CAPÍTULO 1

Marco Teórico y Teoría Básica

1.0.- Introducción

La Teoría de Probabilidades es una rama de las matemáticas que nació como consecuencia de un esfuerzo por modelar juegos de azar; en particular, juegos de dados. Los primeros intentos, documentados, fueron realizados por Cardano y Galileo. Posteriormente, Pascal y Fermat se enfocaron a resolver el problema de determinar apuestas justas o ventajosas en juegos de azar. Estos esfuerzos están relatados amenamente por F. N. David (1962). Los juegos de azar eran conocidos desde la antigüedad,[1] pero el primer tratado matemático sobre juegos de azar, del que tenemos conocimiento, fue publicado en 1663. Notamos que el tiempo que tomó teorizar algo tan familiar como los juegos de azar fue grande y que teorizar o modelar no es una actividad que el común de la gente hagamos.

En el siglo 18 el estudio matemático de los juegos de azar se intensificó y personas como Montmort, formularon problemas que eran resueltos por medio de métodos ingeniosos. En ese tiempo los matemáticos renombrados miraban como mera curiosidad ese tipo de estudio. Esta apreciación cambió radicalmente cuando personas como De Moivre, Gauss y Laplace extendieron la teoría y las aplicaciones. Actualmente, la Teoría de Probabilidades está incluida en los planes de estudio de carreras como Administración de Empresas, Medicina, Economía, Matemáticas, Física, Actuaría y las distintas Ingenierías. Nos preguntamos por qué ¿cómo es posible que una rama del conocimiento sea utilizada en tan variadas actividades? Una respuesta es que los fenómenos *impredecibles* están presentes en todas las actividades humanas y es este tipo de fenómenos los que estudia la teoría de probabilidades.

Un fenómeno es *impredecible* cuando no podemos anticipar lo que va a suceder; solamente cuando ya aconteció sabemos lo que ocurrió. El fenómeno puede cambiar de repente *sin alguna explicación*. Pudimos haber observado el fenómeno en el pasado (como el índice de precios y cotizaciones de la bolsa, el clima, el ritmo cardiaco de una persona, etc.) y aun así, no lo podemos predecir con precisión. En el centro está la *variabilidad* de lo que observamos. No tenemos *explicación* por limitación de información y/o por incapacidad humana; no podemos medir o controlar las variables que *afectan* a los fenómenos. Por ejemplo ¿por qué no podemos decir en qué día, hora y minuto un foco recién fabricado dejará de encender? Podemos pensar que existen demasiadas variables que afectan la longitud de vida de un foco y eso impide hacer una predicción exacta. Si fuera posible medir las variables más importantes que afectan la vida de un foco podríamos hacer una predicción válida bajo ciertas condiciones de uso. Aun así, tendríamos que excluir otras posibilidades como roturas accidentales, cambios repentinos en el voltaje, etc. ya que no están bajo nuestro control. Si pudiésemos controlar y medir todas las variables sería posible, al menos en principio, hacer una predicción razonable del tiempo de vida de un foco (o de una persona, de llantas de automóvil y en general de cualquier aparato o máquina).

Cuando algo acontece por razones desconocidas estamos acostumbrados a decir que sucedió por *azar*. Así, si lanzamos un dado honesto y observamos el número 3, decimos que el azar

---

[1] Suetonio en el libro Los Doce Césares nos relata que el emperador Augusto era jugador empedernido de dados. En el Museo Británico está exhibida una vasija griega antigua que muestra a dos hombres jugando a los dados

produjo el resultado. El azar, nombre elegante para designar a la ignorancia y limitación humanas, produce verbigracia que:

1. Los dueños de los casinos se enriquezcan ya que los apostadores no pueden predecir los resultados del póquer, ruleta, máquinas tragamonedas, etc. En principio, los dueños de los casinos tampoco pueden predecir los resultados individuales de cada apuesta, pero sí pueden predecir, con un margen de error razonable, el comportamiento global de las apuestas. La justificación de esta afirmación está dada por los Teoremas Límite que son discutidos en el capítulo 9.

2. Que los usuarios de computadoras vean paralizadas sus actividades por descomposturas del equipo, fallas en la energía eléctrica, etc. Usuarios importantes son los bancos que sin sistema no pueden, por ejemplo, pagar cheques ni proporcionar saldos de cuentas de tarjetas de crédito. Los ingenieros de sistemas no saben predecir el instante en que deben sustituir piezas desgastadas, cuando va a existir una falla de energía eléctrica, etc.

3. Que los consumidores recibamos productos defectuosos como son planchas, cafeteras, apagadores de luz, ropa, etc. Los fabricantes no pueden controlar al 100% la calidad de los productos fabricados.

4. Que año tras año haya cuantiosas pérdidas materiales y de vidas humanas causadas por fenómenos naturales como terremotos, huracanes, desbordamientos de ríos, deslizamientos de tierra y tornados. La especie humana no sabe predecir con exactitud ni cuándo ni dónde ocurrirá el siguiente fenómeno natural.

En la construcción de modelos matemáticos para describir un fenómeno, podemos apreciar el papel del azar en investigación. Los químicos han utilizado la fórmula

$$h(t;\alpha,\beta) = e^{-t\beta} - e^{-t\alpha} \tag{1.0.1}$$

para predecir la cantidad acumulada de una sustancia B como función del tiempo, t, en una reacción consecutiva de primer orden e irreversible en el que la sustancia A produce la sustancia B que a su vez produce la sustancia C $(A \to B \to C)$; α y β son constantes, por lo general desconocidas, que representan tasas de concentración. Al inicio de la reacción tenemos una unidad de A pero no existe cantidad alguna ni de B ni de C.

Es claro que (1.0.1), al igual que otras leyes físicas y químicas, es sólo una aproximación y por tanto no predice sin error. De hecho, la fórmula (1.0.1) presupone condiciones constantes de temperatura, presión, humedad, etc. y establece que la cantidad existente de la sustancia B después de t unidades de tiempo sólo depende del tiempo; esto es una simplificación grande ¿verdad? Es más, al observar la reacción y medir la cantidad de la sustancia B presente después de $t_1, t_2. \dots, t_n$ unidades de tiempo, los químicos han verificado que (1.0.1), con valores apropiados de α y β, aunque no predice exactamente, sí proporciona aproximaciones razonables ¿Por qué (1.0.1) no predice exactamente? Porque existen variables, no mantenidas constantes y que afectan la cantidad acumulada de la sustancia B, que no están incluidas en la fórmula. Si denotamos por $y(t)$ la cantidad acumulada de la sustancia B después de t unidades de tiempo tendremos que

$$y(t) = h(t;\alpha,\beta) + u(t,\boldsymbol{x}); \text{donde } \boldsymbol{x} = (x_1, \dots, x_m)$$

donde $u(t,\boldsymbol{x})$ es el azar, esto es, el efecto de las variables (representadas por el vector $\boldsymbol{x}$) no incluidas en el modelo y otras funciones de t. He de mencionar que $\boldsymbol{x}$ puede incluir el efecto de la medición (en el ejemplo, el medir la cantidad acumulada de la sustancia B) ya que no siempre es posible medir sin error. Este efecto de la medición es particularmente importante en aplicaciones sociales de la probabilidad.

La investigación científica trata de descubrir tanto la relación funcional, h, como las variables que deben estar en el modelo. Nos damos cuenta de que establecer un modelo adecuado es un proceso iterativo y que requiere ingenio, trabajo y conocimiento (y posiblemente algo más). El proceso es iterativo porque difícilmente al primer intento el científico establecerá el modelo adecuado. Actualmente el uso de *Machine Learning* está facilitando la labor.

En ocasiones el azar es introducido intencionalmente para eliminar preferencias, tendencias o comportamientos sistemáticos que pueden desvirtuar los resultados, así

1. Los juegos de salón involucran el revolver cartas, lanzar dados, mezclar fichas, etc. que tienen el objetivo de hacer que el resultado dependa de la habilidad o suerte del jugador y no de un factor externo (como el hacer trampa).

2. En distintos sorteos como loterías, rifas, etc., los números premiados son extraídos al azar para eliminar sospechas de trampas y dar igual oportunidad de ganar a todos los números.

3. En experimentación agrícola un terreno es dividido en parcelas y luego los tratamientos (variedad de semilla, tipo o cantidad de plaguicida, método de cultivo, tipo o cantidad de abono, etc.) son asignados al azar a las parcelas. El azar es introducido para eliminar la relación entre las parcelas y los tratamientos (es posible que algunas parcelas estén mejor regadas que otras, que existan diferencias no visibles entre las parcelas, etc.) y resulta entonces que algunas son favorables a algún(os) tratamiento(s)).

4. En el muestreo aleatorio los objetos a ser estudiados son seleccionados al azar con el fin de eliminar preferencias o inclinaciones hacia algún tipo de objetos. Si un gerente de producción selecciona una muestra de productos para estimar el porcentaje de defectuosos y resulta que el sueldo del gerente depende del porcentaje hallado en la muestra, sería comprensible, más no justificable, que la muestra fuese seleccionada enfocándola hacia los no defectuosos; obviamente, la estimación del porcentaje de defectuosos sería engañosa. Una selección al azar eliminará dicha preferencia.

5. Las emisoras de televisión por satélite añaden a la señal una onda que resulta aleatoria para los usuarios. De esta manera, no cualquier televisor puede captar la imagen. Las emisoras rentan a los suscriptores decodificadores que eliminan la onda introducida y hacen posible ver la imagen apropiada.

En el pasado, algunas personas se percataron de que varios fenómenos impredecibles presentan una *regularidad estadística* al observarlos repetidamente *bajo las mismas condiciones esenciales*. Esa regularidad estadística es aprovechada por la humanidad para hacer predicciones que eliminan, aunque sea parcialmente la incertidumbre. Verbigracia, las compañías aseguradoras pueden predecir (no sin error) el número de muertes en el grupo de sus asegurados, en un período de doce meses, debido a la regularidad estadística. Con el fin de ilustrar el concepto, consideremos el siguiente fenómeno impredecible: lanzar un dado (de seis caras) honesto y observar el número de la cara superior. Al repetir 750 veces el fenómeno impredecible, el que esto escribe observó el siguiente comportamiento de la *frecuencia* (número de veces que fue observada) y la *frecuencia relativa* (porcentaje de veces que fue observada) de la cara con el número 1; consulte los datos de la tabla 1.1 y la gráfica 1.1. Observamos que en los primeros 300 lanzamientos la frecuencia relativa varía fuertemente pero luego se estabiliza alrededor de

1/6=0.1667. La estabilización de la frecuencia relativa es precisamente la llamada regularidad estadística.

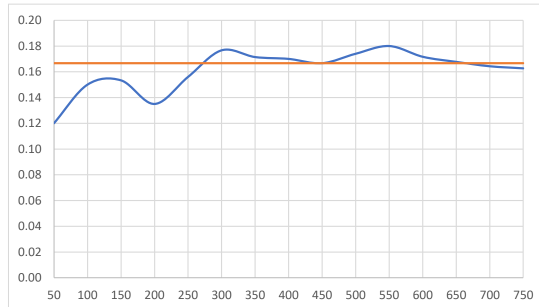
Tabla 1.1

Frecuencia y Frecuencia Relativa de la Cara con el Número Uno

Después de x lanzamientos	Frecuencia	Frecuencia Relativa
50	6	0.120
100	15	0.150
150	23	0.153
200	27	0.135
250	39	0.156
300	53	0.177
350	60	0.171
400	68	0.170
450	75	0.167
500	87	0.174
550	99	0.180
600	103	0.172
650	109	0.168
700	115	0.164
750	122	0.163

Gráfica 1.1

Comportamiento de la Frecuencia Relativa de la Cara con el Número Uno



En Teoría de Probabilidades decimos que *un fenómeno es aleatorio* cuando es impredecible, esto es, cuando es producido por el azar. La Teoría Clásica de Probabilidades modela el comportamiento de los fenómenos aleatorios y cuantifica la incertidumbre con una medida llamada *Probabilidad* y esta última es interpretada como una frecuencia relativa. De aquí que la teoría clásica sea también llamada frecuentista o estadística. Diversos autores no han estado de acuerdo con la interpretación estadística y sugieren que una probabilidad es una medida del grado de credibilidad (en algunos ejemplos de este capítulo discutiremos las dos interpretaciones) y se ha creado una corriente llamada *subjetiva*. El libro de Savage (1972) lo discute a profundidad.

A continuación, introducimos el concepto de Espacio de Probabilidad que es esencial para desarrollar el marco teórico.

1.1.- Espacio de Probabilidad

La Teoría de la Probabilidad plantea un marco teórico que es idéntico al de la *Teoría de la Medida*. Sin embargo, en Probabilidad se presentan aplicaciones que no tienen par en Teoría de la Medida. Adicionalmente, el origen experimental de la Probabilidad genera conceptos, como los de independencia y esperanza condicional y una interpretación de los resultados que no están presentes en Teoría de la Medida. En consecuencia, el estudio de la Probabilidad se ha constituido en una rama de las matemáticas valiosa por si misma. Aunque el presente volumen es de carácter introductorio, presentaremos formalmente el marco teórico. Es claro que el tratamiento no puede ser profundo. Los libros de Ash (1972), Cramér (1999), Kolmogorov (1956) y Loève (1963) sí lo ofrecen; lo mismo que varios sitios de internet.

El estudiante que es expuesto por vez primera al marco teórico, seguramente lo encontrará artificial y desligado de la realidad. Pero, para su tranquilidad, quiero mencionarle que el marco teórico que vamos a presentar fue ideado por Kolmogorov siendo publicado en 1933; después de más de 200 años de existencia del estudio matemático de las probabilidades. Grandes matemáticos como Gauss, Laplace y Lagrange hicieron contribuciones relevantes al cálculo de probabilidades, pero no establecieron el marco teórico como Kolmogorov. Esta situación es comparable con la formalización de la geometría hecha por Euclides, esto es, muchos resultados (teoremas) ya eran conocidos, pero no había un marco teórico (axiomas) que les diera sustento y formalización. Por tanto, la contribución de Kolmogorov formaliza el trabajo de muchos hombres y mujeres. Presentar el marco teórico al inicio del presente volumen es contrario al desarrollo histórico, pero tiene la ventaja de establecer las reglas del juego desde el principio. Por tanto, el lector debe comprender que el material de esta sección es una abstracción genial y por consiguiente no es natural. Empezamos el desarrollo con una definición.

Definición 1.1 – Espacio de Probabilidad

Un *Espacio de Probabilidad,* asociado a un fenómeno aleatorio $\mathcal{E}$, es una terna $(\Omega, \mathcal{A}, \mathrm{P})$ donde: Ω (letra omega mayúscula del alfabeto griego) es el *espacio muestral*, $\mathcal{A}$ (letra a mayúscula manuscrita) es una *sigma álgebra* (σ −álgebra) de subconjuntos de Ω y P es una *medida de probabilidad*. ■

Podemos pensar en que asociamos a cada fenómeno aleatorio $\mathcal{E}$ un espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Esto sugiere que debemos indexar las componentes de la terna con $\mathcal{E}$, $(\Omega_{\mathcal{E}}, \mathcal{A}_{\mathcal{E}}, P_{\mathcal{E}})$, pero con el fin de simplificar la notación no lo hacemos. Sin embargo, mantenga en mente que, si estudiamos otro fenómeno aleatorio, verbigracia $\mathcal{G}$, el espacio de probabilidad cambiará. Hacemos notar que la terna es ordenada, esto es, el primer signo, Ω, siempre representará al espacio muestral; el segundo signo, $\mathcal{A}$, representará la σ −álgebra y el tercero, P, la medida de probabilidad. Es posible asociar más de un espacio de probabilidad a un mismo fenómeno aleatorio $\mathcal{E}$. Por ejemplo, yo puedo proponer $(\Omega, \mathcal{A}, \mathrm{P}_0)$ y alguien más $(\Omega, \mathcal{A}, \mathrm{P}_1)$ donde P_0 y P_1 son medidas de probabilidad distintas.

A continuación, definimos cada una de las componentes de la terna y establecemos qué propiedades deben satisfacer. Mantendremos fijo, pero arbitrario, el fenómeno aleatorio $\mathcal{E}$ bajo estudio.

1.1.1.- Espacio Muestral: Ω

Debido a que los juegos de azar motivaron la teoría del cálculo de probabilidades, el espacio muestral fue originalmente definido como el conjunto de todos los posibles resultados de un fenómeno aleatorio. Aunque debemos aclarar que por conveniencia y simplicidad matemática algunos resultados son excluidos. Verbigracia, al considerar el fenómeno aleatorio $\mathcal{E}$ lanzar una moneda honesta, diremos que hay dos posibles resultados la cara A y la cara S. Entonces, el espacio muestral es $\Omega = \{\mathrm{A}, \mathrm{S}\}$. Observamos que hemos eliminado el resultado posible, la moneda no cae sobre una de las caras (esto no significa que haya caído de canto, sino que la moneda quedó recargada en algo que le impidió reposar sobre una de sus caras). Sin embargo, la eliminación no afecta significativamente las aplicaciones de la teoría ya que el resultado eliminado ocurre rara vez. Es conveniente hacer hincapié que la definición del espacio muestral puede resultar obvia o sencilla para experimentos aleatorios que tienen un número finito de posibles resultados, pero para aquellos que tienen un número infinito de posibles resultados, aunque sea de manera conceptual, la definición puede ser controversial. El espacio muestral es una abstracción que simplifica la realidad pero que debe incluir los posibles resultados relevantes (aunque lo que es relevante para una persona puede no serlo para otra). Verbigracia, al considerar el fenómeno aleatorio $\mathcal{E}$: tiempo, medido en años, que vivirá una persona que va a nacer, algunos estaremos inclinados a proponer $\Omega = [0, 120]$ que representa que una persona vivirá a lo más 120 años. Otros podrán proponer $\Omega = (0{,}100)$ o bien $\Omega = [0, \infty)$. Aunque $\Omega = [0, \infty)$ parece totalmente fuera de orden, puede resultar adecuado si la probabilidad de que la persona sobreviva más de 120 años es muy pequeña. Para concluir, diremos que el espacio muestral Ω es un conjunto que contiene los resultados posibles de interés de un fenómeno aleatorio. Sugiero consultar la definición de espacio muestral en los libros listados en la Bibliografía al final del libro y diversos sitios de internet. A continuación, presentamos algunas instancias.

Ejemplo 1.1

$\mathcal{E}$: Lanzar una vez un dado honesto; entonces podemos proponer, $\Omega = \{1, 2, 3, 4, 5, 6\}$ ya que los resultados posibles son los dígitos del uno al seis. Hemos eliminado los resultados donde el dado no cae sobre uno de sus lados. ∎

Ejemplo 1.2

$\mathcal{E}$: Lanzar dos veces un dado honesto. Si eliminamos los resultados donde el dado no cae sobre uno de sus lados podemos proponer $\Omega = \{(i, j); i, j \in \mathbb{N}, \text{con } i, j \leq 6\}$ ¡Deténgase lector! observe que (1,6) y (6,1) son distintos. Volvamos a Ω; es el conjunto de pares ordenados donde la primera coordenada es un número natural menor o igual a 6 y la segunda coordenada es un número natural menor o igual a 6 ¿Qué significa la primera coordenada? Es el resultado del primer lanzamiento. La segunda coordenada es el resultado del segundo lanzamiento. Ahora ve que (1,6) y (6,1) son distintos ¿verdad? En realidad,

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \equiv \{1, 2, 3, 4, 5, 6\}^2$$

donde $\times$ significa producto Cartesiano. El primer conjunto $\{1, 2, 3, 4, 5, 6\}$ representa los posibles resultados del primer lanzamiento y el segundo los posibles resultados del segundo lanzamiento. Si $\mathcal{E}$ fuese: lanzar tres veces un dado honesto, el espacio muestral sería $\{1, 2, 3, 4, 5, 6\}^3$ y en general con k lanzamientos, $k \in \mathbb{N}$, el espacio muestral sería $\{1, 2, 3, 4, 5, 6\}^k$.

El ejemplo 1.2 pone de manifiesto que si un fenómeno aleatorio $\mathcal{E}$ tiene asociado un espacio muestral Ω entonces, al repetir k veces, $k \in \mathbb{N}$, el fenómeno aleatorio el espacio muestral asociado a las k repeticiones es $\Omega_k = \Omega^k$, el producto Cartesiano de Ω consigo mismo k veces; sin embargo, debemos notar que las repeticiones de $\mathcal{E}$ no deben alterar los posibles resultados de repeticiones posteriores. Verbigracia, si después de la primera ejecución de $\mathcal{E}$, el resultado observado, ya no podrá volver a ocurrir después entonces, ya no será cierto que $\Omega_k = \Omega^k$. Observemos que los elementos de Ω_k son vectores de k coordenadas; siendo la j −ésima coordenada el resultado de la repetición j de $\mathcal{E}$ con $1 \leq j \leq k$. ∎

Ejemplo 1.3

$\mathcal{E}$: Lanzar una vez un par de dados honestos. Con la simplificación mencionada en los ejemplos 1.1. y 1.2 podemos considerar dos situaciones: i) Los dados son distinguibles y ii) Los dados no son distinguibles, esto es, sus características físicas no permiten distinguirlos. i) Si los dados son distinguibles, verbigracia, uno es rojo y el otro azul entonces,

$$\Omega_1 = \{(i, j); i, j \in \mathbb{N}, i, j \leq 6\} = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

donde podemos convenir que la primera coordenada representa el resultado observado en el dado rojo y la segunda coordenada representa el resultado del dado azul. Así, no es lo mismo observar (1,6) que (6,1). ii) Si los dados no son distinguibles, podríamos proponer

$$\Omega_2 = \{(i, j); i, j \in \mathbb{N}, i \leq j \leq 6\}$$

donde la primera coordenada representa el menor valor observado y la segunda el mayor valor observado. En esta situación no habría (1,6) y (6,1) sólo habría (1,6). Es conveniente tener en cuenta las condiciones bajo las cuales propusimos Ω_1 y Ω_2. Ω_1 tiene 36 posibles resultados mientras que Ω_2 tiene 21 y esta diferencia produce discrepancias serias en el momento de calcular probabilidades. La experiencia ha verificado que la primera conceptualización, Ω_1, es más apropiada que la segunda para calcular probabilidades.

Es conveniente mencionar que si los dados no son distinguibles podríamos proponer un espacio muestral diferente; consulte el ejemplo 2.22 y haga $n = 2$.

¿Podría el lector generalizar el ejemplo? esto es, si $\mathcal{E}$ es lanzar una vez k dados honestos ¿cómo definiría el espacio muestral? Tome en cuenta las consideraciones i) y ii). ■

Ejemplo 1.4

$\mathcal{E}$: Examinar una caja con 50 fusibles de la producción de una fábrica y contar el número de defectuosos. Para llevar a cabo $\mathcal{E}$ es indispensable que tengamos una definición clara de fusible defectuoso. Podemos proponer

$$\Omega = \{0,1,2,\dots,50\}.$$

Muchos fabricantes no estarían de acuerdo con la definición de Ω ya que si en 50 fusibles puede haber 50 defectuosos la fábrica quiebra. Posiblemente, los fabricantes estarían dispuestos a aceptar $\Omega = \{0,1,2,3,4,5\}$, que refleja el hecho de que a lo más el 10% de los fusibles es defectuoso. La teoría de probabilidades puede incluir el hecho de que no más del 10% de los fusibles es defectuoso, pero no considera prudente hacerlo en la definición del espacio muestral. En principio, aunque sea de manera conceptual, los 50 fusibles pueden ser defectuosos, que el hecho sea poco probable es otro asunto. ■

Ejemplo 1.5

$\mathcal{E}$: Seleccionar al azar un estudiante de un salón de clases universitario y medirle la estatura. Como el resultado del experimento es la estatura, medida en metros, podemos proponer varias alternativas para el espacio muestral. Verbigracia,

$$\Omega_1 = \{x \in \mathbb{Q}; x = 1 + a_1 10^{-1} + a_2 10^{-2}, \text{con } a_1, a_2 \in \mathbb{N}_0 \text{ y } a_1, a_2 \leq 9\}$$

Donde $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Ω_1 nos dice que la estatura está medida hasta centímetros, con estatura mínima de un metro $(a_1 = 0 = a_2)$ y estatura máxima de un metro con noventa y nueve centímetros $(a_1 = 9 = a_2)$. Este espacio muestral es adecuado para la población mexicana, aunque un metro como estatura mínima puede resultar pequeña. Ω_1 tiene 100 posibles resultados que se derivan de la limitación en la medición de la estatura hasta centímetros. Como en principio, aunque sea de manera conceptual, la estatura es una variable continua también podemos proponer

$$\Omega_2 = [1.00, 1.99] = \{x \in \mathbb{R}; 1.00 \leq x \leq 1.99\}$$

que refleja la posibilidad de medir la estatura con precisión absoluta, con estatura mínima de un metro y máxima de un metro noventa y nueve centímetros. Otras opciones serían: $\Omega_3 = (0, 100) = \{x \in \mathbb{R}; 0 < x < 100\}$ y $\Omega_4 = (0, \infty) = \{x \in \mathbb{R}; 0 < x\}$. Podemos pensar que Ω_3 y Ω_4 son exageraciones, no sólo por la medición con precisión absoluta, pero los dos contienen los resultados posibles de interés. En las aplicaciones del cálculo de probabilidades se ha acostumbrado a proponer espacios muestrales con un número infinito no numerable de posibles resultados, como Ω_2, Ω_3 y Ω_4, para variables que en principio son continuas porque, paradójicamente, los métodos de cálculo resultan más sencillos.

Si deseáramos ejecutar $\mathcal{E}$ tendríamos que saber cómo seleccionar al azar un estudiante; pero ¿qué significa seleccionar al azar un estudiante? En el capítulo 2 discutiremos este espinoso asunto. ■

Ejemplo 1.6

$\mathcal{E}$: Disparar un cañón que apunta al punto (x_0, y_0) y registrar las coordenadas del lugar de impacto. Si suponemos que podemos medir adecuadamente las coordenadas del lugar de impacto es razonable proponer

$$\Omega = \{(x, y); (x - x_0)^2 + (y - y_0)^2 \leq r^2\}$$

que indica que el lugar del impacto distará del punto (x_0, y_0) a lo más en r unidades (verbigracia, metros). Ω tiene un número infinito no numerable de posibles resultados. ■

En los ejercicios del presente capítulo y en los capítulos 2 y 3 daremos más ejemplos. Al repasar las ilustraciones presentadas nos damos cuenta de que los elementos del espacio muestral pueden ser números, vectores, letras (como A o S), etc. y debido a la diversidad los llamaremos en forma genérica *resultados elementales*. De esta manera, $(5, 5)$ es un resultado elemental en el ejemplo 1.2 y 1.81 es un resultado elemental en el ejemplo 1.5. A los resultados elementales los denotaremos con la letra griega omega minúscula, ω.

Diremos que un espacio de probabilidad es *discreto* cuando el espacio muestral tiene un número finito o numerable de resultados elementales. Un espacio de probabilidad es *no discreto* cuando el espacio muestral tiene un número infinito no numerable de resultados elementales. A continuación, presentamos el segundo componente de la terna que define a un espacio de probabilidad.

1.1.2.- $\boldsymbol{\sigma}$ −Álgebra: $\boldsymbol{\mathcal{A}}$

La incorporación de esta componente tiene una razón técnica bien precisa. En principio, uno desearía poderle asignar una probabilidad positiva a cada resultado elemental. Sin embargo, como veremos más adelante, esto no es posible en general. La limitación se produce con los espacios de probabilidad no discretos y es precisamente por ellos que se introdujo la σ −álgebra. El razonamiento fue si no podemos asignarle una probabilidad positiva a cada resultado elemental ¿será posible definir una colección de subconjuntos de Ω a los que sí podamos asignarles probabilidades positivas y resulte útil en la práctica? Pero, por otra parte, la colección de subconjuntos buscada no debería imponer restricciones a los espacios de probabilidad discretos (que no tienen la limitación mencionada anteriormente) a menos que tuviésemos dos tipos de marcos teóricos, uno para espacios de probabilidad discretos y otro para los no discretos. Felizmente, se encontró que sí se puede definir una colección de subconjuntos del espacio muestral que satisfaga las condiciones deseadas y ésta fue la σ −álgebra que definimos a continuación.

Definición 1.2 – Sigma Álgebra

Sean Ω un espacio muestral y $\mathcal{A}$ una colección de subconjuntos de Ω. Decimos que $\mathcal{A}$ es una σ −*álgebra* si satisface las siguientes propiedades

SA1.- $\Omega \in \mathcal{A}$.

SA2.- Si $A \in \mathcal{A}$ entonces, $A^c = \Omega - A$ debe ser un elemento de $\mathcal{A}$, esto es, si $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

SA3.- Si $A_i \in \mathcal{A}$ para $i \in \mathbb{N}$ entonces, $\bigcup_{i=1}^{\infty} A_i$ debe ser un elemento de $\mathcal{A}$, esto es, si $A_k \in \mathcal{A}$ para $k \in \mathbb{N} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$. ■

Los elementos de $\mathcal{A}$ son denominados *eventos* y sólo a ellos les podremos calcular una probabilidad. Para designar eventos utilizaremos generalmente las primeras letras del abecedario (de la A a la H) y siempre en mayúsculas. Si B es un evento entonces $B \in \mathcal{A}$ y necesariamente $B \subset \Omega$. Por otra parte, diremos *que un evento C ocurre* cuando el resultado del experimento o fenómeno aleatorio bajo estudio $\mathcal{E}$ es el resultado elemental ω y ω pertenece a C (esto es, $\omega \in C$). Con estas consideraciones, la definición de la σ −álgebra establece que:

a) Debemos ser capaces de calcular la probabilidad del *evento seguro*, el que siempre ocurre, Ω. Ya que si ω es el resultado de observar o ejecutar $\mathcal{E}$, entonces ω debe ser un elemento de Ω (de aquí que Ω contenga los resultados elementales de interés).

b) Si podemos calcular la probabilidad de que un evento A ocurra entonces, también debemos poder calcular la probabilidad de que *no ocurra* A, esto es, de A^c. En particular, debemos poder calcular la probabilidad de $\Omega^c = \emptyset =$ el *evento imposible*.

c) Si podemos calcular la probabilidad de que ocurra el evento A_k para $k = 1, 2, 3, \ldots$ entonces, también debemos estar en condiciones de calcular la probabilidad de que ocurra *al menos uno* de los A_k, esto es, de $\bigcup_{k=1}^{\infty} A_k$.

Notamos que el álgebra de eventos es igual al álgebra de conjuntos, pero algo que debemos observar es la interpretación de los eventos:

- A^c es interpretado como la no ocurrencia de A.
- $A \cap B$ es entendido como A y B ocurren. Si $A \cap B = \emptyset$ entonces, ocurre A u ocurre B pero no los dos; son eventos *excluyentes*.
- $A \subset B$ significa que siempre que ocurre A ocurre B.
- $A \cup B$ es interpretado como al menos uno de los dos A o B ocurre (o inclusiva no exclusiva).

Enseguida presentamos un caso.

Ejemplo 1.7

Supongamos que un fenómeno aleatorio $\mathcal{E}$ tiene asociado el espacio muestral $\Omega = \{1,2,3,4\}$. En esta situación estamos en posibilidad de calcularle la probabilidad a cualquier subconjunto de Ω. Es más, la colección de todos los subconjuntos de Ω, denotada por 2^{Ω}, está dada por

$$2^{\Omega} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \Omega\}$$

Es sencillo verificar que 2^{Ω} es una σ −álgebra (aunque no es la única). Observe que la tercera propiedad del σ −álgebra afirma que si $A_k \in \mathcal{A}$ para $k \in \mathbb{N}$ entonces, $\bigcup_{k=1}^{\infty} A_k$ debe ser un elemento de $\mathcal{A}$. Para obtener una colección infinita de subconjuntos de $\Omega = \{1,2,3,4\}$ es claro que al menos uno de los subconjuntos se tiene que repetir un número infinito de veces y que la unión a fin de cuentas será un elemento de 2^{Ω}. Otras σ −álgebras son $\mathcal{A}_0 = \{\emptyset, \Omega\}$ y $\mathcal{A}_1 = \{\emptyset, \Omega, A, A^c\}$, donde A es un subconjunto de Ω.

Pensemos ahora en una σ −álgebra que tenga como elementos a dos subconjuntos de Ω que designaremos como A y B; verbigracia, $A = \{1\}$ y $B = \{2,3\}$. La sigma álgebra *generada* por A y B, denotada por $\mathcal{A}(A, B)$, es

$$\mathcal{A}(A, B) = \mathcal{A}(\{1\}, \{2,3\}) = \{\Omega, \emptyset, \{1\}, \{2,3\}, \{1,2,3\}, \{2,3,4\}, \{1,4\}, \{4\}\} \subset 2^{\Omega}.$$

Una pregunta interesante es ¿cómo compara $\mathcal{A}(A, B)$ con $\mathcal{A}(A, B, C)$ donde C es otro subconjunto de Ω? Es claro que $\mathcal{A}(A, B) \subset \mathcal{A}(A, B, C)$, pero ¿podemos afirmar algo más? Más adelante nos enfocamos en esta última pregunta. ■

En el ejemplo 1.7 se estableció que 2^{Ω} (*el conjunto potencia* de Ω que es la colección de todos los subconjuntos de Ω) es una σ −álgebra. Esta afirmación no sólo es válida para el Ω discutido sino para cualquier espacio muestral. La σ −álgebra más *pobre* es $\{\Omega, \emptyset\}$ y la más *rica* es 2^{Ω}; el ejemplo 1.7 mostró que hay otras intermedias (¿cuántas σ −álgebras tendrá un espacio muestral finito?). A continuación, veremos qué tan rica, en términos de elementos, puede ser una σ −álgebra.

Lema 1.1

Si $A_k \in \mathcal{A}$ para $k \in \mathbb{N}$ entonces, $\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$ y en particular, para cualesquiera n eventos $A_{k_1}, A_{k_2}, \dots, A_{k_n}$ tendremos que $\bigcap_{i=1}^{n} A_{k_i} \in \mathcal{A}$.

Demostración

Como $A_k \in \mathcal{A}$, SA2 garantiza que $A_k^c \in \mathcal{A}$ para toda k. La propiedad SA3 nos permite afirmar que $\bigcup_{k=1}^{\infty} A_k^c \in \mathcal{A}$ y al aplicar SA2 obtenemos que $(\bigcup_{k=1}^{\infty} A_k^c)^c \in \mathcal{A}$. Pero $(\bigcup_{k=1}^{\infty} A_k^c)^c = \bigcap_{k=1}^{\infty} A_k$ por las leyes de De Morgan. El caso particular se obtiene de manera análoga y es dejado como un ejercicio. ■

El lema 1.1 junto con SA2 y SA3 establecen que una σ −álgebra es *cerrada* bajo las operaciones complementación, intersección y unión. Por otra parte, si A y B son eventos entonces, $(A - B)$ y $(B - A)$ son también eventos ya que $A - B = A \cap B^c$ y como $B^c \in \mathcal{A}$, el lema 1.1 nos garantiza que $A \cap B^c \in \mathcal{A}$. La *diferencia simétrica* $(A - B) \cup (B - A)$ también es un evento.

Decimos que los eventos $A_1, A_2, \dots$ son *mutuamente excluyentes* cuando $A_i \cap A_j = \emptyset$ para todo par i, j cuando $i \neq j$; esto significa que los eventos A_i y A_j no ocurren. A continuación, mostramos un resultado que será utilizado en diversas partes del texto.

Lema 1.2

Sean $A_1, A_2, \dots$ eventos en un espacio de probabilidad entonces, $\bigcup_{n=1}^{\infty} A_n$ puede ser expresada como la unión de eventos mutuamente excluyentes.

Demostración

Definimos $B_1 = A_1$ y $B_2 = A_2 \cap B_1^c$. Entonces, $B_1 \cap B_2 = \emptyset$ y $B_1 \cup B_2 = A_1 \cup A_2$. Ahora,

$$B_3 = A_3 - (B_1 \cup B_2) = A_3 \cap B_1^c \cap B_2^c$$

$\Rightarrow B_1 \cap B_3 = \emptyset = B_2 \cap B_3$ y $B_1 \cup B_2 \cup B_3 = A_1 \cup A_2 \cup A_3$. En general,

$$B_k = A_k - \bigcup_{j=1}^{k-1} B_j = A_k \cap B_1^c \cap \dots \cap B_{k-1}^c, \text{para } k \geq 2.$$

Donde $B_j \cap B_k = \emptyset$ para $j < k$ y por construcción, $B_i \cap B_j = \emptyset$ para $1 \leq i, j \leq k-1$ con $i \neq j$. Además, $\bigcup_{j=1}^{k-1} B_j = \bigcup_{j=1}^{k-1} A_j$. En consecuencia,

$$\bigcup_{j=1}^{k} B_j = \left[\bigcup_{j=1}^{k-1} A_j\right] \cup B_k = \left[\bigcup_{j=1}^{k-1} A_j\right] \cup \left[A_k - \bigcup_{j=1}^{k-1} A_j\right] = \bigcup_{j=1}^{k} A_j.$$

Por tanto, para $N \in \mathbb{N}$ con $N \geq 2$ tenemos que

$$\bigcup_{k=1}^{N} A_k = \bigcup_{k=1}^{N} B_k \text{, donde } B_j \cap B_i = \emptyset \text{ para } 1 \leq i, j \leq N \text{ con } i \neq j;$$

$$= A_1 \cup \bigcup_{k=2}^{N-1} \left[A_k - \bigcup_{j=1}^{k-1} A_j\right] \text{, por la definición de } B_k;$$

$$= A_1 \cup \bigcup_{k=2}^{N-1} [A_k \cap A_1^c \cap \ldots \cap A_{k-1}^c].$$

Al definir $A_0 = \emptyset$ y tomar el límite cuando N tiende a infinito tenemos que

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} [A_k \cap A_1^c \cap \ldots \cap A_{k-1}^c] = \bigcup_{k=1}^{\infty} B_k \,. \blacksquare$$

Una propiedad adicional de la $\sigma -$álgebra es que cualquier sucesión de eventos convergente, converge a un elemento de la $\sigma -$álgebra (de manera informal podríamos decir que una $\sigma -$álgebra es *compacta*). No desespere lector, enseguida le aclaramos lo que acaba de leer. Empezamos por definir *sucesión de eventos*.

Definición 1.3 – Sucesión de Eventos

Sean $\mathcal{A}$ una $\sigma -$álgebra y n un número natural. Si para cada $n \in \mathbb{N}$, A_n es un elemento de $\mathcal{A}$ entonces $A_1, A_2, A_3, \ldots$ es una *sucesión de eventos*. Utilizaremos la notación $\{A_n\}$ para denotar la sucesión. ∎

Ahora damos la definición de una *sucesión convergente de eventos*.

Definición 1.4 – Sucesión Convergente de Eventos

Sea $\{A_n\}$ una sucesión de eventos. El *límite superior* de la sucesión $\{A_n\}$, denotado por $\text{lím sup}\, A_n$, es definido como

$$\text{lím sup}\, A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \tag{1.1.1}$$

El *límite inferior* de la sucesión $\{A_n\}$,, denotado por $\text{lím inf}\, A_n$, es definido como

$$\text{lím inf}\, A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \tag{1.1.2}$$

Decimos que la sucesión $\{A_n\}$ *converge al* evento A cuando

$$\liminf A_n = A = \limsup A_n \tag{1.1.3}$$

y escribimos $\lim_{n\to\infty} A_n = A$ o alternativamente $A_n \to A$ cuando $n \to \infty$. ■

Observaciones

1) El criterio de convergencia no depende de una *distancia*; consulte el ejercicio 1.48.

2) Si $A_n \in \mathcal{A}$ para toda $n \in \mathbb{N}$, (1.1.1) y (1.1.2) nos permiten afirmar que tanto $\limsup A_n$ como $\liminf A_n$ son elementos de $\mathcal{A}$; esto es, son eventos.

3) Si $A_n \in \mathcal{A}$ es una sucesión que converge a A, $\lim_{n\to\infty} A_n = A$, entonces $A \in \mathcal{A}$; esto es, A es un evento.

4) Si $\omega \in \liminf A_n$ entonces, $\omega \in \bigcap_{k=n}^{\infty} A_k$ para al menos una n natural y esto significa que $\omega \in A_k$ para $k \geq n$. En consecuencia, ω pertenece a todos los elementos de la sucesión salvo quizás a un número finito de eventos; los que tienen un índice menor que n.

5) Si $\omega \in \limsup A_n$esto significa que $\omega \in \bigcup_{k=n}^{\infty} A_k$ para toda $n \geq 1$ por lo que, $\omega \in A_k$ para un número infinito de índices.

6) $\liminf A_n \subseteq \limsup A_n$ ya que $\liminf A_n$ es más restrictivo que $\limsup A_n$. Consulte el ejercicio 1.29.

7) Al aplicar las leyes de De Morgan a (1.1.1) y (1.1.2) deducimos que

$$\limsup A_n^c = (\liminf A_n)^c$$

$$\liminf A_n^c = (\limsup A_n)^c$$

Por tanto, si $\lim_{n\to\infty} A_n = A$ tendremos que $\lim_{n\to\infty} A_n^c = A^c$.

El siguiente resultado será de utilidad más adelante.

Lema 1.3

La intersección de un número arbitrario de σ −álgebras de subconjuntos de un conjunto Ω es una σ −álgebra de subconjuntos de Ω.

Demostración

Sea $\mathcal{J}$ un conjunto arbitrario de índices. Sea $\mathcal{A}_j$ una σ −álgebra de subconjuntos de Ω para cada $j \in \mathcal{J}$. Definamos $\mathcal{A} = \bigcap_{j\in\mathcal{J}} \mathcal{A}_j$ Entonces,

a. Como $\Omega \in \mathcal{A}_j$ para toda $j \in \mathcal{J}$ tenemos que, $\Omega \in \mathcal{A}$.

b. Si $A \in \mathcal{A}$ entonces, $A \in \mathcal{A}_j$ para toda $j \in \mathcal{J}$ y como $\mathcal{A}_j$ es una σ −álgebra, tenemos que $A^c \in \mathcal{A}_j$ para toda $j \in \mathcal{J}$. Por lo que, $A^c \in \mathcal{A}$.

c. Si $A_n \in \mathcal{A}$ para toda $n \in \mathbb{N}$ entonces, $A_n \in \mathcal{A}_j$ para toda $n \in \mathbb{N}$ y para toda $j \in \mathcal{J}$. Como $\mathcal{A}_j$ es una σ −álgebra, tenemos que, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}_j$ para toda $j \in \mathcal{J}$. En consecuencia, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$. ■

En seguida introducimos un concepto nuevo que nos permitirá entender mejor una σ −álgebra particular que es de especial importancia en Teoría de la Probabilidad.

Definición 1.5 – Sigma Álgebra Generada por una Colección de Subconjuntos

Sea $\mathcal{C}$ una colección de subconjuntos de Ω. La *σ −álgebra generada por* $\mathcal{C}$, denotada por $\mathcal{A}(\mathcal{C})$, es una σ −álgebra que satisface las siguientes propiedades

i) $\mathcal{C} \subset \mathcal{A}(\mathcal{C})$; los elementos de $\mathcal{C}$ son los *generadores de* $\mathcal{A}(\mathcal{C})$.

ii) Si $\mathcal{S}$ es otra σ −álgebra que contiene a $\mathcal{C}$ entonces, $\mathcal{A}(\mathcal{C}) \subset \mathcal{S}$; esto significa que $\mathcal{A}(\mathcal{C})$ es *la menor* σ −álgebra que contiene a $\mathcal{C}$. ■

El siguiente resultado establece que $\mathcal{A}(\mathcal{C})$ es única.

Lema 1.4

Sea $\mathcal{C}$ una colección de subconjuntos de Ω entonces, existe una única σ −álgebra generada por $\mathcal{C}$.

Demostración

Definamos $\mathcal{G}$ como

$$\mathcal{G} = \{\mathcal{D} \subseteq 2^{\Omega};\ \mathcal{D} \text{ es una } \sigma - \text{álgebra de subconjuntos de } \Omega \text{ y } \mathcal{C} \subset \mathcal{D}\}$$

Es claro que $\mathcal{G}$ no es vacío ya que $2^{\Omega} \in \mathcal{G}$. Definamos ahora, $\mathcal{A} = \bigcap_{\mathcal{D} \in \mathcal{G}} \mathcal{D}$. $\mathcal{A}$ es una σ − álgebra de acuerdo con el lema 1.3 y como $\mathcal{C} \subset \mathcal{D}$ para toda $\mathcal{D} \in \mathcal{G}$ tenemos que, $\mathcal{C} \subset \mathcal{A}$. También, por definición de $\mathcal{A}$, $\mathcal{A} \subset \mathcal{D}$. Podemos afirmar que $\mathcal{A} = \mathcal{A}(\mathcal{C})$ y que es única. ■

Con el fin de ilustrar las ideas presentadas respecto a la σ −álgebra discutimos ahora la *σ −álgebra de Borel en la recta real*. Como dicha σ −álgebra juega un papel importante le denotaremos con la letra $\mathcal{B}$. Los elementos de $\mathcal{B}$ son llamados *borelianos*. Sugerimos al lector no familiarizado con el tema de las sucesiones de números reales que consulte el apéndice A; allí hallará una exposición concisa.

Definición 1.6- Sigma Álgebra de Borel en la Recta Real

La *σ −álgebra de Borel en la recta real*, $\mathcal{B}$, es la σ −álgebra generada por intervalos de la forma $(-\infty, x]$ donde x es un racional. ■

Que $\mathcal{B}$ está generada por los intervalos de la forma $(-\infty, x]$ con $x \in \mathbb{Q}$ significa que, con esos intervalos, llamados *generadores*, produciremos los demás elementos de la σ −álgebra al utilizar SA2, SA3, el lema 1.1 y límites de sucesiones. Así, un primer paso es generar los intervalos de la forma $(-\infty, y]$ con $y \in \mathbb{R}$. Para lograrlo recurrimos al hecho de que $\mathbb{Q}$ es *denso* en $\mathbb{R}$, esto es, si $y \in \mathbb{R}$ entonces, existe una sucesión de racionales $\{x_n\}$ tal que $x_n \to y$ cuando $n \to \infty$. Es más, podemos hacer que $x_n \searrow y$ o bien que $x_n \nearrow y$ cuando $n \to \infty$. Ahora consideramos la sucesión $\{(-\infty, x_n]\}$; en el lema 1.5 mostraremos que

$$\lim_{n\to\infty} (-\infty, x_n] = \begin{cases} (-\infty, y] & \text{si} \quad x_n \searrow y \text{ cuando } n \to \infty \\ (-\infty, y) & \text{si} \quad x_n \nearrow y \text{ cuando } n \to \infty \end{cases}$$

Por tanto, los intervalos de la forma $(-\infty, y]$ pertenecen a la σ −álgebra de Borel para cualquier $y \in \mathbb{R}$. Como consecuencia de SA2 tenemos que $(y, \infty) \in \mathcal{B}$ para toda $y \in \mathbb{R}$. Similarmente, los intervalos de las formas $(-\infty, y)$ y $[y, \infty)$ pertenecen a la σ −álgebra de Borel para cualquier $y \in \mathbb{R}$. Adicionalmente, el lema 1.1 nos permite afirmar que si $a, b \in \mathbb{R}$ con $a < b$ entonces,

$$(-\infty, b] \cap [a, \infty) = [a, b] \in \mathcal{B}$$

$$(-\infty, b] \cap (a, \infty) = (a, b] \in \mathcal{B}$$

$$(-\infty, b) \cap [a, \infty) = [a, b) \in \mathcal{B}$$

$$(-\infty, b) \cap (a, \infty) = (a, b) \in \mathcal{B}$$

$$(-\infty, b] \cap [b, \infty) = \{b\} \in \mathcal{B}$$

Como $\{y\} \in \mathcal{B}$ para toda $y \in \mathbb{R}$ tenemos que,

$\mathbb{N} \in \mathcal{B}$, ya que $\mathbb{N} = \bigcup_{k=1}^{\infty}\{k\}$ y $\{k\} \in \mathcal{B}$

$\mathbb{Z} \in \mathcal{B}$, ya que $\mathbb{Z} = \mathbb{N} \cup \bigcup_{k=0}^{\infty}\{-k\}$ y $\{-k\} \in \mathcal{B}$

$\mathbb{Q} \in \mathcal{B}$, porque $\mathbb{Q}$ es numerable

$(\mathbb{R} - \mathbb{Q}) \in \mathcal{B}$, por SA2

Debemos tener cuidado de no inferir que si $(\mathbb{R} - \mathbb{Q}) \in \mathcal{B}$ entonces *cualquier* subconjunto de $(\mathbb{R} - \mathbb{Q})$ es un boreliano.

Después de ver la gama de elementos que tiene $\mathcal{B}$ uno podría quedarse con la impresión de que cualquier subconjunto de $\mathbb{R}$ es un boreliano. Sin embargo, este no es el caso; el libro de Royden (1968) en el capítulo 3 exhibe el conjunto de Vitali, un subconjunto de [0,1], (consulte el ejercicio1.28), que no es un boreliano. El conjunto de Vitali ilustra el hecho de que los subconjuntos de $\mathbb{R}$ pueden ser inusuales (o unas monstruosidades).

Ya que vimos para qué nos pueden servir las sucesiones de eventos, discutimos ahora con mayor detalle la convergencia de sucesiones de eventos. Recordamos que, en las sucesiones de números, la convergencia es, por lo general, definida en términos de la distancia entre dos números. Si tratáramos de generalizar esa idea a eventos diríamos que la sucesión $\{A_n\}$ converge al evento A si para $\varepsilon > 0$ dado, A_n *dista* de A en menos de ε para $n \geq N(\varepsilon)$. Nos encontramos aquí con el problema de la definición de *distancia entre eventos*; recuerde que los eventos pueden ser subconjuntos de números reales o subconjuntos del plano, o del espacio. El lector puede intuir que el límite de $A_n = [0,1 + 1/n)$ es $A = [0,1]$ pero ¿cómo definir la distancia de A_n a A? Si quisiéramos trabajar con $A_n - A = (1, 1 + 1/n)$, vemos que podemos hacer la longitud de $A_n - A$ tan pequeña como queramos al incrementar n. Desgraciadamente, $[0,1) = A_*$ tiene la misma propiedad con lo que el conjunto límite no sería único. Sin embargo, el criterio de convergencia dado no depende de una noción de distancia. Ahora mostramos el referido lema 1.5.

Lema 1.5

Sea $\{x_n\}$ una sucesión de números racionales tal que $x_n \to x$ cuando $n \to \infty$. Entonces,

$$\lim_{n\to\infty}(-\infty, x_n] = \begin{cases} (-\infty, x] & \text{si} \quad x_n \searrow x \text{ cuando } n \to \infty \\ (-\infty, x) & \text{si} \quad x_n \nearrow x \text{ cuando } n \to \infty \end{cases}$$

Demostración

Primero, suponemos que $x_n \searrow x$ cuando $n \to \infty$. Esto significa que $(-\infty, x_{n+1}] \subseteq (-\infty, x_n]$ para toda $n \in \mathbb{N}$, esto es, los intervalos $(-\infty, x_n]$ se *contraen* cuando n tiende a infinito. Del inciso 4 del teorema $A.1$ del apéndice A sabemos que $x = \inf_n x_n$.

Notamos que $x \in (-\infty, x_n]$ para toda $n \Rightarrow x \in \liminf(-\infty, x_n]$. Ahora, por definición, $\limsup(-\infty, x_n] = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty}(-\infty, x_k]$. Pero para n fija, $\bigcup_{k=n}^{\infty}(-\infty, x_k] = (-\infty, x_n]$ ya que $x_{m+1} \leq x_m$ para toda $m \in \mathbb{N}$. En consecuencia,

$$\limsup(-\infty, x_n] = \bigcap_{n=1}^{\infty}(-\infty, x_n] = (-\infty, x].$$

Ya que si $y \in \bigcap_{n=1}^{\infty}(-\infty, x_n]$ entonces, $y \leq x_n$ para toda $n \in \mathbb{N}$. Esto significa que y es una cota inferior de $\{x_1, x_2, \dots\}$. Por tanto, $y \leq \inf\{x_n\} =$ máxima cota inferior $= x \Rightarrow y \in (-\infty, x]$. Recíprocamente, si $y \in (-\infty, x]$ entonces, $y \leq x \leq x_n$ para toda $n \in \mathbb{N}$. Por tanto, $y \in (-\infty, x_n]$ para toda $n \in \mathbb{N}$ y esto significa que $y \in \bigcap_{n=1}^{\infty}(-\infty, x_n]$. Concluimos que

$$\lim_{n\to\infty}(-\infty, x_n] = (-\infty, y].$$

Segundo, suponemos ahora que $x_n \nearrow x$ cuando $n \to \infty$. Esto significa que $(-\infty, x_n] \subseteq (-\infty, x_{n+1}]$ para toda $n \in \mathbb{N}$, esto es, los intervalos $(-\infty, x_n]$ se *expanden* cuando n tiende a infinito. Del inciso 4 del teorema A.1 del apéndice A sabemos que $x = \sup_n x_n$.

Notamos que $x \notin (-\infty, x_n]$ para toda $n \in \mathbb{N} \Rightarrow x \notin \liminf(-\infty, x_n]$. Por definición, $\liminf(-\infty, x_n] = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty}(-\infty, x_k]$. Pero para una n fija,

$$\bigcap_{k=n}^{\infty}(-\infty, x_k] = (-\infty, x_n], \text{ ya que } x_{m+1} \geq x_m \text{ para toda } m \in \mathbb{N}.$$

En consecuencia,

$$\liminf(-\infty, x_n] = \bigcup_{n=1}^{\infty}(-\infty, x_n] = (-\infty, x)$$

Ya que si $y \in \bigcup_{n=1}^{\infty}(-\infty, x_n]$ entonces $y \leq x_n$ para al menos una n; esto significa que $y < x_{n+k}$ para $k \geq 1$ por ser la sucesión $\{x_n\}$ creciente. En consecuencia, $y < \sup_n x_n = x$, que significa que $y \in (-\infty, x)$. Recíprocamente, si $y \in (-\infty, x)$ tenemos que $y < x = \sup\{x_n\}$. En consecuencia, para cualquier $\delta > 0$ existe un natural n tal que $(x - \delta) < x_n$. Para, $\delta = (x - y) > 0$, tenemos que $y < x_{n_0}$ para un $n_0 \in \mathbb{N}$. Por tanto, $y \in (-\infty, x_{n_0}]$ y $y \in \bigcup_{n=1}^{\infty}(-\infty, x_n]$.

Por otra parte, $\limsup(-\infty, x_n] = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty}(-\infty, x_n] = \bigcap_{n=1}^{\infty}(-\infty, x) = (-\infty, x)$, por lo deducido en el párrafo anterior. Por tanto,

$$\limsup(-\infty, x_n] = (-\infty, x) = \liminf(-\infty, x_n]$$

y esto significa que

$$\lim_{n\to\infty}(-\infty, x_n] = (-\infty, x). \ \blacksquare$$

El siguiente ejemplo muestra que si $a_n \to a$ cuando $n \to \infty$, con $a_n, a \in \mathbb{R}^+$ para toda $n \in \mathbb{N}$, no podemos afirmar que $[0, a_n) \to [0, a)$ o que $[0, a_n) \to [0, a]$ cuando $n \to \infty$; el límite, si existe, depende de la manera en que $a_n \to a$ cuando $n \to \infty$.

Ejemplo 1.8

Sea $a_n = 1 + (-1/2)^n$ para $n \in \mathbb{N}$. Definamos $A_n = [0, a_n)$. Los primeros ocho intervalos son $A_1 = [0, 0.5)$, $A_2 = [0, 1.25)$, $A_3 = [0, 0.875)$, $A_4 = [0, 1.0625)$, $A_5 = [0, 0.96875)$, $A_6 = [0, 1.015625)$, $A_7 = [0, 0.9921875)$ y $A_8 = [0, 1.00390625)$. Cuando n es par $1 \in A_n$ pero cuando n es impar $1 \notin A_n$. Notamos que la subsucesión $\{A_{2k+1}\}$ se expande mientras que $\{A_{2k}\}$ se contrae.

Definimos $B_n = \bigcup_{k=n}^{\infty} A_n$. Entonces, $B_1 = [0, 1.25)$, $B_2 = [0, 1.25) = A_2$, $B_3 = [0, 1.0626) = A_4$, $B_4 = [0, 1.0625) = A_4$, etc. En general,

$$B_{2r+1} = A_{2r+2} = [0, 1 + (-1/2)^{2r+2}) = B_{2r+2}\text{ , para } r = 0, 1, 2, 3, \ldots$$

Observe que B_n se va contrayendo conforme n crece. Por tanto,

$$\begin{aligned}\text{lím sup}\, A_n &= \bigcap_{n=1}^{\infty} B_n = [0, 1.25) \cap [0, 1.25) \cap [0, 1.0625) \cap [0, 1.0625) \cap \ldots \\ &= [0,1]\text{, de acuerdo con el lema 1.5.}\end{aligned}$$

Definamos ahora, $C_n = \bigcap_{k=n}^{\infty} A_k$. Entonces, $C_1 = [0, 0.5) = A_1$, $C_2 = [0, 0.875) = A_3$, $C_3 = [0, 0.875) = A_3$, $C_4 = [0, 0.96875) = A_5$, etc. En general,

$$C_{2r} = [0, 1 + (-1/2)^{2r+1}) = A_{2r+1} = C_{2r+1}\text{ , para } r = 1, 2, 3, \ldots$$

Note que C_n se va expandiendo conforme n crece. Así,

$$\begin{aligned}\text{lím inf}\, A_n &= \bigcup_{n=1}^{\infty} C_n = [0, 0.5) \cup [0, 0.875) \cup [0, 0.875) \cup [0, 0.96875) \cup \ldots \\ &= [0,1)\text{, de acuerdo con el lema1.5.}\end{aligned}$$

Como lím sup $A_n \neq$ lím inf A_n la sucesión *no converge*, a pesar de que $a_n \to 1$, cuando $n \to \infty$. $\blacksquare$

El ejercicio 1.8 le pide relacionar lím sup$(-\infty, a_n]$ con $(-\infty,$ lím sup $a_n]$ y con $(-\infty,$ lím sup $a_n)$ donde $\{a_n\}$ es una sucesión convergente de números reales.

La σ −álgebra de Borel en la recta real puede ser generalizada al espacio euclidiano de $k \geq 2$ dimensiones y es denotada por $\mathcal{B}^k$. $\mathcal{B}^2$ está generada por rectángulos de la forma $(-\infty, x] \times (-\infty, y]$ con $x, y \in \mathbb{Q}$; se puede mostrar que regiones delimitadas por polígonos, circunferencias, elipses, etc. son borelianos; el procedimiento utilizado es el de aproximar por medio de rectángulos. $\mathcal{B}^k$ será utilizada para $k = 1$ en el capítulo 3 y para $k \geq 2$ en el capítulo 7.

En experimentos aleatorios donde está involucrado un proceso de medición, la σ −álgebra que debe ser utilizada depende de la precisión del instrumento de medición. Verbigracia, si el experimento aleatorio requiere de la medición de una longitud y el instrumento de medición nos permite medir en metros y hasta milímetros entonces, los generadores del $\sigma-$álgebra pueden ser intervalos de la forma $[a, b]$ donde a y b son expresables como

$$x_0 + x_1 10^{-1} + x_2 10^{-2} + x_3 10^{-3}$$

Donde $x_0, x_1, x_2, x_3 \in \mathbb{N}_0$. Debido al instrumento de medición, no podríamos decir si una medición pertenece o no al intervalo [10.1223, 10.1225]. Obviamente, [10.1223, 10.1225] no sería un evento. Pasamos ahora a presentar la tercera componente de la terna que define al espacio de probabilidad.

1.1.3.- Medida de Probabilidad: P

El enfoque para definir P es axiomático y la definimos como una función cuyo dominio es la $\sigma-$álgebra $\mathcal{A}$ y cuyo codominio son los reales no negativos, esto es,

$$\mathrm{P}: \mathcal{A} \to [0, \infty)$$

P debe satisfacer los siguientes *axiomas de Kolmogorov*

P1.- Si $A \in \mathcal{A}$ entonces, $\mathrm{P}(A) \geq 0$.

P2.- $\mathrm{P}(\Omega) = 1$

P3.- Si $A, B \in \mathcal{A}$ con $A \cap B = \emptyset$ entonces, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$. En general, si $A_j \in \mathcal{A}$ para $j \in \mathbb{N}$ son mutuamente excluyentes (esto es, $A_j \cap A_k = \emptyset$ cuando $j \neq k$) entonces

$$\mathrm{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathrm{P}(A_j)$$

A esta última propiedad se le conoce como $\sigma-$*aditividad*. ■

Observaciones

1) P es una función que a un *conjunto* (evento) le *asocia* un número real no negativo; al igual que otras medidas como la cardinalidad, la longitud, la superficie o el volumen. Como $\mathrm{P}(\Omega) = 1$, se dice que P es una *medida normada*.

2) En varios textos se establece el primer axioma de la siguiente manera: si $A \in \mathcal{A}$ entonces, $0 \leq \mathrm{P}(A) \leq 1$. Observe que nosotros eliminamos la desigualdad $\mathrm{P}(A) \leq 1$ debido a que se puede deducir de los otros axiomas. De hecho, si $A \in \mathcal{A}$ entonces,

$$\Omega = A \cup A^c \text{, con } A \cap A^c = \emptyset$$

y al utilizar P2 y P3 obtenemos

$$1 = \mathrm{P}(A) + \mathrm{P}(A^c) \geq \mathrm{P}(A)$$

ya que $\mathrm{P}(A^c) \geq 0$, por P1.

3) Como P está definida en $\mathcal{A}$, sólo a los eventos les podremos calcular P. De esta manera $\mathrm{P}(\Omega)$ está bien definida ya que $\Omega \in \mathcal{A}$. Similarmente para $\mathrm{P}(A \cup B)$, $\mathrm{P}\left(\bigcup_{j=1}^{\infty} A_j\right)$ y $\mathrm{P}(A^c)$ ya que $A \cup B$, $\bigcup_{j=1}^{\infty} A_j$ y A^c son eventos. Recordamos que $\emptyset$ es un evento.

4) Es posible definir más de una medida de probabilidad para un mismo par $(\Omega, \mathcal{A})$.

5) La σ −aditividad *no* es una propiedad que pueda ser deducida por un proceso de toma de límites. Nos podemos convencer fácilmente de que si A_1, A_2, A_3 son mutuamente excluyentes entonces,

$$\mathrm{P}(A_1 \cup A_2 \cup A_3) = \mathrm{P}[A_1 \cup (A_2 \cup A_3)] = \mathrm{P}(A_1) + \mathrm{P}(A_2 \cup A_3), \text{por P3}$$
$$= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3)$$

y que para cualquier número n de eventos mutuamente excluyentes, $\mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathrm{P}\left(A_j\right)$. Si en la última igualdad tomamos el límite cuando $n \to \infty$, tenemos

$$\lim_{n\to\infty} \mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) = \lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{P}\left(A_j\right)$$

Como $\mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq 1$ y $\left\{\sum_{j=1}^{n} \mathrm{P}\left(A_j\right)\right\}$ es una sucesión no decreciente y acotada resulta ser convergente (consulte el inciso 3 del teorema A.1 del apéndice A). Por lo tanto, $\lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{P}\left(A_j\right)$ *existe*. Así,

$$\sum_{j=1}^{\infty} \mathrm{P}\left(A_j\right) = \lim_{n\to\infty} \mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right)$$

Para obtener la afirmación del axioma sería necesario que

$$\lim_{n\to\infty} \mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) = \mathrm{P}\left(\lim_{n\to\infty} \bigcup_{j=1}^{n} A_j\right)$$

Esto es equivalente a pedir que podamos intercambiar P con lím que en general *no* es válido (consulte el ejercicio 1.30). En esencia, el axioma P3 pide que podamos intercambiar P con lím pero únicamente en sucesiones de la forma $\left\{\bigcup_{j=1}^{n} A_j\right\}$ para eventos A_j mutuamente excluyentes. Por tanto, P3 no es redundante.

Enseguida mostramos, a partir de los axiomas, algunas propiedades de la medida de probabilidad; las concentramos en el siguiente

Teorema 1.1

Sea $(\Omega, \mathcal{A}, \mathrm{P})$ un espacio de probabilidad. Sean $A, B, A_1, A_2, \dots, A_n$ eventos. Entonces,

1. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.
2. $\mathrm{P}(\emptyset) = 0$.
3. Si $B \subseteq A$ entonces, $\mathrm{P}(B) \leq \mathrm{P}(A)$ y $\mathrm{P}(A - B) = \mathrm{P}(A) - \mathrm{P}(B)$.
4. $\mathrm{P}(A \cap B) \leq \text{mín}\{\mathrm{P}(A), \mathrm{P}(B)\}$.
5. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$.

6. $\mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{j=1}^{n} \mathrm{P}(A_j)$, conocida como la *Desigualdad de Boole*.

Demostración

La idea es descomponer los eventos de interés como la unión disjunta de eventos para luego utilizar P3.

1. Como $A \cap A^c = \emptyset$ y $A \cup A^c = \Omega$ tenemos que

$$\mathrm{P}(\Omega) = 1 = \mathrm{P}(A) + \mathrm{P}(A^c) \quad \Rightarrow \quad \mathrm{P}(A^c) = 1 - \mathrm{P}(A).$$

2. Al hacer $A = \emptyset$ en el inciso 1, obtenemos $\mathrm{P}(\emptyset^c) = 1 - \mathrm{P}(\emptyset)$, pero $\Omega = \emptyset^c$ por lo que $\mathrm{P}(\emptyset) = 0$.

3. Como por hipótesis $B \subseteq A$, podemos escribir A como

$$A = B \cup (A \cap B^c) \tag{1.1.4}$$

y $B \cap (A \cap B^c) = \emptyset$, al aplicar P3 a (1.1.4) se obtiene que

$$\mathrm{P}(A) = \mathrm{P}(B) + \mathrm{P}(A \cap B^c) \geq \mathrm{P}(B)$$

puesto que $\mathrm{P}(A \cap B^c) \geq 0$. Esta propiedad establece que la medida de probabilidad es *no decreciente*. Además,

$$\mathrm{P}(A \cap B^c) = \mathrm{P}(A - B) = \mathrm{P}(A) - \mathrm{P}(B).$$

4. Como $A \cap B \subseteq A$ y $A \cap B \subseteq B$, al utilizar el inciso 3, obtenemos $\mathrm{P}(A \cap B) \leq \mathrm{P}(A)$ y $\mathrm{P}(A \cap B) \leq \mathrm{P}(B)$, esto es,

$$\mathrm{P}(A \cap B) \leq \text{mín}\{\mathrm{P}(A), \mathrm{P}(B)\}.$$

5. Como $A \cup B = A \cup (A^c \cap B)$ y $A \cap (A^c \cap B) = \emptyset$, P3 nos permite afirmar

$$\begin{aligned}\mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(A^c \cap B)\\ &= \mathrm{P}(A) + [\mathrm{P}(A^c \cap B) + \mathrm{P}(A \cap B)] - \mathrm{P}(A \cap B), \text{al sumar y restar } \mathrm{P}(A \cap B)\\ &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B), \text{ya que } B = (A^c \cap B) \cup (A \cap B)\end{aligned}$$

6. Observe primero que el inciso 5. afirma que

$$\mathrm{P}(A) + \mathrm{P}(B) = \mathrm{P}(A \cup B) + \mathrm{P}(A \cap B) \geq \mathrm{P}(A \cup B)$$

por lo que la desigualdad es válida para dos eventos. Haremos la demostración por inducción sobre n. Sea $n \geq 3$ y supongamos que la desigualdad en 6 es válida para n. Entonces,

$$\mathrm{P}\left(\bigcup_{j=1}^{n+1} A_j\right) = \mathrm{P}\left(\bigcup_{j=1}^{n} A_j \cup A_{n+1}\right) = \mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) + \mathrm{P}(A_{n+1}) - \mathrm{P}\left(A_{n+1} \cap \bigcup_{j=1}^{n} A_j\right)$$

según lo demostrado en el inciso 5. La hipótesis de inducción y el hecho de que $\mathrm{P}\left(A_{n+1} \cap \bigcup_{j=1}^{n} A_j\right) \geq 0$ nos permiten concluir

$$\mathrm{P}\left(\bigcup_{j=1}^{n+1} A_j\right) \leq \sum_{j=1}^{n} \mathrm{P}(A_j) + \mathrm{P}(A_{n+1}) = \sum_{j=1}^{n+1} \mathrm{P}(A_j)\ . \blacksquare$$

Observación

- Los incisos 4, 5 y 6 pueden ser generalizados. De hecho,

i) $\mathrm{P}(A \cap B \cap C) \leq \text{mín}\,\{\mathrm{P}(A), \mathrm{P}(B), \mathrm{P}(C)\}$ ¿puede hacerlo más general?

ii)
$$\begin{aligned}\mathrm{P}(A \cup B \cup C) &= \mathrm{P}[A \cup (B \cup C)] \\ &= \mathrm{P}(A) + \mathrm{P}(B \cup C) - \mathrm{P}[A \cap (B \cup C)] \\ &= \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C) - \mathrm{P}[(A \cap B) \cup (A \cap C)] \\ &= \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) + \mathrm{P}(A \cap B \cap C)\end{aligned}$$

(¿puede hacerlo más general?), consulte el ejercicio 1.3.

iii) La desigualdad de Boole puede generalizarse (consulte el ejercicio 1.9) a

$$\mathrm{P}\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mathrm{P}(A_j)\,.$$

Con el fin de mostrar la utilidad del teorema 1.1 presentamos a continuación dos casos.

Ejemplo 1.9

Sea $(\Omega, \mathcal{A}, \mathrm{P})$ un espacio de probabilidad asociado a un fenómeno aleatorio $\mathcal{E}$. Sean A y B eventos. Definamos al evento C como: ocurre exactamente uno de los dos eventos. Queremos encontrar una expresión para C en términos de A y B y una expresión para $\mathrm{P}(C)$ en términos de $\mathrm{P}(A), \mathrm{P}(B)$ y $\mathrm{P}(A \cap B)$.

Primero, si ω es el resultado del experimento aleatorio, C ocurre cuando $\omega \in C$ y esto sucede si y sólo $\omega \in (A - B) \cup (B - A)$. Ya que si $\omega \in (A - B) \cup (B - A)$ entonces, o bien $\omega \in (A - B)$, que significa A ocurre, pero no B, o bien $\omega \in (B - A)$, que significa B ocurre, pero no A. Por tanto, $C = (A - B) \cup (B - A)$. Una manera alternativa de expresar C es $C = (A \cup B) - (A \cap B)$. Como consecuencia, tenemos que

$$\begin{aligned}\mathrm{P}(C) &= \mathrm{P}[(A \cup B) - (A \cap B)] \\ &= \mathrm{P}(A \cup B) - \mathrm{P}(A \cap B), \text{por el inciso 3 del teorema 1.1} \\ &= \mathrm{P}(A) + \mathrm{P}(B) - 2\mathrm{P}(A \cap B), \text{según el inciso 5 del teorema 1.1}\end{aligned}$$

El ejercicio 1.25 es una generalización de este ejemplo. ■

Ejemplo 1.10

Sea $(\Omega, \mathcal{A}, \mathrm{P})$ un espacio de probabilidad asociado a un fenómeno aleatorio $\mathcal{E}$. Sean A, B y C eventos. Definamos los eventos $D =$ sólo ocurre A, $E =$ sólo A y B ocurren, $F =$ sólo A no ocurre y $G =$ no más de dos ocurren. Encontraremos expresiones para los eventos y sus probabilidades. Sea $\omega \in \Omega$ el resultado de observar $\mathcal{E}$ entonces,

1) D ocurre cuando $\omega \in D$ y esto sucede, si y sólo si $\omega \in A$ pero $\omega \notin (B \cup C)$ (ya que si $\omega \in (B \cup C)$ entonces, B o C o los dos ocurren simultáneamente). Por tanto $D = A \cap (B \cup C)^c$. Para encontrar una expresión para $\mathrm{P}(D)$ utilizamos la siguiente igualdad

$$A = [A \cap (B \cup C)^c] \cup [A \cap (B \cup C)] = D \cup [A \cap (B \cup C)]$$

de donde deducimos, por el inciso 5 del teorema 1.1 que

$$\begin{aligned}\mathrm{P}(D) &= \mathrm{P}(A) - \mathrm{P}[A \cap (B \cup C)], \text{por P3} \\ &= \mathrm{P}(A) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) + \mathrm{P}(A \cap B \cap C)\end{aligned}$$

2) E ocurre cuando $\omega \in E$ y esto sucede si y sólo si $\omega \in A \cap B$ y $\omega \notin \mathrm{C}$. Por tanto, $E = A \cap B \cap C^c$. Al utilizar la igualdad

$$A \cap B = (A \cap B \cap C) \cup (A \cap B \cap C^c) = (A \cap B \cap C) \cup E$$

deducimos que

$$\mathrm{P}(E) = \mathrm{P}(A \cap B) - \mathrm{P}(A \cap B \cap C), \text{de acuerdo con P3}$$

3) F ocurre si y sólo si $\omega \notin A$ y $\omega \in B \cup C$. Por tanto, $F = (B \cup C) \cap A^c$. Al utilizar la igualdad

$$B \cup C = [(B \cup C) \cap A^c] \cup [(B \cup C) \cap A] = F \cup [(B \cup C) \cap A]$$

obtenemos

$$\mathrm{P}(F) = \mathrm{P}(B \cup C) - \mathrm{P}[(B \cup C) \cap A]$$

de donde deducimos, por medio del inciso 5 del teorema 1.1, que

$$\mathrm{P}(F) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C) - \mathrm{P}(B \cap A) - \mathrm{P}(C \cap A) + \mathrm{P}(A \cap B \cap C)$$

4) G es el complemento del evento: más de dos ocurren y como únicamente hay 3 eventos en la discusión, más de dos significa los tres; por tanto, $G = (A \cap B \cap C)^c$. El inciso 1 del teorema 1.1 nos permite afirmar que

$$\mathrm{P}(G) = 1 - \mathrm{P}(A \cap B \cap C). \ \blacksquare$$

Así, como en la discusión de la σ −álgebra presentamos como instancia la σ −álgebra de Borel ahora exhibimos una ilustración de medida de probabilidad que será aplicable a espacios de probabilidad discretos.

Definición 1.7 - Medida de Probabilidad en Espacios Discretos

Consideremos un experimento aleatorio $\mathcal{E}$ que tiene asociado un espacio muestral Ω. Supongamos que Ω es numerable. Si es finito lo podemos escribir como $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ donde N es la *cardinalidad* de Ω, esto es, N es el número de resultados elementales posibles de interés. Si Ω es infinito lo representamos como $\Omega = \{\omega_1, \omega_2, \dots\}$. Sin pérdida de generalidad supondremos que Ω es infinito. Ahora, si a cada resultado elemental ω_j le asociamos un número real p_j tal que $\mathrm{P}(\{\omega_j\}) = p_j$ y se satisfacen las restricciones

$$0 \leq p_j \leq 1 \quad \text{con} \quad \sum_{j=1}^{\infty} p_j = 1\,. \tag{1.1.5}$$

Si Ω es finito y tiene cardinalidad N hacemos $p_j = 0$ para $j > N$. Definimos una medida de probabilidad P *para todo subconjunto* A de Ω de la siguiente manera

$$\mathrm{P}(A) = \sum_{j \in I_A} \mathrm{P}(\{\omega_j\}) \tag{1.1.6}$$

Donde I_A es el conjunto de los índices j tal que $\omega_j \in A$; esto es, $I_A = \{j \in \mathbb{N};\ \ \omega_j \in A\}$ y en particular, $I_\Omega = \mathbb{N}$. P definida en (1.1.6) es una *medida de probabilidad en un espacio de probabilidad discreto* ∎

Enseguida mostramos que P definida según (1.1.6) es una medida de probabilidad.

Lema 1.6

Bajo el contexto de la definición 1.7, P definida en (1.1.6) es una medida de probabilidad.

Demostración

i) Que $0 \leq \mathrm{P}(A)$ para todo subconjunto de Ω es obvio a partir de (1.1.6) y (1.1.5).

ii) Por definición de P tenemos que

$$\mathrm{P}(\Omega) = \mathrm{P}\left(\sum_{j \in I_\Omega} \mathrm{P}(\{\omega_j\})\right) = \sum_{j=1}^{\infty} p_j = 1, \text{según (1.1.5).}$$

iii) Si $A, B \subset \Omega$ con $A \cap B = \emptyset$ entonces,

$$\mathrm{P}(A \cup B) = \sum_{j \in I_{\mathrm{A\cup B}}} \mathrm{P}(\{\omega_j\})$$

Pero, $I_{A\cup B} = I_A \cup I_B$ con $I_A \cap I_B = \emptyset$ ya que A y B son ajenos. Por tanto,

$$\sum_{j \in I_{\mathrm{A\cup B}}} \mathrm{P}(\{\omega_j\}) = \sum_{j \in I_{\mathrm{A}}} \mathrm{P}(\{\omega_j\}) + \sum_{j \in I_{\mathrm{B}}} \mathrm{P}(\{\omega_j\}) = \mathrm{P}(A) + \mathrm{P}(B)$$

En forma más general, si A_k es un subconjunto de Ω para toda k natural donde $A_k \cap A_j = \emptyset$ para $k \neq j$ tenemos que

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{j \in I_{\cup_{k=1}^{\infty} A_k}} \mathrm{P}(\{\omega_j\}) = \sum_{k=1}^{\infty} \sum_{j \in I_{A_k}} \mathrm{P}(\{\omega_j\}) = \sum_{k=1}^{\infty} \mathrm{P}(A_k).$$

Por tanto, P definida en (1.1.6) es una medida de probabilidad. Observe que la σ −álgebra es entonces el *conjunto potencia* de Ω ya que P estuvo definida para cualquier subconjunto de Ω. Cuando Ω es numerable el espacio de probabilidad es $(\Omega, 2^{\Omega}, \mathrm{P})$ con P definida en (1.1.6). ∎

Enseguida estudiamos un caso particular que modela adecuadamente diversos experimentos aleatorios. Supongamos que Ω tiene N elementos; $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ de tal manera que $\mathrm{P}(\{\omega_j\}) = p_j = 1/N$, para $j = 1, \dots, N$; esto es, los *resultados elementales posibles son equiprobables*. De esta manera, si $A \subset \Omega$

$$\mathrm{P}(A) = \sum_{j \in I_A} \mathrm{P}(\{\omega_j\}) = \frac{1}{N}(\text{número de elementos en } A) = \frac{\#A}{\#\Omega} \qquad (1.1.7)$$

donde # significa *cardinalidad*, número de elementos. Entonces, $\mathrm{P}(A)$ es una fracción cuyo numerador es el número de elementos en A y el denominador es el número de elementos en Ω. Se ha acostumbrado a decir $\mathrm{P}(A)$ es *casos favorables entre casos totales* y ésta ha sido llamada la *Definición Clásica de Probabilidad*. (1.1.7) puede ser interpretada como la medida de A entre la medida Ω donde medida significa cardinalidad. Esta última idea del cociente de medidas será utilizada en la sección que trata sobre Probabilidad Geométrica en el capítulo 2, pero por medida entenderemos longitud, área o volumen.

En el inicio de la Teoría de la Probabilidad los investigadores utilizaron (1.1.7) para hacer cálculos numéricos y de hecho definieron probabilidad de esa manera. Es conveniente destacar que Laplace (1951) hace notar que, si los resultados elementales no son equiprobables entonces, se produce la necesidad de asignar a cada resultado elemental ω_j un número p_j y que esto requiere de *gran cuidado*.

Enseguida presentamos instancias de la aplicación de la definición clásica. En la última sección del presente capítulo y en el siguiente capítulo proporcionamos más. Empezaremos con juegos de azar y paulatinamente nos moveremos a otros campos.

Ejemplo 1.11

Supongamos que $\mathcal{E}$ es lanzar una vez un dado honesto. En el ejemplo 1.1 vimos que el espacio muestral Ω puede ser representado por $\{1, 2, 3, 4, 5, 6\}$. Como el dado es honesto (no está cargado) es plausible que los 6 resultados sean equiprobables y esto nos lleva a proponer $\mathrm{P}(\{j\}) = 1/6$ para $j = 1, \dots, 6$.

Definamos los eventos: $A = \{2,4,6\}$, observamos un número par, $B = \{1,3,5\}$, observamos un número impar y $C = \{2,3,5\}$, observamos un número primo. Para calcular probabilidades utilizamos (1.1.7). Como $\#\Omega = 6$ y $\#A = 3 = \#B = \#C$ tenemos que $\mathrm{P}(A) = 1/2 = \mathrm{P}(B) = \mathrm{P}(C)$.

Observe que eventos distintos tienen la misma probabilidad; no esté tentado a concluir que si $\mathrm{P}(A) = \mathrm{P}(B)$ entonces, $A = B$. Por otra parte, $A \cap B = \emptyset$ que nos dice que A y B no pueden ocurrir simultáneamente (que es obvio). Es más, $B = A^c$.

$B \cap C$ es el evento observamos un número impar y primo; tenemos, $\mathrm{P}(B \cap C) = \mathrm{P}(\{3,5\}) = 1/3$. Por otra parte, $\mathrm{P}(B \cup C) = \mathrm{P}(\{1,2,3,5\}) = 2/3$; que también puede ser obtenida al utilizar el inciso 5 del del teorema 1.1,

$$\mathrm{P}(B \cup C) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C) = 2/3\,.$$

¿Cómo interpretamos el que $\mathrm{P}(B \cup C) = 2/3$? La probabilidad de que ocurra A o C, pero no los dos, está dada en el ejemplo 1.9 y es

$$\mathrm{P}(A) + \mathrm{P}(C) - 2\mathrm{P}(A \cap C) = 2/3\,. \blacksquare$$

Ejemplo 1.12

Supongamos que $\mathcal{E}$ es lanzar una vez un par de dados honestos. En el ejemplo 1.3 establecimos dos alternativas para el espacio muestral dependiendo de si los dados son o no distinguibles. Bajo el supuesto de que los dados son distinguibles tenemos que

$$\Omega_0 = \{(k, j); k, j \in \mathbb{N} \text{ con } i, j \leq 6\}.$$

Si los dados son honestos es razonable proponer

$$\mathrm{P}(\{(k, j)\}) = 1/36 \text{, para todo par } (k, j)$$

que refleja el hecho de que todos los resultados posibles son equiprobables. Si definimos los eventos, $A =$ exactamente un dos es observado, $B =$ al menos un dos es observado y $C =$ la suma de los números observados es 7 entonces,

$$A = \{(2,1), (2,3), (2,4), (2,5), (2,6), (1,2), (3,2), (4,2), (5,2), (6,2)\}$$

$$B = A \cup \{(2,2)\}$$

$$C = \{(6,1), (5,2), (4,3), (3,4), (2,5), (1,6)\}$$

Al utilizar (1.1.7) encontramos $\mathrm{P}(A) = 10/36 \cong 0.278$, $\mathrm{P}(B) = 11/36 \cong 0.306$ y $\mathrm{P}(C) = 6/36 \cong 0.167$. El evento B^c =ningún dos es observado, tiene probabilidad

$$\mathrm{P}(B^c) = 1 - \mathrm{P}(B) = 25/36 \cong 0.694.$$

El evento $A \cap C$ es la suma de los números observados es 7 y uno de los sumandos es el dos entonces, $A \cap C = \{(5,2), (2,5)\}$ de donde obtenemos $\mathrm{P}(A \cap C) = 1/18 \cong 0.056$.

Si consideramos que los dados no son distinguibles el espacio muestral, de acuerdo con el ejemplo 1.3 es

$$\Omega_1 = \{(k, j); k, j \in \mathbb{N} \text{ con } k \leq j \leq 6\}$$

y si suponemos que los 21 resultados son equiprobables entonces, $\mathrm{P}(A) = 5/21 \cong 0.238$, $\mathrm{P}(B) = 6/21 \cong 0.286$, $\mathrm{P}(C) = 3/21 \cong 0.143$ y $\mathrm{P}(A \cap C) = 1/21 \cong 0.048$.

Observamos que las probabilidades calculadas bajo este espacio muestral no coinciden con las anteriores ¿cuáles son las correctas? Las dos versiones son correctas matemáticamente, pero la experiencia, esto es, el lanzar repetidamente un par de dados nos dice que el primer espacio muestral, el de los 36 resultados posibles, se apega mejor a la realidad. Podemos verificar lo aquí asentado al lanzar un número *grande* de veces (720 es un número *grande*) un par de dados y comparar las frecuencias relativas observadas de los eventos A, B, C y $A \cap C$ con las probabilidades calculadas bajo los dos supuestos. ■

Ahora consideramos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$ no discreto. Si por analogía con los espacios discretos quisiéramos asignar a cada resultado elemental $\omega \in \Omega$ un número positivo, $\mathrm{P}(\{\omega\})$, de tal manera que si A es un subconjunto de Ω entonces, $\mathrm{P}(A) = \sum_{\omega \in A} \mathrm{P}(\{\omega\})$ nos encontramos, en general, con las siguientes dificultades cuando A tiene la cardinalidad del continuo, verbigracia, un intervalo (a, b) donde $a, b \in \mathbb{R}$ con $a < b$:

a) Asignar $\mathrm{P}(\{\omega\}) > 0$ para cada $\omega \in A$,

b) Calcular $\sum_{\omega \in A} \mathrm{P}(\{\omega\})$ de tal manera que $\sum_{\omega \in A} \mathrm{P}(\{\omega\}) \leq 1$.

Estas dificultades nos impiden asignar una probabilidad a *todo subconjunto* de Ω; estas son algunas de las razones por las cuales se introdujo la sigma álgebra (y la sigma álgebra de Borel). Piense en que el conjunto A es el conjunto de Vitali (definido en el ejercicio 1.28).

Ejemplo 1.13

Supongamos que el espacio de probabilidad es $([0,1], [0,1] \cap \mathcal{B}, \lambda)$ donde

$$[0,1] \cap \mathcal{B} = \{A \subseteq [0,1]; A = [0,1] \cap B \text{ y } B \in \mathcal{B}\}$$

Es la colección de subconjuntos de [0,1] que son expresables como la intersección de [0,1] y un elemento de $\mathcal{B}$ y que resulta ser una σ −álgebra (consulte el ejercicio 1.16). λ es la *medida de Lebesgue* restringida a [0,1]. La medida de Lebesgue es una extensión de la noción clásica de longitud a conjuntos más *complicados*; está definida como: si $\{(a_k, b_k)\}$ es una colección de intervalos disjuntos donde $a_k < b_k$ y $k \in N \subseteq \mathbb{N}$ entonces,

$$\lambda\left(\bigcup_{k \in N} (a_k, b_k)\right) = \sum_{k \in N} (b_k - a_k).$$

Y dado un conjunto cerrado de la forma, $[a, b] - \bigcup_{k \in N} (a_k, b_k)$ tenemos que

$$\lambda\left([a, b] - \bigcup_{k \in N} (a_k, b_k)\right) = b - a - \sum_{k \in n} (b_k - a_k).$$

λ restringida a [0,1] es una medida de probabilidad ya que $\lambda([0.1]) = 1$ y satisface, por definición, P1 y P3. Tenemos que,

$$\lambda(\{a\}) = 0, \text{para todo } a \in [0,1]$$

$$\lambda([a, b]) = b - a = \lambda\big((a, b)\big) = \lambda\big((a, b]\big) = \lambda\big([a, b)\big); \text{donde } 0 \leq a \leq b \leq 1$$

$$\lambda([0,1] \cap \mathbb{Q}) = 0.$$

En el ejercicio 1.35 presentamos el *conjunto ternario de Cantor* que es un boreliano, tiene la cardinalidad del continuo y tiene probabilidad cero. Esta es una instancia de un conjunto *más complicado,* pero al que se le puede asignar una probabilidad. ■

En la siguiente sección desarrollamos propiedades de la medida de probabilidad P relacionadas con sucesiones de eventos.

1.2.- P Es una Función Continua

En esta sección mostramos que la medida de probabilidad P es una *función continua*. Esto es, si $\{A_n\}$ es una sucesión convergente de eventos y $\lim_{n \to \infty} A_n = A$ entonces,

$$\lim_{n \to \infty} \mathrm{P}(A_n) = \mathrm{P}\left(\lim_{n \to \infty} A_n\right).$$

La continuidad de P significa que podemos intercambiar las operaciones lím y P con sucesiones convergentes (esta propiedad es compartida, entre otras funciones, por las funciones reales de variable real que son continuas). Primero mostramos la implicación para sucesiones monótonas de eventos y luego generalizamos a sucesiones de eventos convergentes, pero no necesariamente monótonas. Ahora establecemos una definición.

Definición 1.8 - Sucesión Monótona de Eventos

Una sucesión $\{A_n\}$ es *no creciente* si $A_{n+1} \subseteq A_n$ para toda $n \in \mathbb{N}$ y es *no decreciente* si $A_n \subseteq A_{n+1}$ para toda $n \in \mathbb{N}$. Una sucesión de eventos es denominada *monótona* si es no creciente o no decreciente. ■

El lema 1.5 nos presentó dos casos de sucesiones monótonas. El siguiente resultado establece que las sucesiones monótonas de eventos son convergentes.

Teorema 1.2

Sea $\{A_n\}$ una sucesión monótona de eventos en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Entonces,

i. $\{A_n\}$ converge y

$$\lim_{n\to\infty} A_n = \begin{cases} \bigcup_{n=1}^{\infty} A_n & \text{si } \{A_n\} \text{ es no decreciente} \\ \bigcap_{n=1}^{\infty} A_n & \text{si } \{A_n\} \text{ es no creciente} \end{cases}$$

ii. $\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\lim_{n\to\infty} A_n\right)$.

Demostración

Sólo consideramos las sucesiones no decrecientes; dejamos como ejercicio las no crecientes; vea el ejercicio 1.1.

i. Suponemos que $A_n \subseteq A_{n+1}\ \forall\, n \in \mathbb{N}$ y observamos que $\bigcup_{k=n}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k$, para toda $n \in \mathbb{N}$. por tanto,

$$\text{lím sup}\, A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n.$$

Por otra parte, $\bigcap_{k=n}^{\infty} A_k = A_n$, para toda $n \in \mathbb{N}$ ya que $A_n \subseteq A_{n+1} \subseteq A_{n+2} \subseteq \cdots$. En consecuencia,

$$\text{lím inf}\, A_n = \bigcup_{n=1}^{\infty} A_n.$$

Concluimos que

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \tag{1.2.1}$$

ii. Partimos de (1.2.1) para mostrar el resultado propuesto. Tenemos que mostrar que $\mathrm{P}\left(\lim_{n\to\infty} A_n\right) = \mathrm{P}(\bigcup_{n=1}^{\infty} A_n)$. Ahora expresamos $\bigcup_{n=1}^{\infty} A_n$ como una unión de conjuntos disjuntos; recurrimos al lema 1.2. Como $A_n \subseteq A_{n+1}$ tenemos que $\bigcup_{j=1}^{k-1} A_j = A_{k-1}$, por lo que

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} [A_n - A_{n-1}] \text{, donde } A_0 = \emptyset.$$

En consecuencia,

$$\begin{aligned}
\mathrm{P}\left(\lim_{n\to\infty} A_n\right) &= \sum_{n=1}^{\infty} \mathrm{P}(A_n - A_{n-1}) \text{, ya que P es } \sigma - \text{aditiva} \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} [\mathrm{P}(A_k) - \mathrm{P}(A_{k-1})] \text{, porque } A_{k-1} \subseteq A_k \text{ para toda } k \\
&= \lim_{n\to\infty} \mathrm{P}(A_n) \text{, ya que la suma es telescópica.} \blacksquare
\end{aligned}$$

Notamos que $\liminf A_n$ es el límite de una sucesión no decreciente y $\limsup A_n$ es el límite de una sucesión no creciente. De hecho,

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim_{N\to\infty} \bigcup_{n=1}^{N} B_n \text{, donde } B_n = \bigcap_{k=n}^{\infty} A_k.$$

Así, $\liminf A_n$ es el límite de una sucesión no decreciente.

El siguiente teorema es el resultado más importante de la sección y es una generalización del teorema 1.2.

Teorema 1.3

Sea $\{A_n\}$ una sucesión de eventos en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Si $\lim_{n\to\infty} A_n = A$ entonces,

a) $\mathrm{P}(\limsup A_n) \geq \limsup \mathrm{P}(A_n)$ *y* $\mathrm{P}(\liminf A_n) \leq \liminf \mathrm{P}(A_n)$

b) $\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\lim_{n\to\infty} A_n\right) = \mathrm{P}(A).$

Demostración

a) Por definición, $\limsup A_n = \bigcap_{n=1}^{\infty} B_n$ donde $B_n = \bigcup_{k=n}^{\infty} A_k$; entonces

$$\mathrm{P}(\limsup A_n) = \mathrm{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathrm{P}(B_n) \text{, por el teorema 1.2}$$

y como $A_n \subseteq B_n$ tenemos que $\mathrm{P}(A_n) \leq \mathrm{P}(B_n)$ para toda $n \in \mathbb{N}$. Por tanto,

$$\mathrm{P}(\limsup A_n) \geq \limsup \mathrm{P}(A_n).$$

Por otra parte, $\liminf A_n = \bigcup_{n=1}^{\infty} C_n$ donde $C_n = \bigcap_{k=n}^{\infty} A_k$; entonces

$$\mathrm{P}(\liminf A_n) = \mathrm{P}\left(\bigcup_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} \mathrm{P}(C_n) \text{, por el teorema 1.2}$$

Y como $C_n \subseteq A_n$ tenemos que $\mathrm{P}(C_n) \leq \mathrm{P}(A_n) \leq 1$ para toda $n \in \mathbb{N}$. Por tanto,

$$\mathrm{P}(\liminf A_n) \leq \liminf \mathrm{P}(A_n)$$

Por lo tanto,

$$\liminf \mathrm{P}(A_n) \geq \mathrm{P}(\liminf A_n) = \mathrm{P}(\limsup A_n) \geq \limsup\ \mathrm{P}(A_n)$$

ya que $\liminf A_n = \limsup A_n$ por hipótesis. Pero, sabemos que $\mathrm{P}(\liminf A_n) \leq \mathrm{P}(\limsup A_n)$. En consecuencia, la sucesión $\{\mathrm{P}(A_n)\}$ converge y

$$\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\lim_{n\to\infty} A_n\right) = \mathrm{P}(A). \blacksquare$$

Los resultados de esta sección serán utilizados en los Capítulos 3 y 7 por lo que sugerimos mantenerlos en mente. En la siguiente sección introducimos el concepto de independencia estocástica que juega un papel importante en el resto del texto.

1.3.- Independencia Estocástica

La palabra independencia es utilizada en matemáticas en varios contextos. Verbigracia, independencia lineal de vectores o independencia de una función respecto a una variable y el adjetivo *estocástico* es usado para enfatizar que hablaremos de independencia en un sentido probabilístico. De hecho, la palabra estocástico proviene del griego y significa conjetural; perteneciente o relativo al azar. Debido a que en este texto utilizaremos la palabra independencia principalmente en el sentido probabilístico, omitiremos por lo general el adjetivo estocástico. Pasamos ahora a definir el concepto de independencia (estocástica) entre eventos.

Definición 1.9 – Eventos Mutuamente Independientes

Sean $A_1, A_2, \ldots, A_n$ eventos en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Decimos que los eventos son *mutuamente independientes* si y sólo si para todo subconjunto de tamaño k, $2 \leq k \leq n$, $\{j_1, j_2, \ldots, j_k\}$ de $\{1,2, \ldots, n\}$ se tiene la siguiente igualdad

$$\mathrm{P}\left(\bigcap_{i=1}^{k} A_{j_i}\right) = \prod_{i=1}^{k} \mathrm{P}(A_{j_i}). \blacksquare$$

Cuando $n = 2$, la definición establece que los eventos A_1 y A_2 son independientes si y sólo si

$$\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\,\mathrm{P}(A_2)$$

Observe que para $n = 2$ omitimos el adverbio mutuamente. Cuando $n = 3$, la definición 1.9 nos dice que A_1, A_2, A_3 , son mutuamente independientes si y sólo si

$$\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\,\mathrm{P}(A_2),\ \mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_3),\ \mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2)\,\mathrm{P}(A_3)\ \text{ y}$$

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_2)\,\mathrm{P}(A_3)$$

No es suficiente que sean independientes a pares. De hecho, si el experimento aleatorio $\mathcal{E}$ es lanzar dos veces una moneda honesta entonces,

$$\Omega = \{\omega_1 = (A,A), \omega_2 = (A,S), \omega_3 = (S,A), \omega_4 = (S,S)\}$$

donde A es una de las caras y S es la otra cara. Como la moneda es honesta podemos proponer $\mathrm{P}(\{\omega_j\}) = 1/4$ para $j = 1, 2, 3, 4$. Definamos los eventos $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ y $C = \{\omega_1, \omega_4\}$. Entonces,

$$\mathrm{P}(A) = 1/2 = \mathrm{P}(B) = \mathrm{P}(C)$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(\{\omega_1\}) = 1/4 = \mathrm{P}(A)\,\mathrm{P}(B)$$

$$\mathrm{P}(A \cap C) = \mathrm{P}(\{\omega_1\}) = 1/4 = \mathrm{P}(A)\,\mathrm{P}(C)$$

$$\mathrm{P}(B \cap C) = \mathrm{P}(\{\omega_1\}) = 1/4 = \mathrm{P}(B)\,\mathrm{P}(C)$$

Por lo que los eventos son independientes dos a dos, pero

$$\mathrm{P}(A \cap B \cap C) = \mathrm{P}(\{\omega_1\}) = 1/4 \neq \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C).$$

Concluimos que los eventos A, B, C no son mutuamente independientes. Para verificar la independencia mutua de n eventos tenemos que comprobar la independencia a pares, tercias, etc. en total, $(2^n - n - 1)$ igualdades (en el capítulo 2 justificaremos el número).

Informalmente se acostumbra a decir que los eventos $A_1, A_2, \dots, A_n$ son independientes cuando la ocurrencia de uno de ellos no altera la ocurrencia o no ocurrencia de los demás. Cuando en la discusión entran dos eventos, A y B, algunas personas acostumbran a decir que A y B son independientes cuando *A no tiene que ver con B*. Esta interpretación me parece peligrosa porque *A no tiene que ver con B* es luego interpretada como $A \cap B = \emptyset$ y esto es erróneo. De hecho, si los eventos A y B son mutuamente excluyentes entonces, no pueden ser independientes ya que la ocurrencia de uno implica la no ocurrencia del otro. Es más, si $\mathrm{P}(A) > 0, \mathrm{P}(B) > 0$ y $A \cap B = \emptyset$ tendremos que $\mathrm{P}(A \cap B) = \mathrm{P}(\emptyset) = 0 \neq \mathrm{P}(A)\mathrm{P}(B) > 0$. Por tanto, para que dos o más eventos sean mutuamente independientes es necesario que las intersecciones sean no vacías, esto es, es necesario que puedan ocurrir simultáneamente. La independencia entre eventos tiene consecuencias interesantes; las concentramos en el siguiente

Lema 1.7

Sean A y B dos eventos independientes entonces,

i) A y B^c son independientes (y por simetría A^c y B también son independientes).

ii) A^c *y* B^c son independientes.

Demostración

Utilizamos la igualdad $A = (A \cap B^c) \cup (A \cap B)$. Como $A \cap B^c$ y $A \cap B$ son disjuntos, P3 nos permite afirmar que

$$\begin{aligned}\mathrm{P}(A) &= \mathrm{P}(A \cap B^c) + \mathrm{P}(A \cap B)\\ &= \mathrm{P}(A \cap B^c) + \mathrm{P}(A)\mathrm{P}(B), \text{por el supuesto de independencia.}\end{aligned}$$

En consecuencia,

$$\mathrm{P}(A \cap B^c) = \mathrm{P}(A) - \mathrm{P}(A)\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B^c)$$

Por lo tanto, A y B^c son independientes. Al intercambiar A y B (esto es la simetría) deducimos que A^c y B son independientes. El segundo inciso es dejado como ejercicio. ■

Observación:

- El lema 1.7 es generalizable a más de dos eventos. Verbigracia, para 3 eventos A, B y C mutuamente independientes tendremos que los siguientes eventos son mutuamente independientes

$$A^c, B, C;\ A, B^c, C;\ A, B, C^c;\ A^c, B^c, C;\ A^c, B, C^c;\ A, B^c, C^c \text{ y } A^c, B^c, C^c$$

A manera de instancia, consideremos A^c, B^c, C. Como $C = [C \cap (A \cup B)^c] \cup [C \cap (A \cup B)]$, tenemos que,

$$\begin{aligned} \mathrm{P}(C) &= \mathrm{P}(C \cap A^c \cap B^c) + \mathrm{P}[(C \cap A) \cup (C \cap B)], \text{por P3 y las Leyes de De Morgan} \\ &= \mathrm{P}(C \cap A^c \cap B^c) + \mathrm{P}(C \cap A) + \mathrm{P}(C \cap B) - \mathrm{P}(A \cap B \cap C), \text{por el teorema 1.1} \\ &= \mathrm{P}(C \cap A^c \cap B^c) + \mathrm{P}(A)\mathrm{P}(C) + \mathrm{P}(B)\mathrm{P}(C) - \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C) \end{aligned}$$

al utilizar la hipótesis de independencia mutua. Ahora,

$$\begin{aligned} \mathrm{P}(C \cap A^c \cap B^c) &= \mathrm{P}(C) - \mathrm{P}(A)\mathrm{P}(C) - \mathrm{P}(B)\mathrm{P}(C) + \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C) \\ &= \mathrm{P}(C)[1 - \mathrm{P}(A) - \mathrm{P}(B) + \mathrm{P}(A)\mathrm{P}(B)] \\ &= \mathrm{P}(C)[1 - \mathrm{P}(A)][1 - \mathrm{P}(B)] \\ &= \mathrm{P}(C)\ \mathrm{P}(A^c)\ \mathrm{P}(B^c). \end{aligned}$$

Para deducir la última igualdad tuvimos que utilizar la independencia dos a dos y de los tres ¿Verdad lector que ahora puede generalizar el lema 1.7 a n eventos mutuamente independientes? A continuación, presentamos dos casos para ilustrar la utilidad del lema 1.7.

Ejemplo 1.14

Sea $\mathcal{E}$ lanzar dos veces un dado honesto. El espacio muestral es entonces,

$$\Omega = \{(k, j); k, j \in \mathbb{N}; k, j \leq 6\}.$$

Definamos los eventos $A =$ observamos el 1 en el primer lanzamiento y $B =$ observamos el 1 en el segundo lanzamiento. Al hacer los 36 resultados elementales equiprobables podemos verificar que $\mathrm{P}(A) = 1/6 = \mathrm{P}(B)$, mientras que $\mathrm{P}(A \cap B) = 1/36$. Como, $\mathrm{P}(A \cap B) = 1/36 = \mathrm{P}(A)\mathrm{P}(B)$ concluimos que A y B son independientes. Como consecuencia del lema 1.7 tenemos,

$$\mathrm{P}(A^c \cap B^c) = (1 - 1/6)(1 - 1/6) = 25/36 \cong 0.694$$

$$\mathrm{P}(A^c \cap B) = (1 - 1/6)(1/6) = 5/36 \cong 0.139$$

$$\mathrm{P}(A \cap B^c) = (1/6)(1 - 1/6) = 5/36 \cong 0.139$$

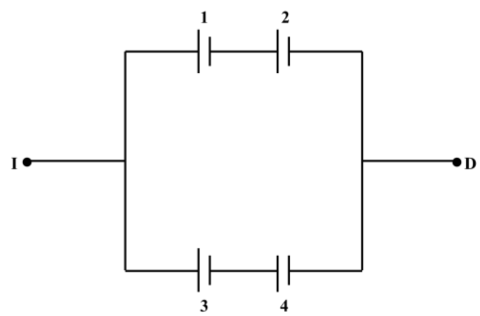
$A^c \cap B^c$ significa no observar el 1 ni en el primer lanzamiento ni en el segundo lanzamiento. $A^c \cap B$ significa no observar el 1 en el primer lanzamiento, pero sí observarlo en el segundo; ¿qué significa $A \cap B^c$? ■

Ejemplo 1.15

En la Gráfica 1.2 mostramos el diagrama de un circuito eléctrico conectado en paralelo. La corriente eléctrica pasará del punto I al punto D siempre y cuando al menos una de las dos conexiones funcione. La conexión superior (inferior) funciona cuando los condensadores 1 y 2 (3 y 4) funcionan simultáneamente. Supongamos que los condensadores funcionan, desde el punto de vista probabilístico, independientemente y que la probabilidad de que cualquiera de ellos se descomponga en un período de t horas es q; $0 < q < 1$ ¿Cuál es la probabilidad de que el flujo eléctrico de I a D no se interrumpa en un lapso de t horas?

Gráfica 1.2

Diagrama de un Circuito Conectado en Paralelo



Definamos los siguientes eventos: $A =$ la corriente de I a D no se interrumpe en t horas, $B =$ la conexión superior trabaja durante t horas y $C =$ la conexión inferior trabaja durante t horas. Entonces, $A = B \cup C$ ya que mientras al menos una de las dos conexiones trabaje, la corriente pasará de I a D. Así, $\mathrm{P}(A) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C)$, ya que B y C no son mutuamente excluyentes. Para calcular $\mathrm{P}(B)$ y $\mathrm{P}(C)$ tenemos que recurrir al *supuesto* que los 4 condensadores operan independientemente. Introducimos el evento $D_j =$ el condensador j funciona t horas, $j = 1, 2, 3, 4$. De donde $B = D_1 \cap D_2$ y $C = D_3 \cap D_4$ y. El lema 1.7 nos permite afirmar que

$$\mathrm{P}(B) = \mathrm{P}(D_1)\mathrm{P}(D_2) = (1-q)^2 = \mathrm{P}(C) \quad \text{y} \quad \mathrm{P}(B \cap C) = (1-q)^4.$$

En consecuencia,

$$\mathrm{P}(A) = 2(1-q)^2 - (1-q)^4 = (1-q)^2[2-(1-q)^2].$$

Si $q = 0.1$, $\mathrm{P}(A) \cong 0.964$. Es interesante calcular la probabilidad de que la corriente pase de I a D si el circuito sólo tuviera la conexión superior. En este caso la probabilidad vale $\mathrm{P}(B) \cong 0.810$. Al tener el circuito dos conexiones en paralelo la probabilidad de que no se interrumpa la corriente en t horas aumentó de 0.810 a 0.964 ¿Cuántas conexiones en paralelo debería tener el circuito para que $\mathrm{P}(A)$ fuese ≥ 0.99? Con tres conexiones tendríamos

$$\mathrm{P}(A) = 3(1-q)^2 - 3(1-q)^4 + (1-q)^6$$

y para $q = 0.1$, $\mathrm{P}(A) \cong 0.993$. Observe que la ganancia ya no fue tan grande. Las probabilidades calculadas están sujetas al *supuesto de independencia* y antes de tomar decisiones más vale verificarlo. Si el supuesto es aproximadamente válido, los cálculos nos sugieren que con 3 conexiones en paralelo podemos estar razonablemente seguros de que la corriente no se interrumpirá en t horas. Si el circuito fuera parte del *cerebro electrónico* de una nave espacial que costó varios miles de millones de dólares es posible que deseáramos $\mathrm{P}(A) \geq 0.9999$ ¿Cuántas conexiones en paralelo se necesitarían? Dejamos al estudiante la responsabilidad de hacer el cálculo. Otra alternativa es producir condensadores donde q sea 0.05 o menor. ■

Ahora que hemos estudiado dos ilustraciones de la aplicación del concepto de independencia, observamos que en el ejemplo 1.14 no hubo necesidad de suponer independencia, *fue consecuencia* de Ω y P. Si los lanzamientos están amañados ni Ω ni P lo reflejan. Es más, Ω y P *presuponen* que los lanzamientos *no* están amañados. Esta observación no es una sutilidad. En otras aplicaciones del cálculo de probabilidades, como el muestreo y el diseño de experimentos, la hipótesis de independencia depende crucialmente de la manera en la que es ejecutado el fenómeno aleatorio bajo estudio. En contraste, el ejemplo 1.15 tuvo que exhibir explícitamente el supuesto de independencia.

Respecto a la interpretación de las probabilidades, el ejemplo 1.15 admite no nada más la interpretación de las frecuencias relativas. Es más, $\mathrm{P}(A) \cong 0.964$ con dos conexiones en paralelo, podría ser interpretada como el *grado de credibilidad* de la afirmación, *la corriente no se interrumpirá en t horas*. Un ingeniero podría decir que 0.964 es la *confiabilidad del circuito*. También podríamos decir que la *incertidumbre* de que pase la corriente de I a D en t horas es $0.036 \cong 1 - \mathrm{P}(A)$.

Cuando los eventos $A_1, A_2, \ldots, A_n$ no satisfacen al menos una de las $(2^n - n - 1)$ condiciones que definen la independencia mutua, decimos que los eventos son *dependientes*. En particular, para dos eventos A y B, decimos que son dependientes cuando $\mathrm{P}(A \cap B) \neq \mathrm{P}(A)\,\mathrm{P}(B)$.

A continuación, definimos el concepto de probabilidad condicional, el cual jugará un papel importante en el resto del texto.

Definición 1.10 - Probabilidad Condicional

Sean A y B eventos en $(\Omega, \mathcal{A}, \mathrm{P})$. La *Probabilidad Condicional* del evento A dado que el evento B ocurrió, denotada por $\mathrm{P}(A|B)$, está definida como

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{1.3.1}$$

siempre y cuando $\mathrm{P}(B) \neq 0$. Por simetría,

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} \tag{1.3.2}$$

siempre y cuando $\mathrm{P}(A) \neq 0$. ■

Vale la pena mencionar explícitamente que ni $A|B$ ni $B|A$ representan eventos. Al escribir $\mathrm{P}(A|B)$ estamos calculando la probabilidad condicional del evento A dado que B ya ocurrió. Una notación alternativa sería $\mathrm{P}_B(A)$ pero no es usual. El meollo para entender la probabilidad condicional es visualizar que en el momento en que B ya ocurrió la $\sigma-$álgebra original $\mathcal{A}$ es modificada a $B \cap \mathcal{A}$, que es la colección de subconjuntos de B que son expresables como $B \cap E$ donde $E \in \mathcal{A}$. El espacio muestral es ahora $\Omega \cap B = B$ y la medida de probabilidad es $\mathrm{P}(\cdot\,|B)$, esto es, el nuevo espacio de probabilidad es $\big(B, B \cap \mathcal{A}, \mathrm{P}(\cdot\,|B)\big)$. Formalmente, debemos *mostrar* que $B \cap \mathcal{A}$ es una $\sigma-$álgebra y que $\mathrm{P}(\cdot\,|B)$ es una medida de probabilidad. Dejamos como ejercicio la demostración de que $B \cap \mathcal{A}$ es una $\sigma-$álgebra, pero mostramos que $\mathrm{P}(\cdot\,|B)$ es una medida de probabilidad definida en $B \cap \mathcal{A}$.

Teorema 1.4

Sea B un evento en $(\Omega, \mathcal{A}, \mathrm{P})$ con $\mathrm{P}(B) > 0$. Entonces, $\mathrm{P}(\cdot\,|B)$ es una medida de probabilidad definida en $B \cap \mathcal{A}$.

Demostración

Debemos verificar que $\mathrm{P}(\cdot\,|B)$ satisface P1, P2 y P3.

1) Sea A un evento entonces en $(\Omega, \mathcal{A}, \mathrm{P})$ entonces,

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \geq 0, \text{ya que } \mathrm{P}(B) > 0 \text{ y } \mathrm{P}(A \cap B) \geq 0.$$

Obviamente $A \cap B \in (B \cap \mathcal{A})$.

2) Como el nuevo espacio muestral es B, que obviamente es un elemento de $B \cap \mathcal{A}$, tenemos

$$\mathrm{P}(B|B) = \frac{\mathrm{P}(B \cap B)}{\mathrm{P}(B)} = 1$$

3) Sean $A_1, A_2, \ldots$ eventos mutuamente excluyentes en B, esto es, $(A_k \cap B) \cap \left(A_j \cap B\right) = \emptyset$ para todo par k, j con $k \neq j$. Como $(A_k \cap B) \in (B \cap \mathcal{A})$ para toda k, tenemos que

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k \,|B\right) = \frac{1}{\mathrm{P}(B)} \mathrm{P}\left[\bigcup_{k=1}^{\infty} (A_k \cap B)\right]$$

$$= \frac{1}{\mathrm{P}(B)} \sum_{k=1}^{\infty} \mathrm{P}(A_k \cap B)\,, \text{por la } \sigma - \text{aditividad de P}$$

$$= \sum_{k=1}^{\infty} \mathrm{P}(A_k|B)\,, \text{por la definición de } \mathrm{P}(\cdot\,|B)$$

Por tanto, $\mathrm{P}(\cdot\,|B)$ satisface P1, P2 y P3 y es una medida de probabilidad. ■

Observación

- Como $\mathrm{P}(\cdot\,|B)$ es una medida de probabilidad podemos utilizar los resultados hasta aquí desarrollados para P; verbigracia,

$$\mathrm{P}(A^c|B) = 1 - \mathrm{P}(A|B)$$

$$\mathrm{P}(E \cup F|B) = \mathrm{P}(E|B) + \mathrm{P}(F|B) - \mathrm{P}(E \cap F|B)$$

$$\mathrm{P}(C|B) \leq \mathrm{P}(D|B) \text{ cuando } C \cap B \subseteq D \cap B$$

para eventos A, C, D, E y F. También podremos utilizar la continuidad de la medida de probabilidad.

Ahora retomamos el cálculo de $\mathrm{P}(A \cap B)$ para eventos dependientes. De (1.3.1) obtenemos

$$\mathrm{P}(A \cap B) = \mathrm{P}(B)\, \mathrm{P}(A|B), \text{cuando } P(B) > 0 \tag{1.3.3}$$

y de (1.3.2)

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\, \mathrm{P}(B|A), \text{cuando } P(A) > 0. \tag{1.3.4}$$

(1.3.3) y (1.3.4) son conocidas conjuntamente como la *regla de la multiplicación*; (1.3.3) nos dice que la probabilidad de que A y B ocurran es el producto de la probabilidad de que B ocurra por la probabilidad condicional de A dado que B ya ocurrió ¿qué dice (1.3.4)? Enseguida establecemos una condición necesaria y suficiente para que dos eventos A y B sean independientes.

Lema 1.8

Sean A y B dos eventos en un espacio probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Si $\mathrm{P}(B) > 0$ entonces, A y B son independientes si y sólo si $\mathrm{P}(A|B) = \mathrm{P}(A)$.

Demostración

$\Longrightarrow$) Si A y B son independientes entonces $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$ por lo que

$$\mathrm{P}(A|B) = \frac{1}{\mathrm{P}(B)}\mathrm{P}(A)\,\mathrm{P}(B) = \mathrm{P}(A).$$

$\Longleftarrow$) Si $\mathrm{P}(A|B) = \mathrm{P}(A)$ entonces la regla de la multiplicación nos permite afirmar

$$\mathrm{P}(A \cap B) = \mathrm{P}(B)\mathrm{P}(A|B) = \mathrm{P}(A)\,\mathrm{P}(B)$$

por lo que concluimos que A y B son independientes. ■

El lema también pudo ser establecido con $\mathrm{P}(B|A) = \mathrm{P}(B)$ y el supuesto $\mathrm{P}(A) > 0$. La regla de la multiplicación puede ser generalizada, bajo supuestos adecuados, a tres o más eventos. Verbigracia, para los eventos A_1, A_2, A_3 tenemos

$$\begin{aligned}
\mathrm{P}(A_1 \cap A_2 \cap A_3) &= \mathrm{P}[A_1 \cap (A_2 \cap A_3)] \\
&= \mathrm{P}(A_1)\,\mathrm{P}(A_2 \cap A_3|A_1), \text{por (1.3.3) cuando } \mathrm{P}(A_1) > 0 \\
&= \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1 \cap A_2), \text{cuando } \mathrm{P}(A_1 \cap A_2) > 0.
\end{aligned}$$

Al permutar el orden de los índices obtenemos cinco igualdades adicionales (¡encuéntrelas!) En total se obtienen 3! igualdades. En el ejercicio 1.4 se le pide generalizar la regla de la multiplicación a n eventos. Enseguida mostramos un caso que ilustra el uso de la regla de la multiplicación.

Ejemplo 1.16

Suponga que al llegar a una fiesta tres mujeres entregan sus abrigos a la anfitriona que los regresa al azar al terminar la fiesta ¿Cuál es la probabilidad de que ninguna de las mujeres reciba su propio abrigo?

Definamos los siguientes eventos $A =$ ninguna mujer recibe su propio abrigo, $E_j =$ la mujer j recibe su propio abrigo para $j = 1,2,3$. Entonces, la probabilidad buscada es

$$\mathrm{P}(A) = \mathrm{P}(E_1^c \cap E_2^c \cap E_3^c) = 1 - \mathrm{P}\left(\bigcup_{j=1}^{3} E_j\right).$$

Al utilizar la fórmula desarrollada en la observación al teorema 1.1, obtenemos

$$\mathrm{P}(A) = 1 - \mathrm{P}(E_1) - \mathrm{P}(E_2) - \mathrm{P}(E_3) + \mathrm{P}(E_1 \cap E_2) + \mathrm{P}(E_1 \cap E_3) + \mathrm{P}(E_2 \cap E_3) - \mathrm{P}(E_1 \cap E_2 \cap E_3)$$

Es claro que $\mathrm{P}(E_j) = 1/3$ ya que hay un caso favorable y 3 posibles. Ahora, $\mathrm{P}(E_i \cap E_j) = \mathrm{P}(E_i)\,\mathrm{P}(E_j|E_i)$, por la regla de la multiplicación. Pero, $\mathrm{P}(E_j|E_i) = 1/2$ ya que al haber entregado el abrigo correcto a la mujer 1 quedan dos por entregar y solamente hay un favorable. Por tanto, $\mathrm{P}(E_i \cap E_j) = (1/3)(1/2) = 1/6$. Similarmente,

$$\mathrm{P}(E_1 \cap E_2 \cap E_3) = \mathrm{P}(E_1)\,\mathrm{P}(E_2|E_1)\,\mathrm{P}(E_3|E_1 \cap E_2) = (1/3)(1/2)(1) = 1/6$$

Así,

$$\mathrm{P}(A) = 1 - 3(1/3) + 3(1/6) - 1/6 = 1/3 \cong 0.333$$

Podemos interpretar la probabilidad calculada de la siguiente manera. Si repetimos un número *grande* de veces el (experimento aleatorio) regresar los abrigos al azar, en aproximadamente un tercio de las veces ninguna mujer recibirá su propio abrigo. También la podemos interpretar como el grado de credibilidad de la afirmación, ninguna mujer recibirá su propio abrigo. Este ejemplo será generalizado en el capítulo 2. ■

Enseguida damos una definición que nos permitirá establecer dos resultados importantes La Fórmula de la Probabilidad Total y el lema de Bayes.

Definición 1.11 - Partición de Ω

Sean $A_1, A_2, \ldots$ eventos mutuamente excluyentes, esto es, $A_k \cap A_j = \emptyset$ para $k \neq j$ tales que $\bigcup_{k=1}^{\infty} A_k = \Omega$. Decimos que $A_1, A_2, \ldots$ es una *partición* de Ω. ■

Observaciones

1) Es posible definir particiones finitas, esto es, con un número finito de eventos. Así, si $A_1, A_2, \ldots, A_n$ son eventos mutuamente excluyentes y $\bigcup_{k=1}^{n} A_k = \Omega$ entonces, $A_1, A_2, \ldots$, A_n forman *una partición* de Ω.

2) Si los eventos $A_1, A_2, \ldots, A_n$ forman una partición decimos que son mutuamente excluyentes y exhaustivos.

3) Si $A_1, A_2, \ldots$ es una partición de Ω y B es un evento entonces,

$$B = B \cap \Omega = \mathrm{B} \cap \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} (A_k \cap B)$$

de donde podemos deducir

$$\begin{aligned} \mathrm{P}(B) &= \sum_{k=1}^{\infty} \mathrm{P}(A_k \cap B)\,, \text{porque } (A_k \cap B) \cap (A_j \cap B) = \emptyset \text{ si } k \neq j \\ &= \sum_{k=1}^{\infty} \mathrm{P}(A_k)\mathrm{P}(B|A_k) \qquad (1.3.5) \end{aligned}$$

al aplicar la regla de la multiplicación. La igualdad (1.3.5) es conocida como la *Fórmula de la Probabilidad Total*. Presentamos ahora dos instancias de aplicación de la fórmula (1.3.5).

Ejemplo 1.17

En algunas encuestas se desea averiguar sobre asuntos *delicados* como comportamiento sexual, consumo de drogas, etc. Como la mayor parte de la gente rehúsa responder o miente en estos tópicos, Warner (1965) ideó la siguiente técnica (conocida como *respuesta aleatorizada*). El entrevistado selecciona al azar una carta de una baraja que tiene dos tipos de cartas. En un tipo hay una pregunta inocua; verbigracia ¿Nació usted un domingo? mientras que en el otro tipo de carta hay una pregunta *delicada*; por ejemplo ¿ha consumido usted mariguana? De esta manera, el entrevistador obtiene una respuesta (sí o no) pero no sabe a qué pregunta; esa es la protección del entrevistado. El investigador, esto es, el interesado en obtener la información debe diseñar las preguntas inocuas de tal manera que conozca la probabilidad de un sí o no. Verbigracia, sin mucho riesgo podemos afirmar que $1/7 \cong \mathrm{P}(\text{nació en domingo})$. También, el investigador conoce la proporción p de cartas que tiene la pregunta inocua.

Definamos el evento A como la pregunta inocua es formulada. Entonces, A^c es el evento la pregunta delicada es formulada. Sea B el evento: la respuesta es sí. Entonces, por la fórmula de la probabilidad total

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(A)\,\mathrm{P}(B|A) + \mathrm{P}(A^c)\,\mathrm{P}(B|A^c)\\ &= p(1/7) + (1-p)\,\mathrm{P}(B|A^c)\end{aligned}$$

Si p_0 es la proporción observada de respuestas sí entonces, el sentido común nos sugiere hacer

$$p_0 \cong p(1/7) + (1-p)\,\mathrm{P}(B|A^c)$$

de donde

$$\mathrm{P}(B|A^c) \cong \frac{p_0 - p(1/7)}{1-p}.$$

desgraciadamente esta aproximación no siempre es adecuada (no tenemos garantía que $0 \leq \mathrm{P}(B|A^c) \leq 1$); sugiero consultar el artículo de Warner para que conocer como *estimar* adecuadamente $\mathrm{P}(B|A^c)$. Si el 20% de las respuestas fueron sí, $p_0 = 0.20$, y el 60% de las cartas tiene la pregunta inocua, $p = 0.60$, entonces, $\mathrm{P}(B|A^c) \cong 0.286$, esto es, estimamos que el 28.6% ha consumido mariguana. ■

Ejemplo 1.18

Consideramos dos urnas U_1 y U_2. Supongamos que la urna U_1 tiene 10 canicas, 6 blancas y 4 negras mientras que la urna U_2 tiene 9 canicas, 3 blancas y 6 negras. De la urna U_1 extraemos secuencialmente al azar, con probabilidades iguales, dos canicas y sin ver los colores las colocamos en la urna U_2. Posteriormente extraemos al azar con probabilidades iguales una canica de U_2 ¿Cuál es la probabilidad de que la canica extraída de U_2 sea blanca?

Definamos los siguientes eventos:

$$\begin{aligned} E &= \text{la canica extraída de } U_2 \text{ es blanca}\\ A_1 &= \text{las canicas extraídas de } U_1 \text{ fueron blancas,}\\ A_2 &= \text{las canicas extraídas de } U_1 \text{ fueron negras}\\ A_3 &= \text{las canicas extraídas de } U_1 \text{ fueron de colores distintos.}\end{aligned}$$

Observemos que A_1, A_2 y A_3 son mutuamente excluyentes y exhaustivos; son las posibles causas del evento E. Entonces, de acuerdo con la Fórmula de la Probabilidad Total

$$\mathrm{P}(E) = \sum_{j=1}^{3} \mathrm{P}(A_j)\,\mathrm{P}(E|A_j). \tag{1.3.6}$$

Ahora bien, $A_1 = D_1 \cap D_2$ donde D_j es el evento la extracción j produce una canica blanca para $j = 1, 2$. La regla de la multiplicación nos permite afirmar que

$$\mathrm{P}(A_1) = \mathrm{P}(D_1)\,\mathrm{P}(D_2|D_1) = (6/10)(5/9) = 1/3 \cong 0.333$$

ya que en la primera extracción hay 6 favorables y 10 totales mientras que en la segunda extracción quedan 5 favorables y 9 totales. Similarmente,

$$\mathrm{P}(A_2) = (4/10)(3/9) = 2/15 \cong 0.133$$

$$\mathrm{P}(A_3) = (6/10)(4/9) + (4/10)(6/9) = 8/15 \cong 0.533$$

En el cálculo de $\mathrm{P}(A_3)$ existen dos posibilidades, la primera extracción fue una blanca y la segunda una negra o la primera extracción fue una negra y la segunda una blanca. Verificamos además que $\mathrm{P}(A_1 \cup A_2 \cup A_3) = 5/15 + 2/15 + 8/15 = 1$.

Por otra parte, $\mathrm{P}(E|A_1) = 5/11$, $\mathrm{P}(E|A_2) = 3/11$ y $\mathrm{P}(E|A_3) = 4/11$. Al sustituir en (1.3.6) las probabilidades calculadas obtenemos, $\mathrm{P}(E) = 21/55 \cong 0.382$. ∎

Ahora pasamos a derivar un resultado conocido como el lema de Bayes o Fórmula de Bayes. El resultado es consecuencia de la definición de la probabilidad condicional y de la fórmula de la probabilidad total, pero tiene especial relevancia porque es la piedra angular de una teoría Estadística llamada Bayesiana.

Lema 1.9 (Lema de Bayes)

Sea $A_1, A_2, \dots, A_n$ una partición de Ω. Sea E un evento con probabilidad positiva. Entonces,

$$\mathrm{P}(A_k|E) = \frac{\mathrm{P}(A_k)\,\mathrm{P}(E|A_k)}{\sum_{j=1}^{n} \mathrm{P}(A_j)\,\mathrm{P}(E|A_j)}\,, 1 \le k \le n. \tag{1.3.7}$$

Demostración

Por definición,

$$\mathrm{P}(A_k|E) = \frac{\mathrm{P}(A_k \cap E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(A_k)P(E|A_k)}{\mathrm{P}(E)}, \tag{1.3.8}$$

por la regla de la multiplicación. Al aplicar la Fórmula de la Probabilidad Total a $\mathrm{P}(E)$ obtenemos

$$\mathrm{P}(E) = \sum_{j=1}^{n} \mathrm{P}(A_j)\,\mathrm{P}(E|A_j) \tag{1.3.9}$$

y al sustituir (1.3.9) en (1.3.8) obtenemos (1.3.7). ∎

El lema de Bayes adquiere un sentido especial cuando pensamos que $A_1, A_2, \dots, A_n$ son las posibles causas, mutuamente excluyentes y exhaustivas, de un evento E y nos preguntamos ¿cuál de esas causas es más probable de haber producido E? Esto es, una vez observado el evento E queremos determinar cuál es la causa más probable. La fórmula (1.3.7) tiene como ingredientes $\mathrm{P}\big(A_j\big)$ y $\mathrm{P}\big(E|A_j\big)$; $\mathrm{P}\big(A_j\big)$ nos indica la credibilidad de la causa A_j mientras que $\mathrm{P}\big(E|A_j\big)$ es la probabilidad del evento E al suponer que A_j es la causa. Presentamos ahora una instancia.

Ejemplo 1.19

Supongamos que en una urna hay seis monedas. Una de las monedas tiene dos caras A, tres son honestas y las dos últimas están cargadas de tal manera que la cara A aparece el 75% de las veces. Seleccionamos al azar con probabilidades iguales una moneda y sin saber cuál fue extraída la lanzamos. Si la cara A es observada ¿cuál es la probabilidad de que la moneda con dos caras A haya sido lanzada?

Definamos los siguientes eventos $E =$ una cara A observada, $A_1 =$ la moneda con dos caras A es seleccionada, $A_2 =$ una moneda honesta es seleccionada y $A_3 =$ una moneda cargada es seleccionada. La probabilidad buscada es $\mathrm{P}(A_1|E)$ y de acuerdo con el lema de Bayes

$$\mathrm{P}(A_1|E) = \frac{\mathrm{P}(A_1)\,\mathrm{P}(E|A_1)}{\sum_{j=1}^{3}\mathrm{P}\big(A_j\big)\,\mathrm{P}\big(E|A_j\big)}$$

Ahora, $\mathrm{P}(A_1) = 1/6$, $\mathrm{P}(A_2) = 3/6$ y $\mathrm{P}(A_3) = 2/6$ ya que la selección se hace con probabilidades iguales. Por otra parte, $\mathrm{P}(E|A_1) = 1$, $\mathrm{P}(E|A_2) = 1/2$ y $\mathrm{P}(E|A_3) = 3/4$. Entonces, $\mathrm{P}(A_1|E) = 1/4 = 0.25$. Similarmente, $\mathrm{P}(A_2|E) = 3/8 = 0.375 = \mathrm{P}(A_3|E)$, por lo que la moneda con dos caras A no es la causa más probable de observar una cara A. De hecho, es más probable que la cara A provenga de una de las tres monedas honestas. Invito al lector a que escriba explícitamente el espacio muestral y que verifique que A_1, A_2 y A_3 son exhaustivos. ■

1.4.- Ejercicios

1.- Complete la demostración del teorema 1.2.

Sugerencia Considere la sucesión $\{A_n^c\}$.

2.- Suponga que $0 < \mathrm{P}(B) < 1$. Si $\mathrm{P}(A|B) = \mathrm{P}(A|B^c)$ ¿será verdad que A y B son eventos independientes? Antes de emplear el álgebra piense en el significado de la condición $\mathrm{P}(A|B) = \mathrm{P}(A|B^c)$ ¿Qué le sugiere el sentido común?

3.- Obtenga una fórmula para $\mathrm{P}(\bigcup_{k=1}^{n} A_k)$, con $n \geq 3$, en términos de $\mathrm{P}(A_i)$; $\mathrm{P}\big(A_i \cap A_j\big), i < j$; $\mathrm{P}\big(A_i \cap A_j \cap A_k\big), i < j < k$; etc.

4.- Obtenga una fórmula para $\mathrm{P}(\bigcap_{k=1}^{n} A_k)$, con $n \geq 3$ en términos de un producto de probabilidades condicionales. Haga los supuestos que necesite.

5.- ¿Será verdad que si $\mathrm{P}(A) < \mathrm{P}(B)$ entonces, $\mathrm{P}(A|C) < \mathrm{P}(B|C)$ donde $\mathrm{P}(C) > 0$?

6.- Encuentre el espacio muestral asociado a los siguientes fenómenos aleatorios:

$\mathcal{E}_1$: lanzar una moneda y un dado honestos.

$\mathcal{E}_2$: lanzar una moneda honesta k veces; $k \in \mathbb{N}$.

$\mathcal{E}_3$: lanzar k monedas una vez; $k \in \mathbb{N}$.

$\mathcal{E}_4$: seleccionar al azar un estudiante de un salón de clases que tiene $k \in \mathbb{N}$ alumnos y registrar la fecha de su nacimiento.

7.- ¿Son verdaderas las siguientes afirmaciones?

7.1.- Si $A \subset \mathbb{N}$ entonces, $A \in \mathcal{B}$

7.2.- Si $B \subset \mathbb{Z}$ entonces, $B \in \mathcal{B}$

7.3.- Si $C \subset \mathbb{Q}$ entonces, $C \in \mathcal{B}$

7.4.- $\{x \in \mathbb{R}; x = a + b\sqrt{2};\ a, b \in \mathbb{Q}\} \in \mathcal{B}$

8.- ¿Bajo qué condición será cierto que:

8.1.- $[0, a_n) \to [0, a)$ cuando $n \to \infty$? si sabemos que $a_n \to a > 0$ cuando $n \to \infty$.

8.2.- $[0, a_n) \to [0, a]$ cuando $n \to \infty$? si sabemos que $a_n \to a > 0$ cuando $n \to \infty$.

9.- Generalice la Desigualdad de Boole de la siguiente manera

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathrm{P}(A_k)$$

donde cada A_k es un evento.

Sugerencias Puede tomar dos caminos, uno por medio de un proceso límite y otro al expresar $\bigcup_{k=1}^{\infty} A_k$ en una unión disjunta. En el segundo, recurra al lema 1.2 y muestre $\mathrm{P}(\bigcup_{m=1}^{\infty} A_m) = \mathrm{P}(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathrm{P}(B_n) \leq \sum_{n=1}^{\infty} \mathrm{P}(A_n)$.

10.- Muestre que la igualdad $\mathrm{P}(A - B) = \mathrm{P}(A) - \mathrm{P}(B)$ es falsa cuando B no es un subconjunto de A.

11.- Para eventos arbitrarios $A_1, A_2, \dots, A_n; n \geq 2$

11.1.- Muestre la desigualdad

$$\text{mín}\{\mathrm{P}(A_k); k = 1,2, \dots n\} \geq \mathrm{P}\left(\bigcap_{k=1}^{n} A_k\right) \geq 1 - \sum_{j=1}^{n} \mathrm{P}(A_j^c)$$

La desigualdad de la extrema derecha es conocida como la Desigualdad de Bonferroni; ampliamente utilizada en Estadística; en particular en comparaciones múltiples.

11.2.- Para $n = 4$ suponga que $\mathrm{P}(A_1) = 0.95, \mathrm{P}(A_2) = 0.90 = \mathrm{P}(A_3)\ y\ \mathrm{P}(A_4) = 0.85$. Acote por abajo la probabilidad de la ocurrencia simultánea de los cuatro eventos.

11.3.- Para $n = 9$ suponga que $\mathrm{P}(A_k) = 0.95$ para toda k. Acote por abajo la probabilidad de la ocurrencia simultánea de los nueve eventos ¿Se imagina la calidad de la desigualdad con 20 eventos todos con probabilidad 0.95 ?

12.- Sean A y B eventos ¿Será cierto que $\mathrm{P}(A \cap B) \leq \mathrm{P}(A)\mathrm{P}(B)$?

13.- Sean E y F dos eventos. Si $\mathrm{P}(E) = 1/3$ y $\mathrm{P}(F^c) = 1/4$ ¿Es posible que E y F sean disjuntos?

14.- Sean A, B y C eventos ¿Será cierto que $\mathrm{P}(A \cap B \cap C) \leq \text{mín}\{\mathrm{P}(A \cap B), \mathrm{P}(A \cap C), \mathrm{P}(B \cap C)\}$?

15.- *Crítica a la definición clásica de Probabilidad*. Si quisiéramos extender la definición clásica cuando $\#\Omega = \infty$ tendríamos que $\mathrm{P}(A)$ es cero o está indeterminada ¿por qué?

16.- Sea $\mathcal{A}$ una $\sigma-$álgebra de subconjuntos de Ω. Sea $B \in \mathcal{A}$. Muestre que

$$B \cap \mathcal{A} = \{E; E = B \cap F \text{ con } F \in \mathcal{A}\}$$

es una $\sigma-$álgebra de subconjuntos de B.

17.- Sean $\mathcal{X}$ un conjunto no vacío y $\boldsymbol{S}$ una colección no vacía de subconjuntos de $\mathcal{X}$. Decimos que $\boldsymbol{S}$ es un *anillo* de subconjuntos de $\mathcal{X}$ si para todo par $A, B \in$ S tenemos que

$A1.$ $A \cup B \in \boldsymbol{S}$

$A2.$ $A - B \in \boldsymbol{S}$

17.1.- Si $A, B \in \boldsymbol{S}$, muestre que $\emptyset \in \boldsymbol{S}$ y que $A \cap B \in \boldsymbol{S}$.

17.2.- Si $A_1, \dots, A_n \in \boldsymbol{S}$ para $n \in \mathbb{N}$, muestre que $\bigcup_{k=1}^{n} A_k \in \boldsymbol{S}$ y que $\bigcap_{k=1}^{n} A_k \in \boldsymbol{S}$.

Note que no hemos supuesto que $\mathcal{X} \in \boldsymbol{S}$.

Un anillo $\boldsymbol{S}$ tal que si $A_k \in \boldsymbol{S}$ para toda $k \in \mathbb{N}$ tenemos que $\bigcup_{k=1}^{\infty} A_k \in \boldsymbol{S}$ es denominado *sigma anillo* o $\sigma -$ *anillo*. Históricamente, se han desarrollado tanto la teoría de integración como la teoría de la medida al utilizar sigma anillos ya que ciertos investigadores no desean imponer la condición de que $\mathcal{X}$ sea un elemento de la sigma anillo.

Un $\sigma -$ anillo que tiene la propiedad de que $\mathcal{X}$ es uno de sus elementos es denominado $\sigma -$ *álgebra* o *sigma campo*.

Sea $\mathcal{A}$ una sigma álgebra de subconjuntos de $\mathcal{X}$. Una función $\mu: \mathcal{A} \to [0, \infty]$ es una *medida* si para toda $A, B \in \mathcal{A}$ con $A \cap B = \emptyset$ tenemos que

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

Y en forma más general, si $A_j \in \mathcal{A}$ para $j \in \mathbb{N}$ y $A_k \cap A_j = \emptyset$ para $k \neq j$ entonces

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)\,.$$

Decimos que la medida es sigma aditiva o completamente aditiva. De esta manera, μ es una función que a un subconjunto de $\mathcal{X}$ (que es un elemento de una sigma álgebra de subconjuntos de $\mathcal{X}$) le asigna un real no negativo o más infinito y que es sigma aditiva.

Decimos que μ es *normada* si $\mu(\mathcal{X}) < \infty$. En particular, si $\mu(\mathcal{X}) = 1$ se dice que μ es una medida de probabilidad.

17.3.- Muestre que la cardinalidad es una medida con $\mathcal{X} = \mathbb{N}$ y $\mathcal{A} = 2^{\mathcal{X}}$. Observe $\mu(\mathcal{X}) = \#\mathbb{N} = \infty$. Si $\mathcal{X} = \{1,2,\dots,n\}$; con $n \in \mathbb{N}$ tenemos que, $\mu(\mathcal{X}) = n$.

17.4.- Con $\mu =$cardinalidad y $\mathcal{X} = \mathbb{N}$; muestre que las igualdades

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

$$\mu(A - B) = \mu(A) - \mu(B), \text{cuando } B \subset A$$

pueden no tener sentido. Si $\mu(\mathcal{X}) < \infty$, las igualdades son válidas; en esta situación, encuentre una fórmula para $\mu(A \cup B \cup C)$.

En el ejercicio 1.35 se afirma que la cardinalidad con $\mathcal{X} = [0,1]$ y $\mathcal{A} = \mathcal{B} \cap [0,1]$ es una medida.

La medida de Lebesgue (en honor del matemático francés Henri Lebesgue) es una generalización de la longitud, área y volumen usuales.

18.- ¿Será cierto que cualquier evento es independiente del evento imposible $\emptyset$? Observe que $\emptyset$ es ajeno a cualquier evento.

19.- Muestre el inciso ii) del lema 1.7.

20.- Muestre que si los eventos A, B y C son mutuamente independientes entonces $A^c, B\ y\ C$ son también mutuamente independientes.

21.- Sean $A_1, A_2, \dots, A_n$ eventos mutuamente independientes. Muestre que

$$\mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right) = 1 - \prod_{j=1}^{n} \left[1 - \mathrm{P}(A_j)\right]$$

Observe que esta manera de calcular $\mathrm{P}\left(\bigcup_{j=1}^{n} A_j\right)$ es más sencilla que la fórmula deducida en el ejercicio 1.3.

22.- Si A y B son independientes ¿será cierto que $\mathrm{P}(A \cap B|D) = \mathrm{P}(A|D)\ \mathrm{P}(B|D)$? ¿Qué aprendió?

23.- Demuestre o dé un contraejemplo de las siguientes proposiciones. Sean A y B dos eventos

23.1.- Si $\mathrm{P}(A) = \mathrm{P}(B^c) \Rightarrow A^c = B$

23.2.- Si $\mathrm{P}(A) = 0 \Rightarrow A = \emptyset$

23.3.- Si $\mathrm{P}(A) = 0 \Rightarrow \mathrm{P}(A \cap B) = 0$ para todo evento B.

24.- Este problema es conocido como *La Ley de Sucesión de Laplace*. Suponga que tenemos $(N+1)$ urnas y que la composición de la j −ésima urna es: j bolas blancas y $(N-j)$ bolas negras; $0 \le j \le N$. Se selecciona al azar, con probabilidades iguales, una urna y se hacen n selecciones con reposición y todas fueron blancas. Calcule la probabilidad de que la extracción $(n+1)$ rinda una bola blanca.

Sugerencias Sean los eventos B_n =las primeras n extracciones producen n bolas blancas y B =la extracción $(n+1)$ produce una bola blanca. Calcule $\mathrm{P}(B|B_n)$. Laplace utilizó el ejercicio para responder la pregunta ¿cuál es la probabilidad de que salga el sol mañana? Las n bolas blancas observadas son las n salidas consecutivas del sol hasta el día de la pregunta ¿Qué son las urnas?

25.- Sean A_1, A_2, A_3 eventos. Sea E_k el evento exactamente k de los A_j's ocurren para $k = 0, 1, 2, 3$.

25.1.- Encuentre una expresión para E_k.

25.2.- Sean $S_1 = \sum_{j=1}^{3} \mathrm{P}(A_j)$, $S_2 = \sum_{j<k}^{3} \mathrm{P}(A_j \cap A_k)$ y $S_3 = \mathrm{P}(A_1 \cap A_2 \cap A_3)$. Muestre que

$$\mathrm{P}(E_0) = 1 - S_1 + S_2 + S_3$$
$$\mathrm{P}(E_1) = S_1 - 2S_2 + 3S_3$$
$$\mathrm{P}(E_2) = S_2 - 3S_3$$
$$\mathrm{P}(E_3) = S_3$$

Sugerencia $\mathrm{P}(A \cap B^c) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$ para A y B eventos.

26.- Un encuestador estima que la probabilidad de entrevistar a una persona en su hogar en la primera visita es 0.4 pero que la probabilidad de realizar la entrevista en una segunda visita dado que no la realizó en la primera es 0.6. Como regla de trabajo el encuestador hace a lo más dos intentos. Calcule la probabilidad de que el encuestador entreviste a una persona ¿Cuál es la probabilidad de que no haga la entrevista?

27.- En un estudio sobre 1,700 compañías se encontró que el 49% realiza estudios sobre la eficiencia de su publicidad, el 61% hace pronósticos de ventas a corto plazo y el 38% hace las dos actividades. Si seleccionamos al azar una compañía de las 1,700.

27.1.- ¿Cuál es la probabilidad de que la compañía seleccionada haga al menos uno de los dos tipos de estudio?

27.2.- ¿Cuál es la probabilidad de que haga pronósticos de ventas a corto plazo si sabemos que hace estudios sobre la eficiencia de su publicidad?

28.- El conjunto de Vitali está definido de la siguiente manera. Consideremos el intervalo $[0,1]$ y definamos la relación $x \sim y \Leftrightarrow x - y \in \mathbb{Q}$.

28.1.- Muestre que la relación es una relación de equivalencia, esto es, muestre que

a) la relación es reflexiva: $x \sim x$,

b) la relación es simétrica: si $x \sim y \Rightarrow y \sim x$

c) la relación es transitiva: si $x \sim y$ y $y \sim z \Rightarrow x \sim z$.

Como consecuencia, $[0,1]$ es la unión disjunta de clases de equivalencia. Denotamos a las clases de equivalencias con E_α donde $\alpha \in A$ y A es un conjunto no numerable de índices. Una de las clases es $[0,1] \cap \mathbb{Q} = \{x \in \mathbb{Q}; 0 \leq x \leq 1\}$ ya que la diferencia de dos racionales es un racional.

Como $\sqrt{2}$ es irracional, tenemos que si $r_1, r_2 \in \mathbb{Q}$ de tal manera que $(\sqrt{2}-1) + r_1, (\sqrt{2}-1) + r_2 \in [0,1]$ entonces, $(\sqrt{2}-1) + r_1$ y $(\sqrt{2}-1) + r_2$ pertenecen a la misma clase de equivalencia ya que su diferencia es un racional. Lo mismo sucedería con $(\pi - 3)$, $(\sqrt{3}-1)$ etc.

28.2.- De acuerdo con el axioma de selección podemos escoger un elemento de cada clase de equivalencia. Sea x_α el elemento seleccionado de la clase de equivalencia E_α. Formamos un conjunto V, un conjunto de Vitali, como $V = \{x_\alpha; \alpha \in A\}$. Es claro que el conjunto depende de los elementos escogidos.

Es posible mostrar que V no es medible bajo la medida de Lebesgue y por tanto no es un boreliano (si lo fuese tendría medida de Lebesgue).

29.- Sea $\{A_n\}$ es una sucesión de eventos en el espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Muestre que $\text{lím inf}\, A_n \subseteq \text{lím sup}\, A_n$.

30.- Sean $f: \mathbb{R} \to \mathbb{R}$ una función continua y no negativa y $[a, b]$ un intervalo $(a < b)$. Entonces, la función $Q([a, b]) = \int_a^b f(x)\, dx$, asocia un real no negativo a cada intervalo (como una medida). Sea $\{f_n(\cdot)\}$ una sucesión de funciones continuas, no negativas tal que $f_n(x) \to f(x)$ para toda $x \in [a, b]$. Definimos $Q_n([a, b]) = \int_a^b f_n(x)\, dx$ para toda n. Nos preguntamos ¿será verdad que

$$\lim_{n\to\infty} Q_n([a, b]) = \lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b \lim_{n\to\infty} f_n(x)\ dx = Q([a, b])?$$

La respuesta es no. El siguiente caso particular exhibe un contraejemplo. Sea $f_n(x) = n^2 x(1 - x^2)^n$ para $x \in [0,1]$. Muestre que

30.1.- $\lim_{n\to\infty} f_n(x) = 0$, para toda $x \in [0,1]$. Esto es, para cada $x \in [0,1]$ dado $\varepsilon > 0$ existe $N(\varepsilon, x)$ tal que $f_n(x) < \varepsilon$ si $n > N(\varepsilon, x)$.

30.2.- $\int_0^1 f_n(x)\, dx = n^2/(2n + 2) \to \infty$ cuando $n \to \infty$.

30.3.- $\lim_{n\to\infty} \int_0^1 f_n(x)\, dx \neq \int_0^1 \lim_{n\to\infty} f_n(x)\, dx$.

31.- Sea $\{A_n\}$ una sucesión de eventos en un espacio de probabilidad. Suponga que $\omega_0 \in A_k$ donde $k = 2^n$ para n es un natural, esto es, $\omega_0 \in A_2$, $\omega_0 \in A_4$, $\omega_0 \in A_8$, etc. Muestre que $\omega_0 \in \text{lím sup}\, A_n$.

32.- Suponga que $(\Omega, \mathcal{A}, \mathrm{P})$ es un espacio de probabilidad donde $\Omega = \mathbb{N}$, $\mathcal{A} = 2^{\Omega}$ y que $\mathrm{P}(\{j\}) = 2^{-j}$ para toda $j \in \mathbb{N}$. Definamos los eventos $A =\{ n \in \mathbb{N}; n$ es impar$\}$, $B =\{ n \in \mathbb{N}; n$ es par$\}$, $C =\{ n \in \mathbb{N}; n$ es múltiplo de $6\}$ y $D =\{ n \in \mathbb{N}; n$ es múltiplo de $5\}$.

32.1.- Verifique que $\mathrm{P}(\Omega) = 1$.

32.2.- Calcule $\mathrm{P}(A), \mathrm{P}(B), \mathrm{P}(C)$ y $\mathrm{P}(D)$.

32.3.- Calcule $\mathrm{P}(C \cup D)$ y $\mathrm{P}(A \cap D)$.

Sugerencias Utilice la igualdad $(A.5)$ del apéndice A. Tenga en mente el ejercicio 1.17.

33.- En una ciudad habitan $2N$ personas, N mujeres y N hombres. Suponga que el 5% de los hombres y 2.5% de las mujeres son daltónicos. Se selecciona al azar una persona y resulta ser daltónica ¿Cuál es la probabilidad de que esta persona sea hombre? ¿mujer?

34.- Suponga que una de cada mil personas contrae la enfermedad E. Suponga además que la Medicina ha desarrollado un análisis de laboratorio que tiene las siguientes propiedades:

a) Si una persona tiene la enfermedad E, con probabilidad 0.99 el resultado del análisis es positivo y

b) Si una persona no tiene la enfermedad E, con probabilidad 0.99 el resultado del análisis es negativo.

Un análisis con esas características puede promocionarse como 99% seguro. Si un médico le informa a Ud. que su análisis dio positivo ¿debe preocuparse?

Sugerencia Calcule la probabilidad condicional de tener la enfermedad E dado que el resultado del análisis fue positivo; es más, calcule la probabilidad condicional de no tener la enfermedad dado que el análisis dio positivo.

35.- **El Conjunto Ternario de Cantor** es un subconjunto del intervalo [0,1] que nos hace reflexionar sobre la frase *sea A subconjunto de* $\mathbb{R}$. Denotemos con C al *conjunto Ternario de Cantor*. C es construido de manera iterativa. En la primera iteración dividimos el intervalo [0,1] en tres subintervalos; dos cerrados en los extremos y el medio abierto, esto es,

$$[0,1] = [0,1/3] \cup (1/3\,,2/3) \cup [2/3\,,1].$$

C_1es obtenido al eliminar el intervalo abierto medio, por lo que, $C_1 = [0,1/3] \cup [2/3\,,1]$, que involucra dos intervalos.

En la segunda iteración dividimos cada intervalo de C_1 en tres subintervalos de la misma manera que dividimos [0,1]. Así,

$$[0,1/3] = [0,1/9] \cup (1/9\,,2/9) \cup [2/9\,,3/9]$$

$$[2/3\,,1] = [6/9\,,7/9] \cup (7/9\,,8/9) \cup [8/9\,,9/9]$$

Para obtener C_2 eliminamos los intervalos abiertos medios y obtenemos $4(=2^2)$ intervalos, esto es,

$$C_2 = [0,1/9] \cup [2/9\,,3/9] \cup [6/9\,,7/9] \cup [8/9\,,9/9].$$

En la tercera iteración dividimos cada intervalo de C_2 en tres subintervalos de la misma manera que dividimos [0,1]. Al eliminar los intervalos abiertos medios obtenemos que C_3 es la unión de $8 = 2^3$ intervalos,

$C_3 = [0,1/27] \cup [2/27\,,3/27] \cup [6/27\,,7/27] \cup [8/27\,,9/27] \cup [18/27\,,19/27] \cup$
$[20/27\,,21/27] \cup [24/27\,,25/27] \cup [26/27\,,27/27]$.

Notamos que los extremos de los intervalos son retenidos de una iteración a la siguiente. En la iteración k tendremos 2^{k+1} extremos de intervalos y C_k será la unión de 2^k intervalos cerrados. Definimos C como,

$$C = \bigcap_{k=1}^{\infty} C_k = \{0,1,1/27\,,2/27\,,3/27\,,6/27\,,7/27\,,\ldots\}.$$

C tiene como elementos los extremos de los intervalos abiertos eliminados más otros elementos. C es cerrado ya que C_k es cerrado para toda k.

35.1.- Muestre que la medida de Lebesgue (recuerde el ejemplo 1.13) de los intervalos eliminados vale uno. Concluya que la medida de Lebesgue de C es cero. Este ejercicio muestra que es posible tener $\mathrm{P}(A) = 0$ pero A no es vacío o finito.

Con el fin de clarificar la definición de C, mencionamos que si x es un número real en [0,1] entonces, x puede ser expresado (en expansión ternaria) como

$$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots + \frac{a_n}{3^n} + \cdots$$

donde $a_n = 0{,}1{,}2$ para cada $n \in \mathbb{N}$. Cada x en $[0.1]$ tiene una única expansión con excepción de aquellos que tienen una expansión finita (esas x's tienen dos expansiones). Es posible mostrar que si $x \in C$ entonces, la expansión ternaria de x no involucra unos (esto es, $a_n \neq 1$ para toda n) a menos que x tenga una expansión finita. Equivalentemente, si $x \in C$ tenemos que la expansión ternaria de x sólo involucra ceros y doses (esto es, $a_n = 0, 2$ para toda n) a menos que la expansión de x sea finita. Es posible mostrar (consulte Goffman (1970) capítulo 6) que C es infinito no numerable (entre otras propiedades). Le recomiendo además consultar el artículo de Galaviz Casas (1996).

35.2.- Muestre que $1/4$ y $1/40 \in C$.

Sugerencia Encuentre la expansión ternaria de $1/4$ $(1/40)$.

Si el estudiante todavía no se inquieta al escuchar *cualquier subconjunto* de $\mathbb{R}$, le recuerdo el conjunto de Vitali y le sugiero consultar el libro de Goffman (1970), capítulo 6, donde se plantea la generalización conjunto k −nario de Cantor (la discusión previa hizo $k = 3$ y se puede pensar en $k = 5$). En opinión del que escribe, el libro de Goffman es un excelente texto de Análisis que desafía constantemente al lector. Recomiendo también leer los dos últimos capítulos del libro de Dunham (1990) en donde encontrará algunos resultados descubiertos por Cantor; estoy seguro de que se impresionará con la originalidad mostrada por este matemático de fines del siglo 19.

36.- Al lanzar repetidamente un dado honesto y registrar los números observados obtenemos una colección de *números aleatorios*. Ahora, dada una sucesión de dígitos $\{x_n\}$ donde x_n es un entero menor o igual a 9 pero mayor o igual a 0 ¿cómo podemos saber si la colección fue producida por un *mecanismo aleatorio*? A.N. Kolmogorov sugirió que si *la Entropía* es próxima a uno tendremos evidencia de que la colección fue producida al azar. La Entropía, para la sucesión mencionada arriba, es definida como

$$H_{10}(p_0, p_1, \ldots, p_9) = -\sum_{j=0}^{9} p_j \log_{10}(p_j)$$

donde: p_j es la *frecuencia relativa* del número $j = n_j/n$, n_j es la frecuencia del número j=número de veces que fue observado el número j para $j = 0{,}1{,}2, \ldots, 9$ y $\log_{10}(p_j)$ es el logaritmo base diez de p_j. El subíndice de H hace referencia al número de valores distintos que pueden asumir los elementos de la sucesión. Hacemos la convención de que si $p_j = 0$ para alguna j entonces, el sumando valdrá cero.

36.1.- Si $\mathcal{C} = \{x_j \in \mathbb{N}_0; x_j \leq K - 1 \text{ para } j = 1, \ldots, n \text{ y } K \text{ un natural} \geq 2\}$ entonces,

$$H_K(p_0, p_1, \ldots, p_{K-1}) = -\sum_{j=0}^{K-1} p_j \log_K(p_j)$$

Muestre que $H_K(1/K, 1/K, \ldots, 1/K) = 1.0$. Observe que la base de los logaritmos varía con K. En el capítulo 4 mostraremos que $H_K(p_0, p_1, \ldots, p_{K-1}) \leq 1$.

36.2.- Muestre la siguiente igualdad

$$H_K(p_0, p_1, \dots, p_{K-1}) = -\sum_{j=0}^{K-1} p_j \ln(p_j)/\ln(K) \equiv H(p_0, p_1, \dots, p_{K-1})/\ln(K)$$

donde $\ln(x)$ es el logaritmo natural (base e) de $x > 0$. Note que la H sin subíndice calcula logaritmos naturales.

36.3.- Al utilizar el inciso 1, muestre que

$$H(p_0, p_1, \dots, p_{K-1}) \leq \ln(K).$$

36.4.- Para $K = 2$, calcule $H_2(0.5, 0.5)$, $H_2(1/3\,, 2/3)$, $H_2(0.1, 0.9)$ y $H_2(0.001, 0.999)$ ¿Cómo interpreta los valores?

La entropía es máxima cuando las probabilidades son iguales y es mínima cuando una de las probabilidades vale uno y las demás valen cero. Recomiendo el libro de Ekeland (1993) en donde se discute, entre otros temas, la generación de números aleatorios.

37.- Suponga que tiene una moneda que produce una cara con probabilidad 0.6 ¿Cómo se las arreglaría para hacer lanzamientos honestos?

Mosteller, diseñó el siguiente procedimiento. La moneda será lanzada no una sino dos veces. Si los resultados son iguales en los dos lanzamientos se cancela la apuesta. Si los resultados son distintos, gana la persona que acierta en el primer lanzamiento ¿Está Ud. de acuerdo en que el procedimiento le da la misma probabilidad de ganar a los dos apostadores?

38.- Considere el experimento aleatorio $\mathcal{E}$ seleccionar al azar un número natural menor o igual a 1000 ¿Cuál es la probabilidad de que el número seleccionado no sea divisible entre 5 o 7?

39.- ¿Cuántos naturales menores o iguales a 1260 son primos relativos de 1260? Este ejercicio utilizará su respuesta al ejercicio 1.17 y está tomado del libro de Yaglom y Yaglom (1964). Sugiero la consulta del libro mencionado si es que le gusta ser desafiado mentalmente.

40.- En un noticiario se anuncia que el Servicio Meteorológico ha dicho que la probabilidad de que llueva el próximo sábado es 50% y que el domingo tiene la misma probabilidad de lluvia. El conductor del noticiario comenta "de seguro lloverá durante el fin de semana" ¿Está Ud. de acuerdo?

41.- En la demostración del teorema 1.3 se utilizan los siguientes resultados:

a) Si $\{x_n\}$ es una sucesión de números reales tal que $a \leq x_n$ para toda n natural donde $a \in \mathbb{R}$ entonces, $a \leq \liminf x_n$

b) Si $\{y_n\}$ es una sucesión de números reales tal que $b \geq x_n$ para toda n natural entonces, $b \geq \limsup y_n$

Muestre que los dos resultados son verdaderos.

Sugerencia Consulte la sección de sucesiones y series del apéndice A.

42.- El tercer axioma sobre la medida de probabilidad, P3, está compuesto de dos partes: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ cuando $A \cap B = \emptyset$, para A y B eventos y la σ $-$aditividad. Si proponemos como axioma que $\mathrm{P}(\emptyset)=0$, P3 se podría enunciar como la sigma aditividad únicamente. Muestre que la primera parte del enunciado es deducible a partir de la sigma aditividad.

Sugerencia Considere una sucesión de eventos donde A_1 y A_2 son arbitrarios pero ajenos y $A_n = \emptyset$ para $n \geq 3$.

43.- Considere el experimento aleatorio $\mathcal{E}$: lanzar dos dados honestos y distinguibles.

43.1.- ¿Cuál es el valor más probable de la suma de los números observados?

43.2.- ¿Aceptaría la siguiente apuesta? Ud. gana un centenario cuando obtiene una suma de 7 y pierde un centenario si la suma no es siete ¿Qué apuesta aceptaría?

43.3.- ¿Cuál es la apuesta cuando compra un seguro de vida?

44.- Remítase al ejemplo 1.13. Sea $D_n = \{x_1, \dots, x_n\}$ con $n \in \mathbb{N}$ y $x_j \in \mathbb{R}\ \forall\, j$. Calcule $\lambda(D_n)$. En los espacios no discretos que estudiaremos tendremos que, si E es un evento entonces, $\mathrm{P}(E)$ será positiva únicamente cuando $\lambda(E) > 0$. Decimos que λ *domina* a P.

45.- Si planteamos como axiomas para P: P1, P2, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ cuando $A, B \in \mathcal{A}$ con $A \cap B = \emptyset$ y P es continua. Muestre que la $\sigma-$aditividad es consecuencia de los axiomas.

46.- Considere el espacio de probabilidad $(\Omega, \mathcal{A}, \mathrm{P})$. Suponga que $A, A_n \in \mathcal{A}$ para toda n natural. Muestre que

$$A \cup \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (A \cup A_n)$$

47.- Sea E el evento la raza humana está en la tierra ¿Cuáles son las posibles causas de nuestra presencia en la tierra? Consideremos las siguientes

C_1 Mutación espontánea y evolución

C_2 Un ser divino creó a la humanidad

C_3 Seres de otros planetas sembraron a los humanos en la tierra

C_4 Otra causa que no sea alguna de las de arriba

Las causas son mutuamente excluyentes y exhaustivas.

47.1.- Asigne una probabilidad a cada causa C_j, esto es, asigne a C_j un número $p_j \in (0,1)$ de tal manera que $p_1 + p_2 + p_3 + p_4 = 1$; note que p_j no puede valer cero o uno.

47.2.- Asigne un número entre cero y uno a $\mathrm{P}(E|C_j)$; estos números no tienen que sumar uno.

Nota Las asignaciones reflejan sus convicciones, conocimientos, prejuicios, ignorancia, etc. Sin embargo, al menos tome en cuenta la dificultad de llegar a la tierra desde otro planeta (Isaac Asimov), la evidencia de la evolución en la tierra (Charles Darwin, Richard Dawkings), sus creencias religiosas (la Biblia, el Corán, etc.), el número de planetas habitados que conoce la humanidad , la edad del universo y de la tierra (Carl Sagan, Yuval Harari, etc.), etc. Recuerde que Laplace nos sugiere *gran cuidado* al hacer las asignaciones.

47.3.- Utilice el lema de Bayes para calcular $\mathrm{P}(C_j|E)$ para $j = 1,2,3,4$. Esas probabilidades sí suman uno ¿Qué opina de sus cálculos?

48.- En diversas partes del texto utilizo en concepto de *distancia* el cual está asociado al de *espacio métrico*. Un espacio métrico es un par $(\mathcal{M}, d)$ que consiste en un conjunto $\mathcal{M}$ y una distancia; una función que asume valores reales no negativos, $d(x, y)$ definida para todo par $x, y \in \mathcal{M}$ y que satisface las siguientes propiedades

a) $d(x, y) = 0$ si y sólo si $x = y$;

b) Simetría: $d(x, y) = d(y, x)$;

c) Desigualdad del triángulo: $d(x, z) \leq d(x, y) + d(y, z)$ para toda $z \in \mathcal{M}$.

48.1.- Si $\mathcal{M} = \mathbb{R}$ y $d(x, y) = |x - y|$ para todo par $x, y \in \mathbb{R}$. Muestre que d es una distancia.

48.2.- Si $\mathcal{M} = \mathbb{R}^2$ y $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ para todo par $\boldsymbol{x} = (x_1, x_2)$, $\boldsymbol{y} = (y_1, y_2)$ en $\mathbb{R}^2$, muestre que d es una distancia. La desigualdad de Cauchy - Schwarz le será de utilidad

$$(x_1 y_1 + x_2 y_2)^2 \leq (x_1^2 + x_2^2)^2 (y_1^2 + y_2^2)^2$$

d así definida es la distancia euclidiana en $\mathbb{R}^2$. Esta distancia es generalizable a $\mathbb{R}^n$ para $n \geq 3$; consulte el apéndice B.