, ponemos *prob* = *TRUE*:

```
hist(observaciones, breaks = 3:18, prob= TRUE,
     main= "Frecuencia relativa con que la suma dio 3,4,...,18",
     sub=paste('Para N=', N), xlab= "Suma de los tres dados",
     ylab= "Frecuencia", col= "pink")
```

e) Repita el paso **c)** varias veces. ¿Ocurre siempre lo mismo? ¿Observa alguna tendencia?

f) Repita ahora el paso **c)** pero aumentando el número de repeticiones N. ¿Observa alguna tendencia?

g) ¿Por qué cree que algunos resultados son más frecuentes que otros?

Para De Méré fue una sorpresa que los resultados de sus experimentos dieran que el 11 era más probable que el 12. Su argumento para justificar que deberían de ser equiprobables fue el siguiente:

> "Sean A y B los eventos de que la suma es 11 y 12 respectivamente.
> A ocurre en 6 formas distintas:
>
> $$6:4:1, \quad 6:3:2, \quad 5:5:1, \quad 5:4:2, \quad 5:3:3, \quad 4:4:3,$$
>
> y B ocurre también en 6 maneras distintas:
>
> $$6:5:1, \quad 6:4:2, \quad 6:3:3, \quad 5:5:2, \quad 5:4:3, \quad 4:4:4,$$
>
> y, al ser equiprobables, la probabilidad coincide".

¿Por qué el argumento es falaz?

Ejercicio 2: Interpretación frecuentista de la probabilidad

La *interpretación frecuentista* de la probabilidad se utiliza cuando la repetición de un experimento (bajo las mismas condiciones) muestra un comportamiento regular destacable (como el experimento de lanzar una moneda).

Supongamos que realizamos el mismo experimento n veces, y sea $n(A)$ el número de veces en que ocurre el evento A. El cociente:

$$\mathrm{F}_n(A) := \frac{n(A)}{n}$$

se denomina *frecuencia relativa* del evento A (que ya definimos informalmente en el ejercicio anterior) en n repeticiones del experimento.

En ciertos experimentos se observa que la frecuencia relativa para diferentes repeticiones se aproxima a un valor fijo cuando n crece. En este caso, se define la probabilidad (*definición frecuentista de la probabilidad*) del evento A como:

$$\mathbb{P}(A) := \lim_{n \to +\infty} \mathrm{F}_n(A).$$

(Acá no pretendemos ser formales con la noción de límite. Sin embargo, veremos más adelante que la *Ley de los Grandes Números* asegura que $\mathrm{F}_n(A)$ converge a $\mathbb{P}(A)$, siendo $\mathbb{P}(A)$ el número proviniente de la definición *clásica* de probabilidad).

a) Considerando esta definición *frecuentista* de probabilidad,

1. Probar que $\mathbb{P}(A) \in [0, 1]$ para cualquier suceso A.
2. Probar que $\mathbb{P}(\Omega) = 1$ y $\mathbb{P}(\emptyset) = 0$.
3. Probar que si A y B son incompatibles, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
4. Probar que si $A \subset B$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.
5. Probar que $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

b) Consideremos el experimento que consiste en lanzar un dado justo.

1. Determinar el espacio muestral
2. Simular el experimento en la computadora utilizando la función *sample*:

```r
dado = 1:6
tirada = sample(dado, 1)
```

3. Consideremos el evento $A = \{2, 3\}$. Para un valor de n fijo, calcule $\mathrm{F}_n(A)$.

```r
#Para un n fijo,
n = 20 # Comencemos con un n chico, de prueba

simulacion = sample(dado, n, replace = TRUE)
simulacion
```

```
##  [1] 1 1 5 5 2 2 2 1 3 2 2 6 2 2 2 4 3 2 5 2
```

```
A = c(2,3)
```

Queremos contar cuántas veces aparece el 2 o el 3 en el vector *simulacion*. Para eso podemos considerar el operador *%in%*, que genera un vector de TRUE o FALSE. Cada vez que se encuentre con un elemento del vector A en el vector *simulacion*, pondrá TRUE:

```
simulacion %in% A
```

```
##  [1] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
## [12] FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE
```

Ahora queremos contar cuántos TRUE hay en el vector anterior. Si le aplicamos la función *sum*, esta contará solo los TRUE (recuerde que TRUE $= 1$, FALSE $= 0$).

```
sum(simulacion %in% A)
```

```
## [1] 12
```

```
# Finalmente,
nA = sum(simulacion %in% A)
F_nA = nA/n
F_nA
```

```
## [1] 0.6
```

4. Repita el paso anterior para varios valores de n y grafique $\mathrm{F}_n(A)$ en función de n.

```
# Consideraremos ahora un vector de n's:
N = 1000
n = seq(10, N, by = 10)

# Creamos un vector que vamos a rellenar con los valores de F_n(A) para cada valor de n
F_nA = rep(0, length(n))

# Repetimos lo de la parte anterior para cada uno de esos valores de n y rellenamos el
# vector F_nA:
for (i in 1:length(n))
{
  simulacion = sample (dado, n[i], replace = TRUE)
  F_nA[i] = sum(simulacion %in% A)/n[i]
}

# Ahora graficamos el vector F_nA (en el eje Oy) en función de n (en el eje Ox):
plot(n, F_nA, main="Fn(A) a medida que aumenta n", xlab= "n", ylab = "Fn(A)")

# Comparamos con 1/3 = P(A)
lines(n, rep(1/3, length(n)), col= "red")
```

5. ¿Qué ocurre al aumentar la cantidad n de tiradas? Si utiliza el script anterior, observe qué ocurre a medida que aumenta N (el valor máximo del vector n).

Ejercicio 3: Observación de la Ley de los Grandes Números.

Consideremos el experimento que consiste en tirar una moneda n veces. Luego, si la moneda es "fiel", el espacio de probabilidad *equiprobable* que modela este experimento es $(\Omega_n, \mathbb{P})$, con:

$$\Omega_n = \{\omega = (\omega_1, ..., \omega_n) : \omega_i \in \{C, N\}\}$$

y $\mathbb{P}$ tal que $\mathbb{P}(\omega) = \frac{1}{2^n}$ $\forall \omega \in \Omega$.

> Lo que dice la *Ley de los Grandes Números* en este caso es que la proporción de veces que resulta cara es *aproximadamente* (veremos en qué sentido) $\frac{1}{2}$.

¿Esto significa que saldrán exactamente $\frac{n}{2}$ caras?. El objetivo de este ejercicio es analizar el significado de la afirmación anterior.

a) Consideremos el evento $A_k^{(n)}$ = "salieron exactamente k caras".

Luego, $A_k^{(n)}$ está formado por $\binom{n}{k}$ elementos de Ω_n (ya que sabemos que hay exactamente k lugares ocupados por C, lo que cambia son las ubicaciones de estos lugares, i.e. en qué tiradas salió C). Entonces:

$$\mathbb{P}(A_k^{(n)}) = \frac{|A_k^{(n)}|}{|\Omega|} = \frac{1}{2^n} \times \binom{n}{k}$$

1. Realice un gráfico de $\mathbb{P}(A_k^{(n)})$ en función de k.

```r
n = 10

k = 0:n # Vector con todos los valores posibles de k

PA_k = choose(n, k)/(2^n) # De una ya creamos un vector con todas las P(A_k)

plot(k, PA_k, main= paste("P(A_k) en función de k para n=", n), xlab = "k",
     ylab = "P(A_k)", type = "s")
```

¿Para qué valor/valores de k el evento $A_k^{(n)}$ es más probable? (en particular, observe qué ocurre cuando considera valores pares o impares para n). Observe qué ocurre para distintos valores de n.

Considere a partir de ahora valores pares para n. ¿Qué ocurre con $\mathbb{P}(A_{\frac{n}{2}}^{(n)})$ cuando $n \to \infty$?

2. Utilizando la *fórmula de Stirling*:

$$n! = e^{-n} n^n \sqrt{2\pi n}(1 + \delta_n), \qquad (\delta_n \to 0 \text{ cuando } n \to \infty)$$

probar que $\mathbb{P}(A_n^{(2n)}) \approx \frac{1}{\sqrt{\pi n}}$.

Es decir que, *la probabilidad de que salgan exactamente la mitad de caras tiende a cero.* ¿Esto contradice la afirmación del comienzo?.

b) Realice una vez el experimento que consiste en tirar la moneda n veces y calcule la frecuencia relativa de "C", es decir:

$$\frac{S_n}{n} = \frac{\text{cant. de veces -dentro de las n- que salió C}}{n}$$

```r
# Experimento: tirar n veces la moneda y contar cuántas veces salió C:
Moneda = c("C", "N")

n = 100

tiradas = sample(Moneda, n, replace = TRUE)

Sn = sum (tiradas %in% "C") # Sn cuenta cuántas veces aparece "C" en el vector "tiradas"

prop_caras = Sn/n
```

Repita el experimento anterior para distintos valores de n. Observe qué ocurre con $\frac{S_n}{n}$ cuando $n \to \infty$.

```r
# Consideraremos ahora un vector de n's:
N = 10000
n = seq(10, N, by = 10)

# Creamos un vector que vamos a rellenar con los valores de Sn/n para cada valor de n
prop_C = rep(0, length(n))

# Repetimos lo de la parte anterior para cada uno de esos valores de n y rellenamos el
# vector prop_C:
for (i in 1:length(n))
{
  tiradas = sample(Moneda, n[i], replace = TRUE)
  prop_C[i] = sum(tiradas %in% "C")/n[i]
}

# Ahora graficamos el vector prop_C (en el eje Oy) en función de n (en el eje Ox):
plot(n, prop_C, main='Una simulación de Sn/n en función de n', xlab= "n", ylab = "Sn/n",
     type = "p")

# Comparamos con 1/2
lines(n, rep(1/2, length(n)), col= "red")
```

Repita el experimento anterior varias veces. ¿Observa alguna tendencia?

c) [Opcional]

La *Ley de los Grandes Números* dice que S_n "tiende" a $\frac{1}{2}$, pero no en el sentido habitual de límite sino en el que describimos a continuación. $\forall \varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) = 1 - \frac{C(\varepsilon, n)}{2^n} \to 0 \text{ cuando } n \to \infty,$$

siendo

$$C(\varepsilon, n) = \#\{\omega \in \Omega_n : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \leq \varepsilon\}$$

Consideremos el suceso $A_n = A_n(\varepsilon) = \{\omega : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \varepsilon\}$. Queremos observar que $\mathbb{P}(A_n) \to 0$ cuando $n \to \infty$.

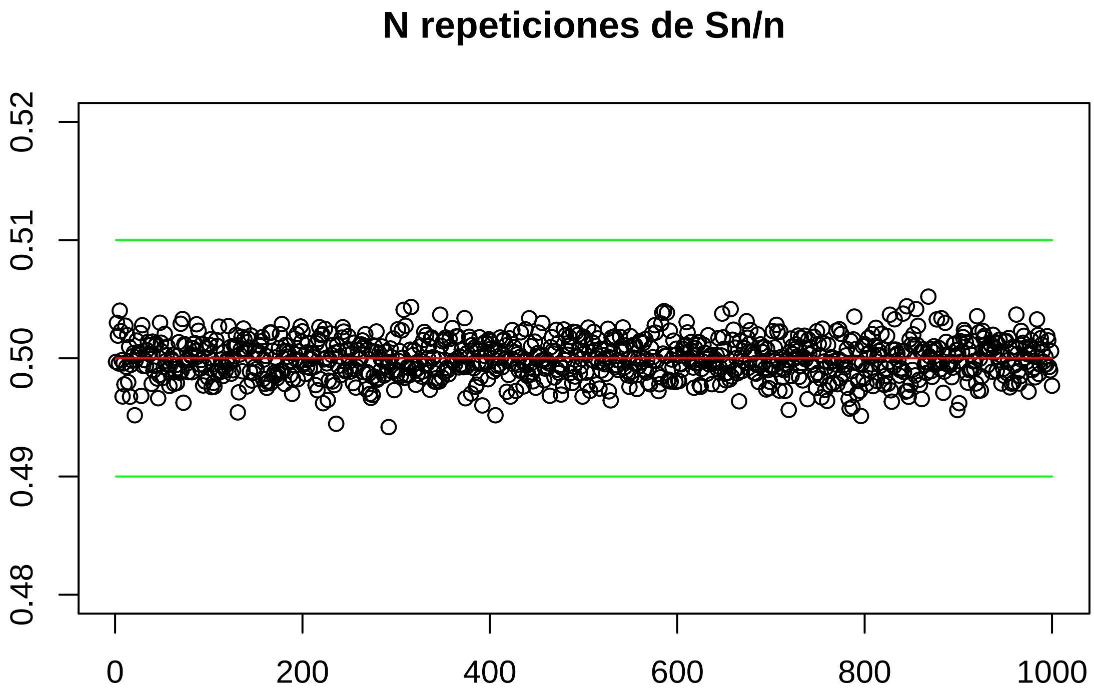
Para eso, repetiremos N veces el experimento de tirar la moneda n veces y observaremos el valor de S_n para cada experimento. Luego, analizaremos la *frecuencia relativa* del suceso A_n en esos N experimentos:

$$\mathrm{F}_N(A_n) = \frac{\text{cant. de veces -dentro de las N- que ocurre } A_n}{N},$$

ya que *sabemos* que $\mathrm{F}_N(A) \to \mathbb{P}(A)$ cuando $N \to \infty$ para cualquier suceso A (esta afirmación, como mencionamos en el ejercicio anterior, también se prueba a partir de la *Ley de los Grandes Números*).

A continuación, trataremos de tomar N y n suficientemente grandes (N grande para que $F_N(A) \to \mathbb{P}(A)$ cuando $N \to \infty$ y n grande para observar que $\frac{S_n}{n}$" $\to$ "$\frac{1}{2}$).

```
N = 1000 # cantidad de veces que vamos a repetir el experimento
Sn = rep(0, N) #vector con el valor de Sn en cada experimento

n = 100000

# Repetimos el experimento anterior N veces:
for (i in 1: N)
{
  tiradas = sample(Moneda, n, replace = TRUE)

  Sn[i] =  sum (tiradas %in% "C")

}

#Fijemos un valor de epsilon:
eps= 0.01 # Observar qué ocurre para distintos valores de eps

plot(1:N, (1/n)*Sn, main="N repeticiones de Sn/n", type = "p", lty=0.1,
     ylim=c(0.5-2*eps,0.5+2*eps), ylab=" ", xlab= " ")

lines(1:N, rep(1/2, N), col="red")

lines(1:N, rep(1/2 + eps, N), col="green")

lines(1:N, rep(1/2 - eps, N), col="green")
```



Realice varias veces el gráfico anterior, aumentando n y para distintos valores de ε.

Observación 1: Si tomamos n grande, ¿cuántas veces (N) tendríamos que repetir el experimento para poder observar que $\frac{S_n}{n}$ se escape del intervalo $[\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]$ en alguno de esos experimentos?.

Es posible probar que para poder observar un experimento en el que $\frac{S_n}{n}$ se escapa de dicho intervalo, habrá que repetir el experimento $N = e^{nk(\varepsilon)}$ veces (en promedio), siendo $k(\varepsilon) > 0$ una constante que depende de ε.

Observación 2: Para los resultados obtenidos en esta simulación, calculemos $F_N(A_n)$:

```
F_NAn =  sum( abs((Sn/n)-0.5) > eps)*(1/N)
```

En este caso, el valor de $F_N(A_n)$ es:

```
F_NAn
```

```
## [1] 0
```

para $\varepsilon =$

```
print(eps)
```

```
## [1] 0.01
```

Observar que $F_N(A_n)$ devolverá valores cercanos a 0 (o 0). Repita el experimento para distintos valores de N y ε teniendo en cuenta la observación 1.

Capítulo 2

Elementos de análisis combinatorio

Regla de conteo, muestreo, permutaciones, combinaciones, arreglos, subpoblaciones, distribución hipergeométrica, problemas de ocupación, tiempo de espera.

[Duración: 2 Semanas]

En esta sección trabajaremos sobre un espacio muestral finito con sucesos elementales equiprobables. Como vimos, en este caso la probabilidad de un evento A es igual al cardinal de A (*casos favorables*) sobre el cardinal de Ω (*casos posibles*).

En este capítulo adquiriremos elementos básicos de conteo para poder analizar modelos de espacio de probababilidad equiprobables. En particular estudiaremos distintas clases de espacios muestrales discretos.

2.1. Conteo básico

Regla fundamental del conteo: Supongamos que queremos hacer r elecciones de manera ordenada donde hay : m_1 posibilidades para la primera elección; m_2 posibilidades para la segunda elección; $\ldots\ldots$; m_r posibilidades para la r-ésima elección;
Entonces el número total de diferentes elecciones es

$$m_1 \cdot m_2 \cdot \cdots \cdot m_r.$$

Ejemplo 2.1.1. En un maso de cartas españolas cada carta tiene asociado uno de cuatro palos distintos y un número del 1 al 12. Concluimos que hay 48 cartas.

Ejemplo 2.1.2. ¿De cuántas maneras podemos ordenar los enteros $1, 2, \ldots, n$?
Para el primer lugar tenemos n posibilidades, para el segundo $n-1$ posibilidades, etc. Concluimos entonces que hay $n \cdot (n-1) \cdots 2 \cdot 1$ maneras distintas de ordenarlos (i.e. factorial de n, que denotamos $n!$).

Ejemplo 2.1.3. Supongamos que tenemos k banderas y las queremos colocar de maneras distintas en n astas alineadas. Acá asumimos que podemos poner varias banderas por asta de manera ordenada. ¿De cuántas maneras podemos izarlas?
Solución: Para la primer bandera tenemos n posibilidades. Dividendo el asta usada en dos (partes superior e inferior) tenemos $n+1$ posibilidades para la segunda bandera. Prosiguiendo de esta manera concluimos que el resultado es $n \cdot (n+1) \cdot \cdots \cdot (n+k-1)$.

A continuación veremos argumentos bastante más refinados que los ejemplos anteriores.

2.2. Muestras ordenadas

Consideremos una "población", conjunto, o urna, de r elementos, de donde tomamos muestras ordenadas de tamaño k. Una forma intuitiva de considerar el orden es pensar que vamos seleccionando de a uno los elementos de la población.

Muchos problemas de conteo pueden ser vistos como subconjuntos de funciones de un conjunto de índices (el dado por el ordenamiento, a saber, k) a un conjunto de posibles resultados. (Recordar la digresión, y sucesivo comentario en página 18). En este sentido sacar una muestra ordenada de longitud k de elementos de un conjunto con r resultados, es equivalente a dar una función $f : \{1, \ldots, k\} \to \{1, \ldots, r\}$ (que abreviaremos por $f : k \to r$).

A continuación veremos clases distintas de muestreos que definen diferentes espacios muestrales.

- *Muestreo con reposición:* cada vez que sacamos un elemento de la urna, lo devolvemos. Cada muestreo de este tipo es equivalente a dar una función $f : k \to r$.

- *Muestreo sin reposición:* Cada vez que sacamos un elemento de la urna, no está permitido usarlo nuevamente (necesitamos $k \leqslant r$). Equivalentemente, tomamos una función $f : k \to r$ inyectiva.

- *Muestreo (de todos los elementos) con reposición:* Es un muestreo con reposición en el que además se requiere que cada elemento de la urna sea elegido por lo menos una vez ($k \geqslant r$).

Equivalentemente, tomamos una función $f : k \to r$ sobreyectiva.

De la regla fundamental podemos concluir el siguiente resultado.

Lema 2.1. *En una población de r elementos y una muestra de tamaño k, existen r^k muestras con reposición y $(r)_k := r \cdot (r-1) \cdot \cdots \cdot (r-k+1) = r!/(r-k)!$ muestras sin reposición.*

Es decir, el cardinal del conjunto de funciones distintas $f : k \to r$ es r^k, y el subconjunto de las funciones inyectivas tiene cardinal $(r)_k$.

◇ **2.1.** Dar una prueba del lema anterior.

Cuando hablemos de lanzamientos de una moneda, o dado, estaremos modelando estos experimentos como un muestreo aleatorio con reposición. Sin embargo cuando barajamos una maso de naipes y hacemos una extracción sucesiva de las mismas estamos considerando el caso de muestreo sin reposición.

Hasta ahora no hablamos de azar ni de probabilidades asociadas a las muestras. En este capítulo, cuando digamos muestras al azar nos estaremos refiriendo a un modelo equiprobable, i.e., asignaremos las mismas probabilidades a cada uno de los posibles resultados. El conjunto de posibles resultados depende del muestreo.

Del Lema 2.1 resulta que la probabilidad de cada tira de largo k, en una población de r elementos, es $1/r^k$ para muestreo con reposición, y $1/(r)_k$ para muestreo sin reposición.

Supongamos que tenemos un conjunto de r elementos del cual tomamos muestras con reposición de largo k.

¿Cuál es la probabilidad del evento de que no aparezcan elementos repetidos?

(Es decir, aquellas muestras que también son muestras del mismo conjunto pero sin reposición.)

Del Lema 2.1 resulta que la probabilidad de ese evento es

$$\frac{(r)_k}{r^k} = \frac{r \cdot (r-1) \cdots (r-k+1)}{r^k}$$

A continuación veremos algunos ejemplos.

Ejemplo 2.2.1 (*Muestreo aleatorio de números*)**.** Consideremos muestras al azar con reposición de longitud 5 del conjunto $\{0, 1, \ldots, 9\}$. La probabilidad que los 5 dígitos aleatorios sean diferentes es $(10)_5/10^5 \approx 0{,}3024$.

♢ **2.2.** Realice el ejercicio 1 de la Práctica 2 en R.

Ejemplo 2.2.2. Supongamos que se colocan aleatoriamente r bolas en r celdas. La probabilidad que cada celda contenga una bola es $r!/r^r$. Por ejemplo, cuando $r = 7$ la probabilidad es 0,00612.

El ejemplo anterior es interesante dado que los resultados no parecen ser intuitivos. Por ejemplo, si uno supone que en una ciudad suceden 7 accidentes de tránsito cada semana, entonces suponiendo un modelo equiprobable, una cada 165 semanas se tendría que ocurren accidentes en cada día de la misma.

♢ **2.3.** Suponga que hay 7 pasajeros en un ascensor en un edificio de 10 pisos. Suponiendo que todas las formas diferentes de bajar los pasajeros son equiprobables. ¿Cuál es la probabilidad de que los pasajeros bajen en diferente piso?

Ejemplo 2.2.3 (*Problema del cumpleaños*)**.** ¿Cuál es la probabilidad p de que los cumpleaños de r estudiantes sean todos en días diferentes? (Acá estamos asumiendo 365 días para todos y que cada posible configuración de cumpleaños es equiprobable.) Según lo visto, la probabilidad es

$$p_r = \frac{(365)_r}{365^r} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{r-1}{365}\right) \tag{2.1}$$

Para $r = 23$ se tiene que $p < 1/2$ y por lo tanto con probabilidad mayor a $1/2$ hay dos estudiantes que cumplen el mismo día. Una aproximación a este resultado es

$$\log p_r \approx -\frac{1 + 2 + \cdots (r-1)}{365} = -\frac{r(r-1)}{2 \cdot 365},$$

de donde resulta que para $r = 23$ se tiene que $p \approx exp\left(-\frac{23 \cdot 22}{2 \cdot 365}\right) \approx 0{,}5$

2.3. Subpoblaciones y particiones

Definición. Una *subpoblación de tamaño r* es un subconjunto de r elementos de una población de tamaño n.

Lema 2.2. *Una población de n elementos tiene*

$$C^n_r := \binom{n}{r} := \frac{n!}{(n-r)!\,r!}$$

subpoblaciones de tamaño $r \leqslant n$.

♢ **2.4.** Dar una prueba del lema. Sugerencia: suponer que las subpoblaciones tienen orden, para luego identificarlas entre ellas.

Lo que dice el resultado es que podemos elegir un subconjunto de tamaño r, de uno de tamaño n, de $\binom{n}{r}$ formas distintas.

♢ **2.5.** Justifica la fórmula del binomio

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2.2}$$

♢ **2.6.** Sea Ω un conjunto finito. Probar que las partes de Ω, $\mathscr{P}(\Omega)$, están en biyección con la funciones de Ω en $\{0,1\}$. Concluir que $\#\mathscr{P}(\Omega) = 2^{\#\Omega}$.

Observar que de la fórmula del binomio (2.2), para $a = b = 1$, se tiene

$$2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Es decir, que $\#\mathscr{P}(\Omega)$ es la suma del número de conjuntos de k elementos, para $0 \leqslant k \leqslant n$ (si $\#\Omega = n$).

Digresión [Acción del grupo de permutaciones] El grupo de permutaciones de n elementos, $\mathcal{S}_n$, *actúa* sobre subconjuntos de $\Omega = \{\omega_1, \dots, \omega_n\}$ de manera natural:

$$\pi \in \mathcal{S}_n, \quad \pi\left(\{\omega_{i_1}, \dots, \omega_{i_k}\}\right) = \{\omega_{\pi(i_1)}, \dots, \omega_{\pi(i_k)}\}.$$

Este grupo actúa *transitivamente* sobre los conjuntos de k elementos (todos los subgrupos se obtienen de la acción sobre uno particular), teniendo como *estabilizador* (las permutaciones que fijan el grupo) a un grupo que es isomorfo a $\mathcal{S}_k \times \mathcal{S}_{n-k}$.

Es sorprendente que $\binom{n}{r}$ es siempre un número entero. En general se calcula por cancelación como el siguiente ejemplo:

$$\binom{11}{6} = \frac{11!}{5! \cdot 6!} = \frac{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7}{5 \cdot 4 \cdot 3 \cdot 2} = 11 \cdot \frac{10}{5 \cdot 2} \cdot \frac{9}{3} \cdot \frac{8}{4} \cdot 7.$$

Observar además que por cada subpoblación de tamaño r tenemos una subpoblación complementaria de tamaño $n - r$. Por lo tanto, tiene que haber tantas subpoblaciones de tamaño r como subpoblaciones de tamaño $n-r$, es decir

$$\binom{n}{r} = \binom{n}{n-r}.$$

(Para que la fórmula tenga sentido, se define $\binom{n}{0} = 1$.)

$\diamondsuit$ **2.7** (*Póquer*). En una mano de póquer, el orden de las cartas no importa. Por lo tanto tenemos $\binom{52}{5}$ que es alrededor de 2,5 millones de manos diferentes. ¿Cuál es la probabilidad de obtener 5 cartas distintas? ¿Cuál es la probabilidad de obtener escalera real, escalera color, póquer, full, color, escalera, trío, doble par, par? (Poner "póquer" en wikipedia para verificar resultados)

$\diamondsuit$ **2.8** (*Problema de ocupación*). Si consideramos el experimento de colocar r bolas en n celdas (recordar Ejemplo 1.3.10), la probabilidad de que una celda específica contenga ℓ bolas es

$$\binom{r}{\ell} \cdot \frac{1}{n^{\ell}} \cdot \left(1 - \frac{1}{n}\right)^{r-\ell}.$$

$\diamondsuit$ **2.9.** Supongamos que tenemos el vector $(\overbrace{\alpha,\cdots,\alpha}^{p},\overbrace{\beta,\cdots,\beta}^{q})$ de longitud $p+q$. ¿De cuántas formas diferentes podemos ordenarlos? (Una estrategia posible es distinguir las α's y las β's para luego no distinguirlas.)

Del ejercicio anterior se concluye que la cantidad de maneras diferentes de ordenarlas coincide con

$$\binom{p+q}{q} = \binom{p+q}{p}.$$

Observar que la formula anterior motiva otra respuesta al ejercicio 2.9. De $p+q$ lugares disponibles, si elegimos p de ellos y ponemos las α's, obtendremos otra forma de contestar nuestra pregunta.

$\diamondsuit$ **2.10.** Una aplicación interesante del resultado anterior es que el número de caminos de longitud minimal que comienzan en una esquina y terminan en otra de un tablero de ajedrez es $\binom{14}{7}$.

El ejercicio 2.9 motiva el siguiente resultado.

Lema 2.3. *Dada una población de n elementos, sean $n_1,\ldots,n_k$ enteros positivos, tales que*

$$n_1 + \cdots + n_k = n.$$

Entonces, hay

$$\binom{n}{n_1 \ldots n_k} := \frac{n!}{n_1!\cdots n_k!}$$

formas distintas de particionar, de manera ordenada, la población en k subpoblaciones de tamaños $n_1,\ldots,n_k$ respectivamente.

◊ **2.11.** Dar una prueba del resultado anterior.

◊ **2.12.** Hallar los coeficientes del miembro derecho de la siguiente igualdad

$$(x_1 + \cdots + x_k)^n = \sum_{j_1,\ldots,j_k:\, j_1+\cdots+j_k=n} c_{j_1,\ldots,j_k} x_1^{j_1} \cdots x_k^{j_k}.$$

2.3.1. Problemas de ocupación

En el Capítulo 1 vimos que la mayoría de los ejemplos se pueden modelar colocando r bolas en n celdas. En muchas oportunidades no es necesario distinguir las bolas (recordar Ejemplo 1.3.10), sino saber la cantidad de bolas que hay en cada celda. En tal caso, un evento que sólo dependa de esto está determinado por los números de ocupación de cada celda. Esto es, por los números

$$r_1, \ldots, r_n \in \mathbb{N}, \quad \text{tal que} \quad r_1 + \cdots + r_n = r, \tag{2.3}$$

donde r_i indica la cantidad de bolas en la i-ésima celda.

Por ejemplo, para el caso de 3 bolas en tres celdas, podemos representar la distribución $(2,0,1)$ como $|**| \quad |*|$ donde las barras delimitan las celdas.

Distribuciones distinguibles se identifican con distintas soluciones de (2.3).

Lema 2.4. *a) El número de distribuciones distinguibles de colocar r bolas en n celdas es:*

$$A_{r,n} := \binom{n-1+r}{n-1} = \binom{n-1+r}{r}.$$

b) Si además asumimos que los enteros $r_i \geqslant 1$, entonces hay

$$\binom{r-1}{n-1}$$

soluciones posibles.

◊ **2.13.** Dar una prueba del Lema 2.4. (Sugerencia: analizar la representación de "$*$" y "$|$" dada más arriba. Para la parte b) poner atención a los espacios entre bolas.)

Una prueba distinta para el caso de $r_i \geqslant 1$, es asumir que hay 1 bola en cada celda, y luego observar que las $r-n$ bolas restantes se distribuyen como el Lema 2.4 (pero para $r-n$ números de bolas).

◊ **2.14.** Dar una prueba alternativa del Lema 2.4 utilizando el Ejemplo 2.1.3 para el caso de banderas de un mismo color.

Ejemplo 2.3.1. Hay $\binom{r+5}{r}$ distribuciones distinguibles de tirar un dado r veces.

Ejemplo 2.3.2. Sea f una función analítica definida en un abierto de $\mathbb{R}^n$. Como las derivadas parciales conmutan (los operadores lineales $\frac{\partial}{\partial x_i}$ conmutan) resulta que el número de derivadas parciales de orden r distintas es $\binom{n-1+r}{r}$

2.3.2. Ejemplos de ocupación: un poco de física

Consideremos un sistema mecánico con r partículas (bolas) que no podemos distinguir. En mecánica estadística el espacio de fases se divide en n regiones (celdas) donde se alojan estas partículas. (A modo de ejemplo, las partículas pueden ser átomos, electrones, fotones; la subdivisión pueden ser niveles de energía.) La descripción del sistema mecánico está dada en la distribución aleatoria de las partículas en cada celda.

Veamos tres tipos de modelos de ocupación distintos conocidos para el sistema mecánico, a saber, las de Maxwell-Boltzmann, Bose-Einstein, y Fermi-Dirac. Todos los modelos son equiprobables, pero al cambiar los espacios muestrales las probabilidades de las distribuciones cambian según el modelo a considerar. Estos modelos se aplican a varios de los ejemplos considerados anteriormente.

- *Mawwell-Boltzmann*: En este modelo se asume que las bolas están numeradas y el espacio muestral es el de todas las posibles configuraciones, i.e., todas las posibles funciones de $f : r \to n$. Para este modelo existen n^r configuraciones posibles.

- *Bose-Einstein*: Las bolas no son distinguibles, y por lo tanto el espacio muestral consiste en el conjunto de todas las distribuciones distinguibles dadas por el Lema 2.4-a), i.e., los vectores $(r_1, \dots, r_n)$ solución de (2.3). Observar que este conjunto está identificado con el cociente del conjunto de funciones $f : r \to n$ cocientadas por la acción del grupo de permutaciones $\mathcal{S}_r$.

- *Fermi-Dirac*: Las bolas son indistinguibles y además no puede haber más de una bola por celda (principio de exclusión).

De lo visto anteriormente podemos concluir el siguiente resultado.

Lema 2.5. *La probabilidad de la distribución donde hay r_1 partículas en la primer celda, r_2 partículas en la segunda celda, ..., r_n partículas en la n-ésima celda, con $r = r_1 + \cdots + r_n$, es:*

- *Maxwell-Boltzmann:* $\dfrac{r!}{r_1! \cdots r_n!} n^{-r}$*;*

- *Bose-Einstein:* ${A_{r,n}}^{-1}$*;*

- *Fermi-Dirac:* $\binom{n}{r}^{-1}$, *siempre que* $r_i = 0$ *o* 1.

□

Ejemplo 2.3.3. Cuando $n = r = 2$ las distribuciones $(2,0)$, $(1,1)$ y $(0,2)$ son equiprobables (e igual a $1/3$) en el modelo de Bose-Einstein. Sin embargo, en el Modelo de Maxwell-Boltzmann la ditribución $(1,1)$ tiene probabilidad $1/2$. Además esta distribución tiene probabilidad 1 en el modelo de Fermi-Dirac por ser la única distribución posible.

Ejemplo 2.3.4. Dado un libro con n símbolos (letras), tiene r errores tipográficos. Un modelo probabilístico para estudiar estos errores es el de Fermi-Dirac dado que los errores ocurren en lugares diferentes.

El siguiente problema consiste en decidir si cierta observación se debe al azar, o se debe a que hay una causa. ¿Es aleatoria la forma en que las personas se sientan en un omnibus? Todos sabemos la respuesta a esta pregunta.

♢ **2.15** (Problema de aleatoriedad)**.** Supongamos que en un mostrador de un bar hay 16 sillas donde hay sentadas 5 personas. ¿Cuál es la probabilidad de que las personas se sienten separadas? (i.e., que siempre exista una silla libre entre medio de dos personas.)

2.4. Distribución hipergeométrica

Supongamos que tenemos una población de N elementos donde N_1 son de color rojo, y $N - N_1$ son de color negro. Si elegimos una muestra de tamaño n, ¿cuál es la probabilidad q_k de encontrar k rojos? Estamos asumiendo que el espacio muestral está formado por todas las subpoblaciones de longitud n, y por tanto el cardinal es $\binom{N}{n}$. Las formas distintas de elegir k rojos de N_1 es $\binom{N_1}{k}$, y la forma de elegir a los restantes elementos no rojos es $\binom{N-N_1}{n-k}$. Además cada elección distinta de rojos se combina con cualquier elección distintas de no rojos para formar una configuración distinta. De esta manera se concluye que

$$q_k = \frac{\binom{N_1}{k} \cdot \binom{N-N_1}{n-k}}{\binom{N}{n}}, \quad k \leqslant \min\{n, N_1\} \tag{2.4}$$

Este sistema de probabilidades se denomina *distribución hipergeométrica*. Desarrollando la ecuación (2.4) obtenemos una nueva expresión para q_k, a saber,

$$q_k = \frac{\binom{n}{k} \cdot \binom{N-n}{N_1-k}}{\binom{N}{N_1}}, \quad \forall k \geqslant 0, \tag{2.5}$$

donde se extiende la definición para todo $k \geqslant 0$, definiendo $q_k = 0$.

Esta distribución se utiliza mucho en estadística para inferir por medio de muestras ciertas poblaciones desconocidas. Veamos algunos ejemplos.

Ejemplo 2.4.1 (*Inspección de calidad*)**.** En el control de calidad de una fábrica se estudian muestras de tamaño n para identificar los items defectuosos. Supongamos que hay N items en total y N_1 son los defectuosos (cantidad desconocida). Tomando una muestra de tamaño n podemos identificar k items defectuosos. Se puede utilizar la fórmula (2.5) para inferir N_1. (Los métodos posibles para estudiar este problema son métodos clásicos de *estadística* que van más allá de los contenidos de este curso.)

Ejemplo 2.4.2 (*Peces en el lago*)**.** Supongamos que queremos estimar la población N de peces en un lago. ¿Cómo lo harían? Un truco que inventaron los estadísticos es el siguiente. "Tomar una muestra de tamaño n", i.e., pescamos n peces, para luego "pintarlos" y devolverlos al mar. Luego se toma otra muestra de tamaño n (representativa) y se identifican cuántos peces están pintados. Es claro que la proporción de peces pintados de nuestra segunda muestra nos brinda información sobre la cantidad posible de peces en el lago.

♢ **2.16.** Pensar en cómo utilizar la distribución hipergeométrica para estimar N en el ejemplo anterior. (Problema para discutir en clase.)

♢ **2.17.** Realizar el ejercicio 2 de la práctica en R.

2.5. Tiempo de espera

Consideremos el experimento de colocar aleatoriamente bolas en n celdas hasta una situación determinada. En esta sección estudiaremos dos situaciones concretas. Colocamos aleatoriamente bolas en n celdas hasta que la siguiente afirmación ocurra:

(i) se coloca una bola en una celda ocupada.

(ii) se coloca una bola en una celda particular (la primera por ejemplo).

El análisis de este experimento se aparta un poco del análisis combinatorio que venimos realizando, pero da un ejemplo distinto de espacio muestral para el cual aplicaremos lo visto de problemas de ocupación. (Recordar Ejemplo 1.4.9.)

Ejemplo 2.5.1. El problema de cumpleaños mencionado en el Ejemplo 2.2.3, puede estudiarse con el modelo (i). Esto es, elijo de a una persona al azar hasta que se repite el día. Si nos preguntamos cuál es la probabilidad que una persona cumpla el mismo día que yo, el modelo para analizarlo es el (ii).

Ejemplo 2.5.2. Supongamos que un hombre quiere abrir una puerta y tienen un manojo de n llaves del cual sólo una llave abre la puerta. El hombre prueba de a una las llaves hasta encontrar la llave que abre la puerta. ¿Cuál es la probabilidad de que el hombre acierte la llave en r-ésimo ensayo? Este problema puede modelarse con (ii).

Para estudiar este modelo de experimento representaremos las celdas por los números $1,\dots,n$, y por $(j_1, j_2, \dots)$ al resultado de que la primer bola se coloca en la celda j_1, la segunda bola la colocamos en j_2, ..., etc.

Motivado por el experimento de colocar r-bolas en n-celdas, asumimos que todas los resultados posibles en r-ensayos son equiprobables, i.e., tiene probabilidad n^{-r}. Por el momento esto es una definición posible (a menos de corroborar que la suma de las probabilidades es 1), pero además veremos que recupera resultados vistos anteriormente.

Caso (i)

Para el experimento (i) resulta que si $(j_1,\dots,j_r)$ es un resultado del experimento entonces $j_1,,\dots,j_{r-1}$ son todos diferentes tomadas de $\{1,\dots,n\}$, y j_r coincide con alguno de los anteriores. El conjunto de estos resultados posibles define el espacio muestral. (Observar que el número r de ensayos puede tomar los valores $\{2,\dots,n+1\}$.)

De esta manera, la probabilidad q_r del evento que "el proceso termina en r-ensayos" es

$$q_r = \frac{(n)_{r-1}(r-1)}{n^r} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)(r-1), \quad (r = 2,3,\dots,n+1),$$

donde definimos de manera natural $q_1 = 0$.

Además, si denotamos por p_r la probabilidad del evento que "el proceso continúa después de r ensayos", se tiene

$$p_r = \frac{(n)_r}{n^r} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right), \qquad (r = 2,3,\dots,n+1). \tag{2.6}$$

En particular, $0 = p_{n+1} = 1 - q_1 - \dots - q_{n+1}$, de donde resulta

$$q_1 + \dots + q_{n+1} = 1.$$

Observar que la fórmula (2.6) coincide con la ecuación (2.1) vista anteriormente para el caso de $n = 365$.

Caso (ii)

A diferencia del caso (i), el espacio muestral del modelo (ii) es infinito. En este caso los resultados posibles son tiras $(j_1, \dots, j_r)$ donde los j_i, $(i = 1, \dots, r-1)$, son diferentes a cierta celda c, y $j_r = c$. Para fijar ideas pensemos en $c = 1$. De esta manera, siguiendo las hipótesis de equiprobabilidad dadas más arriba, la probabilidad del evento que "el proceso termine en el r-ésimo ensayo" es

$$q_r^* = \frac{(n-1)^{r-1}}{n^r} = \left(1 - \frac{1}{n}\right)^{r-1} \frac{1}{n} \qquad (r = 1, 2, \dots, n+1),$$

Observar que realizando la suma geométrica se obtiene $q_1^* + q_2^* + \cdots = 1$. Además no es necesario incluir en el espacio muestral el caso que nunca se detenga el experimento (cf. Ejemplo 1.4.9). De incluirlo, necesitamos que la probabilidad de tal suceso sea 0.

Análogamente al caso anterior, la probabilidad del evento que "el proceso continúa después de r ensayos", es

$$p_r^* = \left(1 - \frac{1}{n}\right)^r \qquad (r = 1, 2, \dots, n+1),$$

$\diamond$ **2.18.** Tomando logaritmos, estimar r para el cual las probabilidades p_r y $p_r{*}$ se aproximen a $1/2$. (Observar que si r es la solución del problema anterior entonces, la probabilidad de que se detenga antes de r, es igual a la probabilidad que continúe después de r.) Observar que la solución anterior aumenta como raíz de n en el modelo (i), y lineal en n en el modelo (ii).

Ejercicios para hacer en R

Ejercicio: Muestreo aleatorio de Números

Consideremos el experimento que consiste en sacar 5 dígitos con reposición del conjunto $\{0, 1, ..., 9\}$. Por lo visto en el capítulo sabemos que:

$$\mathbb{P}(\text{"Los 5 dígitos son diferentes"}) = \frac{(10)_5}{10^5} \approx 0.3024.$$

En este ejercicio aprovecharemos este resultado para analizar estadísticamente la validez de la siguiente afirmación (sacada de Wikipedia):

> *"Todos los ensayos estadísticos realizados sobre la sucesión de los dígitos decimales de π han corroborado su carácter aleatorio".*

Si la afirmación anterior es cierta, debería de cumplirse que si tomamos 5 dígitos consecutivos en los decimales de π, la probabilidad de que sean todos diferentes debería de ser $\frac{(10)_5}{10^5} \approx 0.3024$.

a) Cargar los primeros 10 millones de dígitos decimales de π.

La siguiente página contine los primeros 10 millones de dígitos decimales de π: http://introcs.cs.princeton.edu/java/data/pi-10million.txt.

La siguiente línea carga el .txt que se encuentra en dicho enlace y lo transforma en una palabra de 10 millones de carácteres (no dígitos). Puede ser que demore un poco debido al tamaño del archivo que estamos queriendo cargar.

```
p <- readLines( 'http://introcs.cs.princeton.edu/java/data/pi-10million.txt', warn=F )
```

Ahora transformaremos la 'mega-palabra' p en un vector de 10 millones de dígitos (que es lo que queremos):

```
pii <- as.numeric( unlist( strsplit( p, split='' ) ) )
```

pii es entonces un vector con los primeros 10 millones (10^7) dígitos decimales de π:

```
length(pii)
```

```
## [1] 10000000
```

```
# Por ejemplo, los primeros 10 dígitos decimales de pi son:
pii[1:10]
```

```
##  [1] 1 4 1 5 9 2 6 5 3 5
```

b) Sortear dentro del vector *pii una* ventana de longitud 5. Observar si se repite algún dígito en dicha ventana.

```
# Sorteamos la ubicación del primer dígito de la venta (le llamaremos
# "posición de la ventana" a la posición del primer dígito)

posicion_ventana = sample (1:(10^7-4), 1)

ventana = pii[posicion_ventana: (posicion_ventana+4)]

#Para este caso:
ventana
```

```
## [1] 8 4 1 4 0
```

Ahora queremos determinar si en la ventana sorteada se repite algún dígito. Para eso podemos usar la función *unique()* ya que si v es un vector, *unique(v)* devuelve otro vector en el que se remueven las coordenadas de v que aparecen duplicadas:

```
length((unique(ventana)))
```

```
## [1] 4
```

Si *length((unique(ventana)))* no es 5 es porque en el vector *ventana* aparecen dígitos duplicados, por lo cual no son todos distintos.

Pruebe correr las líneas anteriores varias veces. ¿Con qué frecuencia los dígitos del vector *ventana* son todos distintos?

c) Sortear dentro del vector *pii* N ventanas de longitud 5 y calcular la frecuencia relativa con que los 5 dígitos de esas ventanas son todos distintos:

$$f_N = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\text{En la i-ésima ventana los 5 dígitos son diferentes }\}}$$

Si la afirmación es cierta, debería de cumplirse que $f_N \to \frac{(10)_5}{10^5} \approx 0.3024$ cuando $N \to \infty$.

```
N= 1000

#Sorteamos las posiciones del primer dígito de cada una de las ventanas:
posicion_ventana = sample (1:(10^7-4), N)

# Contaremos cuáles son las ventanas en las que los dígitos son distintos.
# Para eso, creamos un vector de 0's y luego pondremos un 1 en el lugar i si en la
# i-ésima ventana todos los dígitos son distintos:
todos_distintos = rep(0, N)

for (i in 1:N)
{
  ventana = pii[posicion_ventana[i]: (posicion_ventana[i]+4)]
```

```
  if (length(unique(ventana)) == 5) {todos_distintos[i] <- 1}
}

f_N = (1/N)* sum(todos_distintos)

f_N
```

Repetir el experimento anterior muchas veces, variando el tamaño de N. ¿Qué puede concluir de la afirmación del principio?

Ejercicio 2: Problema del cajón de medias

Suponga que su madre fue al barrio de los judíos y le hizo un surtido de medias rojas y negras. Su relajo (*el suyo, no de su madre....ya está grandecit@ como para arreglar su ropero*) es tan grande que no tiene idea de cuántos pares de medias tiene y tampoco las tiene agrupadas de a pares. De hecho, posiblemente haya perdido alguna media así que ni siquiera sabe si tiene números pares de medias rojas y negras.

Así que todas las mañanas elige al azar 2 medias (a ese nivel de vagancia hemos llegado). Se ha dado cuenta -de la empiria- que la probabilidad de que las dos medias elegidas al azar sean rojas es $\frac{1}{2}$.

a) ¿Cuál es el menor número de medias que tiene que haber en el cajón?

Elabore un algoritmo que calcule ese número.

b) Vuelva a responder la pregunta anterior sabiendo ahora que hay un número par de medias negras (es decir que no ha perdido ninguna media negra).

Acomode el algoritmo anterior para calcular este nuevo número.

Capítulo 3

Combinación de eventos

Unión de eventos, principio de inclusión-exclusión, aplicaciones a problemas de ocupación, coincidencias.

[Duración: 1 Semana]

3.1. Unión de eventos

En el Capitulo 1 vimos que si $(\Omega, \mathbb{P})$ es un espacio de probabilidad entonces

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \tag{3.1}$$

El objetivo es hallar la probabilidad de un evento de la forma $A_1 \cup \cdots \cup A_n$. Para tal fin necesitamos un poco de notación.

Definimos

$$p_i := \mathbb{P}(A_i), \quad p_{i,j} := \mathbb{P}(A_i \cap A_j), \quad p_{i,j,k} := \mathbb{P}(A_i \cap A_j \cap A_k) \quad \ldots$$

Observar que el orden de los subíndices es irrelevante, por lo cual podemos asumir que están ordenados de menor a mayor. Además, para tener "unicidad" en la notación, excluimos subíndices iguales.

Denotamos por S_r la suma de los p's con r subíndices. Esto es,

$$S_1 := \sum_{i=1}^{n} p_i, \quad S_2 := \sum_{1 \leqslant i < j \leqslant n} p_{i,j}, \quad S_3 := \sum_{1 \leqslant i < j < k < l \leqslant n} p_{i,j,k} \ldots \tag{3.2}$$

Observar que S_r es la suma de $\binom{n}{r}$ sumandos, donde en particular S_n tiene un único sumando, a saber, $p_{1,2,\ldots,n}$.

Cuando $n = 2$, la expresión (3.1) resulta ser en la nueva notación igual a

$$\mathbb{P}(A \cup B) = S_1 - S_2$$

Teorema 3.1 (Principio de inclusión-exclusión)**.**

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{r=1}^{n} (-1)^{r-1} S_r = S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n-1} S_n.$$

Una forma de justificar esta fórmula es la siguiente:
La suma S_1 sumará todos los ω's en la unión, pero contará doble a aquellos ω's que estén sólamente en dos de los A_i's. Por lo tanto si sutraemos S_2 contaremos bien aquellos ω's que estaban en un sólo A_i, y aquellos que estaban en sólamente dos de los A_i. Sin embargo los ω's que están en 3 de los A_i's se cancelaron y por lo tanto hay que volver a agregarlos.

♢ **3.1.** Dar una prueba del Lema 3.1 por inducción.

♢ **3.2.** (*Matching problem I*) Supongamos que un cartero distraído tiene que poner n cartas en n casilleros pero los pone de forma aleatoria. ¿Cuál es la probabilidad de que al menos una persona reciba la carta correctamente ? Si A_k es el evento que la persona correspondiente al casillero k recibió la carta correcta, entonces lo que queremos calcular es $\mathbb{P}(A_1 \cup \cdots \cup A_n)$.

i) Probar que la probabilidad anterior, cuando n crece, tiende a $1 - e^{-1} = 0{,}632\ldots$. (Hacer una tabla de valores para conocer qué tan rápido se acerca al límite.)

ii) De cuantas maneras podemos colocar las n cartas sin que nadie reciba la carta correcta.

iii) ¿Cuál es la probabilidad $p_{m,n}$ de que exactamente m cartas estén en los casilleros correctos? Probar que
$$\lim_{n \to +\infty} p_{m,n} = \frac{1}{e\,m!}.$$

Es interesante destacar del resultado anterior que la probabilidad de que m personas reciban la carta correcta es casi independiente de n.

Otra forma de atacar el problema anterior es el siguiente. Si B_i es el evento que la i-ésima persona es la primera en recibir la carta correcta, entonces se tiene

$$A_1 \cup \cdots \cup A_n = B_1 \cup \cdots \cup B_n.$$

◇ **3.3.** Hallar $\mathbb{P}(B_1 \cup \cdot \cup B_n)$ usando que los B_i son eventos incompatibles, y verificar que coincide con el resultado de i).

◇ **3.4.** ¿Cuál es el número de funciones sobreyectivas $f : n \to m$ donde $m \leqslant n$?
(Sugerencia: considerar los eventos B_i de que el valor i no es alcanzado.)

3.2. Aplicación al problema de ocupación

Retomemos el problema de colocar r bolas en n celdas de forma aleatoria. (Recordar que estamos asumiendo equiprobabilidad.)
¿Cuál es la probabilidad $p_m(r,n)$ de encontrar exactamente m celdas vacías? Observar que en ejercicio 3.4 probamos que

$$p_0(r,n) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^r$$

◇ **3.5.** Probar que

$$p_m(r,n) = \binom{n}{m} \sum_{i=0}^{n} (-1)^i \binom{n-m}{i} \left(1 - \frac{m+i}{n}\right)^r \tag{3.3}$$

Cuando $n = 10$, la probabilidad $p_m(r,n)$ hallada en (3.3) coincide con la probabilidad de exactamente m de los números $0, 1, \ldots, 9$ no aparezcan en una tira de r dígitos.

◇ **3.6.** Hacer una tabla de valores de $p_m(r,10)$ y de $p_m(r,20)$ para valores de m entre $0,1,\ldots,9$.

La fórmula (3.3) tiene interés cuando n y r crecen. Por ejemplo, nos gustaría saber cuál es la probabilidad de que en 1000 personas no haya ninguna que cumpla el mismo día. La complejidad del cálculo de esta formula se torna difícil cuando n o r son grandes. Sin embargo veremos que existe una aproximación muy precisa de (3.3) en el caso de que n y r crezcan de manera controlada.

(Observar que es necesario que n y r estén relacionados para poder tener una probabilidad no trivial. Por ejemplo, si r/n es muy grande, entonces es “poco probable” (i.e. ocurre con probabilidad casi cero) que existen celdas sin bolas. Por el otro lado si r/n se mantiene menor o igual a una constante menor que 1 es claro que la probabilidad de encontrar m sitios sin bolas tiende a 1.)

Teorema 3.2. *Supongamos que r y n tienden a infinito de manera que $\lambda = ne^{-\frac{r}{n}}$ satisface $0 < a \leqslant \lambda \leqslant b$ para ciertas constantes a y b. Entonces*

$$\lim_{n\to\infty} p_m(r,n) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Demostración. Fijemos $i \in \{0,1,\dots,n\}$. Sea $S_i = \binom{n}{i}\left(1-\frac{i}{n}\right)^r$, de donde resulta $p_0(r,n) = \sum_{i=0}^{n}(-1)^i S_i$.

Desarrollando el factorial se tiene

$$n^i\left(1-\frac{i}{n}\right)^r \leqslant i!\,S_i \leqslant n^i\left(1-\frac{i}{n}\right)^r. \tag{3.4}$$

Usando el desarrollo de la función logaritmo se prueba

$$t \leqslant -\log(1-t) \leqslant \frac{t}{1-t}, \qquad t \in (0,1).$$

Entonces, tomando logaritmos en (3.4) se tiene

$$i\log n - \frac{i}{1-\frac{i}{n}}(r+i) \leqslant \log(i!\,S_i) \leqslant i\log n - r\frac{i}{n},$$

de donde resulta

$$\left(n e^{-\frac{r+i}{n-i}}\right)^i \leqslant i!\,S_i \leqslant \left(n e^{-\frac{r}{n}}\right)^i. \tag{3.5}$$

En lo que sigue suponemos i fijo. Observar que bajo la relación de dependencia entre r y n cuando crecen, se tiene que el cociente de los extremos de esta ecuación tiende a 1, para cada i fijo. Por lo tanto, definiendo $\lambda = ne^{-\frac{r}{n}}$, y tomando límite en n se tiene

$$S_i = \lim_n \frac{\lambda^i}{i!}.$$

para i fijo.

Además, como

$$e^{-\lambda} - p_0(r,n) = \sum_{i=0}^{n}(-1)^i\left(\frac{\lambda^i}{i!} - S_i\right) + \sum_{i=n+1}^{\infty}(-1)^i\frac{\lambda^i}{i!},$$

resulta que tomando límite en n se tiene

$$\lim_n p_0(r,n) = e^{-\lambda}.$$

Del Ejercicio 3.5 sabemos que

$$p_m(r,n) = \binom{n}{m}\left(1-\frac{m}{n}\right)^r p_0(r,n-m)$$

por lo tanto

$$\lim_n p_m(r,n) = \lim_n S_m\, p_0(r,n-m) = \frac{\lambda^m}{m!}e^{-\lambda},$$

concluyendo el teorema. □

Hemos concluido que si n es grande, y r se comporta como en el enunciado del Teorema 3.2 se tiene que $p_m(r,n)$ se puede aproximar por la función

$$p(m;\lambda) = \frac{\lambda^m}{m!}e^{-\lambda},$$

que se denomina *distribución de Poisson* , la cual estudiaremos más adelante.

A menos de estudiar el error, este resultado nos puede ayudar a estudiar el problema de ocupación para n y r grandes.

♢ **3.7.** Dar una aproximación de la probabilidad de que en un grupo de 1900 personas existan m personas que cumplen el mismo día, para m de 0 a 10.

3.3. Realización simultanea de eventos

♢ **3.8.** Dados n eventos $A_1,\ldots,A_n$, probar que la probabilidad $p_{[m]}$ de exactamente m ($m \leqslant n$) eventos ocurriendo de forma simultanea es

$$S_m - \binom{m+1}{m}S_{m+1} + \binom{m+2}{m}S_{m+2} - \cdots + (-1)^{n-m}\binom{n}{m}S_n$$

donde S_i está dado en (3.2).

♢ **3.9.** En el ejercicio 3.2 ¿cuál es la probabilidad de que 5 personas de 10 reciban la carta correcta?

♢ **3.10.** * Se elije un número al azar entre 1 y 1000. ¿Cuál es la probabilidad de que sea una potencia de otro número?

Capítulo 4

Probabilidad condicional e independencia

Probabilidad condicional, sucesos independientes y sucesos no independientes

[Duración: 2 Semanas?]

4.1. Probabilidad condicional

Muchas veces la probabilidad de que un evento ocurra cambia si sabemos que otro ha ocurrido. Un caso extremo es cuando los eventos son incompatibles, pues la ocurrencia de uno de ellos hace imposible la ocurrencia del otro. La probabilidad de un evento bajo la condición de que otro ha ocurrido se llama probabilidad condicional. Esto es lo que estudiaremos en este capítulo.

Motivación

Suponga que el señor de la ferretería quiere *enchufarle* a la gente las lámparas de alto consumo que le quedaron de *clavo* ya que ahora la mayoría de la gente prefiere comprar lámparas de bajo consumo. Para lograrlo, decidió *mechar* dentro de las tres cajas de lámparas de bajo consumo que están en las góndolas (suponga que el señor vende tres marcas de lámparas) algunas lámparas de alto consumo. Además, aclara en la boleta que las compras no tienen cambio (*está bien de vivo*).

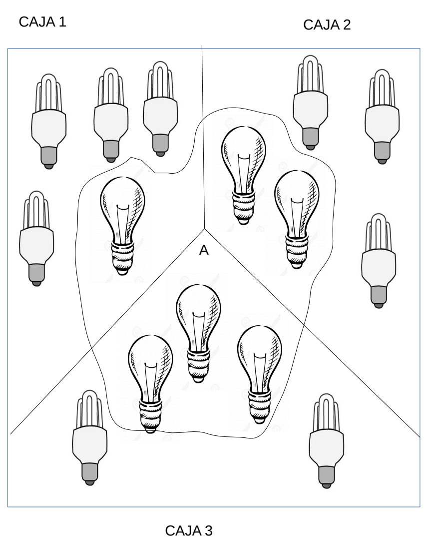
Puesto que el comprador confía en que todas las lámparas son de bajo consumo, solo se fija en la caja de afuera para elegir la marca de la lámpara que quiere comprar. Las tres marcas son de la misma calidad, por lo cual elige al azar la marca de la lámpara que quiere comprar (según como se levante de ánimo).

El señor de la ferretería distribuyó las lámparas de alto consumo dentro de las cajas de la siguiente manera:

- En la *caja 1* dejó 4 lámparas de bajo consumo y metió una de alto consumo;
- en la *caja 2* dejó 3 lámparas de bajo consumo y metió dos de alto consumo;
- y en la *caja 3* dejó solo 2 lámparas de bajo consumo y metió tres de alto consumo.

¿Cuál es la probabilidad de que el primer cliente que llega a la ferretería se lleve (por no revisar antes de pagar) una lámpara de alto consumo?

Para modelar este problema, y calcular dicha probabilidad, podemos pensar en el espacio muestral formado por todas las lámparas que podrían provenir de cada caja, es decir:



Luego, como asumimos que todas las cajas y todas las lámparas tienen la misma probabilidad de ser elegida por el cliente, si le llamamos A al suceso *"el cliente se lleva una lámpara de alto consumo"*,

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{6}{15}$$

> Si ahora el señor de la ferretería observa que el cliente sacó una lámpara de la caja 2, ¿cuál es la probabilidad de que se lleve una lámpara de alto consumo?

Observar que, con esta nueva información, el espacio muestral no es el Ω formado por todas las lámparas, sino que el nuevo espacio muestral pasa a ser $\Omega' = B =$ "elige la caja 2". Por lo tanto, como cambia el espacio muestral, entonces también cambiará la función de probabilidad. Llamémosle $\mathbb{P}_B$ a esta nueva función de probabilidad. Luego,

$$\mathbb{P}_B(\text{“elige una lámpara de alto consumo”}) = \mathbb{P}_B(A) = \frac{2}{5} = \frac{\#(A\cap B)}{\#B} = \frac{\frac{\#(A\cap B)}{\#\Omega}}{\frac{\#B}{\#\Omega}} = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

$\diamondsuit$ **4.1.** Sabiendo que este cliente se llevó una lámpara de alto consumo (pero sin saber de qué caja la sacó), ¿cuál es la probabilidad de que el próximo cliente se lleve una lámpara de alto consumo?, ¿cuál es la probabilidad de que el cliente anterior se haya llevado la única lámpara de alto consumo de la caja 1?.

En general, para referirnos a la nueva función de probabilidad $\mathbb{P}_B(.)$ usaremos la notación (habitual) $\mathbb{P}(.|B)$.

4.1.1. Definición y algunas propiedades

El ejemplo anterior es completamente general, en el sentido de que no es necesario que el espacio de probablidad $(\Omega,\mathbb{P})$ sea equiprobable (más aún, no es necesario que sea discreto).

> **Definición** (*Probablidad Condicional*)**.** Sea $(\Omega,\mathbb{P})$ espacio de probabilidad discreto. Supongamos que ya sabemos que ocurre el suceso $B\subset\Omega$, con $\mathbb{P}(B)>0$. Entonces el nuevo espacio muestral pasa a ser B y sobre los subconjuntos de B (o sobre los subconjuntos de Ω, por cómo la definiremos) definimos una nueva función de probabilidad $\mathbb{P}(\,.\,|B)$, a la que llamaremos *probabilidad condicional dado que ocurre el suceso B*, de la siguiente manera:
>
> $$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}\ \forall A\subset\Omega$$

Observar que la definición ya deja expresada de manera explícita la relación entre $\mathbb{P}(\,.\,|B)$ y $\mathbb{P}(\,.\,)$.

$\diamondsuit$ **4.2.** ¿Tendrá sentido preguntarse qué ocurre con la definición anterior en el caso de que $\mathbb{P}(B)=0$?

♢ **4.3.** Probar que la función $\mathbb{P}(\,.\,|B)$ definida antes es efectivamente una función de probabilidad sobre Ω, es decir que cumple los *axiomas de Kolmogorov*:

- $\mathbb{P}(\varnothing|B) = 0$ y $\mathbb{P}(\Omega|B) = 1$
- $\mathbb{P}(A|B) \in [0,1]\ \forall A \subset \Omega$
- Si $A_1, A_2, \ldots, A_n, \ldots$ es una sucesión de sucesos incompatibles (es decir que $A_i \cap A_j = \varnothing$ si $i \neq j$), entonces:
$$\mathbb{P}\left(\biguplus_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$
donde $A \uplus B$ significa unión disjunta ($A \cap B = \varnothing$).

Luego, si es verdad que $\mathbb{P}(\,.\,|B)$ es una probabilidad, se siguen cumpliendo las propiedades de siempre:

- $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$
- $\mathbb{P}(A \cup C|B) = \mathbb{P}(A|B) + \mathbb{P}(C|B) - \mathbb{P}(A \cap C|B)$
- etcétera.

♢ **4.4.**

- Buscar algún ejemplo para mostrar que, en general, $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$.
- Observar que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$.
- Considere tres sucesos A, B y C. Probar que:
$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(C|A \cap B) \times \mathbb{P}(B|A) \times \mathbb{P}(A)$$
- Probar que:
$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

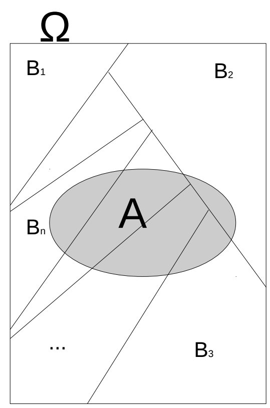
La ecuación anterior, conocida como *fórmula de Bayes (I)*, permite calcular probabilidades del estilo *"pasado" dado "presente"* en función de las probabilidades de *"presente" dado "pasado"* (que son las probabilidades que en general conocemos). Volver a hacer el ejercicio 4.1 utilizando la ecuación anterior.

4.1.2. Fórmula de probabilidad total

Cuando uno quiere calcular la probabilidad de un evento $A \subset \Omega$, muchas veces resulta útil tratar de usar probabilidades condicionales como un paso intermedio (por cómo viene dada la información del espacio muestral, por ejemplo).

Supongamos que tenemos el espacio muestral *partido* en los sucesos disjuntos $B_1, B_2, ..., B_n$, es decir que:

$$\biguplus_{i=1}^{n} B_i = \Omega$$



Podemos escribir entonces al suceso A como

$$A = A \cap \Omega = A \cap \left(\biguplus_{i=1}^{n} B_i \right) = \biguplus_{i=1}^{n} (A \cap B_i),$$

y

$$\mathbb{P}(A) = \mathbb{P}\left(\biguplus_{i=1}^{n} (A \cap B_i) \right) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

A esta fórmula se le llama *fórmula de probabilidad total*.

◊ **4.5.** Probar la *regla de Bayes (II)*: si consideramos uno de los sucesos $\{B_i\}_{i=1,...,n}$, entonces:

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Veamos algunos ejemplos de cómo usar la fórmula de probabilidad total.

Ejemplo 4.1.1. El dilema del prisionero Tres prisioneros A, B y C solicitan libertad condicional. El jurado decide que solo liberará a dos de ellos. Un guardia amigo de A sabe quiénes son pero a A le resulta *anti-ético* (se ve que efectivamente está rehabilitado) preguntarle directamente al guardia si él será liberado. Entonces formula el siguiente plan:

> *"Voy a preguntarle al guardia si B es uno de los que será liberado"*

Pero luego reflexiona:

> *"Ahora tengo probabilidad $\frac{2}{3}$ de ser liberado, pero si le pregunto y me dice que B será liberado, entonces tendré probabilidad $\frac{1}{2}$ de ser liberado...Creo que me voy a enfermar de una gastritis al santo botón, así que mejor no le pregunto nada..."*

♢ **4.6.** ¿Qué está mal en su formulación?

Ejemplo 4.1.2. La ruina del jugador Considere un juego que consiste en tirar la moneda de manera sucesiva. En cada paso, el apostador gana \$1 si acierta a la salida de la moneda y pierde \$1 si no. Suponga que un jugador va decidido a ganar m pesos o *"morir en el intento"*, es decir, jugar hasta ganar \$$m$ o hasta perder toda la plata que haya llevado (quede arruinado).

> Si llega con \$$x$, ¿cuál es la probabilidad de que se vaya arruinado?

Sea $p(x)$ la probabilidad de que el jugador quede arruinado (se vuelva para su casa con las manos vacías) con un capital inicial de \$$x$.

♢ **4.7.** Observar que la probabilidad de quedar arruinado pasa a ser $p(x+1)$ si gana en la primer jugada y $p(x-1)$ si pierde en la primer jugada.

Es decir que, si le llamamos $A = A(x,m)$ al evento *"el pobre tipo se va con las manos vacías"*, B_1 al evento *"gana en la primer partida"* y B_2 a *"pierde en la primer partida"*, entonces:

$$\mathbb{P}(A|B_1) = p(x+1), \qquad \mathbb{P}(A|B_2) = p(x-1)$$

♢ **4.8.** Utilizar la fórmula de probabilidad total para probar que la función $p(x)$ tiene que verificar que:

$$p(x) = \frac{1}{2}[p(x+1)+p(x-1)] \text{ si } 1 \leqslant x \leqslant m-1$$

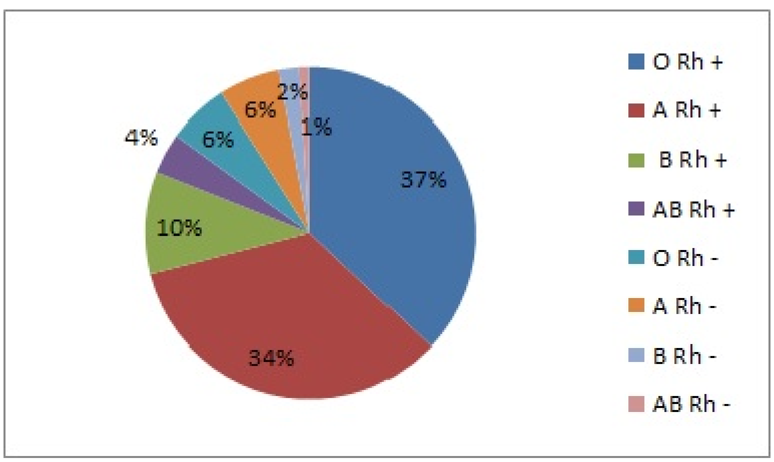
Luego, existen C_1 y C_2 tales que $p(x) = C_1 + C_2 x$. Como además se tiene que verificar que $p(0) = 1$ (*ya que si llega sin plata, se tiene que volver con las manos vacías porque no lo dejan jugar*) y $p(m) = 0$ (*ya que si se da cuenta de que llegó con $m, se vuelve feliz a su casa porque tiene todo lo que quería -no es ambicioso- y no precisa jugar; i.e. es imposible perder*), entonces:

$$p(x) = 1 - \frac{x}{m}.$$

¿Le parece razonable que dé esto?

Ejemplo 4.1.3. Resolviendo crímenes Se ha cometido un delito. El asesino es con seguridad una de las dos personas X e Y (*no podemos nombrarlas porque no tienen antecedentes*). Ambas personas están prófugas de la justicia, y luego de una investigación inicial, ambos fugitivos son igualmente probables de ser el asesino. Al avanzar la investigación se revela que el asesino del crimen tiene sangre tipo "O Rh -". Investigación suplementaria revela que la persona X tiene sangre "O Rh -", pero no ofrece información alguna sobre el grupo sanguíneo de la persona Y.

El siguiente cuadro presenta el porcentaje por grupo sanguíneo en la población (el 6% de la población tiene sangre del tipo "O Rh -").



◊ **4.9.** ¿Cuál es la probabilidad de que la persona X sea el asesino?.

◊ **4.10.** Considere N urnas (numeradas del 1 al N) tales que cada una de ellas contine b bolillas blancas y n bolillas negras. Se elige al azar una bolilla de la urna 1 y se la coloca en la urna 2, luego se elige una bolilla de la urna 2 y se la coloca en la bolilla 3 y así sucesivamente hasta que finalmente se saca una bolilla de la urna N.

¿Cuál es la probabilidad de que esta última bolilla sea blanca?

$\diamondsuit$ **4.11.** Se tienen 10 urnas de las cuales 9 contienen 2 bolillas blancas y 2 negras, mientras que la restante contiene 5 bolillas blancas y una negra. Se elige al azar una urna y de ella saca una bolilla. Si se sabe que salió una bolilla blanca, ¿cuán probable es que haya salido de esta última urna?

4.2. Independencia de sucesos

Cuando decimos que *dos experimentos son independientes* lo que estamos queriendo decir, groseramente, es que *el resultado de un experimento no debería de afectar el resultado del otro experimento*.

Por ejemplo, si digo que voy a tirar un dado y una moneda, entonces parece razonable suponer que lo que salga en el dado no debería de afectar a la salida de *Cara* o *Número* en la moneda. Por lo cual, la probabilidad de salir *Cara* o *Número* en la moneda debería seguir siendo $\frac{1}{2}$. Es decir que:

$$\mathbb{P}(\text{ “sale C” } | \text{ “salió el 5” }) = \frac{1}{2} = \mathbb{P}(\text{“sale C”})$$

4.2.1. Definición y ejemplos

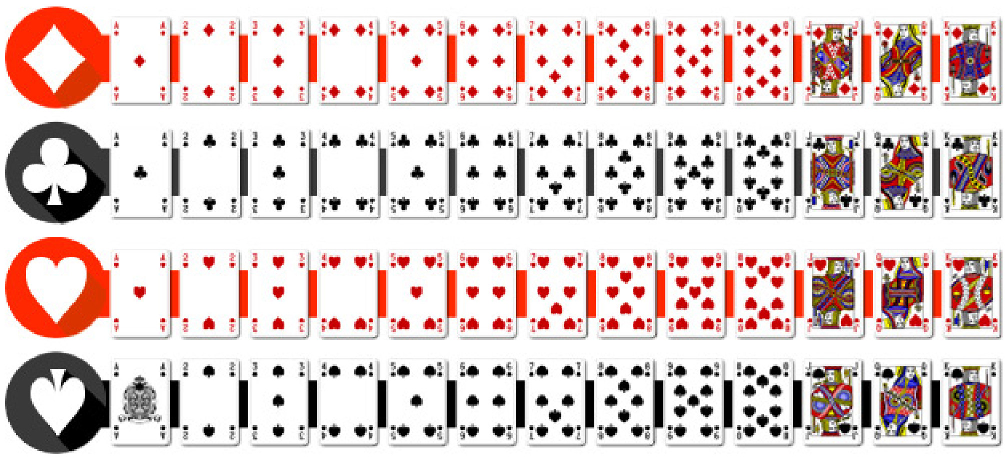
Definición. (*Sucesos independientes*) En general, diremos que dos sucesos A y B en un espacio de probabilidad $(\Omega, \mathbb{P})$ son *independientes* si:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \text{ y } \mathbb{P}(B|A) = \mathbb{P}(B)$$

$\diamondsuit$ **4.12.** Probar que esta definición es equivalente a la siguiente:

A y B son independientes si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$

Ejemplo 4.2.1. Mazo de cartas Considere el experimento que consiste en sacar una carta de un mazo de 52 cartas (*baraja francesa*) y los eventos A = *"sale espada (o pica)"* (♠) y B =*"sale reina"*.



Luego, como $\mathbb{P}(A) = \frac{13}{52} = \frac{1}{4}$ (pues hay 13 espadas por mazo), $\mathbb{P}(B) = \frac{4}{52} = \frac{1}{13}$ (pues hay una reina por mazo, por lo cual hay 4 reinas) y $\mathbb{P}(A \cap B) = \frac{1}{52}$ (pues solo hay una reina de espada), se tiene que:

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

Es decir que los sucesos *"sale espada"* y *"sale reina"* son independientes.

Ejemplo 4.2.2. Dados Considere el experimento que consiste en tirar dos dados:

y los sucesos: A = *"en el primer dado sale un número impar"*, B = *"en el segundo dado sale un número impar"* y C = *"la suma de las salidas de los dos dados es impar"*.

♢ **4.13.**

- Probar que $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$
- Probar que $\mathbb{P}(C|A) = \frac{1}{2}$ y $\mathbb{P}(C|B) = \frac{1}{2}$. Concluir que A y C son independientes y que B y C son independientes.
- ¿Son A y B independientes?

- ¿Son A, B y C independientes? (es decir: ¿se cumple que $\mathbb{P}(A\cap B\cap C) = \mathbb{P}(A)\times\mathbb{P}(B)\times\mathbb{P}(C)$?)

Definición. (*Colección de sucesos independientes*) Diremos que una colección finita de sucesos $A_1, A_2, ..., A_n$ son *independientes* si:

$$\mathbb{P}(A_i\cap A_j) = \mathbb{P}(A_i)\times\mathbb{P}(A_j) \quad \forall\, i\neq j;$$

$$\mathbb{P}(A_i\cap A_j\cap A_k) = \mathbb{P}(A_i)\times\mathbb{P}(A_j)\times\mathbb{P}(A_k) \quad \forall\, i\neq j\neq k;$$

$$\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\ldots\cap A_{i_k}) = \mathbb{P}(A_{i_1})\times\mathbb{P}(A_{i_2})\times\ldots\times\mathbb{P}(A_{i_k}) \quad \forall\, i_j\neq i_l \quad (i_1,i_2,\ldots,i_k\in\{1,2,\ldots,n\}; k\leqslant n).$$

Observar que en el ejemplo anterior los sucesos A, B y C son *independientes dos a dos* pero *no son independientes*.

♢ **4.14.** Probar que si A y B son sucesos independientes, entonces A^c y B^c también son independientes.

♢ **4.15.** Probar que los eventos A y B son independientes si $\mathbb{P}(B|A) = \mathbb{P}(B|A^c)$.

♢ **4.16.** Investigar si la propiedad de *“ser independientes”* cumple la propiedad transitiva, es decir:

Si A y B son independientes y B y C son independientes, ¿son A y C independientes?

♢ **4.17.** Para alentar la promisoria carrera de tennis de Pablito, su padre le ofrece un premio si gana al menos dos sets seguidos en una serie de tres sets a ser jugados con su padre y con el campeón del club alternadamente. Entonces la serie puede ser *Padre - Campeón - Padre* ó *Campeón - Padre - Campeón*. Pablito puede elegir cuál serie jugar. Sabiendo que el campeón del club es mejor jugador que su padre, ¿cuál serie le conviene elegir?

4.2.2. Lema de Borel-Cantelli

Límite superior e inferior de sucesiones de sucesos

Definición. *(Límite superior)* Considere una sucesión de sucesos A_1, A_2,...,A_n,... en un espacio de probabilidad $(\Omega,\mathbb{P})$. Se define el *límite superior* de $\{A_n\}_n$ como:

$$A^\infty = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$$

Se dice que el suceso A^∞ es la ocurrencia de un número infinito de los A_n's. Veamos por qué:

Si $\omega \in A^\infty$, entonces $\omega \in \bigcup_{k=n}^{\infty} A_k$ para todo n.

En particular, $\omega \in \bigcup_{k=1}^{\infty} A_k$, es decir que $\omega \in A_{k_1}$ para algún $k_1 \geqslant 1$.

Pero $\omega \in \bigcup_{k=k_1+1}^{\infty} A_k$, luego $\omega \in A_{k_2}$ para algún $k_2 \geqslant k_1$.

Analogamente, $\omega \in \bigcup_{k=k_2+1}^{\infty} A_k$, luego $\omega \in A_{k_3}$ para algún $k_3 \geqslant k_2 \geqslant k_1$

...

De esta manera se obtiene una sucesión creciente de subíndices $k_1 \leqslant k_2 \leqslant \ldots \leqslant k_n \leqslant \ldots$ tales que $\omega \in A_{k_i}$ para todo i. Entonces ω pertenece a infinitos A_n's.

Definición. *(Límite inferior)* Se define el *límite inferior* de $\{A_n\}_n$ como:

$$A_\infty = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

♢ **4.18.** Probar que A_∞ está formado por el conjunto de los $\omega \in \Omega$ que están en *todos* los A_n's salvo una cantidad finita de ellos. Tratar de visualizar entonces cuál es la diferencia entre A^∞ y A_∞.

♢ **4.19.** *(Sucesiones crecientes y decrecientes de sucesos)*

- Probar que si $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$, entonces:

$$A^\infty = A_\infty = \bigcup_{n=1}^{\infty} A_n.$$

- Probar que si $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$, entonces:

$$A^\infty = A_\infty = \bigcap_{n=1}^{\infty} A_n.$$

Sugerencia: Observe que en los dos casos hay una igualdad que es más fácil de probar y para la otra utilice lo que observamos antes acerca de los sucesos A^∞ y A_∞.

◊ **4.20.** *(Propiedad de "continuidad" de la probabilidad)*

- Probar que si $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$, entonces:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Es decir que $\mathbb{P}(A^{\infty}) = \mathbb{P}(A_{\infty}) = \text{lím}_{n\to\infty}\mathbb{P}(A_n)$. *Sugerencia:* defina $A_0 = \varnothing$ y considere la sucesión de sucesos $\{B_n\}$ tal que $B_n = A_n \setminus A_{n-1}$. Observe que:

$$\bigcup_{n=1}^{\infty} A_n = \biguplus_{n=1}^{\infty} B_n$$

- Probar que si $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$, entonces:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Sugerencia: considere la sucesión de sucesos $\{B_n\}$ tal que $B_n = A_n^c$.

Lema de Borel-Cantelli

Teorema 4.1. *(Lema de Borel - Cantelli)*
Sean A_1, A_2,...,A_n,... sucesos en el espacio de probabilidad $(\Omega, \mathbb{P})$.

a) Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, entonces $\mathbb{P}(A^{\infty}) = 0$.

b) Si los sucesos $\{A_n\}_n$ son independientes y $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, entonces $\mathbb{P}(A^{\infty}) = 1$.

Definición. Diremos que los sucesos $\{A_n\}_n$ son independientes si para todo $N \in \mathbb{N}$ y para toda colección $1 \leqslant j_1 < j_2 < \ldots < j_N$ los sucesos $\{A_{j_1}, A_{j_2}, \ldots, A_{j_N}\}$ son independientes.

Demostración.
a) Como $A^{\infty} \subset \bigcup_{k=n}^{\infty} A_k$ para todo n,

$$\mathbb{P}(A^{\infty}) \leqslant \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) = \sum_{k=n}^{\infty} \mathbb{P}(A_k) \quad \forall\, n.$$

Entonces:

$$\mathbb{P}(A^\infty) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0,$$

por ser la cola de una serie convergente.
b) Recordar que $A^\infty = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Consideremos entonces los sucesos:

$$B_n = \bigcup_{k=n}^{\infty} A_k\,.$$

Observar que $B_1 \supseteq B_2 \supseteq \ldots \supseteq B_n \supseteq \ldots$ y que $A^\infty = \bigcap_{n=1}^{\infty} B_n$. Luego, utilizando la propiedad de *continuidad de la probabilidad*, tenemos que:

$$\mathbb{P}(A^\infty) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = 1 - \lim_{n\to\infty} \mathbb{P}(B_n^c)\,.$$

Calculemos entonces este último límite:

$$\begin{aligned}
\mathbb{P}(B_n^c) &= \mathbb{P}\left[\left(\bigcup_{k=n}^{\infty} A_k\right)^c\right] = \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\
&= \prod_{k=n}^{\infty} \mathbb{P}(A_k^c) = \prod_{k=n}^{\infty} (1 - \mathbb{P}(A_k)) \\
&\leqslant \prod_{k=n}^{\infty} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)} \\
&\to 0
\end{aligned}$$

cuando $n \to \infty$ ya que $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$ (por lo que $\sum_{k=n}^{\infty} \mathbb{P}(A_k) = \infty$ para todo n).

Observar que:

- La segunda igualdad se debe a la *Ley de Morgan*: $(A \cup B)^c = A^c \cap B^c$
- La tercera se debe a que los sucesos $\{A_k^c\}$ son independientes ya que los $\{A_k\}$ lo son (recordar ejercicio).
- La desigualdad sale de que $1 - x \leqslant e^{-x}$ para todo $x > 0$.

Finalmente, acabamos de probar que $\mathbb{P}(A^\infty) = 1$, es decir que con probabilidad uno ocurren una cantidad infinita de los A_n's. □

Aplicaciones del lema de Borel-Cantelli

Ejemplo 4.2.3. Sean $X_1, X_2,...,X_n,...$ una sucesión de variables aleatorias (*v.a*) en $(\Omega,\mathbb{P})$ tales que:

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) = 0\}) = \frac{1}{n^2} \quad \forall n$$

Si elijo al azar un $\omega \in \Omega$ y evalúo cada una de las *v.a.* en en ω:

$$X_1(\omega), X_2(\omega), ..., X_n(\omega), ...$$

¿cuán probable es *agarrar* un ω del siguiente suceso:

$$B = \{\omega \in \Omega : \text{ existen infinitos } j_1 < j_2 < ... < j_n < ... \text{ tales que } X_{n_j}(\omega) = 0\}?$$

Pero esto significa que queremos ver cuán probable es que el suceso $A_n = \{\omega \in \Omega : X_n(\omega) = 0\}$ ocurra para infinitas n's. Es decir que ocurra:

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Pero como:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

el lema de Borel-Cantelli dice que $\mathbb{P}(B) = 0$.

Ejemplo 4.2.4. Se tira sucesivamente una moneda (*"hasta el infinito"*). Podemos modelar los resultados de este experimento con:

$$\Omega = \{\omega = (\omega_1, \omega_2, ..., \omega_n, ...) : \omega_i \in \{C,N\}\}$$

Observar que $\#\Omega = \infty$. Consideremos el suceso $A_n =$ *"en la n-ésima tirada sale Cara"*, es decir que:

$$A_n = \{\omega = (\omega_1, \omega_2, ..., \omega_{n-1}, C, \omega_{n+1}, ...) : \omega_i \in \{C,N\}\}$$

◊ **4.21.** Observar que si la moneda está balanceada (y no se gasta), $\mathbb{P}(A_n) = \frac{1}{2}$ para todo n.

◊ **4.22.** Probar que, en dicho caso, los sucesos $\{A_n\}_n$ son independientes.

Luego, como $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{2} = \infty$ y los sucesos $\{A_n\}_n$ son independientes, el lema de Borel-Cantelli dice que si tiro la moneda infinitas veces, la probabilidad de obtener *infinitas caras* es 1.

Ejemplo 4.2.5. *(Teorema del mono eterno).*

Considere un *mono eterno* (piense por ejemplo en un mono mordido por un vampiro o sucesivas generaciones de monos en la que un nuevo mono reemplaza al mono que se está por morir) que es colocado frente a una computadora o máquina de escribir y destinado a teclear sobre la misma para siempre.

Supongamos que ningún botón le llama más la atención que los otros, así que cada vez teclea cualquiera de los botones con igual probabilidad.

> *¿Cuán probable es que escriba infinitas veces una obra completa de Shakespeare ("Hamlet", por ejemplo)?*

Para responder a esta pregunta, pensemos primero en un caso más sencillo:

> *¿Cuán probable es que escriba infinitas veces la palabra "BANANA"?*

Suponga que el teclado tiene 50 botones (entre letras, dígitos, puntos, comas, etcétera). Luego, podemos modelar los posibles textos escritos por el mono mediante:

$$\Omega = \{\omega = (l_1, l_2, l_3, \ldots., l_n, \ldots.) : l_i \in \{A, B, C, \ldots, Z, 0, 1, 2, \ldots, 9, \text{“.”}, \text{“,”}, \ldots\}\}$$

La probabilidad de que con las primeras seis *tecleadas* escriba la palabra "BANANA" es: $\frac{1}{50} \times \frac{1}{50} \times \frac{1}{50} \times \frac{1}{50} \times \frac{1}{50} \times \frac{1}{50} = \frac{1}{50^6}$.

Pensemos cada $\omega \in \Omega$ partido en *bloques* de longitud 6, es decir, miremos por separado qué ocurre con:

$$[l_1, l_2, l_3, l_4, l_5, l_6]$$
$$[l_{6+1}, l_{6+2}, \ldots, l_{2\times 6}]$$
$$\ldots$$
$$[l_{6n+1}, l_{6n+2}, \ldots, l_{6(n+1)}]$$
$$\ldots$$

Consideremos los sucesos A_n =*"en el n-ésimo bloque escribe BANANA"*.

◊ **4.23.** Observar que los sucesos A_1, A_2,...,A_n,... son independientes y que $\mathbb{P}(A_n) = \frac{1}{50^6}$ para todo n. Observar además que A^∞ está estríctamente incluido en el suceso $B =$*"Escribe BANANA infinitas veces"* ya que, por ejemplo,

$$\omega = (L,B,A,N,A,N,A,B,A,N,A,N,A,B,A,N,A,N,A,B,A,N,A,N,A,B,A,N,A,N,A,....)$$

está en B pero no está en A^∞ ya que nunca entra toda la palabra BANANA dentro de ningún bloque. Es decir que $\mathbb{P}(B) \geqslant \mathbb{P}(A^\infty)$.

Luego, como los sucesos A_1, A_2,...,A_n,... son independientes y

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{50^6} = \infty$$

el lema de Borel-Cantelli implica que $\mathbb{P}(A^\infty) = 1$.

Finalmente, como $1 \geqslant \mathbb{P}(B) \geqslant \mathbb{P}(A^\infty) = 1$, se tiene que:

$$\mathbb{P}(\text{“Escribe BANANA infinitas veces”}) = 1.$$

Observe que lo que acabamos de hacer es completamente general, ya que podemos cambiar la palabra "BANANA" por el texto completo de *Hamlet*. Solo tendremos que contar cuántos carácteres aparecen en total en dicha obra (que es una cantidad grande pero finita) y considerar bloques de ese tamaño.

¿Quién sabe cuántos carácteres tiene la obra?

Es decir que no solo tenemos probabilidad uno de encontrar el texto completo de *Hamlet* en lo que está escribiendo el mono (porque sigue escribiendo hasta el infinito), sino que además (con probabilidad uno) lo escribirá infinitas veces!.

Secuencias aleatorias de palabras: ¿cómo hacer para que parezcan reales?

En esta actividad estudiaremos modelos de secuencias aleatorias de símbolos (ya sean caráteres o palabras). Este tipo de modelos son la base de los sistemas de texto predictivos que se utilizan por ejemplo en los teléfonos celulares. Un mensaje de texto es una secuencia de símbolos, que pueden ser letras o palabras.

Primera parte:

Consideremos para empezar una simplificación del problema y supongamos que nuestro diccionario está compuesto solo por las tres palabras más frecuentes en el idioma español, que son: *de, la*, y *que*. Supongamos además que las probabilidades de aparición de cada una de estas palabras son $\frac{5}{12}$, $\frac{5}{12}$ y $\frac{1}{6}$ respectivamente. Es decir que nuestro diccionario es:

```
Diccionario = c("de", "la", "que")

proba = c(5/12, 5/12, 1/6)
```

Imaginemos que queremos formar una oración de 20 palabras.

a) Si ponemos a hablar a un "mono" (*pobre mono, lo estamos haciendo trabajar como loco*) con estas tres palabras, entonces podemos suponer que cada palabra que use en su oración será *independiente* de lo que haya dicho antes. Por lo cual, podemos simular una frase de largo 20 dicha por este mono mediante la función *sample()*:

```
L=20 # Largo de la oración

palabras_mono = sample(Diccionario, L, prob= proba, replace = T)

palabras_mono
```

```
##  [1] "la"  "de"  "la"  "de"  "la"  "de"  "la"  "que" "que" "de"  "la"
## [12] "la"  "la"  "la"  "de"  "que" "de"  "la"  "de"  "que"
```

Observar que palabras_mono es un vector de 20 palabras, pero no tiene el aspecto de frase. Para que se parezca a una frase, pegaremos las coordenadas del vector (para que desaparezcan las " ") y las separaremos con un espacio:

```
frase_mono = paste(palabras_mono,collapse = " ")

frase_mono
```

```
## [1] "la de la de la de la que que de la la la la de que de la de que"
```

b) Tenga en cuenta que cuando nosotros hablamos, por lo general, no *escupimos palabras al azar* sino que la palabra que uso actualmente está fuertemente vinculada a las palabras que usé antes en la frase.

Para empezar, asumamos que cada vez que uso una palabra nueva, la palabra nueva está sólo relacionada con la palabra que dije inmediátamente antes. Para simular una frase de este tipo, asumamos que uno elige la palabra nueva al azar pero ahora la probabilidad con que la elijo depende de la palabra que usé inmediatamente antes. Esta estructura se puede describir mediante "probabilidades de transición":

$$
\begin{aligned}
p(j|i) = & \text{ probabilidad condicional de que, dado que aparece la palabra } i, \\
& \text{ inmediatamente después aparezca la palabra } j.
\end{aligned}
$$

Verificar que se cumple la siguiente propiedad:

$$\sum_{j \in \text{Diccionario}} p(j|i) = 1$$

.

Supongamos que las probabilidades de transición entre las palabras de Diccionario $= \{de, la, que\}$ son las siguientes:

		j		
		de	la	que
	de	0	4/5	1/5
i	la	4/5	0	1/5
	que	1/2	1/2	0

Es decir que, nunca digo *de de* y un 80% de la veces digo *la* después de haber dicho *de*.

Pongamos ahora a un *mono evolucionado* a hablar con esta nueva regla. Obliguémoslo a comenzar con la palabra *de* (o la palabra que prefiera). Entonces, elegirá la palabra que sigue según las probabilidades definidas en la fila 1 de la tabla anterior. Si dice:

- *la*, entonces elegirá la tercer palabra según la fila 2 de la tabla,
- *que*, entonces elegirá la tercer palabra según la fila 3 de la tabla.

Y así sucesivamente para cada nueva palabra.

```r
# Matriz de transición (la llenamos por filas)
proba_tran <- c(0,4/5,1/5,4/5,0,1/5,1/2,1/2,0)
P <- matrix(proba_tran,ncol = 3,nrow = 3,byrow = TRUE)

#Definimos un vector de palabras vacias ""  y a continuación lo rellenamos elegiendo
# la palabra nueva como dijimos antes:

palabras_mono2 = character(L)

#Elegimos la primer palabra. En este ejemplo elegimos "de".
palabras_mono2[1] = "de"

#Hacemos un loop para ir agregando las palabras de a una:

for(i in 1:(L-1))
{
  #Buscamos el indice en el diccionario de la palabra que dijimos antes
  # (para saber qué fila de la matriz mirar para elegir la próxima palabra)
```

```
  indice <- which(Diccionario == palabras_mono2[i])
  #Sacamos de la matriz de probabilidades de transición las correspondientes a
  # palabras_mono2[i]
  proba <- P[indice,]
  #Elegimos la palabra nueva según las probabilidades anteriores:
  palabras_mono2[i+1] <- sample(Diccionario, 1, prob = proba)
}

# Igual que antes, pegamos las palabras del vector palabras_mono2 para formar el texto final
frase_mono2 <- paste( palabras_mono2, collapse = " ")
```

Luego, el resultado para esta simulación es:

```
frase_mono2
```

```
## [1] "de la de la que la de la de la de que de la de que la de la de"
```

Segunda parte [PRIMERA ENTREGA].

Nota: Este ejercicio será la primer tarea (de las dos) para entregar (opcionalmente) en el curso. Realizar un informe con los resultados obtenidos y adjuntar el *script.* Pueden formarse grupos de, a lo sumo, 2 personas.

Queremos simular ahora un texto más "real" con palabras reales. Para eso,

- Consideraremos como *diccionario* todas las palabras que aparecen en el libro *"El Futbol a sol y sombra (y otros escritos)"* de Eduardo Galiano.
- Estimaremos las probabilidades de transición entre palabras para nuestro idioma a partir de las frecuencias relativas de las transiciones entre las palabras de este texto.
- Con estas probabilidades simularemos una frase más real.

Esto es lo que haremos a continuación:

a) Descargar del EVA el archivo que se llama *GaleanoFutbol.txt.* Observarlo.

Cargar al **R** todas las palabras del libro (del archivo anterior).

- Una opción es cargar todo el texto, partirlo en palabras (consideraremos como palabras todas las secuencias de letras separadas por espacio) y luego crear un *gran vector* con todas esas palabras, en el orden en que aparecen. Para eso utilice el archivo que se llama *GaleanoFutbol.txt*, que está en el EVA.
- Otra opción (más fácil) es cargar a **R** el archivo que se llama *PalabrasGaleano.csv.* En este caso ya se hizo todo el trabajo explicado en el item anterior.

Nota importante: Antes de hacer lo que viene, corra la siguiente línea:

```
options( stringsAsFactors=F )## Hace que los caracteres sean solo eso, caracteres.
```

Recordar que en *Rstudio* se puede importar el archivo con el botón *Import Dataset.* Cargará el .csv como un *data.frame* de una sola columna. Si queremos transformarlo en un vector de palabras, hacemos:

```
PalabrasGaleano = PalabrasGaleano[ ,1 ] # Mira la primer columna
```

b) Defina quién es el nuevo diccionario. Es decir, cuáles son todas las palabras *distintas* que aparecen en este texto (en el orden en que aparecen). Por ejemplo, las primeras 10 palabras del diccionario deberían ser:

```
##  [1] "el"       "futbol"   "a"        "sol"      "y"        "sombra"
##  [7] "otros"    "escritos" "prologo"  "todos"
```

Observar que pasamos de un diccionario de tres palabras a uno de:

```
d = length(Diccionario)
d
```

```
## [1] 4870
```

palabras. Contar cuántas palabras tiene el texto.

c) Estimaremos primero la frecuencia de aparición de cada palabra en este texto (para esto podría resultarle útil usar la función *table()*)

d) Estimaremos ahora las probabilidades de transición entre palabras mediante las frecuencias de las transiciones. Es decir, si i y j son palabras del diccionario, estimaremos $p(j|i)$ mediante:

$$f_N(j|i) = \frac{\# \text{ de veces que ocurre la transición } i-j}{\# \text{ de veces que aparece la palabra } i}$$

- Analice por qué sería razonable estimar las probabilidades de trancisión de esta manera.
- Para cada par de palabras (i, j) del diccionario (observar que podría ser $i = j$) estimar las probabilidades de trancisión $p(j|i)$ y guardarlas en una matriz de trancisión P similar a la de la primera parte.

Primero crear una matriz de transición con 0 y después la rellenamos con las frecuencias relativas.Por ejemplo, para rellenar la primer fila de la matriz, podríamos hacer lo siguiente:

```
d= length(Diccionario) # Cantidad de palabras del diccionario
P=matrix (rep(0, d*d), nrow=d, ncol= d, byrow=T)

# Trancisiones para la primer palabra
i = Diccionario[1] # Le llamo i a la palabra que me interesa (en este caso es la primer
                   # palabra)

pij = rep(0,d)
```

pij funcionará como un contador para las palabras que le sigan a la palabra i.

Ahora recorremos todo el texto (el vector PalabrasGaleano) y cada vez que encontremos la palabra i miraremos la palabra que le sigue. Si la palabra que le sigue es la palabra j, en el lugar j del vector *pij* agregaremos un 1:

```
for (k in 1:(N-1))
{
  if (PalabrasGaleano[k]==i)
  {
pij[which(Diccionario==PalabrasGaleano[k+1])]=pij[which(Diccionario==PalabrasGaleano[k+1])]+1
  }
}
```

Luego, el vector *pij* tiene la cantidad de veces que ocurre la trancisión $i - - > j$. Ahora cambiaremos la fila i (en este caso la fila 1) de la matriz P por el vector *pij* dividido entre la cantidad de veces que aparece la palabra i, es decir:

```
P[1, ] = pij/sum(PalabrasGaleano==i)
```

Por ejemplo, para chequear que todo esté bien, verificar que la fila 1 de P suma 1:

```
sum(P[1,])
```

```
## [1] 1
```

Nota: Sea paciente! Tenga en cuenta que para completar la matriz P tendrá que repetir lo anterior $d = 4870$ veces (P tiene dimensiones 4870×4870) y, puesto que para cada paso tiene que recorrer el texto completo, tendrá que efectuar 4870×19357 operaciones!

¿Habrá alguna manera más eficiente de hacerlo?

Si ve que se cuelga la máquina por más de una hora (por ejemplo), ponga *stop* y pase directo a la parte **e**.

e) Simulación de una oración utilizando las frecuencias de trancisiones "aprendidas" en la parte anterior.

Observe que si queremos simular una frase de $L = 20$ palabras, entonces no necesitamos conocer las 4870 filas de la matriz P sino solo las correspondientes a las palabras del diccionario que van apareciendo en cada paso.

Obliguemos a que la frase comience con la palabra "el":

```
L=20

texto = character(L)

texto[1]="el"
```

Ahora la palabra "el" será la palabra i para el ejemplo anterior, y para elegir la próxima palabra (la segunda palabra) tendremos que:

1. Hallar qué lugar ocupa la palabra "el" en el diccionario. Supongamos que es la i-ésima palabra.
2. Correr el algoritmo de la parte anterior para rellenar la i-ésima fila de la matriz P (por ahora las otras filas no nos importan).
3. Sortear la próxima palabra según las probabilidades definidas antes sobre $P[i,]$.
4. Volver al paso 1 para hallar la próxima palabra.

```
L=20

texto = character(L)

texto[1]= "el"
################################################
#Para hallar la segunda palabra

i=texto[1]

# Construcción de P[i,]
pij = rep(0,d) # Este vector pasará a ser la fila i

for (k in 1:(N-1))
{
  if (PalabrasGaleano[k]==i)
  { pij[which(Diccionario==PalabrasGaleano[k+1])] =
    pij[which(Diccionario==PalabrasGaleano[k+1])]+1 }
}

# Ahora modifico la matriz de transición en la fila correspondiente a la palabra i:
lugar = which(Diccionario == texto[1])
lugar
```

```
## [1] 1
```

```r
P[lugar, ] = pij/sum(PalabrasGaleano==i)

#########################
# Sorteo de la siguiente palabra:
texto[2] = sample(Diccionario, 1, prob = P[lugar, ])
texto[2]
```

```
## [1] "sueno"
```

A partir de las líneas anteriores, vea cómo terminar de construir la frase de 20 palabras.

f) ¿Se le ocurre alguna modificación a lo que hizo en la parte anterior para que el texto obtenido parezca más real?

Capítulo 5

Variables aleatorias en espacios discretos

Variable aleatoria (v.a.); función de probabilidad puntual; esperanza; varianza; esperanza condicional.

[Duración: 2 Semanas]

5.1. Definiciones y notaciones

A continuación estudiaremos más en detalle el concepto, ya introducido en el Capítulo 1, de variable aleatoria (v.a.) definida sobre un espacio muestral discreto.[1] Con ese fin introduciremos algunas definiciones y notaciones.

[1] En la literatura, un variable aleatoria "discreta" es una cierta función definida sobre un espacio de probabilidad no necesariamente discreto que toma valores sobre un conjunto numerable de valores reales. Para formalizar esta noción es necesario definir correctamente qué son los eventos en un espacio muestral no discreto. Todo esto lo veremos más adelante.

Definiciones. Sea $(\Omega, \mathbb{P})$ un espacio de probabilidad discreto.

- Una *variable aleatoria* (v.a.) X es una función $X : \Omega \to \mathbb{R}$.
- Denotaremos por $\Omega_X \subset \mathbb{R}$ el conjunto de valores alcanzados por X.
- Para todo subconjunto $A \subset \Omega_X$, definimos $\mathbb{P}_X(A) := \mathbb{P}(X \in A) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$.
- Para cada $x \in \Omega_X$, definimos $p_X(x) := \mathbb{P}(X = x) = \sum_{\omega : X(\omega) = x} \mathbb{P}(\{\omega\})$.
- A la función $p_X : \Omega_X \to \mathbb{R}$ se le llama *función de probabilidad puntual* de la v.a. X.

La terminología "variable aleatoria" no es del todo adecuada dado que en realidad no es ni una variable y ni aleatoria. Sin embargo, el término hace referencia a que el dominio de X es un espacio muestral, de donde se puede pensar que $X(\cdot)$ toma valores aleatorios según la realización $\omega \in \Omega$ de cierto experimento aleatorio.

◇ **5.1.** Sea X una v.a. definida en un espacio de probabilidad discreto $(\Omega, \mathbb{P})$. Entonces la función de probabilidad puntual $p_X : \Omega_X \to \mathbb{R}$ es una función de probabilidad para el espacio muestral Ω_X (recordar definición en Sección 1.4.2). En este sentido una v.a. X induce un espacio de probabilidad discreto, a saber, $(\Omega_x, \mathbb{P}_X)$.

Comentario 4. La función de probabilidad puntual p_X puede extenderse a toda la recta de manera natural; se extiende $p_X(x) = 0$ para $x \notin \Omega_X$.

Ejemplo 5.1.1. (*Constante*) El caso más sencillo, y menos interesante, es el de una variable aleatoria constante. Sea $c \in \mathbb{R}$, definimos la v.a. $X : \Omega \to \mathbb{R}$ por $X(\omega) = c$, para todo $\omega \in \Omega$. Es claro que $\Omega_X = \{c\}$, y la función de probabilidad puntual $p_X(c) = 1$.

Ejemplo 5.1.2. (*Indicatriz de un conjunto*) Sea $A \subset \Omega$ un evento particular. Entonces definimos la *función indicatriz* o *indicatriz* de A, a la v.a. $\mathbb{1}_A : \Omega \to \mathbb{R}$ definda por

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in \Omega \setminus A. \end{cases}$$

Teorema 5.1. *Sea $\{x_i\}_{i \in \mathbb{N}}$ una sucesión de números reales distintos, y sea $\{p_i\}_{i \in \mathbb{N}}$ una sucesión de números reales no negativos tales que $\sum_i p_i = 1$. Entonces existe un espacio de probabilidad discreto $(\Omega, \mathbb{P})$ y una v.a. X sobre dicho espacio tal que la función de probailidad puntal p_X de X satisface $p_X(x_i) = p_i$.*

◇ **5.2.** Dar una prueba del Teorema 5.1.

Comentario 5. El Teorema 5.1 es de mucha utilidad para la Probabilidad. Muestra la irrelevancia del espacio de probabilidad específico a considerar (espacio muestral, eventos, medida de probabilidad). Basta con decir que tenemos una v.a. X que toma los valores x_i, con probabilidad p_i, para saber que existe un modelo de un experimento con esta información. (Dicho de manera más vulgar, nos ahorramos la perorata "considere el experimento de ..."
Además, este teorema da una evidencia de la filosofía de la teoría de Probabilidad. La forma (topología, geometría) del espacio muestral Ω es irrelevante desde el punto de vista probabilístico. En este sentido para un probabilista la función de probabilidad puntual p_X codifica la información relevante.

5.2. Ejemplos

Veamos algunos ejemplos frecuentes de variables aleatorias discretas. Como vimos en el Teorema 5.1 no es necesario especificar el espacio muestral subyacente, y por lo tanto evitaremos hacer mención al mismo.

Ejemplo 5.2.1. (*Distribución de Bernoulli*) Decimos que X tiene *distribución de Bernoulli con parámetro* p, y se denota por $X \sim \mathcal{B}(p)$, si X toma los valores $\{0,1\}$ con función de probabilidad puntual

$$p_X(0) = p, \quad p_X(1) = 1-p.$$

La distribución Bernoulli es la v.a. más simple (no trivial). Modela un experimento con dos resultados, "éxito-fracaso", donde la probabilidad de éxito es p, y la de fracaso es $1-p$. (Por ejemplo el experimento de lanzar una moneda.)

Una clase particular de distribución de Bernoulli es la función indicatriz dada en el Ejemplo 5.1.2. De esta manera, la indicatriz de A es una v.a. de Bernoulli con parámetro $\mathbb{P}(A)$.

Ejemplo 5.2.2. (*Distribución Binomial*) Decimos que X tiene *distribución binomial con parámetros* n *y* p, y se denota $X \sim B(n,p)$, si X toma los valores $\{0,1,\dots,n\}$, y la función de probabilidad puntual es

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (k = 0,1,\dots,n). \tag{5.1}$$

Ejemplo 5.2.3. (*Distribución Poisson*) Si una v.a. X toma los valores $0,1,\dots$ con distribución de probabilidad

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad (k = 0,1,\dots), \tag{5.2}$$

para $\lambda > 0$ se dice que tiene *distribución de Poisson con parámetro* λ, y escribimos $X \sim \mathcal{P}(\lambda)$.

Ejemplo 5.2.4. (*Distribución Geométrica*) Decimos que X es una v.a. con *distribución geométrica de parámetro* $p \in (0,1)$ si toma los valores $\{1,2,...\}$ y

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad (k = 1,2,...) \tag{5.3}$$

◇ **5.3.** Verificar que las fórmulas (5.1), (5.2), (5.3) definen efectivamente funciones de probailidad puntual de variables aleatorias (en el sentido del Teorema 5.1).

Ejemplo 5.2.5. En este ejemplo veremos cómo el modelo de lanzar una moneda induce varias de las distribuciones de probabilidad mencionadas arriba. Supongamos que lanzamos una moneda n veces y la probabilidad de que resulte cara en cada tirada sea p. Motivado por la Subsección 1.5.1, definimos

- el espacio muestral Ω_n está formado por las sucesiones $\omega = (\omega_1, \dots, \omega_n)$, donde cada ω_i representa "C" (cara) o "N" (número);
- la función de probabilidad

  $$p(\omega) = p^{S_n(\omega)} q^{n-S_n(\omega)},$$

  donde $S_n(\omega)$ cuenta la cantidad de caras en la tira ω, y $q = 1-p$. Sea $\mathbb{P}$ la medida de probabilidad inducida.

Para $i = 1, \dots, n$, sean las variables aleatorias $X_i : \Omega_n \to \mathbb{R}$ definidas por

$$X_i(\omega) = \begin{cases} 1 & \omega_i = C \\ 0 & \omega_i = N \end{cases} \tag{5.4}$$

donde $\omega = (\omega_1, \dots, \omega_n) \in \Omega_n$. Es decir, $X_i(\omega)$ toma el valor 1 cuando la i-ésima coordenada de ω es cara. (En particular $S_n = X_1 + \cdots + X_n$.)

◇ **5.4.** Probar que X_i, para $i = 1,2,\dots,n$, son variables aleatorias con distribución de Bernoulli de parámetro p (y de hecho independientes).

◇ **5.5.** Probar que S_n es una v.a. con $\mathcal{B}(n,p)$.

◇ **5.6.** * Para n grande y p pequeño de manera tal que $np = \lambda$, entonces S_n puede ser aproximado por una Poisson, esto es, $\mathbb{P}(S_n = k) \sim \frac{1}{k!}\lambda^k e^{-\lambda}$.

5.3. Esperanza

♢ **5.7.** Si tenemos un dado de seis caras donde cinco de ellas tienen un 3 y una de ellas un 6. Si tiramos el dado 6000 veces. ¿cuál sería el valor esperado del promedio?

♢ **5.8.** Considere el siguiente juego. Lanzamos dos dados. Si el resultado suma 2 entonces el jugador gana \$1000. De lo contrario pierde \$100. ¿jugarías a este juego?

Supongamos que $y_1, \dots, y_N$ son los resultados numéricos de repetir un experimento aleatorio N veces de manera independiente. (Obviamente se pueden repetir los valores.) El valor medio, o promedio de los respectivos valores es

$$\text{prom} = \frac{y_1 + \cdots + y_N}{N}.$$

Si pensamos que estos resultados numéricos son muestras independientes de una v.a. X con función de probabilidad puntual p_X sobre $\Omega_X = \{x_1, \dots, x_n\}$, (donde $n \leqslant N$), resulta de reordenar la última ecuación que

$$\text{prom} = \frac{c_1 x_1 + \cdots + c_n x_n}{N},$$

donde $c_i = \#\{\ell : y_\ell = x_i\}$. De nuestra definición frecuentista de la probabilidad resulta que $p_X(x_i) \sim \frac{c_i}{N}$, de donde resulta

$$\text{prom} \sim \sum_{i=1}^{n} x_i p_X(x_i).$$

Esto motiva la siguiente definición.

Definición. El *valor esperado*, o *esperanza*, de una v.a. discreta X se define como

$$\mathbb{E}(X) := \sum_{x \in \Omega_X} x p_X(x).$$

siempre que la suma sea absolutamente convergente.

Si la serie $\sum_{x \in \Omega_X} x p_X(x)$ no es absolutamente convergente podemos convenir que la esperanza es $+\infty, -\infty$, o no existente, dependiendo de los términos positivos y negativos de la serie. Por ejemplo, si $\Omega_X \subset \mathbb{R}^+$, entonces $\mathbb{E}(X)$ es un número positivo o $+\infty$.

Comentario 6. Otra forma de entender la esperanza es mirar su analogía física con la noción de centro de masa. Si colocamos masas de peso $p_X(x_i)$ sobre puntos con coordenadas x_i, entonces el centro de masa del sistema es

$$\frac{\sum_i x_i p_X(x_i)}{\sum_i p_X(x_i)}$$

que resulta ser $\mathbb{E}(X)$.

A continuación calcularemos la esperanza de variables aleatorias con distintas distribuciones de probabilidad.

♢ **5.9.** Si X es una v.a. constante, entonces $\mathbb{E}(X)$ coincide con el valor de la constante.

♢ **5.10.** Si X es una v.a. con distribución de Bernoulli de parámetro p, entonces $\mathbb{E}(X) = p$.

♢ **5.11.** Si $\mathbb{1}_A$ es la indicatriz de cierto evento $A \subset \Omega$, entonces $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

♢ **5.12.** Si $\lambda > 0$ y $X \sim \mathcal{P}(\lambda)$ entonces $\mathbb{E}(X) = \lambda$.

♢ **5.13.** Si $X \in B(n, p)$, entonces $\mathbb{E}(X) = np$.

La esperanza cumple algunas propiedades.

Lema 5.2. *Sea X una v.a. sobre un espacio muestral discreto* $(\Omega, \mathbb{P})$. *El valor esperado* $\mathbb{E}$ *cumple:*

a) Si $a, b \in \mathcal{R}$, *entonces* $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

b) si $X \geqslant 0$, *entonces* $\mathbb{E}(X) \geqslant 0$ *(positivo)*

c) si $X = \mathbb{1}_\Omega$, *entonces* $\mathbb{E}(X) = 1$.

d) si $g : \mathbb{R} \to \mathbb{R}$ *es una función, entonces* $\mathbb{E}(g(X)) = \sum_{y \in \Omega_X} g(x)\mathbb{P}(X = x)$.

e) Si $\mathbb{P}(X \geqslant 0) = 1$, *y* $\mathbb{E}(X) = 0$ *entonces* $\mathbb{P}(X = 0) = 1$.

♢ **5.14.** Dar una prueba del lema anterior.

♢ **5.15.** Si $X \sim \mathcal{P}(\lambda)$ entonces $\mathbb{E}(e^X) = e^{\lambda(e-1)}$.

♢ **5.16.** Si $X \sim \mathcal{P}(\lambda)$, entonces $\mathbb{E}(X(X-1)\cdots(X-k)) = \lambda^{k+1}$.

5.4. Varianza

La esperanza de una variable aleatoria da información sobre el "centro" de la variable aleatoria. Sin embargo, la esperanza no brinda información sobre cómo están dispersados los valores alrededor de su centro. Para tener una medida de la dispersión se define la *varianza*.

Definición. La *varianza* $\mathrm{var}(X)$ de una variable aleatoria discreta X se define por

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

Del Lema 5.2 se obtiene que si $\mathbb{E}(X) = \mu$, entonces

$$\mathrm{var}(X) = \sum_{x \in \Omega_X} (x - \mu)^2 \mathbb{P}(X = x).$$

◇ **5.17.** Desarrollar la fórmula anterior para probar que

$$\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{5.5}$$

◇ **5.18.** Probar los siguientes resultados.

- Si X tiene distribución geométrica de parámetro p, entonces $\mathrm{var}(X) = (1-p)p^{-2}$.
- Si $X \sim \mathcal{B}(n,p)$ entonces $\mathrm{var}(E) = npq$.
- Si $X \sim \mathcal{P}(\lambda)$, entonces $\mathrm{var}(X) = \lambda$.

◇ **5.19.** Mostrar que $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$, para cualesquiera $a, b \in \mathcal{R}$.

◇ **5.20.** Sea X v.a. en espacio de probabilidad discreto. Si $\mathbb{E}(X^2) = 0$ entonces $\mathbb{P}(X = 0) = 1$. Deducir que si $\mathrm{var}(X) = 0$ entonces $\mathbb{P}(X = \mathbb{E}(X)) = 1$ (siempre que la esperanza sea finita).

5.5. Esperanza Condicional

Supongamos que X es una v.a. sobre un espacio de probabilidad discreto $(\Omega, \mathbb{P})$, y sea B un evento con probabilidad positiva. El conocimiento de la ocurrencia del evento B altera la distribución de la v.a. X. Esto es, la probabilidad $\mathbb{P}(X = x_i)$ es remplazada por

$$\mathbb{P}(X = x_i | B) = \frac{\mathbb{P}(\{X = x_i\} \cap B)}{\mathbb{P}(B)}.$$

Esta observación motiva la siguiente definición.

Definición. Sea X una v.a. sobre un espacio muestral discreto $(\Omega,\mathbb{P})$ con $\mathbb{P}(B)>0$. La *esperanza condicional de X dado B*, se denota por $\mathbb{E}(X|B)$, y se define por

$$\mathbb{E}(X|B)=\sum_{x\in\Omega_X} x\,\mathbb{P}(X=x|B),$$

siempre que la suma sea absolutamente convergente.

Teorema 5.3. *Sea X una v.a. sobre un espacio muestral discreto* $(\Omega,\mathbb{P})$*, y* $\{B_1,\dots\}$ *una partición de* Ω *por conjuntos de probabilidad positiva. Entonces,*

$$\mathbb{E}(X)=\sum_i \mathbb{E}(X|B)\,\mathbb{P}(B_i).$$

◊ **5.21.** Dar una prueba de este teorema.

◊ **5.22.** Supongamos que lanzamos una moneda varias veces, donde la probabilidad de que resulte cara es p. Calcular la esperanza de la longitud de la racha inicial. (Una racha inicial es una colección de tiradas donde se repite el mismo lado de la moneda que la primer tirada.)

◊ **5.23.** Sea X una v.a. en un espacio de probabilidad discreto, y sea $f:\mathbb{R}\to\mathbb{R}$. Si para $x\in\mathbb{R}$ se tiene que $\mathbb{P}(X=x)>0$, entonces probar que

$$\mathbb{E}(g(X)|X=x)=g(x).$$

Deducir que $\mathbb{E}(g(X))=\sum_{x\in\Omega_X} g(x)\,\mathbb{P}(X=x)$.

◊ **5.24.** Probar, usando esperanza condicional y condicionando por resultado de primer tirada, que la esperanza del tiempo de espera hasta que salga cara en la tirada de una moneda es 2. Hacer los mismo para calcular la esperanza del tiempo de espera del problema de ocupación (ii) en Sección 2.5.

5.6. Ejercicios complementarios

◊ **5.25** (Pérdida de memoria)**.** Probar que si X tiene distribución geométrica con parámetro p, entonces

$$\mathbb{P}(X>m+n|X>m)=\mathbb{P}(X>n),$$

para $m,n\in\mathbb{N}$. Además probar que la distribución geométrica es la única distribución concentrada en los naturales positivos tales que tiene la propiedad de pérdida de memoria.

$\diamondsuit$ **5.26.** Si X es una v.a. que toma valores no negativos, entonces se tiene

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

Capítulo 6

Vectores aleatorios en espacios discretos

Vectores aleatorios, distribución conjunta, independencia, esperanza, funciones generatriz de probailidad, sumas de v.a.

[Duración: 1 Semana]

En este capítulo extenderemos lo visto en el capítulo anterior al caso de varias variables aleatorias definidas sobre un mismo espacio de probabilidad discreto. Además estudiaremos sumas de variables aleatorias con valores naturales mediante la técnica de funciones generatrices de probabilidad.

6.1. Distribuciones multivariadas

Sean X e Y variables aleatorias definidas sobre un mismo espacio de probabilidad discreto $(\Omega, \mathbb{P})$. Nos concentraremos en estudiar el *vector aleatorio* (X,Y) en $\mathbb{R}^2$. Comencemos por definir la función de probabilidad puntual de un vector aleatorio.

Definiciones. Sean X e Y variables aleatorias definidas sobre un mismo espacio de probabilidad discreto $(\Omega, \mathbb{P})$, definimos la *(función de) probabilidad puntual conjunta* $p_{X,Y}$ del vector (X,Y) a la función $p_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$, dada por

$$p_{X,Y}(x,y) := \mathbb{P}(X = x, Y = y) := \mathbb{P}(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y, \})$$

$\diamondsuit$ **6.1.** Observar que $p_{X,Y}(x,y) \geqslant 0$, y

$$\sum_{x\in\Omega_X, y\in\Omega_Y} p_{X,Y}(x,y) = 1$$

Sean X e Y v.a. sobre un mismo espacio de probabilidad discreto. Conociendo la probabilidad puntual conjunta $p_{X,Y}$ podemos obtener las funciones de probabilidad puntuales p_X y p_Y de la siguiente manera

$$p_X(x) = \sum_{y\in\Omega_Y} p_{X,Y}(x,y), \quad p_Y(y) = \sum_{x\in\Omega_X} p_{X,Y}(x,y). \tag{6.1}$$

Estas probabilidades puntuales se denominan comúnmente las *marginales* de la probabilidad puntual conjunta.

$\diamondsuit$ **6.2.** Consideremos un par de v.a. (X,Y) con valores sobre $\mathbb{N}^2$ tal que la probabilidad puntual conjunta satisface

$$p_{X,Y}(i,j) = \begin{cases} \theta^{i+j+1} & i,j = 0,1,2 \\ 0 & \text{en otro caso} \end{cases}$$

Probar que θ satisface la ecuación $\theta + 2\theta^2 + 3\theta^3 + 2\theta^4 + \theta^5 = 1$. Encontrar las marginales.

6.2. Esperanza multivariada

Supongamos que tenemos el vector aleatorio (X,Y) definido en un espacio de probabilidad discreto $(\Omega,\mathbb{P})$. Sea $g:\mathbb{R}^2 \to \mathbb{R}$ una función. Entonces

$$Z := g(X,Y) : \Omega \to \mathbb{R}$$

define una nueva variable aleatoria en Ω, donde $Z(\omega) = g(X(\omega),Y(\omega))$.

Teorema 6.1. *Resulta*

$$\mathbb{E}(g(X,Y)) = \sum_{x\in\Omega_X, y\in\Omega_Y} g(x,y)p_{X,Y}(x,y)$$

$\diamondsuit$ **6.3.** Probar el resultado anterior.

$\diamondsuit$ **6.4.** Supongamos que el vector aleatorio (X,Y) satisface $p_{X,Y}(i,j) = \theta^{i+j+1}$, para $i,j = 0,1,2$. Probar que $\mathbb{E}(XY) = \theta^3 + 4\theta^4 + 4\theta^5$, y $\mathbb{E}(X) = \theta^2 + 3\theta^3 + 3\theta^4 + 2\theta^5$.

6.2.1. Linealidad de la esperanza

◊ **6.5.** La esperanza $\mathbb{E}$ se puede pensar como un operador sobre el conjunto de v.a. definidas en un espacio de probabilidad discreto $(\Omega, \mathbb{P})$. Probar que la esperanza actúa de manera lineal. Esto es, si X e Y son v.a. sobre un mismo espacio, entonces

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y), \quad a, b \in \mathcal{R}.$$

Este resultado parece inofensivo pero es de suma utilidad para los cálculos. Observar que no es necesario pedir que las variables aleatorias sean independientes.

Ejemplo 6.2.1. ¿Cuál es la esperanza de la suma del resultado de tirar dos dados? Si X_1 y X_2 son los resultados de cada dado, entonces $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 3{,}5 + 3{,}5 = 7$. Pero ¿qué sucede si los dados son dependientes? Nada, el resultado es el mismo. Por ejemplo, en el caso extremo de dependencia en que el segundo dado repite el resultado del primero, i.e. $X_2 = X_1$.

Lo anterior siempre es verdad asumiendo el comportamiento individual de cada dado es un dado fiel.

Una aplicación importante de la linealidad, para el caso continuo, se puede ver en la prueba del problema de la Aguja de Buffon en Sección 11.5. Otra aplicación inmediata es ver que el cálculo de la esperanza de una binomial no requiere del cálculo explícito de las probabilidades intermedias. Sólo basta saber la esperanza de una Bernoulli.

Ejemplo 6.2.2. Recordando el matching problem I dado en el ejercicio 3.2. Cada persona recibe su misma carta con probabilidad $1/n$. ¿Cuál es el valor esperado de personas que reciben su carta? Si definimos por X_i la v.a. que toma el valor 1 si la persona i recibe su carta (i.e. $X_i = \mathbb{1}_{A_i}$, donde A_i está dado en el ejercicio mencionado.) Entonces lo que buscamos es $\mathbb{E}(X_1 + \cdots + X_n)$. Dado que $\mathbb{E}(X_i) = \mathbb{P}(A_i) = 1/n$, concluimos de la linealidad que el número esperado de personas que reciban su carta es 1.

Comentario 7. El problema anterior es un problema de *permutaciones* y cantidad de *puntos fijos* de las permutaciones. Es interesante mencionar que hacer el ejemplo anterior por medio de la expresión de esperanza puede ser complicado. Además, tampoco estamos asumiendo que la distribución de las permutaciones sea uniforme. Lo único que requerimos que la probabilidad de que el el lugar i sea fijo $1/n$.

6.3. Distribución conjunta e independencia

Recordemos del Capítulo 4 que dos eventos A y B en un espacio de probabilidad $(\Omega, \mathbb{P})$ son independientes si la ocurrencia de uno no altera la probabilidad de que el segundo ocurra. Esto es, A y B son independientes si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Extendiendo esta definición a variables aleatorias definidas sobre un espacio de probabilidad discreto, diremos que las v.a. X e Y son independientes si la realización de alguno de los valores de X no altera la distribución de probabilidad de Y.

Definición. Decimos que dos v.a. discretas X e Y en un espacio de probabilidad son *independientes* si los eventos $\{X = x\}$ e $\{Y = y\}$ son independientes para cualesquiera $x \in \Omega_X$, $y \in \Omega_Y$.

Es decir, X e Y son independientes si $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.

♢ **6.6.** Probar que si X e y Y son v.a. independientes, y $f, g : \mathbb{R} \to \mathbb{R}$, entonces $f(X)$ y $g(Y)$ son independientes.

Observar que si X e Y son independientes entonces

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad x, y \in \mathbb{R},$$

donde p_X y p_Y son las probabilidades marginales definidas en (6.1).

Teorema 6.2. *Sean X e Y v.a. sobre un espacio de probabilidad discreto. Se tiene que X e Y son independientes si y sólo si existen funciones $f, g : \mathbb{R} \to \mathbb{R}$ tales que la función de probabilidad puntual conjunta satisface*

$$p_{X,Y}(x,y) = f(x)g(y), \quad x, y \in \mathbb{R}.$$

♢ **6.7.** Dar una prueba del teorema anterior.

Observar que en el teorema anterior $\sum_{x \in \Omega_X} f(x)$ y $\sum_{y \in \Omega_Y} g(y)$ no tiene por que tomar el valor 1. Esto resulta de que la factorización de la probabilidad conjunta no es única. Sin embargo, las marginales quedan únivocamente determinadas.

Ejemplo 6.3.1. Sean X e Y v.a. que toman valores naturales tales que la función de probabilidad puntual conjunta es

$$p_{X,Y}(i,j) = \left(3\frac{1}{i!}\lambda^i\right)\left(\frac{1}{3j!}\mu^j e^{-\lambda-\mu}\right).$$

Del Teorema 6.2 concluimos que X e Y son independientes.

♢ **6.8.** ¿Podrían identificar las marginales en el ejemplo anterior?

Teorema 6.3. *Si X e Y son v.a. independientes con esperanzas $\mathbb{E}(X)$ y $\mathbb{E}(Y)$, entonces*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{6.2}$$

♢ **6.9.** Probar el resultado anterior.

♢ **6.10.** Probar que si X e Y son independientes entonces $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$ (siempre que éstas existan).

En general la linealidad de la varianza no es válida. En este sentido se define una magnitud, denominada *covarianza*, que nos dice cuán distante están dos variables aleatorias de ser independientes. La covarianza es proporcional a $\text{var}(X+Y) - \text{var}(X) - \text{var}(Y)$, y está dada por el siguiente enunciado.

Definición. Se define la *covarianza* entre v.a. X e Y como

$$\text{cov}(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)$$

♢ **6.11.** Probar que la covarianza cumple las siguientes propiedades:

- $\text{cov}(X,c) = 0$; (con c una constante)
- $\text{cov}(X+c,Y) = \text{cov}(X,Y)$;
- $\text{cov}(X,Y) = \text{cov}(Y,X)$;
- $\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
- $\text{cov}(X+Z,Y) = \text{cov}(X,Y) + \text{cov}(Z,Y)$;
- $\text{cov}(X,X) = \text{var}(X)$;
- $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y)$;
- Si X e Y independientes, entonces $\text{cov}(X,Y) = 0$.

(¿Les recuerda a algo estas propiedades?)

La condición (6.2) no es suficiente para que las v.a. X e Y sean independientes. Veamos esto en el siguiente ejemplo.

Ejemplo 6.3.2. Supongamos que X es una v.a. que toma los valores $-1,0,1$ con la misma probabilidad puntual, y sea Y la v.a.

$$Y = \begin{cases} 0 & \text{si } X = 0, \\ 1 & \text{si } X \neq 0, \end{cases}$$

Veamos que podemos construir un espacio de probabilidad que realice estas variables aleatorias. Sea $\Omega = \{-1,0,1\}$ y con la medida de probabilidad $\mathbb{P}$ uniforme. Sea X la función identidad, esto es $X(\omega) = \omega$, y sea $Y(\omega) = |\omega|$.

◊ **6.12.** Probar que X e Y satisfacen las condición (6.2) pero sin embargo no son independientes.

◊ **6.13.** Probar que dos eventos son independientes si las funciones indicatrices respectivas son v.a. independientes.

Todo lo visto en este capítulo se puede extender de manera natural a más variables aleatorias. Esto es: $X_1,\dots,X_n$ son v.a. *independientes* si $\mathbb{P}(X_1 = x_1,\dots,X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i)$. Equivalentemente, utilizando la función de probabilidad puntual conjunta asociada, se tiene

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^n p_{X_i}(x_i),$$

donde p_{X_i} es la marginal asociada a la v.a. X_i.

Además se tiene que si $X_1,\dots,X_n$ son v.a. independientes entonces

$$\mathbb{E}(X_1 \cdots X_n) = \prod_{i=1}^n \mathbb{E}(X_i).$$

6.4. Función generatriz de probabilidad

Una forma de codificar una sucesión de números $\{a_n\}_{n\in\mathbb{N}}$ es considerar la *función generatriz* de la sucesión, a saber,

$$a_0 + a_1 s^1 + a_2 s^2 \cdots$$

Definición. Sea X una v.a. que toma valores naturales $0,1,2,\ldots$, y sea p_X la función de probabilidad puntual asociada. La *función generatriz de probabilidad de X*, se define por

$$G_X(s) := \sum_{k=0}^{\infty} p_X(k)s^k,$$

para s donde la serie converja absolutamente.

Observar que si $-1 \leqslant s \leqslant 1$, se tiene $\sum_{k=0}^{\infty} p_X(k)|s|^k \leqslant \sum_{k=0}^{\infty} p_X(k) = 1$, y por lo tanto G_X converge al menos en el intervalo $[-1,1]$ (criterio de Abel).

◊ **6.14.** Observar que $G_X(s) = \mathbb{E}(s^X)$, y que además $G_X(0) = p_X(0)$ y $G_X(1) = 1$.

Si G_X es una función generatriz de probabilidad, entonces podemos obtener la función de probabilidad puntual tomando sucesivas derivadas en 0. Esto es,

$$p_X(k) = G^{(k)}(0)/k!, \quad k = 0,1,2,\ldots$$

De esta manera, las funciones generatrices de probabilidad determinan unívocamente la distribución de la variable aleatoria. Es decir, si $G_X(s) = G_Y(s)$ en un entorno de cero, entonces las distribuciones de X e Y son las mismas.

Veamos algunos ejemplos de funciones generatrices de probabilidad. (Recordar que dado $p \in (0,1)$, denotamos $q = 1-p$.)

Ejemplo 6.4.1 (Distribución geométrica)**.** Si X es una v.a. con distribución Geo(p),

$$G_X(s) = \frac{ps}{1-ps}.$$

Ejemplo 6.4.2 (Distribución Bernoulli)**.** Si X es una v.a. con distribución $\mathcal{B}(p)$,

$$G_X(s) = q + ps.$$

Ejemplo 6.4.3 (Distribución Binomial)**.** Si X es una v.a. con distribución $\mathcal{B}(n,p)$,

$$G_X(s) = (q+ps)^n.$$

Ejemplo 6.4.4 (Distribución Poisson)**.** Si X es una v.a. con distribución Geo(p),

$$G_X(s) = e^{\lambda(s-1)}.$$

Una de las virtudes de las funciones generatrices de probabilidad es que permiten obtener propiedades de la distribución de una v.a. de manera más elegante y sencilla.

Una aplicación importante en este sentido es conocer los *momentos* de una variable aleatoria definida en un espacio de probabilidad discreto.

Definición. Sea X una v.a. sobre un espacio de probabilidad discreto. Dado $k = 1,2,\ldots$, definimos el *momento de orden k* de X a la cantidad $\mathbb{E}(X^k)$.

Por ejemplo, para conocer la varianza de X es suficiente conocer los dos primeros momentos (recordar (5.5) en Sección 5.4).

◊ **6.15.** Si X es una v.a. que toma valores naturales, con función de generatriz de probabilidad G_X. Entonces

$$G_X^{(k)}(1) = \mathbb{E}(X(X-1)\cdots(X-k+1)), \quad k = 1,2,\ldots.$$

(Se puede utilizar el siguiente resultado de Abel: si $A(s) = \sum_{k=0}^{\infty} a_k s^k$ es convergente en $|s| < 1$, entonces $\sum_k a_k = \lim_{s\to 1} A(s)$ (pudiendo ser ambos infinitos).

◊ **6.16.** Si X es un v.a. que toma valores naturales, dar una expresión de la varianza de X en función de las derivadas de G_X.

◊ **6.17.** Rehacer cálculos de esperanza y varianza de las distribuciones geométricas y Poisson mediante este método.

6.5. Suma de variables aleatorias

Como ya hemos visto a lo largo del curso, la teoría de Probabilidad tiene especial interés en entender suma de variables aleatorias independientes. En esta sección estudiaremos estas sumas de v.a. independientes definidas sobre un mismo espacio de probabilidad discreto, y utilizaremos la poderosa herramienta de funciones generatrices para el caso de v.a. que toman valores naturales.

Si X e Y son v.a. sobre un mismo espacio, ¿cómo se calcula la función de probaiblidad puntual de la v.a. $Z = X + Y$. Dado que Z toma el valor z si $X = x$ y $Y = z - x$, se tiene que

$$p_Z(z) = \mathbb{P}\left(\bigcup_{x\in\Omega_X} \{X = x\} \cap \{Y = z - x\}\right) = \sum_{x\in\Omega_X} \mathbb{P}(X = x, Y = z - x).$$

En el caso de ser independientes se obtiene la siguiente fórmula

$$p_Z(z) = \sum_{x \in \Omega_X} p_X(x) p_Y(z-x), \quad z \in \mathbb{R}. \tag{6.3}$$

Esta fórmula se conoce como la *convolución* de las funciones p_X y p_Y. Sin embargo, cuando tenemos más variables aleatorias, la fórmula anterior para encontrar p_Z resulta engorrosa dado que debemos realizar $n-1$ convoluciones consecutivas.

Para el caso de v.a. que toman valores naturales, podemos utilizar la poderosa herramienta de funciones generatrices de probabilidad.

Teorema 6.4. *Sean $X_1, \dots, X_n$, v.a independientes tomando valores en los naturales, entonces*

$$G_{X_1+\cdots+X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s). \tag{6.4}$$

Demostración. Se tiene

$$\begin{aligned} G_{X_1+\cdots+X_n}(s) &= \mathbb{E}(s^{X_1+\cdots+X_n}) \\ &= \mathbb{E}(s^{X_1} \cdots s^{X_n}) \\ &= \mathbb{E}(s^{X_1}) \cdots \mathbb{E}(s^{X_n}) = G_{X_1}(s) \cdots G_{X_n}(s) \end{aligned}$$

donde en la penúltima igualdad usamos la independencia de las v.a. s^{X_i}. □

◇ **6.18.** Hallar distribución de la suma de v.a. indpendientes con distribución $\mathcal{B}(n,p)$ y $\mathcal{B}(m,p)$ respectivamente. Concluir que la distribución de la suma de n variables aleatorias de Bernoulli de parámetro p tiene distribución Binomial de parámetro n y p.

◇ **6.19.** Mostrar que la suma de v.a. independientes con distribución $\mathcal{P}(\lambda)$ y $\mathcal{P}(\mu)$ respectivamente, tiene distribución $\mathcal{P}(\lambda+\mu)$.

6.6. Problemas complementarios

En esta sección de problemas asumiremos que trabajamos sobre un espacio de probabilidad discreto.

◇ **6.20.** Observarndo que $\mathbb{1}_{A\cup B} = 1 - \mathbb{1}_{A^c \cap B^c} = 1 - (1-\mathbb{1}_A)(1-\mathbb{1}_B)$, dar una nueva prueba del principio de inclusión y exclusión utilizando funciones indicatrices.

◇ **6.21.** Sean X e Y v.a. independientes con distribución Geo(p). Mostrar que

$$\mathbb{P}(X = k | X+Y = n) = \frac{1}{n+1}.$$

♢ **6.22.** Sean $X_1,\ldots,X_n$ v.a. independientes tales que

$$\mathbb{P}(X_i = k) = \frac{1}{n}, \quad k = 1,2,\ldots,n.$$

Encontrar la función de probabilidad puntual de

$$m_n = \text{mín}\{X_1,\ldots,X_n\}, \quad M_n = \text{máx}\{X_1,\ldots,X_n\}.$$

♢ **6.23.** Sean $X_1,X_2,\ldots$ v.a. donde cada una tiene esperanza μ, y sea N un v.a. que toma valores naturales y que es independiente con cada X_i. Sea $S = X_1 + \cdots + X_N$. Mostrar, condicionando en valores de N, que

$$\mathbb{E}(S) = \mu\mathbb{E}(N).$$

♢ **6.24.** Sean $X_1,X_2,\ldots$ v.a. independientes con igual distribución, y sea $S_n = X_1 + \cdots + X_n$. Mostrar que $\mathbb{E}(S_n/S_m) = n/m$ si $m \leqslant n$, y $\mathbb{E}(S_m/S_n) = 1 + (m-n)\mu\mathbb{E}(1/S_n)$ si $m > n$, donde $\mu = \mathbb{E}(X_1)$.

♢ **6.25.** N bolas se tiran sobre M celdas donde se permiten ocupar una celda con varias bolas. Mostrar que el número esperado de celdas vacías es $(M-1)^N/M^{N-1}$.

♢ **6.26.** Consideremos las n raíces de la unidad $\zeta_1,\ldots,\zeta_n$ (i.e. soluciones de $z^n = 1$, $z \in \mathbb{C}$.). Sobre cada par i, j (sin importar órden) se lanza una moneda con probabilidad de cara p. Si el resultado es cara unimos por el segmento que conecta esas dos raíces. Probar que el número esperado de aristas es $\frac{1}{2}n(n-1)p$. Probar que elnúmero esperado de triángulos es $\frac{1}{6}n(n-1)(n-2)p^3$.

♢ **6.27.** Sean N y $X_1,X_2,\ldots$ v.a. que toman valores naturales, y son independientes. Si las X_i tienen igual distribución con función generatriz de probabilidad G_X. Probar que $G_S(s) = G_N(G_X(s))$.

♢ **6.28.** Un dado es lanzado de manera independiente siete veces. Hallar probabilidad de que la suma total sea 14.

Capítulo 7

Función de distribución de v.a. discretas

Definición de función de distribución para v.a. discretas; ejemplos y propiedades. Función de distribución de vectores aleatorios en espacios discretos.

[Duración: 1 clase]

Consideraremos este breve capítulo como una especie de *conexión mental* entre las v.a. discretas que hemos estado estudiando y otro tipo de v.a. que comenzaremos a estudiar a partir del próximo capítulo.

7.1. Función de distribución de variables aleatorias discretas

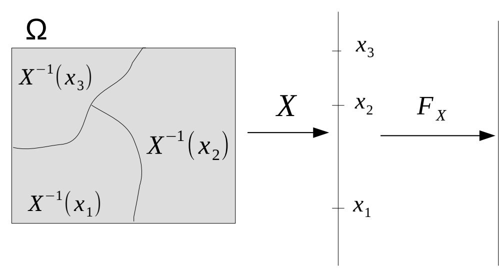
Sea X una variable aleatoria (v.a.) en un espacio de probabilidad discreto $(\Omega, \mathbb{P})$. En el capítulo anterior observamos que la función generatriz de X, $G_X : D_X \subset \mathbb{R} \to \mathbb{R}$ tal que $G_X(s) = \mathbb{E}(s^X)$, caracteriza la *distribución* ("qué tipo de v.a. es X") de X y además permite obtener propiedades de la distribución de X de manera más elegante y sencilla. En este capítulo definiremos otra función que también caracteriza a la v.a. X, a esta le llamaremos justamente *función de distribución de la v.a. X*.

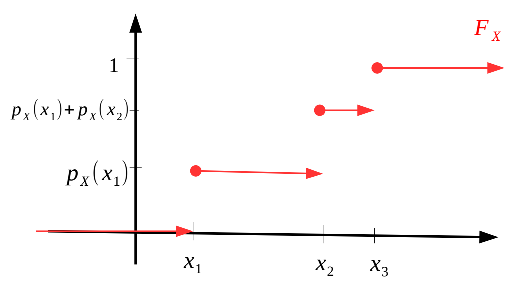
Definición. Sea $X : \Omega \to \mathbb{R}$ una v.a. en un espacio de probabilidad discreto $(\Omega, \mathbb{P})$, le llamamos *función de distribución de X* a $F_X : \mathbb{R} \to \mathbb{R}$ tal que $F_X(x) = \mathbb{P}(X \leqslant x)$ para todo $x \in \mathbb{R}$.

Observar que estamos definiendo F_X en todo $\mathbb{R}$, no solo en el recorrido de X (Ω_X). A veces a la función F_X se le llama también la *función de distribución acumulada de* X.

Ejemplo 7.1.1. Consideremos una v.a. X sobre $(\Omega, \mathbb{P})$ tal que $Rec(X) = \Omega_X = \{x_1 < x_2 < x_3\}$, entonces $F_X : \mathbb{R} \to \mathbb{R}$ queda definida así:

- Si $x < x_1$, entonces $F_X(x) = \mathbb{P}(X \leqslant x_1) = \mathbb{P}(\varnothing) = 0$.
- Si $x_1 \leqslant x < x_2$, entonces $F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X = x_1) = p_X(x_1)$.
- Si $x_2 \leqslant x < x_3$, entonces $F_X(x) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) = p_X(x_1) + p_X(x_2)$
- Si $x_3 \leqslant x$, entonces $F_X(x) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \mathbb{P}(X = x_3) = p_X(x_1) + p_X(x_2) + p_X(x_3) = 1$





◇ **7.1.** Probar que si X es una v.a. sobre un espacio $(\Omega,\mathbb{P})$ discreto, entonces su función de distribución F_X verifica:

1. $F_X(x) \in [0,1]$ para todo $x \in \mathbb{R}$.

2. F_X es una función creciente.

3. F_X es una *función càdlàg* (del francés: *"continue à droite, limite à gauche"*, continua por derecha con límite por izquierda).

4. $\lim_{x\to-\infty} F_X(x) = 0$ y $\lim_{x\to+\infty} F_X(x) = 1$.

5. $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$ para todo $a < b$.

6. Si $x_i \in \Omega_X$, entonces $p_X(x_i) = F_X(x_i) - \lim_{x\nearrow x_i} F_X(x)$.

7. Si X e Y son dos v.a. en $(\Omega,\mathbb{P})$ tales que $F_X(x) = F_Y(x)$ para todo $x \in \mathbb{R}$, entonces $\Omega_X = \Omega_Y$ y $p_X(x_i) = p_Y(x_i)$ para todo $x_i \in \Omega_X$. Es decir que X e Y *tienen la misma distribución*.

Luego, gracias a esta última propiedad, sabemos que podemos caracterizar a la v.a. X a partir de su función de distribución F_X. Esto nos permitirá muchas veces probar de manera elegante algunas propiedades de una v.a..

Más aún, es posible probar que cualquier función $F : \mathbb{R} \to \mathbb{R}$ que verifique las propiedades 1, 2, 3 y 4 del ejercicio anterior se corresponde con una v.a. X. Es decir que si F cumple estas cuatro condiciones entonces existe un espacio de probabilidad $(\Omega,\mathbb{P})$ y una v.a. X definida sobre este espacio tal que $F_X = F$. Omitiremos la prueba de este resultado por ahora (pero tener en mente el **teorema 5.1**). Lo bueno de este resultado es que en muchas circunstancias nos permite evitar el trabajo tedioso de tener que definir explícitamente el espacio de probabilidad y la variable aleatoria.

Ejemplo 7.1.2. Consideremos dos v.a. $X_1 \sim Geo(p_1)$ y $X_2 \sim Geo(p_2)$ independientes en $(\Omega,\mathbb{P})$ y definimos en $(\Omega,\mathbb{P})$ una nueva v.a. discreta: $X^* = \min\{X_1, X_2\}$ (i.e. $X^*(\omega) = \min\{X_1(\omega), X_2(\omega)\} \in \mathbb{R}$ para todo $\omega \in \Omega$). Resulta natural preguntarse qué tipo de v.a. es X^* o, en otras palabras, qué distribución tiene esta v.a.

◇ **7.2.** Determine cuál es el recorrido de la v.a. X^*. Observe que el cálculo de $p_{X^*}(k)$ para todo $k \in Rec(X^*)$ puede ser engorroso.

Estudiemos entonces cómo es $F_{X^*}:\mathbb{R}\to\mathbb{R}$. Observar que para todo $x\in\mathbb{R}$ se cumple:

$$\begin{aligned}
F_{X^*}(x) &= \mathbb{P}(\text{mín}\{X_1,X_2\}\leqslant x) = 1-\mathbb{P}(\text{mín}\{X_1,X_2\}>x)\\
&= 1-\mathbb{P}(\{X_1>x\}\cap\{X_2>x\}) = 1-\mathbb{P}(X_1>x)\mathbb{P}(X_2>x)\\
&= 1-(1-F_{X_1}(x))(1-F_{X_2}(x))\\
&= F_{X_1}(x)+F_{X_2}(x)-F_{X_1}(x)F_{X_2}(x)
\end{aligned}$$

◊ **7.3.**

1. Probar que si $X\sim Geo(p)$, entonces $F_X(x)=1-(1-p)^{[x]}$.

2. Mostrar que $F_{X^*}(x)=1-(1-(p_1+p_2-p_1p_2))^{[x]}$ para todo $x\in\mathbb{R}$.

3. Concluir que $X^*\sim Geo(p^*)$, siendo $p^*=p_1+p_2-p_1p_2$. Observar que $p^*\in[0,1]$ para todo $p_1,p_2\in[0,1]$.

Es decir que en este caso pudimos "descubrir" cuál es la *distribución* de X^* a partir de identificar que su función de distribución coincide con el de una v.a. conocida.

◊ **7.4.** Considere las v.a. discretas $X_1,X_2,\dots,X_n$ i.i.d. (independientes con idéntica distribución) definidas sobre $(\Omega,\mathbb{P})$. Sean:

$$m^*=\text{mín}\{X_1,X_2,\dots,X_n\} \text{ y } M^*=\text{máx}\{X_1,X_2,\dots,X_n\}$$

1. Halle la función de distribución de m^*.

2. Halle la función de distribución de M^*.

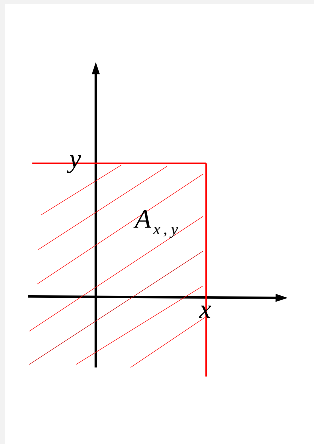
3. Halle F_{m^*} y F_{M^*} para el caso particular en que $X_1\sim Geo(p)$.

7.2. Función de distribución de vectores aleatorios discretos

Definición. Consideremos un espacio de probabilidad (discreto) $(\Omega, \mathbb{P})$ y un vector aleatorio $(X, Y) : \Omega \to \mathbb{R}$ definido sobre este espacio. Definimos la *función de distribución conjunta de* (X, Y) como $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ tal que:

$$F_{X,Y}(x, y) = \mathbb{P}(A_{x,y}) \text{ para todo } (x, y) \in \mathbb{R}^2,$$

siendo $A_{x,y}$ la región sombreada en el siguiente dibujo:



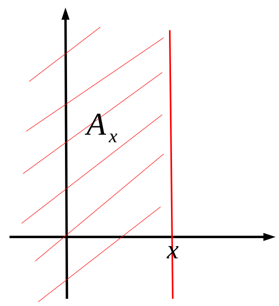
Observar que $\mathbb{P}(A_{x,y}) = \mathbb{P}(\{X \leqslant x\} \cap \{Y \leqslant y\})$.

◇ **7.5.** Considere la v.a. X que toma los valores -1, 0 y 1 con igual probabilidad y la v.a. Y tal que:

$$Y = \begin{cases} 0 \text{ si } X = 0 \\ 1 \text{ si } X \neq 0 \end{cases}$$

Halle la función de distribución conjunta del vector (X, Y).

Definición. Si (X, Y) es un vector aleatorio sobre un espacio de probabilidad (discreto) $(\Omega, \mathbb{P})$ y $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ es su función de distribución, le llamamos *distribución marginal de* X a la función $F_X : \mathbb{R} \to \mathbb{R}$ tal que $F_X(x) = \mathbb{P}(A_x)$, siendo A_x el siguiente conjunto:



Observar que $\mathbb{P}(A_x) = \mathbb{P}(\{X \leqslant x\} \cap \{Y \in \mathbb{R}\})$. Análogamente se define la función de distribución marginal de Y.

7.3. Función de distribución de v.a. no discretas

Hasta ahora hemos trabajado con v.a. en espacios de probabilidad discretos. Dijimos que si Ω es un conjunto discreto, entonces una v.a. sobre $(\Omega, \mathbb{P})$ es *cualquier* función $X : \Omega \to \mathbb{R}$.

Si Ω no es discreto, todavía no sabemos qué significa tener una medida de probabilidad $\mathbb{P}$ sobre Ω (eso es lo que comenzaremos a estudiar en el próximo capítulo). Veremos más adelante que una v.a. sobre $(\Omega, \mathbb{P})$ es una función $X : \Omega \to \mathbb{R}$ que cumple ciertas *"condiciones técnicas"* (que definiremos más adelante como *"ser una función medible"*).

Sin embargo, las definiciones que vimos antes de función de distribución de variables aleatorias y vectores aleatorios son completamente generales, en el sentido de que todo está bien definido aún cuando el espacio muestral no sea discreto. Más aún, el trabajar con la función de distribución de la v.a. X en lugar de lidear con Ω, $\mathbb{P}$ y la propia v.a. es lo más razonable ya que F_X (como la definimos antes) *aguantará bien* todas las *tecnicidades* que haya que exigirles a Ω, $\mathbb{P}$ y X.

Con esto en mente, pasaremos a estudiar v.a. sobre espacios muestrales no discretos, teniendo en cuenta que a partir de ahora las *vedette* de nuestro análisis serán las funciones de distribución (que son las que se *comportan "bien"*).

Parte II

Espacios de probabilidad continuos

Capítulo 8

Variables aleatorias continuas

Motivación, definición de v.a. continua; función densidad; formalización de probabilidad; ejemplos; función de distribución; funciones de v.a.; esperanza; varianza.

[Duración: 1 Semanas]

8.1. Motivación y definiciones

Hasta ahora hemos trabajado con espacios de probabilidad discretos, esto es, un par $(\Omega, \mathbb{P})$ donde Ω es numerable, y $\mathbb{P}$ es una medida de probabilidad (función definida sobre los subconjuntos de Ω satisfaciendo las propiedades definidas en la Sección 1.4.2).

En esta segunda parte de las notas estamos interesados en experimentos donde los resultados posibles son un conjunto "continuo"; por ejemplo: altura de las personas, peso exacto de un pote de dulce de leche (que por lo general no coincide con lo que dice el pote), lugar exacto en el que frena un auto (teniendo en cuenta que por lo general no conocemos todas las condiciones del ambiente que determinarían en lugar exacto en el que va a frenar), etcétera. Es decir que ahora Ω_X deja de ser numerable, por lo que $\Omega = X^{-1}(\Omega_X)$ tampoco podrá ser numerable ($X : \Omega \to \Omega_X \subset \mathbb{R}$ sigue siendo una función). Como mencionamos al final del capítulo anterior, esta función X tendrá que cumplir ciertas condiciones técnicas. Para hacer un estudio riguroso en este caso, es necesario enfrentar ciertos problemas técnicos que comentaremos en la Sección 8.2. Por ahora, haremos *la vista gorda* y tomaremos un punto de vista más pragmático.

Supongamos que ahora queremos estudiar la aguja de un reloj que se mueve continuamente a velocidad angular constante. Podemos parametrizar el movimiento del puntero con el espacio $\Omega = \{\omega : \omega \in [0, 2\pi]\}$.

Si queremos modelar la probabilidad de que el puntero esté en cierta región del reloj, sería razonable considerar que la probabilidad de un arco sea proporcional al tamaño del arco. Esto es

$$\mathbb{P}(\alpha \leqslant \omega \leqslant \beta) = \frac{\beta - \alpha}{2\pi},$$

para cualesquiera ángulos $0 \leqslant \alpha \leqslant \beta \leqslant 2\pi$.

Si consideramos la variable aleatoria $X : \Omega \to \mathbb{R}$ tal que X mide la posición del puntero, i.e. $X(\omega) = \omega$, entonces resulta

$$\mathbb{P}(\alpha \leqslant X \leqslant \beta) = \mathbb{P}(X \in [\alpha, \beta]) = \frac{\beta - \alpha}{2\pi} = \int_\alpha^\beta \frac{1}{2\pi} \mathrm{d}x, \quad 0 \leqslant \alpha \leqslant \beta \leqslant 2\pi.$$

En particular, $\mathbb{P}(X = \alpha) = 0$. Luego, en términos de la función de distribución de X, F_X, que definimos en el capítulo anterior para cualquier v.a., tenemos que:

$$\mathbb{P}(\alpha \leqslant X \leqslant \beta) = \mathbb{P}(\alpha < X \leqslant \beta) = F_X(\beta) - F_X(\alpha) = \int_\alpha^\beta \frac{1}{2\pi} \mathrm{d}x$$

Observar que si consideramos la función $f : \mathbb{R} \to \mathbb{R}$ tal que $f(x) = \frac{1}{2\pi}$ si $x \in [0, 2\pi]$ y vale 0 en otro caso, entonces:

$$F_X(b) - F_X(a) = \int_a^b f(x) \mathrm{d}x.$$

Este ejemplo motiva la siguiente definición.

Definición. Diremos que una v.a. $X : \Omega \to \mathbb{R}$ es una *variable aleatoria continua (absolutamente continua)* si su función de distribución $F_X : \mathbb{R} \to \mathbb{R}$ verifica que existe una función $f : \mathbb{R} \to \mathbb{R}$ tal que:

1. $f(t) \geqslant 0$ para todo $t \in \mathbb{R}$;
2. $\displaystyle\int_{-\infty}^{\infty} f(t)\, \mathrm{d}t = 1$;
3. $F_X(x) = \displaystyle\int_{-\infty}^{x} f(t)\, \mathrm{d}t$ para todo $x \in \mathbb{R}$

En este caso, a la función $f = f_X$ se le denomina *función densidad de probabilidad* (o *densidad*) de la v.a. X.

En el ejemplo anterior la densidad de X está dada por la función

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & 0 \leqslant x \leqslant 2\pi \\ 0 & \text{en otro caso} \end{cases}$$

Cuando la f_X de una v.a. continua X es como la de antes (constante en un intervalo y 0 fuera de ese intarvalo) se dice que X tiene *distribución uniforme* en ese intervalo (en este caso, en $[0, 2\pi]$).

Intuitivamente, si tomamos varias muestras de la v.a. continua X (llamémosle $x_1, x_2, \ldots, x_n$ a estas muestras), y miramos la proporción de realizaciones que caen en un intervalo $[x, x+\delta x]$, resulta que:

$$\frac{\#\{i : x_i \in [x, x+\delta]\}}{n} \approx \mathbb{P}(x < X \leqslant x+\delta x) = \int_x^{x+\delta x} f(t)\,\mathrm{d}t \approx f(x)\delta x,$$

es decir que la "probabilidad infinitesimal" es $f(x)\delta x$. En este momento es cuando uno se ve tentado a identificar a la función de densidad $f(x)$ con la función de probabilidad puntual de una v.a. discreta. Lo peligroso de esta identificación es que en realidad $f(x)$ no tiene por qué ser una probabilidad (observar que no pedimos que $f(x) \in [0,1]$, por lo que $f(x)$ puede ser mayor que 1). Es decir que la verdadera analogía no es entre $f(x)$ y $p_X(x)$ sino entre $f(x)\delta x$ y $p_X(x)$.

◊ **8.1.** Probar que si X es una v.a. continua sobre $(\Omega, \mathbb{P})$, entonces su función de distribución X verifica las cuatro condiciones que mencionamos en el capítulo anterior que tenía que verificar una función de distribución, es decir:

1. $F_X(x) \in [0,1]$ para todo $x \in \mathbb{R}$;
2. F_X es una función creciente;
3. F_X es una función càdlàg (más aún, es continua);
4. $\lim\limits_{x\to-\infty} F_X(x) = 0$ y $\lim\limits_{x\to+\infty} F_X(x) = 1$.

Probar que además verifica que $\mathbb{P}(X = x) = 0$ para todo $x \in \mathbb{R}$ (i.e. no tiene sentido hablar de probabilidad puntual). Luego,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leqslant X < b) = \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(t)\mathrm{d}t$$

Sugerencia: Considere una sucesión $\{\varepsilon_n\}_n$ tal que $\varepsilon_n \searrow 0$ y escriba el suceso $\{X = x\}$ como:

$$\{X = x\} = \bigcap_{n=0}^{\infty} \{x - \varepsilon_n < X \leqslant x\}$$

y utilizar la *continuidad de la función de probabilidad* $\mathbb{P}$ que probamos en el **capítulo 4** (ejercicio **4.20**). Observar que en este caso la sucesión de sucesos $A_n = \{x - \varepsilon_n < X \leqslant x\}$ es decreciente.

◊ **8.2.** Probar que si F_1 y F_2 son funciones de distribución (i.e. verifican las condiciones 1. al 4. del ejercicio anterior), entonces $G:\mathbb{R}\to\mathbb{R}$ tal que $G(x)=\alpha F_1(x)+(1-\alpha)F_2(x)$ es también una función de distribución, para todo $\alpha\in[0,1]$

◊ **8.3.** Sea $F:\mathbb{R}\to\mathbb{R}$ tal que:

$$F(x)=c\int_{-\infty}^{x}e^{-|t|}\mathrm{d}t \text{ para todo } x\in\mathbb{R}.$$

1. ¿Cuánto tiene que valer c para que F sea una función de distribución?
2. ¿Se corresponde F con una v.a. continua)?

8.2. Un problema técnico*

(Esta sección puede ser omitida en una primera lectura.)

Hemos pasado por arriba un problema técnico en nuestro estudio de variables aleatorias continuas que comentaremos a continuación.

Para simplificar, tomaremos como referencia el caso de X con distribución uniforme en $[0,1]$.

Con la definición de probabilidad dada anteriormente, si $A\subset[0,1]$ es un cierto "evento" se tiene que

$$\mathbb{P}(X\in A)=\mathrm{long}(A),$$

donde por $\mathrm{long}(A)$ entendemos la longitud del conjunto A. Si A es un intervalo, la longitud está bien definida. Pero, ¿qué podemos decir de la longitud para un conjunto arbitrario?

Es un problema difícil y técnico describir cuáles conjuntos tienen definida una medida de probabilidad en este sentido. Una descripción y solución completa de este problema se ve en el curso de Teoría de la Medida donde se estudia la medida de Lebesgue.

Pero para los estudiantes interesados, acá les dejamos un "abre boca" sobre este problema.

En la Parte I de las notas trabajamos con espacios muestrales numerables, $\Omega=\{\omega_1,\omega_2,\ldots\}$, donde para cada $\omega_i\in\Omega$ podíamos definir una función de probabilidad puntual $p_i=p(\omega_i)\geqslant 0$, tal que $\sum_i p_i=1$. De esta manera construímos un espacio de probabilidad discreto $(\Omega,\mathbb{P})$: donde $\mathbb{P}(E):=\sum_{i:\,\omega_i\in E}p(\omega_i)$, para cualquier evento $E\subset\Omega$, define una *medida de probabilidad* que satisface las siguientes propiedades:

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

(ii) $\mathbb{P}(E)\geqslant 0$, para todo $E\subset\Omega$;

(iii) $\mathbb{P}(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$, para eventos disjuntos $\{E_i\}_{i=1,2,\dots}$, $E_i \subset \Omega$.

Ahora, si Ω es un conjunto no-numerable, el enfoque tomado para el caso discreto no es viable dado que la probabilidad puntual debería ser cero salvo para una cantidad numerable de ω's en Ω (de lo contrario existirían conjuntos con medidas de probabilidad infinito). (En particular, observar que la probabilidad uniforme en $[0,1]$ no puede obtenerse de esta manera.)

Esto motiva que para definir una medida de probabilidad sobre Ω es necesario tomar un enfoque distinto. Un enfoque natural es precisamente el que tomamos más arriba en nuestra construcción de la probabilidad uniforme. Esto es, definir la probabilidad para ciertos conjuntos básicos, para luego extenderla a un sistema de conjuntos más complejos donde las propiedades (i), (ii) y (iii) sigan valiendo.

Por ejemplo, en el caso $\Omega = [0,1]$ podríamos tomar la colección de todos los intervalos (para los cuales la noción de longitud está definida), sus complementos, uniones numerables y sus complementos, y así sucesivamente. Para todos ellos podríamos definir la "longitud" de esos conjuntos. La colección construida por este proceso se denomina *σ-álgebra de Borel*. Una pregunta natural es saber si con esta construcción podemos llegar a todos los subconjuntos de $[0,1]$, y en tal caso poder definir su longitud.

La respuesta a esta pregunta es no. Una idea de la prueba es que podemos escribir el intervalo $[0,1]$ como unión disjunta de conjuntos que son el trasladado de uno fijo (usando cierta relación de equivalencia y el axioma de elección). Por lo tanto, los trasladados miden lo mismo y la condición (i) y (iii) no son compatibles.

La colección $\mathcal{F}$ de conjuntos que resultan de la construcción mencionada más arriba, a partir de piezas básicas, verifican las siguientes propiedades:

1. $\varnothing, \Omega$ pertenecen a $\mathcal{F}$;

2. cerrada por complementos: $E \in \mathcal{F}$ entonces $E^c \in \mathcal{F}$;

3. cerrada por uniones numerables: $\{E_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ entonces $\bigcup_{i\in\mathbb{N}} E_i \in \mathcal{F}$.

Una colección que verifica las propiedades 1., 2. y 3. se dice una σ-álgebra.

Una σ-álgebra puede ser muy sencilla, como el conjunto $\mathcal{F} = \{\Omega, \varnothing, E, E^c\}$ para cierto conjunto $E \subset \Omega$, o puede ser una colección muy rica de conjuntos como la σ-álgebra de Borel de $[0,1]$ mencionada más arriba (que es no-numerable).

8.2.1. Axiomas de Probabilidad

Esta discusión motiva las siguientes definiciones que se enmarcan en la axiomatización de la teoría de Probabilidad que realizó A. Kolmogorov en la década de 1930.

Un espacio de probabilidad es una terna $(\Omega,\mathcal{F},\mathbb{P})$ donde Ω es un conjunto arbitrario, $\mathcal{F}$ es una σ-álgbera de subconjuntos de Ω, y $\mathbb{P}$ es una medida de probabilidad sobre $\mathcal{F}$: esto es, una función $\mathbb{P}:\mathcal{F}\to[0,1]$ que satisface las propiedades (i), (ii), y (iii) dadas anteriormente.

Todo los espacios de probabilidad discretos vistos hasta el momento pueden definirse de esta manera, simplemente tomando $\mathcal{F}$ como las partes de Ω, y $\mathbb{P}$ inducido por una función de probabilidad puntual.

Para el caso $\Omega=[0,1]$, podemos tomar como $\mathcal{F}$ la σ-álgebra de Borel que es la "generada" por los intervalos. En particular, y evitando tecnicismos, $\mathcal{F}$ contiene al menos todos los intervalos abiertos, cerrados, semiabiartos, conjuntos numerables, uniones e intersecciones numerables y complementos de todos estos.[1] La medida de probabilidad $\mathbb{P}$ existe y extiende a la longitud en los casos donde es claro definirlo. (La extensión a toda la σ-álgebra se denomina *medida de Lebesgue* en $[0,1]$.

8.2.2. Formalización de variables aleatorias

Ahora que tenemos definido con rigor nuestro espacio de probabilidad, podemos definir correctamente las variables aleatorias (indpendientemente si el espacio es discreto, o no).

Definición. Sea $(\Omega,\mathcal{F},\mathbb{P})$ un espacio de probabilidad. Una variable aleatoria X es una función $X:\Omega\to\mathbb{R}$ tal que $X^{-1}((-\infty,x])\in\mathcal{F}$, para todo $x\in\mathbb{R}$ (a esta condición le llamaremos ser una *función medible*).

Con esta definición, está perfectamente definida la función de distribución de X, a saber,

$$F_X(x)=\mathbb{P}(X^{-1}((-\infty,x]))$$

para *cualquier* v.a. (no solo las discretas y continuas, que son las que conocemos hasta ahora).

8.3. Ejemplos de v.a. continuas

Presentamos a continuación algunas v.a. continuas "con nombre".

Ejemplo 8.3.1 (*Uniformes*)**.** Diremos que X es una v.a. uniforme $[a,b]$ ($X\sim U[a,b]$) si:

$$f_X(x)=\begin{cases}\frac{1}{b-a} \text{ si } x\in[a,b]\\ 0 \text{ en otro caso}\end{cases}$$

$\diamondsuit$ **8.4.** Probar que en este caso $F_X=\begin{cases}0 \text{ si } x<a\\ \frac{x-a}{b-a} \text{ si } a\leqslant x\leqslant b\\ 1 \text{ si } x>b\end{cases}$

[1]La σ-álgebra de Borel contiene mucho más conjuntos, pero para nuestros fines alcanza con conocer sólo algunos de los conjuntos de la σ-álgebra.

Ejemplo 8.3.2 (*Exponencial*)**.** Diremos que X es una v.a. exponencial de parámetro ("tasa") λ ($X \sim Exp(\lambda)$) si:

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

◊ **8.5.** Probar que $F_X(x) = \begin{cases} 1 - e^{-\lambda x} \text{ si } x \geqslant 0 \\ 0 \text{ si } x < 0 \end{cases}$.

◊ **8.6.** Considere la v.a. discreta T que definiremos a partir de la v.a. X como:

$$T := [X],$$

donde $[X]$ indica la parte entera de X. Probar que T es una v.a. geométrica y calcular el parámetro p. Es decir que la v.a. exponencial es la *versión continua* de la v.a. geométrica. En general, se la usa para modelar el *tiempo de espera HASTA que ocurra algo*.

◊ **8.7.** Probar que si $X \sim Exp(\lambda_1)$ e $Y \sim Exp(\lambda_2)$ son independientes, entonces:

$$m^* = \text{mín}\{X, Y\}$$

es también una v.a. exponencial. Determinar cuál es la tasa. (Recordar que en **capítulo 7** probamos que el mínimo de v.a. geométricas independientes es también una v.a. geométrica).

Aplicación: Si el tiempo de vida de mi televisor se puede modelar como una v.a. X_{tv} exponencial de parámetro λ_{tv}, el tiempo de vida de mi heladera como una v.a. $X_h \sim Exp(\lambda_h)$ y el del microondas como una $X_m \sim Exp(\lambda_m)$, ¿cuál es la probabilidad de que tenga que cambiar alguno de esos electrodomésticos en los próximos dos años?

Ejemplo 8.3.3 (*Doble exponencial*)**.** Es X tal que:

$$f_X(x) = \frac{1}{2}e^{-|x|} \text{ para todo } x \in \mathbb{R}$$

◊ **8.8.** Hallar F_X

Ejemplo 8.3.4 (*Normal o gaussiana*)**.** Diremos que la v.a. X tiene distribución normal de parámetros (μ, σ) ($X \sim N(\mu, \sigma^2)$) si:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \text{ para todo } x \in \mathbb{R}$$

Al parámetro μ le llamaremos *"la media"* y a σ le llamaremos *"el desvío"* (más adelante veremos por qué). Cuando $\mu = 0$ y $\sigma = 1$, se dice que X es una v.a. *normal estándar* (en general, se nota $Z \sim N(0,1)$).

♢ **8.9.** Probar que si $X \sim N(\mu, \sigma^2)$, entonces:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \text{ para todo } x \in \mathbb{R}$$

siendo $\Phi(x) = F_Z(x)$ (es la notación habitual).

Observar que, puesto que la función e^{-x^2} no tiene primitiva, no es posible determinar escribir explícitamente a la función Φ (sin usar la notación con la integral). Es por esto que lo habitual, si no se usa una computadora, es considerar aproximaciones para $\Phi(x)$ a partir de aproximaciones de $\Phi(x)$ para algunos valores de x. Esto es lo que se presenta habitualmente en lo que se le llama la *tabla de distribución normal (estándar)*, que presentamos en la próxima página.

♢ **8.10.** Probar que la función de distribución Φ verifica:

$$\Phi(-x) = 1 - \Phi(x) \text{ para todo } x \in \mathbb{R}$$

Por lo cual solo precisamos conocer los valores de $\Phi(x)$ para todo $x \geqslant 0$.

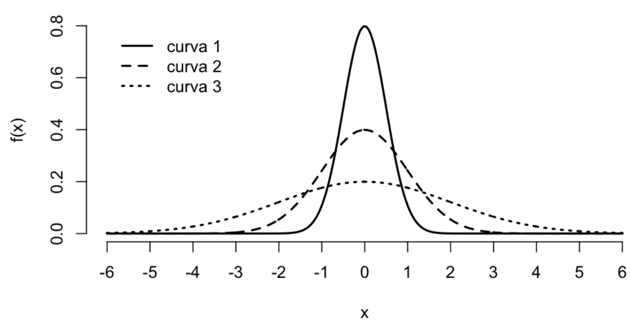
♢ **8.11.** 1. Sea $Z \sim N(0,1)$. Mostrar que:

$$\mathbb{P}(Z \in [-1,1]) \simeq 0{,}68, \quad \mathbb{P}(Z \in [-2,2]) \simeq 0{,}95 \quad \text{y} \quad \mathbb{P}(Z \in [-3,3]) \simeq 0{,}997.$$

2. Sea $X \sim N(\mu, \sigma^2)$. Calcular las siguientes probabilidades:

$$\mathbb{P}(X \in [\mu-\sigma, \mu+\sigma]), \quad \mathbb{P}(X \in [\mu-2\sigma, \mu+2\sigma]) \quad \text{y} \quad \mathbb{P}(X \in [\mu-3\sigma, \mu+3\sigma]).$$

3. La siguiente figura muestra la densidad de tres variables aleatorias normales centradas ($\mu = 0$). Indicar aproximadamente cuánto vale σ en cada caso.



♢ **8.12.** En una población se definen las siguientes categorías en base al índice de Presión Sistólica en Sangre (PSS):

- Normal si PSS $\leqslant 120$,
- Pre-hipertenso si $120 < \text{PSS} \leqslant 140$,
- Presión alta si PSS > 140.

Se asume que la PSS sigue una distribución normal con parámetros $\mu = 125$ y $\sigma^2 = 144$.

1. Hallar la probabilidad de que un individuo pertenezca a cada uno de los grupos definidos.
2. ¿Cuál es la probabilidad de no tener presión alta?
3. Hallar el intervalo centrado en la media que contiene al 68 % de la población. Idem para el 95 % de la población.
4. Hallar x_1 y x_2 tal que $\mathbb{P}(\text{PSS} \leqslant x_1) = 0{,}25$ y $\mathbb{P}(\text{PSS} > x_2) = 0{,}25$.
5. Se sabe que la probabilidad de sufrir un infarto para un individuo de la categoría Normal es de 0.15, mientras que dicha probabilidad se eleva a 0.55 dentro de la categoría Presión alta. Finalmente, la probabilidad de infarto de un individuo pre-hipertenso es de 0.25. Se pide:

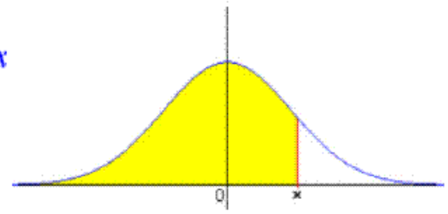
    a) Calcular la probabilidad de que un individuo elegido al azar de la población sufra un infarto.

    b) Dado que un individuo sufre un infarto, ¿Cuál es la probabilidad de que pertenezca a la categoría Normal?

TABLA DE DISTRIBUCIÓN

NORMAL

$$F(x)=P(X \leq x)=\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$



	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0,1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0,2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0,3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0,4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0,5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0,6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0,7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0,8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0,9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1,0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1,1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1,2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1,3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1,4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1,5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1,6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1,7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1,8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1,9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2,0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2,1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2,2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2,3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2,4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2,5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2,6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2,7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2,8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2,9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3,0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Ejemplo 8.3.5 (*Cauchy*)**.** Diremos que la v.a. X tiene distribución Cauchy o de Lorentz de parámetros M y γ ($X \sim C(M,\gamma)$) si su función de densidad es:

$$f_X(x) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-M}{\gamma}\right)^2\right)} \text{ para todo } x \in \mathbb{R}$$

Cuando $M = 0$ y $\gamma = 1$, se dice que X tiene distribución Cauchy estándar.

8.4. Funciones de v.a. absolutamente continuas

Sea X una v.a. en el espacio muestral $(\Omega,\mathcal{F},\mathbb{P})$ y $g : \mathbb{R} \to \mathbb{R}$ una función. En esta sección estudiaremos la función que se obtiene al componer la v.a. X con la función g. Sea $Y = g(X) : \Omega \to \mathbb{R}$. Observar que, en principio, Y no tiene por qué ser una v.a. (teniendo en cuenta los comentarios técnicos de la Sección 8.2) y menos una v.a. continua.

En esta sección supondremos que la función g es tal que Y es una v.a. (¿qué tendría que verificar g para que esto se cumpla?), pero no tiene por qué ser una v.a. absolutamente continua.

Ejemplo 8.4.1. Supongamos que X es una v.a. continua en $(\Omega,\mathcal{F},\mathbb{P})$ con densidad f_X y $g : \mathbb{R} \to \mathbb{R}$ tal que $g(x) = ax + b$ con $a > 0$. ¿Cuál es la función de distribución de la v.a. Y?

Puesto que:

$$F_Y(x) = \mathbb{P}(aX + b \leqslant x) = \mathbb{P}(X \leqslant \frac{x-b}{a}) = F_X\left(\frac{x-b}{a}\right)$$

y F_X, tenemos que F_Y es derivable. i.e. Y es una v.a. continua con función de densidad:

$$f_Y(x) = \frac{d}{dx}F_Y(x) = f_X\left(\frac{x-b}{a}\right)\frac{1}{a}$$

Observar que en el ejemplo anterior era necesario que $a > 0$ para que los sucesos $\{aX + b \leqslant x\}$ y $\{X \leqslant \frac{x-b}{a}\}$ sean iguales (se mantenga la desigualdad).

◊ **8.13.** ¿ Cómo quedan F_Y y f_Y si $a < 0$?

Es posible extender estos resultados para los casos en que la función g es estrictamente creciente (o decreciente). Esto es lo que enunciamos a continuación.

Teorema 8.1. *Sea X una v.a. continua con función de densidad f_X y $g:\mathbb{R}\to\mathbb{R}$ una función estrictamente creciente y diferenciable, entonces $Y = g(X)$ es una v.a. continua con función de densidad:*

$$f_Y(x) = f_X\left(g^{-1}(x)\right)\frac{d}{dx}(g^{-1}(x))$$

para todo $x\in\mathbb{R}$.

◊ **8.14.** Sorteando la prueba de que Y es una v.a., probar el teorema anterior.

◊ **8.15.** Enunciar y probar un teorema similar al anterior para el caso en que g es diferenciable y estrictamente decreciente.

◊ **8.16.** Sea $X \sim N(\mu,\sigma^2)$:

1. Hallar la función de distibución de $Y = X^2$ y su densidad.

2. Hallar la función de distribución de $Z = |X|$ y su densidad.

◊ **8.17.** Sea $X = \log(M) \sim \mathcal{N}(\mu,\sigma^2)$. Hallar la distribución de M y su densidad

◊ **8.18.** Sea $X \sim \text{Exp}(\lambda)$

1. Hallar la función de distibución de $Y = aX$ ($a > 0$)y su densidad.

2. Hallar la función de distribución de $Z = ae^X$ ($a > 0$) y su densidad. Esta distribución se llama Pareto de parámetros a y λ.

3. Hallar la función de distribución de $Y = \frac{1}{1+X}$ y su densidad.

◊ **8.19. Aplicación: simulación de variables aleatorias**

Sea F una función de distribución y $U \sim \mathcal{U}[0,1]$. Probar que la función de distribución de $X = F^{-1}(U)$ es F, esto es $F_X = F$.
Observación: Puesto que no estamos pidiendo que F sea estrictamente creciente, no tendría por qué ser invertible. En este caso, F^{-1} se refiere a la *inversa generalizada de F*, que se define así:

$$F^{-1}(y) = \text{ínf}\{x\in\mathbb{R} \text{ tal que } F(x) = y\}$$

para todo $y\in Im(F)$. (En nuestro caso, como F es una distribución, F^{-1} está definida en $[0,1]$).

8.5. Esperanza y Varianza de v.a. continuas

Motivados por la definición de esperanza en el caso de las v.a. discretas y por la similitud que existe entre $p_X(x)$ y $f_X(x)\mathrm{d}x$, definiremos la *esperanza o valor esperado* de una v.a. discreta así:

Definición. Sea X una v.a. continua definida en $(\Omega, \mathcal{F}, \mathbb{P})$ con función de densidad f_X. Definimos la *esperanza* de X como:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x,$$

si esta integral converge absolutamente (i.e. $\int_{-\infty}^{+\infty} |x| f_X(x)\mathrm{d}x < \infty$)

◇ **8.20.** Calcular la esperanza de las v.a. absolutamente continuas que definimos en la Sección 8.3 de una v.a. uniforme, exponencial, normal y Cauchy.

◇ **8.21.** Probar que si la v.a. continua X solo toma valores positivos (i.e. $X \geqslant 0$), entonces:

$$\mathbb{E}(X) = \int_{0}^{+\infty} (1 - F_X(x))\mathrm{d}x$$

8.6. Esperanza de funciones de v.a. continuas

Puesto que en la sección anterior definimos v.a. nuevas a partir de una v.a. continua y una función $g : \mathbb{R} \to \mathbb{R}$, resulta natural preguntarse cómo calcular la esperanza de una v.a. $Y = g(X)$.

Opción 1: Si Y es continua

Si $g : \mathbb{R} \to \mathbb{R}$ es diferenciable y estrictamente creciente, vimos que la v.a. Y es continua con función de densidad:

$$f_Y(x) = f_X\left(g^{-1}(x)\right) \frac{d}{dx}(g^{-1}(x)).$$

Luego, para calcular $\mathbb{E}(Y)$ usamos la definición que vimos antes:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_Y(x)\mathrm{d}x,$$

◊ **8.22.** Probar que en este caso:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x = \int_{-\infty}^{+\infty} g(x) f_X(x)\mathrm{d}x,$$

Se puede probar análogamente que si g es diferenciable y estríctamente decreciente, entonces:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} g(x) f_X(x)\mathrm{d}x,$$

Opción 2: si Y no es una v.a. continua

El resultado anterior es cierto aún cuando Y no sea una v.a. continua. Lo presentamos como un teorema, cuya demostración omitiremos (por ser larga, no por ser difícil).

Teorema 8.2. *Si X es una v.a. continua con densidad f_X y $g : \mathbb{R} \to \mathbb{R}$ es tal que $Y = g(X)$ es una v.a., entonces:*

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} g(x) f_X(x)dx,$$

(recordar que en el caso de v.a. discretas habíamos probado que:

$$\mathbb{E}(X) = \sum_x g(x) p_X(x)$$

).

◊ **8.23.** Si X es una v.a. con distribución normal estándar, calcular la esperanza de la v.a. $Y = e^{2X}$.

8.6.1. Varianza de v.a. continuas

En particular, si tomamos la función $g(x) = (x-\mu)^2$, siendo $\mu = \mathbb{E}(X)$, tenemos que la *varianza* de la v.a. X queda bien definida:

Definición. Si X es una v.a. continua, definimos la *varianza* de X como:

$$var(X) = \mathbb{E}\left(X - \mathbb{E}(X)\right)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\mathrm{d}x,$$

si esta integral converge absolutamente.

◊ **8.24.** Probar que se sigue verificando que $var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

◊ **8.25.** Calcular la varianza (si existe) de las v.a. uniformes, exponenciales, normales y Cauchy.

Capítulo 9

Vectores aleatorias continuos

Vectores aleatorios, distribución conjunta, marginales, independencia, densidad conjunta, aplicaciones geométricas, transformaciones, densidad condicional, esperanza, esperanza condicional.

[Duración: 1 Semanas]

9.1. Distribución conjunta y marginales

Al igual que como hicimos en el caso discreto, para poder entender varias variables aleatorias sobre un mismo espacio se define la noción de vector aleatorio.

En este capítulo omitiremos algunas pruebas dado que tienen una mecánica similar a la del capítulo de vectores aleatorios discretos. Haremos mención de los detalles de alguna prueba en el caso que sea necesario.

Definición. Sean X e Y v.a. sobre un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Consideramos el *vector aleatorio* $(X, Y) : \Omega \to \mathbb{R}^2$. La *función de distribución conjunta* del par X, Y, es la función $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ dada por

$$F_{X,Y}(x,y) := \mathbb{P}(X \leqslant x, Y \leqslant y).$$

Esta función verifica propiedades análogas a las propiedades de las funciones de distribución dadas en Capítulo 7. Por ejemplo

- $\lim_{x,y\to-\infty} F(x,y) = 0$;
- $\lim_{x,y\to+\infty} F(x,y) = 1$;
- $F(x_1,y_1) \leqslant F(x_2,y_2)$, si $x_1 \leqslant x_2$ y $y_1 \leqslant y_2$.

Las distribuciones marginales se pueden obtener de manera análoga a lo ya visto. Por ejemplo, para encontrar la función de distribución de X basta con hallar $\mathbb{P}(X \leqslant x) = \mathbb{P}(X \leqslant x, Y \in \mathbb{R})$. Esto motiva motiva la siguiente definición.

Definición. Dado v.a. X e Y con distribución conjunta $F_{X,Y}$, las *distribuciónes marginales* de X e Y respectivas, están dadas por

$$F_X(x) := \lim_{y\to+\infty} F_{X,Y}(x,y), \qquad F_Y(y) := \lim_{x\to+\infty} F_{X,Y}(x,y).$$

9.2. Independencia

Hemos visto la definición de sucesos independientes, y de v.a. discretas independientes. La siguiente definición de independencia abarca la definición de v.a. discretas también.

Definición. Diremos que las v.a. X e Y, definidas sobre el mismo espacio de probabilidad, son *independientes* si los eventos $\{X \leqslant x\}$ e $\{Y \leqslant y\}$ son independientes.

Utilizando las distribuciones marginales podemos concluir lo siguiente.

Teorema 9.1. *Las v.a. X e Y son independientes si y sólo si $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, para todo $x, y \in \mathbb{R}$.*

◊ **9.1.** Probar el teorema anterior.

Comentario 8. Todo lo anterior puede ser generalizado a una cantidad finita de variables aleatorias $X_1,\dots,X_n$ definidas sobre un mismo espacio de probabilidad. Omitiremos esta extensión natural y lo dejamos a cargo del lector.

Comentario 9. De ahora en adelante, siempre que hablemos de distribución conjunta de variables aleatorias damos por sobreentendido que las variables viven en un mismo espacio de probabilidad.

◊ **9.2.** Sea X e Y v.a. con distribución conjunta:

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{si } x, y \geqslant 0 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar distribuciones marginales. ¿Son X e Y independientes?

9.3. Densidad conjunta, marginal, e independencia

Generalizando la noción de densidad de un v.a. continua, definimos lo siguiente.

Definición. Diremos que las v.a. X e Y tienen *distribución conjunta continua (o absolutamente continuo)* si se verifica que su función de distribución conjunta satisface

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt\, ds, \qquad x, y \in \mathbb{R},$$

para cierta función $f : \mathbb{R}^2 \to [0, +\infty)$. En este caso, la función f se denomina *función de densidad conjunta*, y la denotamos por $f_{X,Y}$.

Estamos omitiendo qué regularidad debe tener la función f (al menos debe ser integrable), pero para nuestros fines alcanza con que sea suficientemente regular como para que las siguientes propiedades se verifiquen.

- (Intercambio de orden)

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} ds \int_{-\infty}^{y} f(s,t)\, dt = \int_{-\infty}^{y} dt \int_{-\infty}^{x} f(s,t)\, ds$$

- (Diferenciación)

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y),$$

  en los puntos donde $F_{X,Y}$ es diferenciable.

En el curso de Teoría de la Medida se estudiará cuán regular debe ser f para que las propiedades anteriores sean válidas. (En este curso podemos suponer que f es continua salvo en un conjunto de puntos de “medida cero” en $\mathbb{R}^2$, como puntos, rectas, o curvas planares.)

◊ **9.3.** Si X e Y tienen distribución conjunta continua, entonces $f_{X,Y} \geqslant 0$, y $\int\int_{\mathbb{R}^2} f_{X,Y}(x,y)\, dx\, dy = 1$.

Recíprocamente, si f es una función que verifica las dos propiedades de ◇ 9.3, entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y un vector aleatorio definido sobre este espacio de manera tal que f sea la función de densidad conjunta. Para los estudiantes interesados, les dejamos un desafío a continuación.

◇ **9.4.** * Dar una prueba del comentario anterior? (Sugerencia: recordar el Teorema 5.1.)

Al igual que discutimos en el capítulo anterior, la función densidad conjunta da una medida de probabilidad infinitesimal, esto es, la probabilidad de que el vector aleatorio tome valores en el cuadrado infinitesimal $[x, x+\delta x] \times [y, y+\delta y]$ satisface

$$\mathbb{P}(x < X \leqslant x+\delta x, y < Y \leqslant y+\delta y) \approx f_{X,Y}(x,y)\,\delta x\,\delta y. \tag{9.1}$$

Sumando estos aportes infinitesimales sobre un conjunto “regular” $A \subset \mathbb{R}^2$, podemos concluir el siguiente resultado.

Teorema 9.2. *Sea $A \subset \mathbb{R}^2$ que puede ser “aproximado” por uniones de rectángulos. Entonces*

$$\mathbb{P}((X,Y) \in A) = \int\int_{(x,y)\in A} f_{X,Y}(x,y)\,dx\,dy.$$

Para probar este teorema basta aproximar el conjunto A por una unión disjunta de cuadraditos (por defecto y por exceso) para luego sumar las probabilidades de cada cuadradito. Luego se observa que la suma anterior es una suma de Riemann de la función $f_{X,Y}$.

Ejemplo 9.3.1. (Distribución uniforme) Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función indicatriz del cuadrado unidad, y sea (X,Y) un vector aleatorio con f como función densidad. Entonces decimos que el vector aleatorio tiene *distribución uniforme* sobre $[0,1]^2$. Si A es un subconjunto “regular” del plano, entonces resulta $\mathbb{P}((X,Y) \in A) = \text{área}(A \cap [0,1]^2)$.

◇ **9.5.** La distribución uniforme sobre un rectángulo de lados a y b es la distribución que tiene como densidad una constante sobre el rectángulo, y cero en el complemento. ¿Cuál es esa consante?

◇ **9.6.** Probar que si X e Y son v.a. con densidad conjunta $f_{X,Y}$, entonces, X e Y son v.a. continuas con funciones densidad igual a

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy, \qquad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx.$$

Las funciones anteriores se denominan *densidades marginales* de las v.a. X e Y respectivamente.

$\diamondsuit$ **9.7.** Probar que si X e Y son v.a. con densidad conjunta $f_{X,Y}$, entonces se tiene que X es independiente a Y si y sólo si $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, donde f_X y f_Y son las densidades marginales.

Más en general vale el siguiente resultado (útil), que es el análogo al Teorema 6.2 para el caso continuo.

Teorema 9.3. *Las v.a. X e Y con distribución conjunta continua son independientes si y sólo si existen funciones $g, h : \mathbb{R} \to \mathbb{R}$ tales que la densidad conjunta $f_{X,Y}$ se factoriza*

$$f_{X,Y}(x,y) = g(x)h(y), \quad x, y \in \mathbb{R}.$$

La prueba del resultado anterior es análoga a la prueba del Teorema 6.2, cambiando el símbolo de suma por el símbolo de integral.

Ejemplo 9.3.2 (Distribución Gaussiana en $\mathbb{R}^2$)**.** La *distribución normal estándar* (o gaussiana) en el plano tiene densidad conjunta

$$\gamma_2(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad x, y \in \mathbb{R}.$$

$\diamondsuit$ **9.8.** Probar que efectivamente γ_2 define una densidad en el plano. Hallar densidades marginales, y concluir que la distribución normal se obtiene de tomar la primera y segunda coordenada con distribución normal estándar independientes.

9.4. Aplicaciones geométricas

En esta sección veremos dos aplicaciones al cálculo de probabilidades geométricas.

$\diamondsuit$ **9.9** (*Aguja de Buffon*)**.** Supongamos que tiramos un aguja de longitud ℓ sobre la hoja de un cuaderno a rayas. ¿Cuál es la probabilidad que la aguja tirada al azar intersecte alguna de las lineas? La respuesta dependerá de los parámetros dado por la longitud de la aguja, y por la distancia entre lineas. Obviamente es necesario imponer hipótesis sobre el modelo de aleatoriedad a elegir.

$\diamondsuit$ **9.10.** Supongamos que tenemos una varilla longitud 1, y la partimos en los puntos con coordenadas X e Y tomados de la distribución uniforme en $[0,1]^2$. ¿Cuál es la probabilidad de que los tres partes de la varilla formen un triángulo?

9.5. Transformaciones de vectores aleatorios

En muchos casos queremos estudiar la distribución de un vector aleatorio que resulta de aplicar una transformación al plano. Por ejemplo, si (X,Y) es un vector aleatorio con distribución gaussiana estándar. ¿Cuál es la distribución del vector aleatorio (R,Θ) donde $R = \sqrt{X^2+Y^2}$ y $\Theta = \mathrm{Arg}(X,Y)$, donde $\mathrm{Arg}(\cdot)$ es la función argumento principal definida en $\mathbb{R}^2 \setminus \{0\}$?

Teorema 9.4. *Sean X e Y v.a. con densidad conjunta $f_{X,Y}$, y sea $S = \{(x,y) \in \mathbb{R}^2 : f_{X,Y}(x,y) > 0\}$. Si $\varphi : S \to S'$, $S' \subset \mathbb{R}^2$ es un difeomorfismo, entonces el vector aleatorio $(U,V) = \varphi(X,Y)$ tiene distribución continua y su densidad conjunta*

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v),y(u,v))\,|\det(D\varphi^{-1}(u,v)|, \quad (u,v) \in S',$$

y toma el valor 0 en otro caso.

Idea de la prueba: Si para $B \subset S'$ se prueba que

$$\mathbb{P}((U,V) \in B) = \int\int_B g(u,v)\,du\,dv,$$

para cierta función g, entonces, resulta que (U,V) es un vector aleatorio continuo con densidad conjunta $f_{U,V} := g$. Para obtener una fórmula como la anterior se sugiere escribir el primer miembro en función de los valores de (X,Y). $\square$

◇ **9.11.** Terminar los detalles en la demostración anterior.

◇ **9.12.** Hallar la densidad conjunta del vector coordenadas polares de un vector gaussiano estándar en $\mathbb{R}^2$. ¿Qué se puede decir sobre la dependencia entre la v.a. radial y la v.a. angular?

En muchas situaciones no tenemos definido un mapa biyectivo de manera natural, pero sí un mapa k a 1: esto es un mapa que la cantidad de preimágenes es k. Por ejemplo, la función $x \mapsto x^2$ es un mapa 2 a 1 para cada real positivo x^2. Aunque las hipótesis del teorema se pide que el mapa de cambio de variable sea un difeomorfismo sobre S, el teorema se puede extender a mapas diferenciables k a 1.

◇ **9.13.** Sean (X,Y) un vector gaussiano estándar en $\mathbb{R}^2$. Considere el nuevo vector aleatorio (Z,W), donde $Z = X^2+Y^2$, y $W = Y/X$. Probar que Z tiene distribución *chi-cuadrado con dos grados de libertad* ($Z \sim \chi^2(2)$) y que W tiene distribución de Cauchy (ver Ejemplo 8.3.5). Esto es,

$$f_Z(z) = \frac{1}{2}e^{-z/2}, (z>0) \qquad f_W(w) = \frac{1}{\pi(w^2+1)}, (w \in \mathbb{R}).$$

(Sugerencia: partir el plano en los conjuntos $\{(x,y) : y > 0\}$ y $\{(x,y) : y < 0\}$).

◊ **9.14.** * Generalizar el teorema anterior para el caso de funciónes k a 1. (Para los interesados, existe una extensión del teorema de cambio de variable que se aplica para este tipo de transformaciones. La fórmula es conocida como *fórmula del área*, que es un caso particular de la *fórmula de co-área*.)

9.6. Densidad condicional

Sean X e Y v.a. con densidad conjunta $f_{X,Y}$. En esta sección estudiaremos cómo una de las variable aleatorias es afectada al conocerse que la otra variable toma cierto valor. (Recordar Sección 5.5).

Supongamos que conocemos que la variable X toma el valor x. ¿Cómo afecta esto a la distribución de la variable Y?

Para responder a esta pregunta basta con ver cómo el evento $\{X = x\}$ altera la distribución marginal de Y. Esto es, conocer $\mathbb{P}(Y \leqslant y | X = x)$. Sin embargo, este cálculo no lo podemos hacer dado que estamos condicionando respecto a un evento con probabilidad 0.

Al igual que hicimos con la función densidad, la manera correcta de estimar la cantidad anterior es sustituir la condición $\{X = x\}$ por la condición infinitesimal $\{X \in [x, x + \delta x]\}$. De esta manera obtendremos que la *densidad condicional de Y dado $X = x$* resulta

$$f_{Y|X=x}(y) := \lim_{\delta x \to 0} \frac{\mathbb{P}\big(Y \leqslant y | X \in [x, x + \delta x]\big)}{\delta x}.$$

La aproximación anterior motiva la siguiente definición.

Definición. La *densidad condicional de Y dado $X = x$*, que denotamos por $f_{Y|X=x}(y)$, se define por

$$f_{Y|X=x}(y) := \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad y \in \mathbb{R},$$

siempre que $f_X(x) > 0$.

◊ **9.15.** Justificar la definición anterior. Corroborar que la función definida es una función densidad.

Observar que si X e Y son independientes entonces $f_{Y|X=x}(y) = f_Y(y)$.

◊ **9.16.** Sean X e Y v.a. con densidad conjunta $f_{X,Y}(x,y) = e^{-x^2 y}$, para $x > 1$ e $y > 0$. Encontrar y describir la distribución condicional de Y dado $X = x$.

♢ **9.17.** Sean X e Y v.a. con distribución conjunta uniforme sobre el disco unidad. Encontrar la densidad marginal de X y la densidad condicional de Y dado $X = x$. Observar que la distribución marginal es distinta a la condicional.

La densidad condicional puede ser muy útil para hallar densidades conjuntas. Veamos el siguiente ejemplo.

♢ **9.18.** Sea Y un v.a. con distribución uniforme en $[0,1]$, y sea X una v.a. con distribución uniforme sobre $[0,Y]$. Hallar la densidad conjunta del vector (X,Y).

9.7. Esperanza

Supongamos que tenemos un vector aleatorio continuo (X,Y) definido sobre $(\Omega,\mathcal{F},\mathbb{P})$, y sea $g:\mathbb{R}^2\to\mathbb{R}$ una función. De esta manera $Z = g(X,Y)$ define una nueva variable aleatoria sobre el mismo espacio de probabilidad. Para calcular la esperanza de dicha variable, podemos hacer uso de la representación infinitesimal (9.1), y del Teorema 6.1 al igual que hicimos en la Sección 8.1 utilizando la analogía entre masa puntual y densidad.

Teorema 9.5. *Se tiene*

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy,$$

siempre que la integral sea absolutamente convergente.

(Omitiremos la prueba de este resultado.)

Del Teorema 9.5 resulta que la esperanza es un operador lineal sobre el espacio de variables aleatorias continuas, (siempre que estas tengan distribución conjunta continua).

En particular, si X e Y son independientes, resulta $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

9.7.1. Esperanza condicional

En esta subsección estudiaremos la esperanza de variables aleatorias condicionadas. (Ver Sección 5.5 para el caso discreto.)

Sean X e Y v.a. con densidad conjunta $f_{X,Y}$. Supongamos que se tiene la información de que el evento $\{X = x\}$ ocurre.

Definición. La *esperanza condicional de Y dado $X = x$*, denotada $\mathbb{E}(Y\,|\,X = x)$, se define por

$$\mathbb{E}(Y\,|\,X = x) := \int_{-\infty}^{+\infty} y\, f_{Y\,|\,X=x}(y)\, dy,$$

donde lo anterior es válido para $x \in \mathbb{R}$ tal que $f_X(x) > 0$.

Al igual que vimos en el caso discreto, el siguiente teorema brinda una herramienta importante para calcular esperanzas.

Teorema 9.6. *Sean X e Y v.a. con distribución conjunta continua. Entonces*

$$\mathbb{E}(Y) = \int \mathbb{E}(Y\,|\,X = x) f_X(x)\, dx, \tag{9.2}$$

donde la integral es sobre todos los x tales que $f_X(x) > 0$.

♢ **9.19.** Probar el teorema anterior.

♢ **9.20.** Si X e Y tienen distribución conjunta uniforme sobre el disco unidad, hallar $\mathbb{E}(R)$ y $\mathbb{E}(R^2)$, donde $R = \sqrt{X^2 + Y^2}$.

Comentario 10. La fórmula (9.2) se puede abreviar por

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)). \tag{9.3}$$

En esta expresión se considera $\mathbb{E}(Y|X)$ como la función que a cada realización x, de la variable aleatoria X, le corresponde el valor de la esperanza condicional de Y dado $X = x$. Esto es, $\mathbb{E}(Y|X)$ es la función

$$\mathbb{E}(X|Y) : \Omega'_X \to \mathbb{R}, \quad \mathbb{E}(X|Y)(x) := \mathbb{E}(Y|X = x).$$

(Aquí $\Omega'_X := \{x \in \mathrm{Im}(X) : f_X(x) > 0\}$. Luego, la expresión (9.3) cobra sentido, donde el segundo miembro debe entenderse que es la esperanza de una función de la variable aleatoria X.

Esta función se conoce como *esperanza condicional de Y dado X*.

Parte III

Teoremas límites

Capítulo 10

Ley de los grandes números

Desigualdad de Chebyshev, ley de los grandes números, teorema de aproximación de Weierstrass, tipos de convergencia.

[Duración: 1 Semanas]

10.1. Introducción

La *Ley de los grandes números (LGN)* y el *teorema central del límite (TCL)* son dos de los teoremas límites más conocidos en probabilidad. En esta sección estudiaremos el LGN, el cuál ya fue abordado en un caso particular en la Sección 1.5. En dicha sección estudiamos la relación entre la versión frecuentista de la probabilidad, y su contrapartida axiomática, para el caso de tiradas (independientes) de una moneda.

La LGN dice que si promediamos los valores de tomar muestras independientes de v.a. con igual distribución, (con esperanza μ y varianza finita) entonces la v.a. resultante "tiende" a μ. En particular, si las v.a. son la indicatriz de cierto evento A, (i.e. toma valor 1 si A ocurre, y 0 si no), entonces la LGN dice que la frecuencia relativa tiende a la probabilidad de A.

10.2. Desigualdad de Chebyshev

Lo que veremos a continuación es una formalización de la noción de límite anterior, y la correspondiente prueba de la LGN.

Lema 10.1 (Desigualdad de Chebyshev)**.** *Sea X una v.a. tal que $\mathbb{E}(X^2) < \infty$, y sea $\varepsilon > 0$. Entonces*

$$\mathbb{P}(|X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(X^2)}{\varepsilon^2}.$$

Demostración. Usando la monotonía de la esperanza, la prueba resulta de la desigualdad: $\varepsilon^2 \mathbb{1}_{\{|X| \geqslant \varepsilon\}} \leqslant X^2$. (Recordar Comentario 3 en página 29.) □

Comentario 11.

- Se puede probar que la desigualdad es fina, en el sentido que existen v.a. que toman la igualdad en el caso anterior.
- Si $\mathbb{E}(X) = \mu$, entonces, para todo $\varepsilon > 0$ se tiene

$$\mathbb{P}(|X - \mu| \geqslant \varepsilon) \leqslant \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

10.3. Ley débil de los grandes números

Ya estamos en condiciones de enunciar y probar la ley (débil) de los grandes números.

Teorema 10.2 (Ley débil de los grandes números)**.** *Sean $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes con igual esperanza μ e igual varianza σ^2. Sea*

$$S_n = X_1 + \cdots + X_n, \quad n \in \mathbb{N}.$$

Entonces, para todo $\varepsilon > 0$, se tiene

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

♢ **10.1.** Probar el teorema. □

♢ **10.2.** Mostrar con un ejemplo que si las variables son dependientes el resultado no siempre es cierto.

Una de las aplicaciones más importantes de la ley débil de los grandes números es en el caso de que las v.a. X_i son independientes con igual distribución, lo que comunmente se escribe como *sucesión de v.a. i.i.d.*. Si $\{X_i\}$ son v.a. i.i.d con esperanza μ y varianza finita σ^2, entonces S_n/n tiende a μ en el sentido anterior.

Ejemplo 10.3.1. Al igual que hicimos en el caso de la moneda, podemos utilizar la LGN para recuperar la interpretación frecuentista de la probabilidad. Supongamos que lanzamos una moneda de manera indpendiente, y sabemos que la probabilidad de que resulte cara es p. Si definimos las v.a. $X_i = \mathbb{1}_{A_i}$, donde A_i es el evento de que en la tirada i-ésima resultó cara, resulta que

$$\frac{S_n}{n} = \frac{\text{número de caras}}{\text{número de tiradas}}.$$

Observar que $\mu = \mathbb{E}(X_1) = \mathbb{P}(\text{“sale cara”}) = p$. Entonces de la LGN se concluye que para todo $\varepsilon > 0$, se tiene

$$\mathbb{P}\left(\left|\frac{\text{número de caras}}{\text{número de tiradas}} - p\right| > \varepsilon\right) \to 0, \quad \text{cuando} \quad n \to +\infty.$$

10.4. Algunos tipos de convergencia

Existen varias nociones distintas de convergencia de una sucesión de v.a. $\{X_i\}$. En esta sección definiremos distintos tipos de convergencia que nos facilitará la exposición en lo que sigue.

Comencemos primero con la convergencia en *probabilidad* que es exactamente la que ocurre en la ley débil de los grandes números.

Definición. Decimos que una sucesión de v.a. $\{X_i\}$ converge a X en *probabilidad* si para todo $\varepsilon > 0$, se tiene

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0, \quad \text{cuando} \quad n \to \infty.$$

En tal caso escribimos $X_n \xrightarrow{\text{P}} X$.

Esta condición implica que para cualquier $\varepsilon > 0$ dado, y $\delta > 0$ pequeño, podemos encontrar un n_0 donde $|X_n - X| \leqslant \varepsilon$ en un conjunto de probabilidad $1 - \delta$.

Con esta definición, podemos reparafrasear una consecuencia de la ley débil de los grandes números de la siguiente manera.

Si $\{X_i\}$ son v.a. i.i.d. con esperanza μ y varianza finita, entonces $S_n/n \xrightarrow{\text{P}} \mu$.

Una noción más fuerte de convergencia es la siguiente.

Definición. Decimos que una sucesión de v.a. $\{X_i\}$ converge a X en L^2, si

$$\mathbb{E}((X_n - X)^2) \to 0, \quad \text{cuando} \quad n \to \infty.$$

En tal caso escribimos $X_n \overset{L^2}{\to} X$.

♢ **10.3.** Probar que si $X_n \overset{L^2}{\to} X$, entonces $X_n \overset{\mathrm{P}}{\to} X$.

Comentario 12. Si miramos la prueba realizada de la ley débil de los grandes números, en realidad se prueba un resultado más fuerte: bajo las mismas hipótesis resulta $S_n/n \overset{L^2}{\to} \mu$.

Comentario 13. Observar que para tener convergencia en probabilidad no es necesario pedirle a las v.a. que tengan varianza finita. Sin embargo, lo anterior sí es un requisito en el caso de convergencia en L^2. Es por esta razón, y por razones históricas, que la LGN se enuncia en estas dos modalidades por separado.

Existe otro tipo de convergencia que está relacionada a la convergencia puntual en un conjunto de medida de probabilidad uno.

Definición. Decimos que una sucesión de v.a. $\{X_i\}$ converge a X *casi seguramente* si

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\right\}\right) = 1.$$

En tal caso escribimos $X_n \overset{\text{c.s.}}{\to} X$.

La ley *fuerte* de los grandes números refiere a condiciones que tiene que tener la sucesión $\{X_i\}$ para que se tenga convergencia casi segura a μ. De hecho, las hipótesis dadas para la ley débil son suficientes para probar la ley fuerte, pero la prueba es más complicada. Es más, uno puede relajar las hipótesis para seguir teniendo convergencia en probabilida por ejemplo, pero estas versiones se escapan de este curso. Con esto queremos al menos motivar que existen distintas versiones de la LGN que se aplican a distintas modelos de variables i.i.d.

10.5. Aplicaciones

Teorema 10.3 (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. □

♢ **10.4.** Sean $\{X_i\}$ unsa sucesión de v.a. i.i.d. con distribución $\mathcal{B}(p)$. Sea S_n las sumas parciales.

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

En el siguiente ejemplo veremos cómo nuestra intuición geométrica empieza a fallar a gran escala.

Ejemplo 10.5.1. Sean $\{X_i\}$ v.a. i.i.d con distribución uniforme en $(-1,1)$. Sean $Y_i=X_i^2$.

♢ **10.5.** Probar que
$$\frac{X_1^2+\cdots+X_n^2}{n}\xrightarrow{\mathrm{P}}\frac{1}{3}\quad\text{cuando}\quad n\to\infty.$$

♢ **10.6.** Si $A_{n,\varepsilon}=\{x\in\mathbb{R}^n:(1-\varepsilon)\sqrt{n/3}\leqslant\|x\|\leqslant(1+\varepsilon)\sqrt{n/3}\}$, entonces, para todo $\varepsilon>0$ se tiene
$$\lim_n\frac{\mathrm{vol}(A_{n,\varepsilon}\cap(-1,1)^n)}{2^n}=1.$$

Esto último significa que la mayor parte del volumen del hipercubo $(-1,1)^n$ está concentrado en el borde de la bola de radio $\sqrt{n/3}$.

10.6. Ley fuerte de los grandes números

La idea de esta sección es probar una *ley fuerte de los grandes números*. (Ejercicio tomado del segundo parcial).

Teorema 10.4 (Ley fuerte de los grandes números)**.** *Si $\{X_i\}$ es una sucesión de v.a. i.i.d. con $\mathbb{E}(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2 < +\infty$, y cuarto momento finito, i.e. $\mathbb{E}(X_1^4) < \infty$, entonces si $S_n = X_1 + \cdots + X_n$, se tiene*

$$\mathbb{P}\left(\lim_n \left|\frac{S_n}{n} - \mu\right| = 0\right) = 1. \tag{10.1}$$

Es decir $S_n/n \overset{\text{c.s.}}{\to} \mu$.

A continuación guiaremos la prueba. Primero observar que podemos asumir $\mu = 0$.

1. Sea X una v.a. positiva, con cuarto momento finito. Probar que se tiene $\mathbb{P}(X > \varepsilon) \leqslant \frac{\mathbb{E}(X^4)}{\varepsilon^4}$.
2. Probar que existe una constante $C > 0$ tal que $\mathbb{E}(S_n^4) \leqslant Cn^2$.
3. Fijemos $\varepsilon > 0$. Sea $A_n^\varepsilon = \{\left|\frac{S_n}{n}\right| > \varepsilon\}$. Probar que $\mathbb{P}(A_n^\varepsilon) \leqslant \frac{C'}{n^2}$, para cierta constante $C' > 0$.
4. Probar que para todo $\varepsilon > 0$ se tiene $\mathbb{P}(\overline{\lim_n} A_n^\varepsilon) = 0$. Concluir la prueba de la ley fuerte de los grandes números.

Capítulo 11

Teorema central de límite

Teorema central del límite, funciones características, aplicaciones.

[Duración: 1 Semana]

11.1. Introducción

En esta sección veremos uno de los teoremas más importantes de la matemática, a saber, *el teorema central del límite* (TCL).

Consideremos una sucesión $\{X_n\}$ de v.a. i.i.d. con esperanza μ y varianza σ^2. La LGN dice que las sumas parciales $S_n = X_1 + \cdots + X_n$ se aproximan $n\mu$ cuando n crece. Dado que la varianza de S_n es $n\sigma^2$, podemos decir que

$$S_n = n\mu + O(\sqrt{n}\sigma),$$

donde la anterior igual debe interpretarse que los valores de S_n son cercanos a $n\mu$ pero con una dispersión del orden de $\sqrt{n}$.

Para estudiar el comportamiento asintótico de S_n es razonable normalizarlo, para que esté centrado en cero y su varianza no dependa de n. En este sentido se considera la v.a.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}, \quad n \in \mathbb{N},$$

donde Z_n es la normalización de S_n que satisface

$$\mathbb{E}(Z_n) = 0, \qquad \mathrm{Var}(Z_n) = 1.$$

Observar que la normalización Z_n es lineal en S_n, i.e., $Z_n = aS_n + b$, para ciertos $a, b \in \mathbb{R}$.

Ahora podemos preguntarnos,

> *¿Es esperable que las v.a. Z_n tengan una distribución límite conocida? En tal caso, ¿será la misma que la de X_1?*

La respuestas a estas preguntas son: SÍ y NO, (en ese orden). Esto es, Z_n tiene una distribución límite, pero no tiene por qué ser la misma que distribución de X_1. Lo sorprendente es que, sin importar la distribución de X_1, la distribución límite es siempre la misma: es la distribución normal estándar $\mathcal{N}(0,1)$!

Este resultado se conoce como el *teorema central del límite* y es uno de los teoremas más importantes y bellos de la matemática.

Teorema 11.1 (Teorema central del límite)**.** *Sean $\{X_n\}$ una sucesión de v.a. i.i.d. con esperanza μ y varianza σ^2. Si $S_n = X_1 + \cdots + X_n$, y*

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}, \quad n \in \mathbb{N},$$

entonces

$$\lim_n \mathbb{P}(Z_n \leqslant x) = \Phi(x), \quad x \in \mathbb{R},$$

donde $\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \quad x \in \mathbb{R}.$

(El TCL motiva otra noción de convergencia distinta a la vista en el capítulo anterior. En este caso se dice que Z_n tiende en *distribución* a distribución normal estándar, y se escribe $Z_n \overset{\mathcal{D}}{\to} \mathcal{N}(0,1)$.)

Este teorema fue probado, para el caso particular de v.a. Bernoulli, por de Moivre y Laplace a mediados del siglo XVIII. La versión general del teorema la probó Lyapunov en 1901.

Hay muchas versiones y pruebas distintas de este teorema. En este curso seguiremos técnicas basadas en *análisis de Fourier*.

11.2. Función Característica

En esta sección daremos un pantallazo sobre *funciones características* de variables aleatorias, las cuales nos dan una versión alternativa para estudiar la distribución de las mismas. Veremos cómo determinan la distribución de una v.a., para luego ver cómo la función característica nos permite estudiar límites en distribución de sucesiones de variables aleatorias.

Definición. La *función característica* de una v.a. X se define como la función $\varphi_X : \mathbb{R} \to \mathbb{C}$ definida por
$$\varphi_X(t) := \mathbb{E}(e^{itX}) := \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)), \quad t \in \mathbb{R},$$

Aquí “i” es la unidad imaginaria de los números complejos, y extendemos naturalmente la definición de esperanza de variables aleatorias que toman valores complejos. Esto es, si $Z = X + iY$, donde X e Y son v.a., entonces definimos $\mathbb{E}(Z) := \mathbb{E}(X) + i\mathbb{E}(Y)$.

Como la función e^{itX} está acotada, entonces $\varphi_X(t)$ existe para cualquier v.a.

Si $X = \sum_k p_k \delta_k$ es una v.a. discreta que toma el entero k con probabilidad p_k, entonces resulta $\varphi_X(t) = \sum_k p_k e^{itk}$. Esto implica lo siguiente.

Lema 11.2. *Si X es una v.a. discreta tal que $\Omega_X \subset \mathbb{N}$, entonces $\varphi_X(t) = G_X(e^{it})$, donde G_X es la función generatriz de probabilidad.*

Observar que en este sentido la función característica permite extender de manera natural la noción de función generatriz (la cual tenía propiedades interesantes sobre suma de variables aleatorias).

En el caso continuo también tenemos una fórmula sencilla. Si X es una v.a. continua con densidad f, resulta
$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx, \tag{11.1}$$
donde entendemos la integral anterior como el número complejo con parte real $\int_{-\infty}^{+\infty} \cos(tx) f(x)\, dx$ y parte imaginaria $\int_{-\infty}^{+\infty} \sin(tx) f(x)\, dx$.

La fórmula (11.1) se conoce como la *transformada de Fourier* de la función f.

Veamos algunas propiedades sencillas de la función característica.

♢ **11.1.** $\varphi_X(0) = 1$

♢ **11.2.** $|\varphi_X(t)| \leqslant 1$.

♢ **11.3.** $\varphi_X(-t) = \overline{\varphi_X(t)}$, donde $\overline{z}$ es el conjugado de $z \in \mathbb{C}$.

♢ **11.4.** $\varphi_{aX+b}(t) = \varphi_X(at)e^{itb}$.

También se puede probar, usando la regla de Leibnitz de intercambio entre derivada y signo de integral, que las sucesivas derivadas satisfacen

$$\varphi_X'(0) = \mathbb{E}(iX), \quad \varphi_X''(0) = -\mathbb{E}(X^2) \quad ,\ldots, \quad \varphi_X^{(n)}(0) = i^n\mathbb{E}(X^n), \tag{11.2}$$

siempre que los momentos $\mathbb{E}(X^k)$ existen.

Al igual que ocurrió con las funciones generatrices de probabilidad, las funciones características determinan unívocamente las distribuciones de probabilidad. (Lo asumimos sin demostración.)

Lema 11.3. *Si X e Y tienen iguales funciones características, entonces X e Y tienen la misma distribución.*

La prueba del lema anterior se basa en una "fórmula de inversión" donde se puede recuperar la distribución a través de la función característica. (Otra forma posible es similar al caso de funciones generatrices, donde tomando derivadas en cero podemos recuperar los momentos. Esto motiva una problema interesante, a saber, cuándo los momentos determinan la distribución de un v.a.)

El siguiente resultado es crucial, y es el que sugiere la utilización de funciones características para estudiar sumas de variables aleatorias.

Lema 11.4. *Sean X e Y v.a. independientes. Entonces $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$, para todo $t \in \mathbb{R}$.*

Veamos ahora la propiedad de estabilidad de la distribución gaussiana, a saber, la función caracacterística de una v.a. gaussiana es una función de densidad gaussiana.

Teorema 11.5. *Si X tiene distribución $\mathcal{N}(0,1)$, entonces*

$$\varphi_X(t) = e^{-t^2/2}, \quad t \in \mathbb{R}.$$

Demostración. Completando cuadrados resulta

$$\begin{aligned}\varphi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2}\, dx \\ &= \frac{e^{-t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x-it)^2/2}\, dx\end{aligned}$$

donde la última integral resulta que toma el valor uno por ser la densidad de una normal centrada en it.[1] □

[1] Para evitar justificar correctamente los pasos anteriores se da una prueba formal. Derivando bajo el signo de integral (regla de Leibnitz), se prueba que $\varphi_X'(t) = -t\varphi_X(t)$. Luego resulta $\frac{d}{dt}(\varphi_X(t)\exp^{t^2/2}) = 0$. Luego se concluye el resultado usando el hecho que $\varphi_X(0) = 1$.

Comentario 14. En otras palabras, el resultado anterior dice que la densidad gaussiana es un punto fijo (a menos de una constante) de la transformada de Fourier (11.1).

♢ **11.5.** Utilizando las propiedades de la función característica se concluye lo siguiente.

$$X \sim \mathcal{N}(\mu,\sigma^2), \quad \text{entonces} \quad \varphi_X(t) = e^{it\mu-\sigma t^2/2}.$$

Esto nos permite concluir un resultado interesante.

Teorema 11.6. *Si $X_1,\ldots,X_n$, son v.a. gaussianas independientes, entonces $X_1+\cdots+X_n$ es nuevamente una v.a. gaussiana.*

♢ **11.6.** Probar el teorema anterior. ¿Cuál sería la esperanza y varianza?

Comentario 15. Una prueba alternativa, y con menos prerequisitos, es hacerla por convolución. Pero es muy complicada. El resultado sigue siendo cierto para el caso de v.a. gaussianas dependientes. Sin embargo la varianza y la esperanza no serán lineales en sus miembros. (Cf. distribución gaussiana multivariada.)

A continuación enunciaremos un resultado crucial sobre la relación entre convergencia en distribución y convergencia de las funciones características. (Lo asumimos sin demostración.)

Teorema 11.7 (Teorema de continuidad)**.** *Sea Z una v.a. y $\{Z_n\}$ una sucesión de v.a. tales que φ_Z es continua en 0, y se satisface*

$$\lim_n \varphi_{Z_n}(t) = \varphi_Z(t),$$

para todo t punto de continuidad de φ_Z.

Entonces, resulta

$$\lim_n \mathbb{P}(Z_n \leqslant x) = \mathbb{P}(Z \leqslant x),$$

para todo punto x de continuidad de la distribución $x \mapsto \mathbb{P}(Z \leqslant x)$. □

Corolario 11.8. *Si $Z_1,Z_2,\ldots$, son v.a. tales que sus funciones características $\varphi_{Z_1},\varphi_{Z_1},\ldots$, satisfacen*

$$\lim_n \varphi_{Z_n}(t) = e^{-t^2/2},$$

para todo $t \in \mathbb{R}$, entonces,

$$\lim_n \mathbb{P}(Z_n \leqslant x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du, \quad x \in \mathbb{R}.$$

En otras palabras, si las funciones características de unas sucesión de v.a. tienden puntualmente a la función característica de una v.a. $\mathcal{N}(0,1)$, entonces, la sucesión tiende en distribución a una $\mathcal{N}(0,1)$.

Demostración del Teorema 11.1 . Primero, por simplicidad, observemos que podemos escribir $Z_n = (Y_1 + \cdots + Y_n)/\sqrt{n}$, donde cada una de estas nuevas v.a. Y_k son i.i.d, centradas y con varianza 1. Entonces usando Lema 11.4 se tiene

$$\varphi_{Z_n}(t) = \varphi_{Y_1/\sqrt{n}}(t)\cdots\varphi_{Y_n/\sqrt{n}}(t) = \left(\varphi_{Y_1}(t/\sqrt{n})\right)^n, \quad t \in \mathbb{R}. \tag{11.3}$$

En las hipótesis del teorema resulta que φ_{Y_1} tiene primera y segunda derivada. Haciendo el desarrollo de Taylor de φ_{Y_1}, en un entorno de cero, se tiene

$$\varphi_{Y_1}(u) = 1 + \varphi'_{Y_1}(0)u + \varphi''_{Y_1}(0)u^2/2 + o(u^2).$$

Usando (11.2), resulta

$$\varphi_{Y_1}(u) = 1 - u^2/2 + o(u^2), \tag{11.4}$$

para u en un entorno de 0.

Fijemos $t \in \mathbb{R}$. De (11.3) y (11.4), (para n grande tal que t^2/n esté en el entono de cero del desarrollo (11.4)), se tiene

$$\varphi_{Z_n}(t) = \left(1 - t^2/n + o(t^2/n)\right)^n,$$

donde resulta

$$\lim_n \varphi_{Z_n}(t) = e^{-t^2/2}.$$

□

Una forma de justificar el límite anterior es la siguiente. Para t fijo, si escribimos $a_n = 1 - t^2/n$ y $b_n = 1 - t^2/n + o(1/n)$, entonces

$$\begin{aligned} |{a_n}^n - {b_n}^n| &= |(a_n - b_n)({a_n}^{n-1} + {a_n}^{n-2}b_n + \cdots + {a_n}^1{b_n}^{n-2} + {b_n}^{n-1})| \\ &\leqslant |(a_n - b_n)|n = o(1/n)n, \end{aligned}$$

y por lo tanto ambas sucesiones tienen igual límite cuando $n \to +\infty$. Luego se observa que

$$\lim_n {a_n}^n = e^{-t^2/2}.$$

◊ **11.7.** Asumiendo que la función característica de una v.a con distribución de Cauchy es $\varphi_X(t) = e^{|t|}$, $t \in \mathbb{R}$, probar que el TCL no es cierto para suma de v.a. i.i.d. con distribución de Cauchy. ¿Qué distribución tiene la suma normalizada?

Proposición 11.9. *Observar que se puede recuperar la ley débil de los grandes números usando el TCL.*

11.3. Aplicaciones del TCL

En esta última sección de estas notas veremos algunas aplicaciones del TCL. La idea básica en cada ejemplo es aproximar las sumas normalizadas Z_n por una v.a. con distribución normal. Es posible tomar esta aproximación con todo rigor, ya que se puede conocer la velocidad de convergencia y por lo tanto cuantificar el error en esta aproximación. Existen varios teoremas en este sentido, como el de *Berry-Essen*, que dice que si el tercer momento de las v.a. es finito, entonces la aproximación tiene orden $1/\sqrt{n}$, (donde la constante depende del segundo y tercer momento). Este tipo de resultados se pueden ver en cursos de *Procesos Estocásticos*.

11.3.1. Aproximación normal de la binomial

Si S_n tiene distribución binomial de parámetros n y p (i.e. $S_n \sim \mathcal{B}(n,p)$), entonces cuando n es grande, y p se mantiene constante, se tiene del TCL que

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant x\right) \approx \Phi(x), \quad x \in \mathbb{R},$$

Donde Φ es distribución de una $\mathcal{N}(0,1)$.

Este resultado se conoce como la *aproximación normal a la binomial*. Es posible probar este resultado mediante cálculos elementales (ver Feller [F1] por ejemplo.)

Ejemplo 11.3.1. Supongamos que en un estadio quieren hacer un concierto para 10 mil personas, y deciden utilizar la cancha para colocar sillas. Como hay dos entradas independientes, y por razones de seguridad, deciden hacer dos zonas divididas por una valla. Se distribuyen las sillas en igual cantidad a ambos lados. Si asumimos (la no tan real idea de) que las personas elijen al azar con igual probabilidad cada entrada (a la entrada tiran una moneda fiel y dependiendo del resultado elijen en qué puerta entrar). Si quieren estar seguros en un 99 % de que que nadie quede sin silla. ¿Cuántas sillas deben poner en cada lado?

Sea $n = 10000$. Con las hipótesis dadas, si S es el número de personas que entra a la primer entrada, podemos asumir que S es una v.a. con distribución $\mathcal{B}(n,1/2)$.

Sea s la cantidad de sillas necesarias a cada lado. Lo que queremos es que s satisfaga

$\mathbb{P}(S \leqslant s) = 0{,}99$. Entonces usando la aproximación a la normal se tiene

$$\begin{aligned} 0{,}01 = \mathbb{P}(S > s) = \mathbb{P}\left(\frac{S - n/2}{\sqrt{n/4}} > \frac{s - n/2}{\sqrt{n/4}}\right) \\ \approx 1 - \Phi\left(\frac{2s-n}{\sqrt{n}}\right). \end{aligned}$$

Se tiene $\Phi(2{,}34) \approx 0{,}99$, por lo cual debemos resolver $\frac{2s-n}{\sqrt{n}} = 2{,}34$, y de de donde resulta

$$s = \frac{n}{2} + 1{,}17\sqrt{n}.$$

Esto es, para $n = 10000$, $s = 5117$. Esto es, se necesitan comprar 234 sillas extras. Sorprendente, no?

Lo anterior resulta más interesante cuando n es más grande aún. Observar que $s \approx n/2 + \sqrt{n}$, y por lo tanto el cociente $s/(n/2)$ es aproximadamente $2/\sqrt{n}$, y por lo tanto tiende a cero a cuando n crece. Esto significa que cuando n crece la proporción de sillas extras tiende a cero. Este fenómeno de cancelación se conoce como *concentración de la medida*.

11.4. Muestra estadística

Hay un porcentaje p de electores que votan en blanco en las elecciones. Queremos estimar p, haciendo una encuesta, con un error menor a 0.005. ¿qué tan grande tiene que ser la muestra?

Sea n el tamaño de la muestra (variable a determinar). Sea S_n la v.a. suma de votantes que votan en blanco. Podemos asumir que $S_n \sim \mathcal{B}(n, p)$. Por una convención utilizada en *estadística*, denotamos por $\widehat{p}$ al promedio anterior

Queremos encontrar n tal que $|\widehat{p} - p| \leqslant 0{,}005$. Es claro que esto no puede ocurrir siempre, dado que la v.a. $|S_n/n - p|$ puede tomar valores mayores(por ejemplo si todos los de la encuesta votaban en blanco). La forma correcta de plantear el problema es dar una probabilidad para que ocurra ese error. Por ejemplo, hallar n para que el error del $|\widehat{p} - p| \leqslant 0{,}005$ ocurra con probabilidad mayor al 95 %. Esto es,

$$\mathbb{P}(|\widehat{p} - p| \leqslant 0{,}005) \geqslant 0{,}95.$$

Usando la aproximación dada por el TCL obtenemos

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\leqslant 0{,}005\right) &= \mathbb{P}\left(\left|\frac{S_n-np}{\sqrt{np(1-p)}}\right|\leqslant\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right)\\
&\approx \Phi\left(\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right)-\Phi\left(-\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right)\\
&= 2\Phi\left(\frac{0{,}005\sqrt{n}}{\sqrt{p(1-p)}}\right)-1\\
&\geqslant 2\Phi\left(0{,}005\sqrt{4n}\right)-1,
\end{aligned}$$

(donde la última desigualdad es consecuencia de que $p(1-p)\leqslant 1/4$. Resulta que si $0{,}005\sqrt{4n}\geqslant 1{,}96$, entonces resulta $2\Phi\left(0{,}005\sqrt{4n}\right)-1\geqslant 0{,}95$. Esto es $n>40000$.

Sin embargo, si sólo demandamos un error del 3 %, se prueba que $n\approx 1000$ funciona. Este número es el que generalmente se utiliza en las encuestas que vemos a diario.

11.5. Aguja de Buffon reconsiderada

En esta sección mostraremos cómo estimar π con el problema visto de la Aguja de Buffon (ver Ejercicio 9.9).

Supongamos que tiramos un aguja de longitud ℓ sobre la hoja de un cuaderno a rayas. Si r es la distancia entre lineas, y $\ell\leqslant r$, resulta que la probabilidad p que la aguja tirada al azar intersecte alguna de las lineas es igual a

$$p=\frac{2\ell}{\pi r}.$$

Veamos una prueba distinta a la que realizaron los estudiantes en clase. A diferencia de la prueba vista, que se basaba en el cálculo de una integral, esta prueba utiliza argumentos probabilísticos y geométricos. La principal herramienta a usar se basa en la linealidad de la esperanza.

Existen varias referencias sobre la versión de la prueba que daremos. Recomendamos la biblia sobre geometría integral de Santaló [S1], y en particular la introducción del libro de Klain y Rota [KR].

Supongamos por el momento que tiramos al azar una aguja de longitud ℓ_1 (sin la restricción $\ell_1\leqslant r$). Entonces si definimos la v.a. X_1 que cuenta la cantidad de cruces con las lineas, resulta

$$\mathbb{E}(X_1)=0p_0+p_1+2p_2+\cdots,$$

donde p_i es la probabilidad de que la aguja corte las lineas en exactamente i lugares. Por lo tanto si $\ell_1 \leqslant r$ resulta

$$\mathbb{E}(X_1) = p_1,$$

que es la cantidad buscada. Por lo tanto reduciremos nuestro problema a calcular $\mathbb{E}(X_1)$.

Consideremos otra aguja de longitud ℓ_2 que tiramos también al azar. Sea X_2 la v.a. que cuenta los puntos de intersección de esta nueva aguja. Supongamos que la trasladamos para luego unirla a un extremo de la otra aguja. De esta manera podemos pensar en una“visagra” aleatoria. El número de puntos de intersección de la visagra con las lineas es $X_1 + X_2$. Observar que ambas v.a. son dependientes, pero sin embargo, por la linealidad de la esperanza se tiene

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2.$$

(Observar que es natural asumir que la ley de X_2 no tiene ninguna preferencia, y por tanto puede considerarse que tiene el mismo azar que la aguja original. También se puede trazar un argumento más convincente de la siguiente manera. Si asumimos que un extremo de la aguja tiene distribución en altura uniforme, y el ángulo de la aguja con distribución uniforme e independientes, entonces, al unir la nueva aguja tirada con el mismo azar por el extremo que conocemos, obtenemos una nueva aguja que individualmente tiene el azar mencionado.)

Este razonamiento se puede extender a una “cadena” de k agujas que son unidas por uno de sus extremos a otra.

Es interesante, y llamativo, observar que si las agujas se unen de manera que quede un segmento de recta, entonces la esperanza del números de cortes total sigue siendo igual a la esperanza del número de cortes de la cadena. Es decir, la forma no importa (sólo que sean segmentos de rectas unidos).

Es claro que $\mathbb{E}(X_1)$ es una función que sólo depende de la longitud ℓ_1. Sea f la función dada por $f(\ell_1) = \mathbb{E}(X_1)$.

<u>Afirmación 1:</u> Resulta que f satisface

$$f(\ell_1 + \ell_2) = f(\ell_1) + f(\ell_2).$$

Esto resulta de que una aguja de longitud $\ell = \ell_1 + \ell_2$ puede ser pensada como la unión de dos agujas unidas por los extremos de longitudes ℓ_1 y ℓ_2 respectivamente, y entonces se tiene

$$f(\ell_1 + \ell_2) = \mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 = f(\ell_1) + f(\ell_2)$$

<u>Afirmación 2:</u> f es lineal sobre $\mathbb{Q}$:

Basta ver que la Afirmación 1 implica el resultado tomando segmentos unitarios. Se tiene $f(k\ell) = kf(\ell)$, y además $f(1/n)n = f(1)$. Entonces se tiene $f(k/n) = f(1)k/n$.

Afirmación 3: existe $k \in \mathbb{R}$, tal que $f(\ell) = k\ell$, para todo $\ell \in \mathbb{R}$.

La prueba resulta de ser lineal en $\mathbb{Q}$ y de ser monótona creciente.

Afirmación 4: Sea $\mathcal{C}$ es un alambre rígido de longitud ℓ (una curva rectificable de longitud ℓ)[2] que se tira al azar. Entonces si X es la v.a. que cuenta la cantidad de intersecciones, resulta

$$\mathbb{E}X = k\ell$$

La prueba sigue de aproximar $\mathcal{C}$ por poligonales. Esto es, si $\mathcal{C}_n$ es una poligonal fija de longitud ℓ_n que aproxima $\mathcal{C}$, entonces, el número de cortes de la poligonal (al azar) $\mathcal{C}_n$ tiende al número de cortes de $\mathcal{C}$ tirados al azar. Dado que la esperanza de número de cortes de esta poligonal es $f(\ell_n)$, resulta tomando límite en n que $\mathbb{E}X = f(\ell) = k\ell$.

Afirmación 5: Se tiene $k = 2/(\pi r)$.

Basta tomar una curva donde podamos computar la esperanza de cortes. Si consideramos $\mathcal{C}$ el círculo de diámetro r, resulta que el número de cortes, que denotamos por X, es una v.a. con distribución constante igual a 2. Entonces de la Afirmación 5 se tiene

$$2 = \mathbb{E}(X) = k\mathrm{long}(\mathcal{C}) = k\pi r, \quad \text{entonces} \quad k = \frac{2}{\pi r}.$$

Hemos concluido que la probabilidad p de intersección de una aguja de longitud ℓ tirada al azar es

$$p = \frac{2\ell}{\pi r}.$$

Comentario 16. Otro hecho interesante es que como no importa la forma del alambre, podemos tirar al azar un "fideo" de longitud ℓ y sobre la grilla para tener el mismo resultado. Claro está que para esto es un poco más difícil entender qué quiere decir tirar al azar.

Comentario 17. La técnica usada en esta prueba está basada en la *geometría integral*, la cual tiene como fin estudiar propiedades geométricas de objetos a través de integrar ciertas cantidades numéricas. A modo de ejemplo, se puede probar de manera similar a lo anterior, que la longitud de una curva encajada en la esfera S^2 se puede recuperar integrando sobre todos los "ecuadores" la cantidad de puntos de intersección del ecuador con la curva. (Una forma de considerar esta integral, es observar que podemos parametrizar los "ecuadores al azar" con el ecuador ortogonal a un vector con distribución uniforme sobre la esfera.

[2] Que su longitud se pueda aproximar por poligonales.

11.5.1. Estimación de π

Supongamos que lanzamos la aguja n veces. Queremos encontrar n tal que el error en la aproximación de π sea menor 0,001, con probabilidad mayor a 95 %.

Sea N el número de veces que la aguja intersecta las lineas. Podemos suponer que N tiene distribución $\mathcal{B}(n,p)$.

Del TCL tenemos que

$$\frac{N-np}{\sqrt{np(1-p)}} \approx \gamma,$$

donde $\gamma \sim \mathcal{N}(0,1)$. (Acá estamos abusando de la notación, donde $Z_n \approx \gamma$ se lee, Z_n tiende en distribución a γ.) Entonces

$$N \approx np + \sqrt{np(1-p)}\gamma$$

i.e. N tiende en distribución a una v.a. con distribución $\mathcal{N}(np, np(1-p))$.

La probabildiad "empírica" es $\widehat{p} = N/n$, y sea $\widehat{\pi}$ la v.a. que aproxima a π. Esto es

$$\widehat{\pi} = \frac{2\ell}{r\widehat{p}}a = \pi\frac{p}{\widehat{p}} \quad \text{lo cual implica} \quad \widehat{\pi} - \pi = \pi\frac{p-\widehat{p}}{\widehat{p}},$$

Acá hacemos una aproximación nuevamente para poder simplificar los cálculos. Esto es considerar

$$\widehat{\pi} - \pi \approx \pi\frac{p-\widehat{p}}{p}.$$

Resulta entonces del TCL que

$$\widehat{\pi} - \pi \approx \mathcal{N}\left(0, \frac{\pi^2}{n}\frac{(1-p)}{p}\right).$$

Al igual que en la sección anterior, queremos independizarnos de p, por lo cual podemos tomar una aguja de longitud ℓ que minimice la varianza anterior. El mínimo ocurre cuando p es maximizado, y esto es si tomamos $\ell = r$. En tal caso se tiene $2/\pi$, y

$$\widehat{\pi} - \pi \approx \mathcal{N}\left(0, \frac{\pi^2}{n}\left(\frac{\pi}{2} - 1\right)\right).$$

Ahora estamos en condiciones de resolver nuestro problema.

Con las aproximaciones en cuestión se tiene que

$$\mathbb{P}(|\widehat{\pi} - \pi| \leqslant 0{,}001) \geqslant 0{,}95 \quad \text{si y sólo si} \quad \frac{0{,}001}{\frac{\pi}{\sqrt{n}}\sqrt{\left(\frac{\pi}{2}-1\right)}} \geqslant 1{,}96,$$

donde utilizamos la aproximación de la sección anterior. Basta tomar

$$n > \left(\frac{1,96}{0,001}\right)^2 \pi^2 \left(\frac{\pi}{2} - 1\right)$$

Basta tomar n del orden de 21 millones. Es decir, mejor busquemos otro método para estimarlo :)

11.6. Estimación de integrales mediante el método de Monte Carlo

Esta sección es la segunda entrega del curso.

El objetivo de esta tarea es dar una estimación de una integral $\int_a^b f(x)\mathrm{d}x$ que no seamos capaces de calcular utilizando las herramientas que aprendimos en el curso de Cálculo I, a partir de la *Ley (débil) de los Grandes Números*.

1. Probar que si la sucesión de v.a. $\{X_n\}_n$ y X son tales que $X_n \to_{\mathbb{P}} X$ y $g : \mathbb{R} \to \mathbb{R}$ es una función continua, entonces:
$$g(X_n) \to_{\mathbb{P}} g(X)$$
Sugerencia: suponga primero que g es uniformemente continua

2. Demostrar que si $X_1, X_2, \ldots, X_n$ son v.a. i.i.d. con $\mathbb{E}(X_1) = \mu$ y $\mathrm{var}(X_1) = \sigma^2$ finitos, entonces la v.a.
$$\sigma_n^2 := \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2$$
converge en probabilidad a σ^2.

3. Demostrar que si $U_1, U_2, \ldots, U_n$ son v.a. i.i.d. con distribución $U(a,b)$ y $f : [a,b] \to \mathbb{R}$ es integrable, entonces:
$$\frac{1}{n}\sum_{i=1}^n f(U_i) \to_{\mathbb{P}} \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x.$$

4. Utilizar el resultado anterior para dar una estimación de $\int_2^7 e^{-x^2}\mathrm{d}x$. Comparar con el resultado que se obtiene en *Wolfram Alpha* (por ejemplo) a medida que aumenta n.

5. Considere el *error (aleatorio)* que cometemos por aproximar la integral $\int_a^b f(x)\mathrm{d}x$ por $(b-a)\times\frac{1}{n}\sum_{i=1}^n f(U_i)$, llamémosle:

$$E_n := \frac{1}{n}\sum_{i=1}^n f(U_i) - \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x.$$

   a) Acotar la probabilidad de que E_n sea más grande que $\varepsilon > 0$ utilizando la *desigualdad de Chebyshev*. Indique a partir de esta cota cuán grande tiene que ser n si quiero estar seguro de que con probabilidad 0,98 el verdadero valor de la integral dista de mi estimación menos de 0,01.

   b) Suponiendo que:

$$\sigma^2 := \frac{1}{b-a}\int_a^b (f(x))^2\mathrm{d}x - \left(\frac{1}{b-a}\int_a^b f(x)\mathrm{d}x\right)^2$$

   fuera conocido, probar que la función de distribucoón de E_n puede ser aproximada por $\Phi\left(\frac{\sqrt{n}x}{\sigma}\right)$.

   Observar que si $f(x) = e^{-x^2}$, como antes, entonces σ^2 sigue siendo desconocida. No se preocupe, ya veremos cómo arreglamos esto.

   c) ¿Cómo estimaría σ?

6. *a*) Utilizando Φ y esa estimación de σ, calcule cuán grande tiene que ser n si quiero asegurar que el error que cometo al aproximar $\int_2^7 e^{-x^2}dx$ por $\frac{5}{n}\sum_{i=1}^n e^{-u_i{}^2}$ es menor que 0.01 con probabilidad (aproximada) 0.98.

   b) Para ese valor de n, realice varias simulaciones. ¿Puede dar un intervalo tal que pueda asegurar que con probabilidad (aproximada) 0.98 contiene al verdadero valor de la integral? Para esto podría resultarle conveniente graficar en el eje de las y el valor obtenido para cada simulación y en el eje de las x numerar las repeticiones de la simulación. Observe si para alguna simulación la estimación de la integral se escapa de dicho intervalo.
   [3]

11.6.1. Solución

A continuación agregamos una solución a la entrega realizada por el estudiante Facundo Almeida.

[3] simulamos $u_1, u_2, \ldots, u_n$ M.A.S. provinientes de $U_1, U_2, \ldots, U_n$ i.i.d. $U(a,b)$ con la función **runif()** de **R**.

1. Lo que queremos demostrar es que dado $\varepsilon_0 > 0$ se cumple que para todo $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que

$$\mathbb{P}(|g(X_n) - g(X)| > \varepsilon_0) < \varepsilon \qquad n \geqslant n_0.$$

Comencemos tomando $\varepsilon_0, \varepsilon > 0$ arbitrarios. Considérese la función de distribución $F_X : \mathbb{R} \to \mathbb{R}$. Como sabemos, se cumple que $\lim_{x\to-\infty} F_X(x) = 0$ y $\lim_{x\to+\infty} F_X(x) = 1$. Por lo tanto, podemos encontrar números reales α y β, con $\alpha < \beta$, tales que $F_X(\alpha) < \varepsilon/4$ y $F_X(\beta) > 1 - \varepsilon/4$.

Sea $A = [\alpha, \beta]$. Entonces

$$\begin{aligned} \mathbb{P}(X \notin A) &= 1 - \mathbb{P}(X \in A) \\ &= 1 - (F_X(\beta) - F_X(\alpha)) \\ &< 1 - (1 - \varepsilon/4 - \varepsilon/4) \\ &= \varepsilon/2. \end{aligned} \tag{11.5}$$

A continuación observemos que A es un subconjunto compacto de $\mathbb{R}$. Como g es una función continua, es uniformemente continua en A. Luego existe $\delta > 0$ tal que $|g(x) - g(y)| < \varepsilon_0 \;\; \forall x, y \in A : |x - y| \leqslant \delta$. Además por hipótesis $X_n \xrightarrow{\mathbb{P}} X$, así que podemos tomar $n_0 \in \mathbb{N}$ tal que $\mathbb{P}(|X_n - X| > \delta) < \varepsilon/2$ para todo $n \geqslant n_0$. Veamos que esta elección de n_0 es adecuada.

Tomemos $n \geqslant n_0$ cualquiera. Si $\omega \in \{|X_n - X| \leqslant \delta, X \in A\}$, entonces $|X_n(\omega) - X(\omega)| \leqslant \delta$ y $X(\omega) \in A$, así que $|g(X_n(\omega)) - g(X(\omega))| \leqslant \varepsilon_0$, i.e. $\omega \in \{|g(X_n) - g(X)| \leqslant \varepsilon_0\}$. Por lo tanto,

$$\begin{aligned} &\{|X_n - X| \leqslant \delta, X \in A\} \subseteq \{|g(X_n) - g(X)| \leqslant \varepsilon_0, X \in A\} \\ \Longrightarrow\; &\mathbb{P}(|X_n - X| \leqslant \delta, X \in A) \leqslant \mathbb{P}(|g(X_n) - g(X)| \leqslant \varepsilon_0, X \in A) \end{aligned} \tag{11.6}$$

Por otro lado,

$$\begin{aligned} \mathbb{P}(|X_n - X| \leqslant \delta) &= \mathbb{P}(|X_n - X| \leqslant \delta, X \in A) + \mathbb{P}(|X_n - X| \leqslant \delta, X \notin A) \\ &\leqslant \mathbb{P}(|X_n - X| \leqslant \delta, X \in A) + \mathbb{P}(X \notin A) \\ \Longrightarrow \mathbb{P}(|X_n - X| \leqslant \delta, X \in A) &\geqslant \mathbb{P}(|X_n - X| \leqslant \delta) - \mathbb{P}(X \notin A). \end{aligned} \tag{11.7}$$

Entonces

$$
\begin{aligned}
\mathbb{P}(|g(X_n)-g(X)|\leqslant \varepsilon_0) &\geqslant \mathbb{P}(|g(X_n)-g(X)|\leqslant \varepsilon_0, X\in A)\\
&\geqslant \mathbb{P}(|X_n-X|\leqslant \delta, X\in A) \qquad \text{(por 2)}\\
&\geqslant \mathbb{P}(|X_n-X|\leqslant \delta)-\mathbb{P}(X\notin A) \qquad \text{(por 3)}\\
&> \mathbb{P}(|X_n-X|\leqslant \delta)-\varepsilon/2 \qquad \text{(por 1)}\\
\Longrightarrow \mathbb{P}(|g(X_n)-g(X)|> \varepsilon_0) &< 1-\mathbb{P}(|X_n-X|\leqslant \delta)+\varepsilon/2\\
&< \mathbb{P}(|X_n-X|> \delta)+\varepsilon/2\\
&< \varepsilon/2+\varepsilon/2=\varepsilon.
\end{aligned}
$$

2. Para esta parte será conveniente contar con un resultado que nos hable de lo que pasa con el "límite en probabilidad" de la suma de dos sucesiones de variables aleatorias.

Lema 11.10. *Si $X_n \xrightarrow{\mathbb{P}} X$ e $Y_n \xrightarrow{\mathbb{P}} Y$, entonces $X_n+Y_n \xrightarrow{\mathbb{P}} X+Y$.*

Demostración. Sea $\varepsilon>0$. Entonces

$$
\begin{aligned}
\mathbb{P}(|X_n+Y_n-X-Y|>\varepsilon) &\leqslant \mathbb{P}(|X_n-X|+|Y_n-Y|>\varepsilon) \qquad \text{(desigualdad triangular)}\\
&= 1-\mathbb{P}(|X_n-X|+|Y_n-Y|\leqslant\varepsilon)\\
&\leqslant 1-\mathbb{P}(|X_n-X|\leqslant\varepsilon/2,\ |Y_n-Y|\leqslant\varepsilon/2)\\
&= \mathbb{P}(|X_n-X|>\varepsilon/2 \text{ o } |Y_n-Y|>\varepsilon/2)\\
&\leqslant \mathbb{P}(|X_n-X|>\varepsilon/2)+\mathbb{P}(|Y_n-Y|>\varepsilon/2).
\end{aligned}
$$

Por hipótesis $\lim_{n\to+\infty}\mathbb{P}(|X_n-X|>\varepsilon/2)=\mathbb{P}(|Y_n-Y|>\varepsilon/2)=0$, así que $\lim_{n\to+\infty}\mathbb{P}(|X_n+Y_n-X-Y|>\varepsilon)=0$, y el lema queda demostrado. □

Supongamos ahora que $X_1,\dots,X_n$ son como en la letra. Como $X_1,\dots,X_n$ tienen idéntica distribución, es claro de la definición de función distribución que $X_1^2,\dots,X_n^2$ tienen idéntica distribución. Además vimos en el teórico que la independencia de $X_1,\dots,X_n$ implica la independencia de $X_1^2,\dots,X_n^2$ (ya que para cada i se cumple $X_i^2=g(X_i)$, donde $g:\mathbb{R}\to\mathbb{R}$ está dada por $g(x)=x^2$). Entonces, por la ley débil de los grandes números, $\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)_n$ converge en probabilidad a $\mathbb{E}(X_1^2)$. Por otro lado, la LGN también implica que $\left(\frac{1}{n}\sum_{i=1}^n X_i\right)_n$ converge en probabilidad a $\mathbb{E}(X_1)$; por la parte 1, $\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right)_n$ converge en probabilidad a $\mathbb{E}(X_1)^2$. Entonces, por el lema 1, σ_n converge en probabilidad a $\mathbb{E}(X_1^2)-\mathbb{E}(X_1)^2=\sigma^2$.

3. Para cada i sea $X_i = f(U_i)$. Argumentando como en la parte anterior, deducimos que las variables aleatorias $X_1,\dots,X_n$ son i. i. d.. Entonces, por la LGN, la sucesión $(\frac{1}{n}\sum_{i=1}^n X_i)_n$ converge en probabilidad a

$$\begin{aligned}\mathbb{E}(X_1) &= \mathbb{E}(f(U_1))\\ &= \int_{-\infty}^{\infty} f(x) f_{U_1}(x)\,dx\\ &= \frac{1}{b-a}\int_a^b f(x)dx \qquad (U_1 \text{ tiene distribución uniforme en } [a,b]).\end{aligned}$$

4. El resultado anterior nos dice que muestras de la variable aleatoria $(7-2)\frac{1}{n}\sum_{i=1}^n e^{-U_i^2} = \frac{5}{n}\sum_{i=1}^n e^{-U_i^2}$ sirven como aproximación para $\int_2^5 e^{-x^2}\,dx$, donde tomamos $f:[2,7]\to\mathbb{R}$ dada por $f(x)=e^{-x^2}$ y consideramos a las variables aleatorias U_i con distribución $\mathcal{U}(2,7)$.

   Para la estimación utilizamos la función **`runif`** de R (que simula muestras de una variable aleatoria con distribución uniforme) como sigue: para un valor fijo de $n \geqslant 1$ tomamos n muestras aleatorias de una v.a. con distribución $\mathcal{U}(2,7)$ con ayuda de la función **`runif`**; evaluamos la función f en el valor obtienido de cada muestra; sumamos todos los resultados; multiplicamos todo por $\frac{5}{n}$. En resumen, ejecutamos la línea

```
5/n * sum(exp(-runif(n, 2, 7)^2))
```

   (`exp` denota la función exponencial). En la siguiente tabla están registrados los resultados para distintos valores de n junto con el error de la aproximación con respecto al valor proporcionado por Wolfram Alpha: $\int_2^7 e^{-x^2}\,dx \approx 0{,}0041553$.

n	`5/n * sum(exp(-runif(n, 2, 7)^2))`	error
5	0.0107149	0.0065596
10	0.0034393	0.0007160
100	0.0037871	0.0003682
500	0.0038945	0.0002608
1000	0.0040949	0.0000604
10000	0.0043993	0.0002440
100000	0.0041864	0.0000311

5. a) Notar que $\mathbb{E}(\frac{1}{n}\sum_{i=1}^n f(U_i)) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(f(U_i)) = \mathbb{E}(f(U_1)) = \frac{1}{b-a}\int_a^b f(x)\,dx$ (a esto último

ya lo probamos en la parte 3). Si ponemos $Y_n := \frac{1}{n}\sum_{i=1}^n f(U_i)$, entonces $E_n = Y_n - \mathbb{E}(Y_n)$, y

$$\begin{aligned}
\mathbb{P}(|E_n| > \varepsilon) &\leqslant \frac{\mathbb{E}(E_n^2)}{\varepsilon^2} \qquad \text{(desigualdad de Chebyshev)}\\
&= \frac{\mathbb{E}((Y_n - \mathbb{E}(Y_n))^2)}{\varepsilon^2}\\
&= \frac{\operatorname{var}(Y_n)}{\varepsilon^2}\\
&= \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^n \operatorname{var}(f(U_i)) \qquad (*)\\
&= \frac{1}{\varepsilon^2 n}\operatorname{var}(f(U_1))\\
&= \frac{1}{\varepsilon^2 n}(\mathbb{E}(f(U_1)^2) - \mathbb{E}(f(U_1))^2)\\
&= \frac{1}{\varepsilon^2 n}\left(\frac{1}{b-a}\int_a^b f(x)^2\,dx - \left(\int_a^b f(x)\,dx\right)^2\right) \qquad (**)
\end{aligned}$$

(en $(*)$ usamos la independencia de las v.a. U_i para desarrollar la suma; en $(**)$ usamos la parte 3).

Lo que se pide es encontrar un valor de n para el cual $\mathbb{P}(|E_n| > 0{,}01) < 0{,}02$. Tomando $\sigma^2 = \frac{1}{b-a}\int_a^b f(x)^2\,dx - \left(\int_a^b f(x)\,dx\right)^2$, la inecuación a resolver es $\frac{1}{0{,}01^2 n}\sigma^2 < 0{,}02$. Bastará entonces con tomar $n > \frac{\sigma^2}{0{,}01^2 \cdot 0{,}02} = 5\cdot 10^5\sigma^2$.

b) Para cada n sea $S_n = \sum_{i=1}^n f(U_i)$, y sea $\mu = \frac{1}{b-a}\int_a^b f(x)\,dx$. Notar que $E_n = \frac{1}{n}(S_n - n\mu)$ (parte 3) y que σ^2 es la varianza de $f(U_1)$ (parte 5-a)). Por el teorema central del límite, la sucesión $\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right)_n = \left(\frac{\sqrt{n}}{\sigma}E_n\right)_n$ converge en distribución a $\mathcal{N}(0,1)$, esto es,

$$\begin{aligned}
&\lim_{n\to+\infty}\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}E_n \leqslant x\right) = \Phi(x) \quad \forall x\in\mathbb{R}\\
\Longrightarrow\ &\lim_{n\to+\infty}\mathbb{P}(E_n \leqslant x) = \Phi\left(\frac{\sqrt{n}}{\sigma}x\right) \quad \forall x\in\mathbb{R}.
\end{aligned}$$

En otras palabras, la función distribución de E_n se puede aproximar por $\Phi\left(\frac{\sqrt{n}}{\sigma}x\right)$.

c) Lo que nos interesa es obtener una cota superior razonable para σ, ya que vamos a usar el valor de σ para acotar por debajo el valor de $\Phi(\frac{\sqrt{n}}{\sigma}x)$, y la función Φ es creciente. Para eso podemos buscar una cota superior para $\int_a^b (f(x))^2\,dx = \int_2^7 e^{-2x^2}\,dx$ y una cota inferior para $\int_a^b f(x)\,dx = \int_2^7 e^{-x^2}\,dx$. Bastará con acotar los integrandos; esto es sencillo porque ambos son funciones decrecientes. Tenemos

$$\int_2^7 e^{-2x^2}\,dx \leqslant (7-2)e^{-2\cdot 2^2} = 5e^{-8},$$
$$\int_2^7 e^{-x^2}\,dx \geqslant (7-2)e^{-7^2} = 5e^{-49}.$$

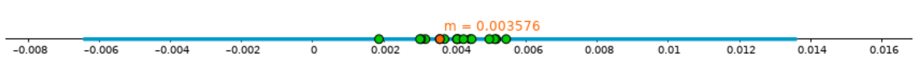
Así, $\sigma^2 = \frac{1}{5}\int_2^7 e^{-2x^2}\,dx - \left(\frac{1}{5}\int_2^7 e^{-x^2}\,dx\right)^2 \leqslant e^{-8} - (e^{-49})^2$. Luego $\sigma \leqslant \sqrt{e^{-8}-(e^{-49})^2} \approx$ 0,01832.

6. a) Queremos encontrar n lo suficientemente grande para que $\mathbb{P}(|E_n| \leqslant 0{,}01) \geqslant 0{,}98$. Sustituyendo la función distribución de E_n por $\Phi(\frac{\sqrt{n}}{\sigma}x)$, la inecuación queda $\Phi(\frac{\sqrt{n}}{\sigma}0{,}01) - \Phi(-\frac{\sqrt{n}}{\sigma}0{,}01) \geqslant 0{,}98$, esto es, $\Phi(\frac{\sqrt{n}}{\sigma}0{,}01) \geqslant 0{,}99$. Ahora reemplazamos σ por la aproximación que calculamos en el punto anterior[4] para obtener $\Phi(0{,}5459\sqrt{n}) \geqslant 0{,}99$. Observando la tabla del capítulo 8 de las notas, vemos que será suficiente con que $0{,}5459\sqrt{n} \geqslant 2{,}33$. Despejando resulta que nos servirá n tal que $n \geqslant 18{,}2$, o sea, $n \geqslant 19$.

   b) Reiterando las simulaciones ahora con $n = 19$, obtenemos las estimaciones:

$$0{,}001864, 0{,}003707, 0{,}004052, 0{,}000938, 0{,}00357, 0{,}002575, 0{,}003084,$$
$$0{,}001786, 0{,}003576, 0{,}003170, 0{,}005159, 0{,}003053, 0{,}004337, 0{,}003338.$$

   Para dar con un intervalo como el que se pide, consideremos una muestra cualquiera m de las obtenidas con $n = 19$. Sabemos que la distancia entre m y $\int_2^7 e^{-x^2}\,dx$ es, con probabilidad aproximada 0.98, menor a 0.01. Esto es lo mismo que decir que $\int_2^7 e^{-x^2}\,dx \in (m-0{,}01, m+0{,}01)$ con probabilidad 0.98. Tomando por ejemplo $m = 0{,}003576$ obtenemos el intervalo $(-0{,}006423, 0{,}013576)$. A continuación graficamos varias muestras y nos fijamos si caen en dicho intervalo:



   En la imagen se aprecia que efectivamente todas las muestras tomadas en este caso caen dentro del intervalo elegido.

[4] Acá queda claro por qué acotamos σ *inferiormente*: para la resolución de la inecuación, la aproximación elegida σ' de σ debe ser una que haga que $\frac{1}{\sigma'}0{,}01$ no sea menor que el valor real, ya que de lo contrario el valor que tomemos para n podría no ser lo suficientemente grande.

Parte IV

Apéndice

Capítulo 12

Ensayos independientes de un experimento

En este capítulo formalizaremos la noción de ensayos independientes de un experimento. Dado dos espacios de probabilidad, la idea es construir un nuevo espacio de probabilidad donde "vivan" cada uno de los modelos anteriores, de manera independiente. Esta construcción nos permitirá ver la existencia de v.a. independientes sobre un mismo espacio (tema que con cierta complicidad pasamos por arriba sin mencionarlo a lo largo del curso).

> **Definición.** Diremos que un espacio $(\Omega', \mathcal{F}', \mathbb{P}')$ *extiende* un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, si existe un mapa $\pi : \Omega' \to \Omega$ tal que es medible, i.e., $\pi^{-1}(A) \in \mathcal{F}', \quad A \in \mathcal{F}$, y preserva la medida, esto es, $\mathbb{P}'(\pi^{-1}(A)) = \mathbb{P}(A), \quad A \in \mathcal{F}$.
> (En este caso diremos $((\Omega', \mathcal{F}', \mathbb{P}'), \pi)$ es una extensión de $(\Omega, \mathcal{F}, \mathbb{P})$.)

De esta manera, el evento $A \in \mathcal{F}$ puede ser identificado con el nuevo evento $\pi^{-1}(A) \in \mathcal{F}'$.

El objetivo de este capítulo es encontrar una extensión de dos espacios de probabilidad donde los nuevos eventos asociados sean independientes.

12.1. Medida producto: caso finito

Consideremos dos espacios de probabilidad $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, y $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$.

Como comentamos nuestro objetivo es diseñar un nuevo espacio de probabilidad donde podamos estudiar el experimento en conjunto donde cada uno de los experimentos se realiza de manera independiente.

12.1.1. Espacio muestral producto

El espacio natural a considerar es el espacio producto $\Omega = \Omega_1 \times \Omega_2$, que viene provisto de las proyecciones canónicas $\pi_i : \Omega \to \Omega_i$, con $i = 1, 2$.

12.1.2. σ-álgebra producto

El conjunto de eventos de la nueva σ-álgebra a definir debería tener los conjuntos de la forma $\pi_1^{-1}(A_1) = A_1 \times \Omega_2$, y $\pi_2^{-1}(A_2) = \Omega_1 \times A_2$, donde $A_i \in \mathcal{F}_i$. Esto sería razonable para poder hablar de que ocurrió el evento A_1 en el primer experimento, y A_2 en el segundo.

Definición. Sean $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, y $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ dos espacios de probabilidad. Definimos la σ-álgebra producto $\mathcal{F} := \mathcal{F} \times \mathcal{F}_2$ en $\Omega := \Omega_1 \times \Omega_2$ como la mínima σ-álgebra que contiene los conjuntos $A_1 \times \Omega_2$, y $\Omega_1 \times A_2$, donde $A_i \in \mathcal{F}_i$.

Observar que con esta definición $\mathcal{F}$ contiene a los “rectángulos” de la forma $\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) = A_1 \times A_2$.

12.1.3. Probabilidad producto

Ahora nos resta construir la medida de probabilidad $\mathbb{P}$ definida en $(\Omega, \mathcal{F})$.

La medida $\mathbb{P}$ a construir debe tener ciertas propiedades que nos permita recupera las leyes de cada experimento (recordar la definición de extensión de un espacio de probabilidad). Esto es,

$$\mathbb{P}(\pi_1^{-1}(A_1)) = \mathbb{P}_1(A_1), \quad \mathbb{P}(\pi_2^{-1}(A_2)) = \mathbb{P}_2(A_2),$$

para $A_i \in \mathcal{F}_i$.

Además, queremos que la medida $\mathbb{P}$, sea consistente con la noción de independencia. Esto es, que los eventos $\pi_1^{-1}(A_1)$ y $\pi_2^{-1}(A_2)$ sean independientes. De esta propiedad y la anterior resulta que $\mathbb{P}$ debe satisfacer

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

Teorema 12.1. *Existe una única medida $\mathbb{P}$ en Ω, denominada* medida producto*, que satisface la condición*

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2),$$

donde $A_i \in \mathcal{F}_i$ *para* $i = 1,2$.

Corolario 12.2. *El espacio* $(\Omega, \mathcal{F}, \mathbb{P})$ *así construido es una extensión de los espacios* $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ *y* $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$*, para las proyecciones canónicas* $\pi_i : \Omega \to \Omega_i$*,* $i = 1,2$.

Para el caso en que Ω_1 y Ω_2 sean espacios discretos la prueba del Teorema 12.1 es muy fácil.

- Primero observar que en este caso, en los espacios Ω_1 y Ω_2 podemos tomar como familia de eventos el conjunto de las partes de cada uno.
- Luego basta considerar la familia de todos los subconjuntos del producto $\Omega_1 \times \Omega_2$ como σ-álgebra $\mathcal{F}$, y como medida, la dada por
$$\mathbb{P}(A) = \sum_{(\omega_1,\omega_2)\in A} p_1(\omega_1) p_2(\omega_2),$$
donde $A \subset \Omega_1 \times \Omega_2$.

Es muy sencillo ver que esto define una medida en el producto.

La prueba del teorema en el caso general implica el *teorema de extensión de Carathedory*, la cual es un teorema básico del curso *teoría de la medida*. No haremos la prueba, pero al menos justificaremos que se requiere algo de trabajo.

Quizás sirva para tener un modelo en la cabeza de como se construye la medida uniforme en el cuadrado unidad.

Lo primero a observar es que la σ-álgebra generada por los rectángulos $A_1 \times A_2$, con A_1 y A_2 Borel medibles, contiene otras cosas que no son rectángulos. Por ejemplo el disco unidad $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\} \subset [0,1]^2$.

Como es claro que $\mathbb{D}$ no es un rectángulo, se necesita un argumento límite para calcularlo.

12.1.4. Construcción de finitas variables aleatorias independientes

Teorema 12.3. *Dadas dos funciones de distribución* F_1 *y* F_2*, existe un espacio de probabilidad, y variables aleatorias* X_1 *y* X_2 *independientes, con esas distribuciones.*

Demostración. Sea $\Omega = \mathbb{R}^2$, $\mathcal{F}$ la σ-álgebra de Borel. Definimos las medidas de probabilidad en $\mathbb{R}$ que satisfacen
$$\mathbb{P}_1((a,b]) = F_1(b) - F_1(a), \quad \mathbb{P}_2((a,b]) = F_2(b) - F_2(a).$$

Sea $\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2$ la medida producto de $\mathbb{P}_1$ y $\mathbb{P}_2$, y sea $X : \Omega \to \mathbb{R}^2$ es el vector aleatorio definido por $X = (X_1, X_2)$ donde $X_i(\omega_1, \omega_2) = \omega_i$, con $i = 1, 2$. Entonces, X_1 y X_2 satisfacen que son independientes con las distribuciones buscadas.. $\square$

12.2. Medida producto: caso infinito

En esta sección se sugiere la construcción de la medida producto en el producto infinito de espacios de probabilidad.

Un estudio completo de este problema se hace en el curso de *Procesos estocásticos*.

12.2.1. Construcción de infinitas variables aleatorias independientes

El objetivo de esta sección es construir una sucesión infinita de variables aleatorias $X_1, X_2, \dots$ que sean independientes con una distribuciones $F_1, F_2 \dots$ dadas.

Procedamos como en el caso finito.

Sean $\{\mathbb{P}_i\}_{i \in \mathbb{N}}$ medidas de probabilidad inducidas en $\mathbb{R}$ por las distribuciones $\{F_i\}_{i \in \mathbb{N}}$. Esto es, medidas de probabilidad $\mathbb{P}_i$ que satisfacen $\mathbb{P}_i((a, b]) = F_i(b) - F_i(a)$.

Consideremos el espacio producto

$$\Omega := \mathbb{R}^{\mathbb{N}} := \{\omega : \mathbb{N} \to \mathbb{R}\},$$

donde escribimos $\omega_i := \omega(i)$, $i \in \mathbb{N}$. (Recordar digresión sobre espacios productos en página 18.)

Sea $\mathcal{F}$ la σ-álgebra producto generado por los rectángulos finito-dimensionales

$$\{\omega \in \Omega : \omega_i \in A_i, i \in I, \#I < \infty\},$$

para borelianos $A_i \subset \mathbb{R}$, $i \in I$. En otras palabras, un rectángulo es dar condiciones en finitas coordenadas de ω. (Recordar la Sección 8.2, para la noción de σ-álgebra generada.)

Motivado por lo anterior, buscamos una medida $\mathbb{P}$ definida en $\mathcal{F}$ que satisfaga

$$\mathbb{P}(\{\omega \in \Omega : \omega_i \in A_i, i \in I, \#I < \infty\}) = \prod_{i \in I} \mathbb{P}_i(A_i), \tag{12.1}$$

para borelianos $A_i \subset \mathbb{R}$, $i \in \mathbb{N}$.

Para probar la existencia de una medida con esta propiedad se utiliza el siguiente resultado conocido como *teorema de extensión de Kolmogorov*

Teorema 12.4. *Sean μ_n medidas de probabilidad definidas en $\mathbb{R}^n$ para la σ-álgebra de Borel. Supongamos que las medidas verifican la siguiente condición de consistencia*

$$\mu_{n+1}((a_1,b_1]\times\cdots\times(a_n,b_n]\times\mathbb{R}) = \mu_n((a_1,b_1]\times\cdots\times(a_n,b_n]).$$

Entonces, existe una única medida de probabilidad $\mathbb{P}$ en $\mathbb{R}^{\mathbb{N}}$, con σ-álgebra generada por rectángulos finito-dimensionales, que verifica

$$\mathbb{P}(\{\omega\in\Omega:\ \omega_i\in(a_i,b_i],\ i\in I,\ \#I<\infty\}) = \prod_{i\in I}\mu_i((a_i,b_i]).$$

Por lo tanto, si definimos $X_i:\Omega\to\mathbb{R}$, con $X_i(\omega):=\omega_i$, $i\in\mathbb{N}$, y definimos medidas finito-dimensionales que satisfagan la igualdad (12.1), resulta del Teorema 12.4, que existe $\mathbb{P}$ medida de probabilidad en $\mathbb{R}^{\mathbb{N}}$, tal que las v.a. X_i están definidas sobre Ω, con medida de probabilidad inducidas $\mathbb{P}_i$, que son independientes.

Observar que nuevamente hemos construido una extensión $(\Omega,\mathcal{F},\mathbb{P})$ de espacios de probabilidad $(\mathbb{R},\mathcal{B}_{\mathbb{R}},\mathbb{P}_i)$, para $i\in\mathbb{N}$.

Capítulo 13

Esperanza abstracta e integral de Lebesgue

En este apéndice daremos algunas definiciones y enunciados (sin demostración) sobre la definición de abstracta de esperanza.

Recordar que hemos visto dos definiciones de *esperanza* de una v.a., a saber

$$\begin{aligned} \mathbb{E}(X) &= \sum x p_X(x), \quad \text{cuando } X \text{ es discreta;} \\ \mathbb{E}(X) &= \int x f_X(x)\, dx \quad \text{cuando } X \text{ es continua.} \end{aligned}$$

La idea es tener una definición de esperanza que se pueda extender a variables aleatorias en general, que no tienen porqué ser continuas o discretas. Por ejemplo, si $X \sim \mathcal{B}(1/2)$, y U tiene distribución uniforme en $[0,1]$, y son independientes, entonces la v.a.

$$Y = \begin{cases} 1 & \text{si } X = 1, \\ U & \text{si } X = 0, \end{cases}$$

no es ni continua ni discreta. Esto resulta de que toma una cantidad no numerable de valores, pero sin embargo $\mathbb{P}(Y = 1)$ es $1/2$. (La v.a. Y puede pensarse como el resultado de tirar una moneda al azar, y según el resultado tirar, o no, una uniforme.)

13.1. Integración abstracta, y esperanza

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad.

Definición. Una v.a. $X : \Omega \to \mathbb{R}$ se dice *simple* si toma sólo una cantidad finita de valores. Esto es, X puede escribirse por

$$X = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k},$$

donde $A_k \in \mathcal{F}$, $k = 1, \ldots, n$.

En la siguiente sección veremos que las v.a. simples en un espacio de probabilidad son exactamente las v.a. discretas.

Definición. Sea $X = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}$, una v.a. simple. Definimos la *integral de Lebesgue* de la v.a. simple X, (o su *esperanza*), como el número dado por

$$\mathbb{E}(X) := \int_{\Omega} X \, d\mathbb{P} := \sum_{k=1}^{n} x_k \mathbb{P}(A_k).$$

Esta definición es consistente, en el sentido que no depende de la representación de X como combinación lineal de indicatrices.

Ahora haremos algunas afirmaciones, en forma de lemas, que daremos las ideas de la prueba pero evitando justificar los detalles.

Lema 13.1. *Toda v.a. X positiva, i.e. $X : \Omega \to [0, +\infty)$, puede ser aproximada, por debajo, por una sucesión de v.a. simples X_n. Esto es, para todo $\omega \in \Omega$, se tiene que $\{X_n(\omega)\}$ es una sucesión creciente tal que,*

$$\lim_n X_n(\omega) = X(\omega),$$

Demostración. Sea

$$X_n = \min\left\{\frac{[2^n X]}{2^n}, n\right\}, \quad n \in \mathbb{N}.$$

Observar que X_n tiene la siguiente descripción explícita,

$$X_n = \sum_{i=0}^{n2^n} \frac{i}{2^n} \mathbb{1}_{X \in [\frac{i}{2^n}, \frac{i+1}{2^n})}.$$

y por tanto es una v.a. simple. Es fácil ver que además es creciente, con límite puntual X.

□

Definición. Dada una v.a. $X : \Omega \to [0,\infty)$, definimos su *integral de Lebesgue de* X, (o *esperanza de* X), como

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

donde X_n es alguna sucesión creciente de v.a. simples que converge X.

Se puede probar que la definición es consistente en el sentido de que la expresión anterior es independiente de la sucesión creciente de simples que uno tome. Esto es, si $\{Y_n\}$ es otra sucesión de v.a. simples que convergen a X, entonces se tiene $\lim_n X_n(\omega) = \lim_n Y(\omega)$.

Observar también que puede suceder que $\mathbb{E}(X) = +\infty$.

Lema 13.2. *Toda v.a.* $X : \Omega \to \mathbb{R}$ *puede escribirse como* $X = X^+ - X^-$*, donde* X^+ *y* X^- *son v.a. positivas.*

Demostración. Para probar lo anterior basta tomar $X^+(\omega) := \max\{X(\omega), 0\}$ y $X^- := \max\{-X(\omega), 0\}$. □

Las v.a. X^+ y X^-, se llaman *parte positiva* y *parte negativa* respectivamente.

Definición. Sea $X : \Omega \to \mathbb{R}$ una v.a., y sean X^+ y X^- sus partes positiva y negativa respectivas. Si ambas tienen integral de Lebesgue (o esperanza) finita, entonces se define

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

Esta definición de esperanza verifica las siguientes propiedades que hemos venido usando a lo largo del curso.

- Si $X = c$ entonces $\mathbb{E}(X) = c$.
- Si $X \leqslant Y$ entonces $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.
- Si $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
- Si $A \in \mathcal{F}$ es un evento, entonces $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

13.1.1. Relación con esperanza discreta y continua

Para finalizar veamos que esta definición extiende a la definición de esperanza que hemos visto.

Veamos primero el caso discreto.

Sea X una v.a. discreta definida sobre $(\Omega, \mathcal{F}, \mathbb{P})$. Supongamos X toma los valores x_k con probabilidad p_k. Entonces definiendo los eventos $A_k = X^{-1}(x_k)$, resulta que X es la v.a. simple

$$X = \sum_k x_k \mathbb{1}_{A_k},$$

donde $\mathbb{P}(A_k) = \mathbb{P}(X = x_k) = p_k$.

Entonces, la integral de Lebesgue de X resulta

$$\int_\Omega X \, d\mathbb{P} = \sum_k x_k \mathbb{E}(\mathbb{1}_{A_k}) = \sum_k x_k \mathbb{P}(A_k) = \sum_k x_k p_k,$$

lo cual coincide con nuestra definición de esperanza para v.a. discretas. Entonces

El caso continuo requiere un poco más de trabajo. Lo que queremos probar es lo siguiente.

Proposición 13.3. *Si X es una v.a. continua con función densidad f_X se tiene que*

$$\int_\Omega X \, d\mathbb{P} = \int_{-\infty}^{+\infty} x f_X(x) \, dx.$$

Demostración. Para probar esto supongamos que $X \geqslant 0$. (De lo contrario, se trabaja con su parte positiva y negativa.)

Sea

$$X_n = \text{mín}\left\{ \frac{[2^n X]}{2^n}, n \right\}, \quad n \in \mathbb{N}.$$

Entonces, de la definición de la integral de Lebesgue resulta

$$\int_\Omega X \, d\mathbb{P} = \text{lím}_n \int_\Omega X_n \, d\mathbb{P} = \text{lím}_n \sum_{i=0}^{n2^n} \frac{i}{2^n} \mathbb{P}\left(X \in \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) \right).$$

Cuando n es grande, podemos aproximar la probabilidad anterior de la siguiente manera:

$$\mathbb{P}\left(X \in \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) \right) = \int_{i/2^n}^{(i+1)/2^n} f_X(x) \, dx \approx f_X\left(\frac{i}{2^n} \right) \frac{1}{2^n}.$$

Luego resulta

$$\begin{aligned} \text{lím}_n \int_\Omega X_n \, d\mathbb{P} &= \text{lím}_n \sum_{i=0}^{n2^n} \frac{i}{2^n} f_X\left(\frac{i}{2^n} \right) \frac{1}{2^n} \\ &= \int_0^{+\infty} x f_X(x) \, dx. \end{aligned}$$

Concluimos de esta manera nuestro resultado. □

13.1.2. Un resultado general

El resultado que vimos es un caso particular de un teorema más fuerte que enunciaremos a continuación para cerrar este apéndice.

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, y una v.a. $X : \Omega \to \mathbb{R}$, podemos definir un nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, donde $\mathcal{B}$ es la σ-álgebra de Borel, y $\mathbb{P}_X$ es la medida de probabilidad inducida, i.e., $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$, para todo $B \in \mathcal{B}$.

Teorema 13.4. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y sea $X : \Omega \to \mathbb{R}$ una v.a. Entonces, si $g : \mathbb{R} \to \mathbb{R}$ es una función medible Borel, se tiene*

$$\int_\Omega g(X)\, d\mathbb{P} = \int_\mathbb{R} g\, d\mathbb{P}_X,$$

donde la última integral debe entenderse como la integral de Lebesgue respecto a la medida $\mathbb{P}_X$.

Proposición 13.5. *Si X es una v.a. continua, con densidad f_X, y $g : \mathbb{R} \to \mathbb{R}$ es una función medible Borel, entonces se tiene*

$$\int_\mathbb{R} g\, d\mathbb{P}_X = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx.$$

Las pruebas de estos resultados son similares y todas pasan por estudiar primero qué sucede si g es una función simple, para luego pasar al límite. La razón de esta estrategia es simplemente que así se define la integral de Lebesgue.

Observar que la Proposición 13.3 es un corolario del Teorema 13.4 y de la Proposición 13.5, tomando como caso particular g la función identidad, i.e. $g(x) = x, x \in \mathbb{R}$. Observar además que nuestra prueba se basó en aproximar la identidad por la función simple $g_n(x) = \text{mín}\{[2^n x]/2^n, n\}$.

Bibliografía

[LB] L. Breiman, *Probability*.

[F1] Feller, *Introducción al a teoría de las probabilidades y sus aplicaciones, Vol 1*.

[F2] ———, *Introducción al a teoría de las probabilidades y sus aplicaciones, Vol 2*.

[GW] Grimmet and Welsh, *Probability, an introduction*.

[K] Keller, *The probability of Heads.*, American Mathematical Monthly, 1986.

[M] Mcmullen, *Probability theory*, Harvard course notes, 2011, (http://www.math.harvard.edu/~ctm/math101/www/home/text/class/harvard/154/11/html/home/course/course.pdf).

[MP] Mordecki and Petrov, *Teoría de probabilidades*.

[KR] Klain and Rota, *Introduction to geometric probability*.

[R] Rozanov, *Probability theory: a concise course*.

[S1] Santalo, *Integral geometry and geometric probability*.

[S2] Shiryaev, *Probability*.

[SK] Suhov and Kelbert, *Probability and Statistics by Example, Vol 1*.

[W] Weber, *Probability*, Cambridge course notes, 2014 (http://www.statslab.cam.ac.uk/~rrw1/prob/prob-weber.pdf).

Índice alfabético