LA CONSTRUCCIÓN DE LA DEFINICIÓN AXIOMÁTICA DE PROBABILIDAD

Mario Antonio Gneri*

26 de agosto de 2012

Depto de Estatística, IMECC-UNICAMP, C. Universitária Zeferino Vaz, Barão Geraldo, 13083-859, Campinas, SP, Brasil

Resumen

Mostramos aquí una forma de abordar la definición axiomática de probabilidad. Haremos una breve reseña histórica, comenzando por la versión original (restricta al ámbito de la teoria de los juegos) y culminando en la definición axiomática (ya en en el marco de la teoría de la medida), mostrando la evolución del concepto. También exhibiremos algunas demostraciones simples a partir de la última definición para ilustrar el funcionamiento el método axiomático.

Palabras llave: teoría axiomática, función de conjunto, σ-álgebra, σ-aditividad, probabilidad.

1. Introducción

Todo estudiante de matemática o de ciencias se topó alguna vez con la definición axiomática de probabilidad. Esta es una definición que cabe en 3 líneas, pero que llevó más de 3 siglos de trabajo para ser escrita. El objetivo principal de este trabajo es mostrar como se llegó a ella, como ejemplo de la forma en que se van lapidando los conceptos en matemática. Otro objetivo es mostrar el uso y la importancia del lenguaje matemático.

*Corresponding author. E-mail adress: gneri@ime.unicamp.br

El trabajo está organizado en la forma descripta a seguir.
En la Sección 2 se presenta una reseña histórica de la evolución del concepto de probabilidad, siguiendo el rastro de diversas definiciones de este concepto.
En la Sección 3 hacemos una breve y personal reflexión sobre el lenguaje matemático y sobre las dificultades en el aprendizaje de la matemática.
En la Sección 4 presentamos formalmente las definiciones clásica y frecuentista y una versión preliminar de la definición axiomática. Las dos primeras son del siglo XVII y la tercera del siglo XX. Todas ellas son expresadas en el lenguaje matemático actual (que es sólo contemporáneo de la definición axiomática), para su mejor compresión. También demostramos unas pocas propiedades a partir de la última definición, con los objetivos de mostrar el funcionamiento del método axiomático y el uso del lenguaje. Las demostraciones en todo el texto serán muy pocas, siempre con los objetivos que acabamos de explicitar.
La Sección 5 es reservada a la versión final de la definición axiomática.
En el Apéndice (Sección 6) enunciamos el Teorema de Bernoulli y definimos los conjuntos borelianos.

2. Breve historia

En los comienzos de la era moderna la nobleza dedicaba buena parte de su tiempo ocioso a los juegos de azar. El objetivo inicial de la teoría de las probabilidades fue la descripción y estudio de estos juegos para producir estrategias útiles a los jugadores.

Algunos problemas particulares sobre juegos habían sido resueltos por matemáticos italianos en los siglos XV y XVI, pero la teoría propiamente dicha se inició con el trabajo de los matemáticos franceses Pierre de Fermat (1601?-1664) y Blaise Pascal (1623-1662) en la mitad del siglo XVII. Pascal vivía en París y Fermat en Toulouse y se comunicaban por correspondencia. Estos documentos fueron conservados en parte. En ellos aparece, aunque no de forma explícita, la después denominada definición clásica de probabilidad. En realidad, el primer tratado formal sobre probabilidades, intitulado *De Ratiotiniis in Ludo Alea* fue escrito en 1657 por el físico, geómetra y astrónomo holandés Christiaan Huygens (1629-1695), utilizando como base la correspondencia entre Fermat y Pascal (casi toda del año 1654).

Comenzaremos presentando la definición clásica, cuya formalización veremos posteriormente. Consideremos, a manera de ejemplo, el juego más simple: dos personas apuestan a acertar qué faz de una moneda será visible al detenerse en el piso la misma después de arrojada hacia arriba. Es imposible prever el resultado de cada lance. Por más que el lanzamiento sea siempre repetido en las mismas condiciones, algunas veces el resultado será *cara* y en otros *cruz*. Esta imposibilidad de predicción es lo que se define como aleatoriedad. Lo mismo ocurriría si jugásemos a acertar el resultado del lanzamiento de un dado. Pero, a pesar de la imposibilidad de predecir el próximo resultado, en ambos juegos podemos exhibir el conjunto de todos los resultados posibles: en el caso de la moneda seria el conjunto $\{cara, cruz\}$ y en el caso del dado el conjunto $\{1, 2, 3, 4, 5, 6\}$. Saber todo lo que puede ocurrir es un primer paso para lidiar con la incertidumbre.

Volvamos al caso de la moneda. El problema ahora es ponderar o dar un peso (o probabilidad) a cada uno de los casos posibles o a cualquier subconjunto de los mismos. Así, por ejemplo, se pensó que si un jugador apostase en *cara* tendría un resultado a su favor entre los dos posibles (*cara*, *cruz*). Así se definió $\mathit{probabilidad}(\{cara\}) = 1/2$ y por simetría también $\mathit{probabilidad}(\{cruz\}) = 1/2$. La simetria se tradujo en la equiprobabilidad de los dos casos posibles.

Análogamente, en el juego del dado, tendremos que $\mathit{probabilidad}(\{j\}) = 1/6$ para cualquier natural j tal que $1 \leq j \leq 6$. Consideremos o caso de apostar en un conjunto no unitario, por ejemplo $\{2, 3\}$. En este caso el jugador tiene a favor 2 de los 6 casos posibles y así resulta natural afirmar que $\mathit{probabilidad}(\{2, 3\}) = 2/6$. También, atribuyendo a la probabilidad la denominada propiedad aditiva, podemos decir que
$\mathit{probabilidad}(\{2, 3\}) = \mathit{probabilidad}(\{2\}) + \mathit{probabilidad}(\{3\}) = 1/6 + 1/6$, llegando al mismo resultado.

Generalizando los argumentos de los dos parágrafos anteriores se llega a la definición clásica: si el jugador apuesta en un subconjunto cualquiera del conjunto de los casos posibles (llamado el subconjunto de los casos favorables). La probabilidad del jugador ganar el juego se define como el cociente entre el número de casos favorables al jugador (numerador) y el número de casos posibles (denominador).

Casi simultánemente a la aparición de la definición clásica se desarrolló una forma experimental de calcular probabilidades (que daría origen al posterioremente llamado enfoque frecuentista) a través de lo que hoy denominaríamos simulación. El procedimiento era el siguiente: se repetía el juego un número muy grande de veces y se consideraba el cociente entre el número de veces que el jugador habría ganado (numerador) y el número total de tentativas (denominador). Existía la intuición de que este número, denominado frecuencia relativa del jugador ganar, sería una buena aproximación de la probabilidad.

Por ejemplo: considere la siguiente secuencia de 20 resultados obtenidos al arrojar un dado: 2,4,5,2,3,6,6,1,4,3,5,2,2,1,6,2,3,5,1,2. Las estimativas de las probabilidades (o sea las frecuencias relativas) de 1, 2, 3, 4, 5 y 6 serían 3/30, 6/20, 3/20, 2/20, 3/20 y 3/20, respectivamente. Todas estas probabilidades, calculadas por la definición clásica son iguales a 1/6. Se suponía que si el número de repeticiones fuese aumentado considerablemente, todas las frecuencias relativas se aproximarían a 1/6. Este tipo de conjetura fue testado empíricamente para algunos juegos simples.

En Gnedenko (1978) encontramos la siguiente tabla, relativa a lances con una moneda.

Experimentador	Número de lances	Número de caras	Frecuencia relativa
Buffon	4040	2048	0,5080
Karl Pearson	12000	6019	0,5016
Karl Pearson	24000	12.012	0,5005

El siguiente ejemplo fue extraído de Cramer (1964). Se realizaron 5 series de 10 lances de una moneda y se anotaron la mayor y menor frecuencia relativa obtenidas para las caras. Luego se repitió el experimento con 5 series de 100 lances y finalmente con 5 series de 1000 lances. Los resultados están resumidos en la siguiente tabla.

Número de lances en cada serie	Máxima	Mínima	Diferencia
10	0,600	0,300	0,300
100	0,550	0,480	0,070
1000	0,507	0,496	0,011

Los resultados exhibidos en las dos tablas precedentes muestran que a medida que aumenta el número de lances las frecuencias relativas quedan cada vez más próximas del valor 1/2, que es la probabilidad de *cara* según la definición clásica. A partir de ejemplos como estos se conjeturaba que el limite de la frecuencia relativa cuando el número de lances tiende a infinito sería 1/2. Más tarde Bernoulli probaría que, bajo ciertas condiciones, esta intuición era correcta.

La necesidad de utilizar la frecuencia relativa como estimativa de la probabilidad se debió a la existencia de juegos mucho más complicados que los ejemplos de la moneda y el dado aquí citados. En estos juegos más complejos puede ser muy difícil contar los casos favorables y los posibles. Debe tenerse en cuenta que en la segunda mitad del siglo XVII no estaban suficientemente desarrolladas las técnicas de conteo. Por ello muchos casos eran olvidados o ignorados y al no ser contados provocaban muchos errores. A través del uso de la frecuencia relativa era posible obviar la enumeración o el conteo de los casos favorables y de los posibles.

Pensemos ahora en un problema donde no se pueda partir de la hipótesis de equiprobabilidad de los casos posibles. En Punta Loma, en la Península de Valdés, provincia de Chubut, un grupo de lobos marinos tiene su habitat. Estudios afirman que nace un macho a cada 5 nacimientos. Obviamente estos estudios fueron hechos via frecuencia relativa y las propias frecuencias relativas se encargaron de mostrar que la definición clásica no funcionaba. O sea, el enfoque frecuentista permite estimar probabilidades en casos en que no hay equilibrio entre los casos posibles y en los que por lo tanto la definición clásica no se puede aplicar en forma directa.

También fue notado que las frecuencias relativas permitían abordar problemas como el siguiente, en los que hay que considerar las llamadas variables continuas: cual es la probabilidad de que el peso de un bebé al nacer esté entre 2,5 y 3,2 kilogramos? Por otro lado, es claro que en este caso la definición clásica llevaria a una indeterminación del tipo $\frac{\infty}{\infty}$. Vemos así que para poder aplicar la definición clásica es necesario que el conjunto de todos los resultados posibles sea finito. Esta cuestión estaría implícita en el trabajo de Fermat y Pascal.

En el siglo XVIII los nombres más importantes fueron Jakob Bernoulli (1654-1705) y Abraham de Moivre (1667-1754).

Bernoulli, en su obra *Ars Conjectandi* (publicada en 1713) provó que, bajo determinadas condiciones, las frecuencias relativas de ocurrencia de un evento convergen (en un sentido que especificaremos después en el Apéndice, vea 6.1) a la probabilidad del mismo. Este resultado, que fue llamado posteriormente Teorema de Bernoulli y es la primera versión de la llamada Ley Débil de los Grandes Números, dio consistencia al enfoque frecuentista de la probabilidad.

De Moivre (hugonote francés radicado en Inglaterra por causa de persecuciones religiosas) introdujo la que más tarde sería llamada distribución normal o gaussiana, al aproximar mediante la fórmula de Stirling los factoriales que aparecen en los coeficientes de la distribución binomial, desarrollando de esta forma la primera y particularísima versión del Teorema Central del Límite, uno de los resultados más importantes de la teoria de las probabilidades. Estas contribuciones y muchas otras constan en su obra *The Doctrine of Chances*, que se editó tres veces (1718, 1738 y 1756). Esta versión primitiva del Teorema Central del Limite sería posteriormente completada y explicitada por Pierre de Laplace (1749-1827).

El Teorema de Bernoulli será comentado en el Apéndice (Sección 6). Ya los resultados de De Moivre están fuera del alcance de este trabajo. Una excelente introducción al estudio del Teorema Central del Limite se encuentra en Feller (1968). También el lector interesado puede consultar Todhunter (1965).

También en el siglo XVIII se verificó que las reglas del cálculo de probabilidades podían aplicarse fuera del ámbito de los juegos de azar, como por ejemplo em demografía, cálculo actuarial y mecánica estadística. En 1812 Laplace, en su libro *Théorie Analityque des Probabilités*, además de presentar una exposición sistemática y completa de la teoría matemática de los juegos de azar, documentó las nuevas aplicaciones e introdujo nuevas ideas y técnicas matemáticas. Según Apostol (1969): *Antes de Laplace las probabilidades estaban asociadas exclusivamente al análisis de los juegos de azar. Laplace aplicó probabilidades en la solución de diversos problemas científicos y prácticos.*

Laplace afirmaba que era posible aplicar siempre la definición clásica, particionando el conjunto de todos los casos posibles en conjuntos equiprobables. La construcción de estos conjuntos nunca quedó muy clara.

Cramer (1964) afirma: *La obra de Laplace ejerció una profunda influencia sobre el desarrollo subsiguiente de esta materia. En vista del imponente aparato matemático y de los importantes resultados prácticos ya obtenidos o fácilmente susceptibles de alcanzar, era tentador pasar por alto la debilidad fatal de los fundamentos conceptuales. Como resultado, el campo de la aplicación de la teoría de la probabilidad se expandió rápidamente y sin interrupción durante el siglo XIX, mientras la teoría matemática en sí mostró durante el mismo tiempo una señalada tendencia al estancamiento.*

Mencionaremos brevemente algunos de los tipos más importantes de aplicaciones que se originaron en este período. Gauss (1777-1855) y Laplace discutieron independientemente uno del otro las aplicaciones de la teoría de la probabilidad matemática al análisis numérico de los errores de medición en las observaciones físicas y astronómicas. Esta teoría y el método de los cuadrados mínimos, íntimamente relacionados con ella, han adquirido gran importancia teórica y práctica.

El enorme desarrollo de los seguros de vida desde comienzos del siglo XIX ha sido posible gracias a un correspondiente desarrollo de la matemática actuarial, que a su vez está basada en la aplicación de la probabilidad a las estadísticas de mortalidad. Quetelet (1796-1874) y su escuela hicieron nuevas aplicaciones a la demografía y otras ramas de las ciencias sociales.

En la física matemática, la teoría de la probabilidad fue introducida por la obra de Maxwell (1831-1879), Boltzmann (1844-1906) y Gibbs (1839-1903) sobre mecánica estadística, que ha sido de fundamental importancia para determinadas partes de la ciencia física moderna.

Durante el siglo XX este desarrollo ha continuado a ritmo acelerado. Los métodos de la estadística matemática se utilizan en un número creciente de actividad científica y práctica. La teoría matemática en que se basan estos métodos se apoya esencialmente sobre el fundamento de la probabilidad matemática. En el momento actual, las aplicaciones de la teoría de la probabilidad abarcan campos tan diferentes como la genética, la economía, la psicología y la ingeniería.

El trabajo de Gauss en la teoría de los errores, citado en el texto de

Cramer que acabamos de reproducir, merece un breve comentario. Se trataba del modelo de medición (muy usado entre los astrónomos de la época), en el que se suponía que el resultado de una medición (X) podia ser expresado de la forma: $X = M + e$, donde la constante M era la "verdadera medida" del objeto en estudio y e un error aleatorio. Gauss propuso representar la distribución del error e por el que luego se llamaría modelo normal o gaussiano. Interesa resaltar aquí que en este caso las probabilidades se calculan a partir de lo que hoy se llama una función de densidad. Se trata de una función $f : R \to R$, no negativa y tal que $\int_{(-\infty,\infty)} f(t)dt = 1$. La probabilidad de un intervalo I de R es igual a $\int_I f(t)dt$.

Dado este panorama, se hacía necesaria una definición coherente de probabilidad. Esta definición debía ser muy clara y precisa, para que las conclusiones basadas en ella fuesen irreprochables y por otro lado lo suficientemente amplia como para abarcar toda la variedad de problemas y las aplicaciones que acabamos de mencionar. Richard von Mises, figura principal del frecuentismo juntamente con R.A. Fisher, desarrolló el concepto de espacio de resultados, que sirvió de soporte a la construcción de una teoría de las probabilidades basada en la teoría de la medida. Hubo bastantes esfuerzos en esta dirección en la década de 1920 que culminaron con la definición axiomática de probabilidad de Kolmogorov en 1933 (Ver Kolmogorov (1956)). En las secciones 4 y 5 nos ocuparemos con cierto detalle de esta definición, cuyo estudio es el principal objetivo de este trabajo.

3. El lenguaje matemático y las dificultades en el aprendizaje de la matemática

La matemática fue construyendo su lenguaje desde el comienzo de la historia escrita hasta nuestros días. Su lenguaje actual es sintético y muy preciso. Algunos lo encuentran hermético y atribuyen a este hecho la imposibilidad de gran parte de la población mundial acceder al conocimiento matemático. En realidad ocurre lo contrario, el lenguaje absolutamente formal y preciso a ultranza de la matemática contribuye a la socialización de su conocimiento, permitiendo que no apenas los genios puedan entenderla o producirla.

Pensemos en la fórmula del cuadrado de un binomio de números reales, que algunos aprendimos a recitar de la siguiente forma: “el cuadrado de un binomio es igual al cuadrado del primer término más e cuadrado del segundo más el doble producto del primero por el segundo“. En la actualidad tal discurso se tornó obsoleto, porque puede ser substituído por la simple expresión $(a+b)^2 = a^2 + b^2 + 2ab$. Incluso no sería necesario recordar esta fórmula, ya que es fácil demostrarla conociendo apenas las definición de cuadrado y las propiedades conmutativas de la suma y el producto y la propiedad distributiva: $(a+b)^2 = (a+b)(a+b) = a(a+b) + b(a+b) = aa + ab + ba + bb = a^2 + ab + ab + b^2 = a^2 + b^2 + 2ab$.

Muchos atribuyen los fracasos escolares a la ”dificultad intrínseca” de la matemática. Nadie atribuye a la historia, por ejemplo, tal nivel de dificultad. Aprendimos a saber de memoria por qué calles del antiguo Buenos Aires intentaron entrar las tropas británicas en las invasiones inglesas de 1806 y 1807. Nunca se nos explicó porqué, a pesar de la derrota inglesa en ambas oportunidades, pocos años después Inglaterra conseguía dominar la economía argentina sin necesidad de enviar siquiera un soldado a los mares del sur (Peña (1972) cita que el British Comercial Room o Sala de Comercio Británica se estableció en Buenos Aires en 1811).

O sea, se enseña a repetir algunos hechos, generalmente batallas con sus fechas bien precisas, sin la menor posibilidad de entender los procesos históricos en su plenitud. Se puede obtener nota 10 en una prueba sobre la edad media ignorando por completo la existencia del imperio romano. Ya en matemática es más difícil engañar: es imposible saber sumar sin saber contar. Una buena pregunta: que es más difícil de entender, la revolución francesa o las cuatro operaciones aritméticas? Que la revolución francesa no fue bien entendida lo demuestran las castas monárquicas que todavía subsisten ociosamente en la culta Europa, sustentadas con recursos del erario público.

Es imposible probar que la matemática es más difícil que la história o la biología, por ejemplo. No afirmamos aquí que la matemática sea fácil. Su construcción ha llevado milenios. Hoy debemos absorber en pocos años parte de esa inmensa construcción colectiva, por lo tanto es necesario modestia y trabajo para conocerla. En todo ese tiempo la matemática fue construyendo su lenguaje claro y preciso. Este lenguaje nos da una enorme ventaja sobre nuestros antepasados que debieron enfrentar los problemas sin poder expresarse con la precisión adecuada, corriendo riesgo de confundirse no sólo en las demostraciones sino también en la fase previa de formular conjeturas. Es posible entender matemática desde que sea estudiada con cariño y apropiándose de su lenguaje.

En lo siguiente nos dedicaremos a traducir al lenguaje matemático actual las ideas y definiciones precedentes, producidas en un período de más de tres siglos, tomando como punto de partida a Fermat y Pascal.

4. Formalización de las definiciones

4.1. Notación, definiciones y algunas observaciones

- Un *fenómeno aleatorio* es un fenómeno que se caracateriza por la siguiente propiedad: al observarlo bajo cierto conjunto de condiciones supuestamente invariantes no siempre se obtiene el mismo resultado. Cuando el interesado tiene en sus manos la posibilidad de repetir el fenómeno (como ocurre en el lanzamiento de una moneda) diremos que estamos en presencia de un *experimento aleatorio.* La teoría de las probabilidades puede ser considerada como el estudio de los modelos matemáticos de fenómenos aleatorios.

- El *espacio de resultados* de un experimento aleatorio es el *conjunto* de *todos* sus resultados posibles. Será denotado por la letra griega Ω. Esta definición permite asociar el espacio de resultados al concepto de universo de la teoría de conjuntos. Siempre será Ω no vacío. Consideraremos Ω finito o enumerable hasta aviso en contrario.

- Un *evento* es, por definición, cualquier subconjunto de Ω.

- Si $A \subset \Omega$, $\quad n^o(A)$ denota o número de elementos de A y se llama función de *conteo.* Decimos que $n^o(A)$, por estar definida en los subconjuntos de Ω es una *función de conjunto* (en inglés *set function*). La función de conjunto n^o satisface las siguientes propiedades:
  i) para todo $A \subset \Omega$ vale que $n^o(\emptyset) = 0 \leq n^o(A) \leq n^o(\Omega)$;
  ii) si $A \subset \Omega$, $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $n^o(A \bigcup B) = n^o(A) + n^o(B)$.

4.2. La definición clásica

4.2.1. Definición

Sea Ω um conjunto finito no vacío. Para todo subconjunto A de Ω se define

$$Probabilidad(A) = \frac{n^o(A)}{n^o(\Omega)}.$$

La función de conjunto *Probabilidad*, que será denotada por P, satisface las siguientes propiedades:
I) para todo $A \subset \Omega$ está definida $P(A)$ y vale que $0 \leq P(A) \leq 1$;
II) $P(\Omega) = 1$;
III) si $A \subset \Omega$, $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $P(A \bigcup B) = P(A) + P(B)$.

(La tercera propiedad se llama propiedad aditiva de la probabilidad).

4.2.2. Ejercicio

Demuestre las propiedades de la función P a partir de su definición y de las propiedades de la función n^o enunciadas anteriormente. Recuerde siempre que Ω es finito y no vacío (por qué esto es necesario para el funcionamiento de la definición clásica?).

4.2.3. Observaciones

La definición clásica parte de dos supuestos siguientes (que obviamente limitan su aplicabilidad):
i) El espacio de resultados Ω debe ser finito;
ii) Todos los elementos de Ω deben tener probabilidad igual a $(n^o(\Omega))^{-1}$.

4.3. La frecuencia relativa

4.3.1. Frecuencia absoluta y relativa

Supongamos que repetimos n veces (n natural), siempre en las mismas condiciones, un experimento aleatorio cuyo espacio de resultados sea el conjunto Ω. Sea A um subconjunto de Ω. Denominamos $frecuencia\ absoluta\ de\ A$ o simplemente $frecuencia\ de\ A$ (notación $f(A, n)$) al número de ocurrencias del evento A en las n repeticiones del experimento.

La función de conjunto f tiene las siguientes propiedades:
I) para todo $A \subset \Omega$ vale que $0 \leq f(A,n)) \leq n$;
II) $f(\Omega,n) = n$;
III) si $A \subset \Omega$, $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $f(A \bigcup B, n) = f(A,n) + f(B,n)$ (propiedad aditiva de la frecuencia absoluta).

También podemos definir *frecuencia relativa de A* (notación $fr(A,n)$) como el cociente $\frac{f(A,n)}{n}$. La función de conjunto fr tiene las siguientes propiedades (que pueden ser demostradas a partir de las propiedades de f):
I) para todo $A \subset \Omega$ vale que $0 \leq fr(A,n) \leq 1$;
II) $fr(\Omega,n) = 1$;
III) si $A \subset \Omega$, $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $fr(A \bigcup B, n) = fr(A,n) + fr(B,n)$ (propiedad aditiva de fr).

4.3.2. Observaciones

Poco después de la aparición de la definición clásica la frecuencia relativa comenzó a ser usada como una estimativa de la probabilidad. Este enfoque permitió superar problemas generados por la definición clásica, pero a su vez introdujo otros.

Entre las ventajas podemos observar que el cálculo de frecuencias relativas independe de la finitud de Ω, que no requiere la equiprobabilidad de los elementos de Ω y que exime al experimentador de la enumeración de los casos favorables y posibles.

Por otro lado, aparecem nuevas dificultades.
a) Se torna necesaria la estabilidad de las frecuencias relativas del conjunto de interés A, condición necesaria para la convergencia a un valor que sería la probabilidad de A. Esta propiedad es denominada *regularidad estadística*;
b) Mismo admitiendo la existencia del límite mencionado en a), cuando habría que parar? En el Apéndice (vea 6.1) se da una respuesta a esta pregunta.

4.4. Versión preliminar de la definición axiomática

Toda la matemática pasó por un proceso de axiomatización a partir de la segunda mitad del siglo XIX y la definición de Kolmogorov es parte de ese proceso. El afirmó que la teoría de las probabilidades podría ser desarrollada a partir de axiomas, de la misma forma que el álgebra y la geometría. En estos axiomas se establecen los entes matemáticos que serán estudiados y sus relaciones. Toda la teoría se construye a partir de los axiomas, independientemente de cualquier interpretación de los mismos o de sus consecuencias. Como veremos en seguida, en la definición de Kolmogorov aparecen como axiomas propiedades comunes a las nociones clásica y frecuentista, que de esta forma se tornaron caso particular de la definición axiomática. En realidad Kolmogorov tomó esta definición de la teoria de la medida.

Dada la complejidad de esta definición, la misma será alcanzada en tres etapas. Las dos primeras serán llamadas provisorias y la definición propiamente dicha será llamada definitiva.

4.4.1. Primera Definición Provisoria (restricta a espacios de resultados finitos o enumerables)

Sea Ω un conjunto no vacío finito o enumerable. Una probabilidad en Ω es una función de conjunto P que asocia a cada subconjunto A de Ω un número real $P(A)$, de tal forma que:
I) para todo $A \subset \Omega$ vale que $0 \leq P(A) \leq 1$;
II) $P(\Omega) = 1$;
III) si $A \subset \Omega$ y $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $P(A \bigcup B) = P(A) + P(B)$.

4.4.2. Ejemplos

Para ilustrar la definición daremos los siguientes ejemplos:
i) Considere o experimento consistente en arrojar una moneda hacia arriba y observar la faz superior. El espacio de resultados $\Omega = \{cara, cruz\}$. Vamos a definir la probabilidad P de la siguiente forma:
$P(\{cara\}) = p$ y $P(\{cruz\}) = 1 - p$, donde $0 \leq p \leq 1$;
Observe que cada valor de p define una función de probabilidad diferente; el caso $p = 1/2$ corresponde a una moneda equilibrada;
ii) Ahora el experimento es arrojar un dado hacia arriba y observar la faz superior. En este caso $\Omega = \{1, 2, 3, 4, 5, 6\}$. Sean $p_1, p_2, p_3, p_4, p_5, p_6$ números

reales no negativos cuya suma vale 1. Definimos $P(\{j\}) = p_j$ para $1 \leq j \leq 6$. Se puede verificar que P define una probabilidad en Ω. La extensión de P a la familia de todos los subconjuntos de Ω se hace aplicando la propiedad aditiva, por ejemplo, $P(\{1,5\}) = p_1 + p_5$;
iii) Este ejemplo generaliza los dos anteriores, ya que comprende todas las funciones de probabilidad que pueden ser definidas en un conjunto finito. Sean $\Omega = \{w_1, w_2, ..., w_n\}$ un conjunto de n elementos (n natural) y $p_1, p_2, ...p_n$ números reales no negativos cuya suma es 1. Definimos la probabilidad P sobre los conjuntos unitarios mediante la fórmula $P(\{j\}) = p_j$ para $0 \leq j \leq n$. La extensión de la definición de P a cualquier A de la familia de todos los subconjuntos de Ω se hace via aditividad, o sea $P(A) = \sum_{\{j/w_j \in A\}} p_j$;
iv) En forma análoga al caso anterior podemos considerar un espacio de resultados enumerable. Sean $\Omega = \{w_1, w_2, ..., w_n, ...\}$ y $p_1, p_2, ...p_n....$ una sucesión de números reales no negativos cuya suma es 1 ($\sum_{j=1}^{\infty} p_j = 1$). Definimos la probabilidad P sobre los conjuntos unitarios mediante la fórmula $P(\{w_j\}) = p_j$ para todo j natural. La extensión de la definición de P a cualquier A de la familia de todos los subconjuntos de Ω se hace en forma análoga al caso anterior, o sea $P(A) = \sum_{\{j/w_j \in A\}} p_j$. Cabe observar que los índices que aparecen en esta suma corresponden a cualquier subconjunto de los naturales (un subconjunto de los naturales para cada subconjunto A de Ω). No hay problemas en cuanto a la sumabilidad, porque estaríamos sumando algunos términos de una serie absolutamente convergente.
Un caso particular importante, llamado distribución geométrica, ocurre cuando Ω es el conjunto de los naturales y P está definida por $P(\{j\}) = p(1-p)^{j-1}$ para todo j natural, donde p es un número real tal que $0 \leq p \leq 1$.
Otro caso particular muy utilizado es el llamado modelo de Poisson de parámetro λ, donde $\lambda > 0$. Aqui $\Omega = Z^+$, el conjunto de los enteros no negativos. P se define por $P(\{j\}) = e^{-\lambda} \frac{\lambda^j}{j!}$ para todo $j \in Z^+$.

4.5. Breve contacto con la teoría axiomática de probabilidades

4.5.1. Observación

Los resultados condensados en el teorema siguiente tienen por objetivo mostrar como se construye la teoría a partir de los axiomas. Una vez demostrada cualquier propiedad a partir de los axiomas, ésta puede ser utilizada como base para otras demostraciones. Denominamos axiomas a las propiedades I, II y III de la Primera Definición Provisoria (4.4.1). Las propiedades que serán demostradas en el teorema a seguir serán numeradas a partir de IV. Verifique la importancia del axioma III (propiedad aditiva) en las demostraciones.

4.5.2. Teorema

Sea Ω un conjunto no vacío finito o enumerable y P una probabilidad en Ω. Son verdaderas las siguientes afirmaciones.
IV) $P(\emptyset) = 0$;
V) $P(A^c) = 1 - P(A)$, donde A^c denota el complemento de A ($A^c = \omega \in \Omega / \omega \notin A$);
VI) si $A \subset B$ entonces $P(A) \leq P(B)$ y $P(B - A) = P(B) - P(A)$, donde $B - A = B \bigcap A^c$;
VII) $P(B - A) = P(B) - P(B \bigcap A)$;
VIII) $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$.

Demostración
IV) $\emptyset = \emptyset \bigcup \emptyset$ y también $\emptyset \bigcap \emptyset = \emptyset$, entonces podemos aplicar el axioma III: $P(\emptyset) = P(\emptyset \bigcup \emptyset) = P(\emptyset) + P(\emptyset)$; esto es, vimos que $P(\emptyset) = P(\emptyset) + P(\emptyset)$, por lo tanto $P(\emptyset) = 0$;
V) el axioma II afirma que $P(\Omega) = 1$; por otro lado $\Omega = A \bigcup A^c$ y $A \bigcap A^c = \emptyset$, luego se puede aplicar también el axioma III: $1 = P(\Omega) = P(A \bigcup A^c) = P(A) + P(A^c)$; provamos entonces que $1 = P(A) + P(A^c)$ y de aquí sale V;
VI) siempre vale que $A \bigcap (B - A) = \emptyset$, además tenemos que si $A \subset B$ entonces $B = A \bigcup (B - A)$; una vez más aplicamos el axioma III obteniendo: $P(B) = P(A \bigcup (B - A)) = P(A) + P(B - A)$, o sea se probó que $P(B) = P(A) + P(B - A)$ y de aquí resulta que $P(B - A) = P(B) - P(A)$ y, dada la positividad de P, que $P(B) \geq P(A)$;
VII) basta observar que $B \bigcap A \subset B$ y la igualdad de conjuntos $B - A = B - B \bigcap A$ y aplicar la propiedad VI que acabamos de probar;
VIII) siempre vale que $A \bigcup B = A \bigcup (B - A)$ y que $A \bigcap (B - A) = \emptyset$, entonces la propiedad VIII es consecuencia del axioma III y de la propiedad VII recién demostrada.

4.5.3. Ejercicio

Estudie atentamente las demostraciones dadas en el teorema anterior, completando los pasos omisos e identificando en cada momento los axiomas o propiedades anteriormente demostradas que fueron utilizados. Verifique que la propiedad VII es una generalización de la VI y que la VIII generaliza el axioma III (o equivalentemente, VI es caso particular de VII y III es caso particular de VIII).

4.5.4. Observación

Es usual evitar la redundancia en el conjunto de los axiomas, o sea, no debe haber un axioma o parte del mismo que pueda ser deducido a partir de otros. Nosotros violamos aquí esta norma, ya que no sería necesario colocar en el Axioma I que para todo $A \subset \Omega$ sea $P(A) \leq 1$. Observe que esta afirmación nunca fue utilizada y que podria ser deducida de las propiedades II y VI. La decisión de incluir la afirmación en el Axioma I fue para dejar bien claro que jamás una probabilidad puede superar el valor 1.

5. Construcción de la definición axiomática

5.1. Observaciones

La Primera Definición Provisoria para espacios de resultados finitos o enumerables no puede extenderse a espacios de resultados con cardinalidad mayor. Por ejemplo si Ω es el conjunto de los números reales no existe ninguna función de conjunto P definida *para todos los subconjuntos de* Ω que satisfaga la Primera Definición Provisoria. No demostraremos este hecho aquí, nos limitaremos apenas a explicar porqué ocurren estas cosas y después veremos una forma de superar este problema.

Para esto recurriremos al siguiente ejemplo elemental: consideremos sistemas de ecuaciones lineales en el plano R^2. Una ecuación lineal es del tipo $a.x + b.y = k$, onde x e y son las coordenadas cartesianas y a, b y k son números reales fijos. Sabemos que si $a \neq 0$ o $b \neq 0$ cada ecuación representa una recta en el plano y un sistema de n ecuaciones del tipo descripto representa la intersección de n rectas. Si descartamos la posibilidad de tener rectas coincidentes (o sea la misma recta repetida) sabemos que 2 rectas o son paralelas (y no coincidentes) o se cortan en un punto. Si son paralelas y no coincidentes la intersección es el conjunto vacío y el sistema no tiene solución. Supongamos ahora que se cortan en un punto (x_0, y_0) y agreguemos al sistema una ecuación más. Si la recta representada por la nueva ecuación no pasa por

(x_0, y_0) la intersección de las 3 rectas es vacía y el sistema no tiene solución. Si la nueva recta pasa por (x_0, y_0) la intersección de las 3 vuelve a ser (x_0, y_0).

Continuando de esta forma vemos que, si todas las rectas a partir de la cuarta pasan por (x_0, y_0) la única solución del sistema es (x_0, y_0), bastando una única recta que no pase por tal punto para que la intersección sea vacía y el sistema no tenga solución.

Podemos observar que a medida que agregamos ecuaciones a un sistema, más se restringe el conjunto de soluciones, pudiendo ocurrir que no haya ninguna solución.

Ahora volvamos a la definición de probabilidad y consideremos el Axioma III: si $A \subset \Omega$ y $B \subset \Omega$ y $A \bigcap B = \emptyset$ entonces $P(A \bigcup B) = P(A) + P(B)$.

Esta ecuación escrita en una línea representa en realidad muchas ecuaciones, tantas cuanto pares de subconjuntos disjuntos existan.

Observemos que si Ω tiene n elementos existen 2^n subconjuntos disjuntos con $\emptyset$ (recuerde que un conjunto de n elementos tiene 2^n subconjuntos). Claro que hay muchos otros pares de subconjuntos disjuntos.

Por lo tanto, en un conjunto de n elementos hay por lo menos de 2^n pares de subconjuntos disjuntos. En conjuntos de cardinalidad numerable hay infinitos pares de conjuntos disjuntos (puede probarse que tantos cuanto la $cardinalidad\{reales\}$), pero aún es posible definir probabilidades como definidas en la Primera Definición Provisoria. (como vimos en el ejemplo iii) de 4.4.2). Ya si la cardinalidad es mayor esto no es posible.

Una forma de superar esta dificultad es, en vez de definir la función de probabilidad para todos los subconjuntos de Ω, restringirla a una familia menor de subconjuntos. En este caso se exigiría que el Axioma III fuese válido apenas para pares de subconjuntos de la familia. Sería bueno también que estas familias seam adecuadas a las propiedades pretendidas de la función de probabilidad. O sea que exista entre las familias de subconjuntos y la función de probabilidad una compatibilidad del tipo de la que existe entre los espacios vectoriales y las funciones lineales, por ejemplo.

Un tipo de familia razonable podría ser un *álgebra* (o también *álgebra de Boole*) de conjuntos, cuya definición y algunas propiedades daremos a seguir como preámbulo de la que llamaremos Segunda Definición Provisoria de Probabilidad.

5.2. Segunda definición Provisoria de Probabilidad

Definición

Sea Ω un conjunto no vacío. Un álgebra en Ω es una familia β de subconjuntos de Ω que satisface:
i) $\Omega \in \beta$;
ii) si $A \in \beta$ entonces $A^c \in \beta$;
iii) si $A \in \beta$ y $B \in \beta$ entonces $A \bigcup B \in \beta$.
Es equivalente expresar las condiciones precedentes de la siguiente forma: Ω está en β y la familia β es *cerrada* para complementos y uniones.

Propiedades

Sea β un álgebra en Ω. Afirmamos que:
iv) si $A \in \beta$ y $B \in \beta$ entonces $A \bigcap B \in \beta$;
v) si $A \in \beta$ y $B \in \beta$ entonces $A - B \in \beta$;
vi) $\emptyset \in \beta$.
Las propiedades que acabamos de enunciar pueden expresarse diciendo que un álgebra es una familia cerrada para intersecciones y diferencias y además $\emptyset$ está en cualquier álgebra.
Demostración:
iv) y v) siguen de las igualdades $A \bigcap B = A^c (\bigcup B^c)^c$ y $A - B = A \bigcap B^c$;
vi) es consecuencia de v) pues $\emptyset = \Omega - \Omega$.

Definición (Segunda definición Provisoria)

Sea Ω un conjunto no vacío y β un álgebra de sunconjuntos de Ω. Una probabilidad en (Ω, β) es una función de conjunto P definida en β a valores reales tal que:
I) para todo $A \in \beta$ vale que $0 \leq P(A) \leq 1$;
II) $P(\Omega) = 1$;
III) si $A \in \beta$ y $B \in \beta$ y $A \bigcap B = \emptyset$ entonces $P(A \bigcup B) = P(A) + P(B)$.

5.3. Flexibilidad de la nueva definición

Vimos que la Segunda Definición Provisoria de Probabilidad fue introducida para superar las dificultades referidas en 5.1. Veremos ahora que, además de eso, permite encarar problemas que la Primera Defınción Provisoria no conseguía. Ilustraremos esta afirmación con un ejemplo.

5.3.1. Ejemplo

Supongamos que tenemos un dado, o sea $\Omega = \{1, 2, 3, 4, 5, 6\}$ y que nuestras informaciones a respecto de la estructura probabilística del mismo se reducen al conocimiento de las probabilidades de los siguientes conjuntos: $A = \{1, 2, 3\}$, $B = \{4, 5\}$ y $C = \{6\}$, siendo las mismas $P(A) = 0,28$, $P(B) = 0,42$ y $P(C) = 0,30$. Es obvio que el dado no es equilibrado. Por otro lado sabemos que siempre $P(\emptyset) = 0$ y $P(\Omega) = 1$. Además podemos averiguar las probabilidades de otros conjuntos usando las propiedades de una probabilidad: $P(A \bigcup B) = P(A)+P(B) = 0,70$, $P(A \bigcup C) = P(A)+P(C) = 0,58$ y $P(B \bigcup C) = P(B)+P(C) = 0,72$. El cálculo de estas probabilidades puede hacerse también de otras formas. Por ejemplo, dado que $A \bigcup B = C^c$ se deduce que $P(A \bigcup B) = P(C^c) = 1 - P(C) = 1 - 0,30 = 0,70$.

Entonces agora sabemos las probabilidades de los siguientes conjuntos (y sólo de ellos): $\emptyset$, $\{1, 2, 3\}$ $\{4, 5\}$, $\{6\}$, $\{1, 2, 3, 4, 5\}$, $\{1, 2, 3, 6\}$, $\{4, 5, 6\}$ y Ω. Puede verificarse fácilmente que la familia formada por estos ocho conjuntos es un álgebra β_1 sobre la cual está definida la función de conjunto P, que es una probabilidad.

También se puede ver que a partir de la información disponible no es posible calcular la probabilidad de ningún conjunto más. Por ejemplo, es imposible saber la probabilidad de $\{1, 2\}$ o de $\{3\}$ (si supiésemos la da uno sabríamos la del otro calculando probabilidades de diferencias a partir de $\{1, 2, 3\}$; verifique esta afirmación).

En resumen, tenemos el álgebra β_1 en $\Omega = \{1, 2, 3, 4, 5, 6\}$, donde $\beta_1 = \{\emptyset, \{1, 2, 3\}, \{4, 5\}, \{6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 6\}, \{4, 5, 6\}, \Omega\}$ y P es definida en β_1 de la siguiente forma: $P(\emptyset) = 0$, $P(\{1, 2, 3\}) = 0,28$, $P(\{4, 5\}) = 0{,}42$, $P(\{6\}) = 0,30$, $P(\{1, 2, 3, 4, 5\}) = 0,70$, $P(\{1, 2, 3, 6\}) = 0,58$, $P(\{4, 5, 6\}) = 0,72$ y $P(\Omega) = 1$.

Con la Primera Definición Provisoria este ente no sería una probabilidad. Con la Segunda Definición Provisoria lo es y podemos aplicar todas las reglas del cálculo de probabilidades que conocemos sin necesidad de conocer las probabilidades de los conjuntos unitarios. La diferencia fundamental es que sólo trabajaremos con conjuntos de la familia β_1, ya que P está definida apenas para los conjuntos de β_1.

Si nos llegase ahora la información adicional de que $P(\{4\}) = 0,22$, podríamos también calcular $P(\{5\}) = P(\{4,5\}) - P(\{4\}) = 0,20$ y a partir de esto aumentar considerablemente el álgebra de los conjuntos cuya probabilidad conocemos usando las operaciones permitidas. En efecto, ahora podemos agregar a los 8 conjuntos de β_1 los siguientes: $\{4\}$, $\{5\}$, $\{1,2,3,4\}$, $\{1,2,3,5\}$, $\{4,6\}$, $\{5,6\}$, $\{1,2,3,4,6\}$ y $\{1,2,3,5,6\}$, obteniendo así la familia β_2.

5.3.2. Ejercicio

Verifique que la familia β_2, compuesta por todos los conjuntos de β_1 más los 8 que agregamos en el parágrafo anterior, forman una nueva álgebra que denominaremos β_2. Verifique que $\beta_1 \subset \beta_2$ y calcule las probabilidades de los nuevos conjuntos (los que no estaban en la familia β_1).

5.3.3. Notación y algunas observaciones

Formalmente, en la situación original del Ejemplo 5.3.1, diremos que estamos en el espacio de probabilidad (Ω, β_1, P). Al agregar los nuevos conjuntos debemos considerar el espacio de probabilidad (Ω, β_2, Q), donde la probabilidad Q coincide con P para todos los conjuntos de β_1. Es necesario utilizar una letra diferente para denotar la "nueva" probabilidad porque las funciones de conjunto P y Q, a pesar de coincidir sobre β_1, tienen dominios diferentes: P está definida sólo sobre β_1 y Q está definida sobre β_2.

Por definición, diremos que el álgebra β_2 es más fina que el álgebra β_1 si y sólo si $\beta_1 \subset \beta_2$. Vemos que no hay ninguna contradicción entre las estructuras (Ω, β_1, P) y (Ω, β_2, Q), lo que es claro es que en la segunda situación tenemos más información que en la primera. De esta forma vemos que, cuando más fina sea el álgebra, habrá más información.

El álgebra menos fina de todas es la llamada trivial: $\{\emptyset, \Omega\}$. La más fina es la denominada Partes de de Ω, que consiste en la familia de todos los subconjuntos de Ω.

No siempre dos álgebras son comparables, en el sentido de ser una más fina que la otra. Por ejemplo consideremos el siguiente conjunto Ω de tres elementos, $\Omega = \{a, b, c\}$. Sean β_1 y β_2 las siguientes álgebras:

$\beta_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ y $\beta_2 = \{\emptyset, \{a, b\}, \{c\}, \Omega\}$.

Aquí $\beta_2 \not\subset \beta_1$, pues $\{c\} \in \beta_2$ y $\{c\} \notin \beta_1$.

Como $\{a\} \in \beta_1$ y $\{a\} \notin \beta_2$ resulta $\beta_1 \not\subset \beta_2$.

En este momento es bueno revisar la Primera Definición Provisoria, en la que siquiera se menciona la palabra álgebra. Pero como en este caso la probabilidad P está definida para todos los subconjuntos de Ω esto significa que el álgebra es justamente Partes de Ω. Vimos ya que la enorme cantidad de conjuntos de esta álgebra en conjuntos de cardinalidad grande tornaba incompatibles las condiciones que definían probabilidad en el sentido de la Primera Definición Provisoria. Este fue el motivo de la introducción de la Segunda Definición Provisoria.

5.3.4. Observaciones

i) Volvamos al espacio de probabilidad (Ω, β_1, P), donde $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\beta_1 = \{\emptyset, \{1, 2, 3\}, \{4, 5\}, \{6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 6\}, \{4, 5, 6\}, \Omega\}$ y P es definida en β_1 de la siguiente forma: $P(\emptyset) = 0$, $P(\{1, 2, 3\}) = 0,28$, $P(\{4, 5\}) = 0{,}42$, $P(\{6\}) = 0,30$, $P(\{1, 2, 3, 4, 5\}) = 0,70$, $P(\{1, 2, 3, 6\}) = 0,58$, $P(\{4, 5, 6\}) = 0,72$ y $P(\Omega) = 1$.

Podríamos identificar todas las funciones de probabilidad definidas em Partes de Ω que coinciden con P en β_1 con el siguiente subconjunto convexo de R^6:

$$\{\ Q = (q_1, q_2, q_3, q_4, q_5, q_6)\ /\ q_j \geq 0,\ j\ natural,\ 1 \leq j \leq 6; \\ q_1 + q_2 + q_3 = 0,28;\ q_4 + q_5 = 0,42;\ q_6 = 0,30\ \}.$$

ii) La introducción de la σ-álgebra β_1 en el ejemplo anterior puede parecer artificial. Y en cierta forma lo es, ya que podríamos haber pensado de la siguiente forma: pintar las caras $1, 2$ y 3 del dado de azul, las caras 4 y 5 de rojo y la 6 de amarillo. Entonces consideraríamos $\Omega' = \{azul,\ rojo,\ amarillo\}$ y la probabilidad P' definida en partes de Ω' por:

$$P'(\{azul\}) = 0,28\ ,\quad P'(\{rojo\}) = 0,42\ \ y\ \ P'(\{amarillo\}) = 0,30.$$

De esta forma conseguiríamos encajarnos en el molde de la Primera Definición Provisoria. El hecho que motivó el primer abordaje, introduciendo el álgebra β_1, fue la necesidad de presentar el concepto de álgebra sobre un ejemplo bien simple.

5.4. La definición definitiva

Veremos aquí la definición de probabilidad de la forma que se encuentra en textos no elementales. La definición se tornará un poco más restrictiva, porque el álgebra y la propiedad aditiva de la Segunda Definición Provisoria serán substituídas por σ-álgebra y propiedad σ-aditiva, conceptos que definiremos oportunamente. Antes de eso mostraremos uno de los problemas típicos cuya solución requiere la substitución del álgebra por la σ-álgebra.

5.4.1. Ejemplo: sucesiones de variables aleatorias

Además de estudiar los espacios de probabilidades es interesante estudiar ciertas funciones definidas en ellos llamadas variables aleatorias. Con precisión, sea (Ω, β, P) un espacio de probabilidad denotemos por R al conjunto de los números reales. En lo que sigue, cuando digamos intervalo de R entenderemos intervalo de cualquier tipo: abierto, semiabierto, cerrado, semicerrado, con extremos finitos o infinitos.

Por definición, una variable aleatoria es una función $X : \Omega \longrightarrow R$ tal que para todo intervalo $I \subset R$ sea verdad que $X^{-1}(I) = \{\omega \in \Omega / X(\omega) \in I\} \in \beta$.

El interés de la definición es bastante evidente, se trata de poder calcular la probabilidad de que los valores de la función X estén dentro de ciertas fajas (intervalos), de ahí la exigencia de que $X^{-1}(I) \in \beta$.

Un concepto importante, cuando se trata de funciones a valores reales es el de convergencia. Por ejemplo, si $\{X_n, n \in N\}$ (donde N denota el conjunto de los números naturales) es una sucesión de variables aleatorias, es importante saber en qué subconjunto de Ω la misma converge puntualmente a una variable aleatoria X. Puede provarse, por aplicación de la definición de límite, que el subconjunto de Ω donde la convergencia ocurre puede escribirse como:

$\{\omega \in \Omega \ / \ X_n(\omega) \to X(w)\} = \bigcap_{\epsilon>0} \bigcup_{n\in N} \bigcap_{k\geq n} \{\omega \in \Omega / |X_k(\omega) - X(\omega)| < \epsilon\}$;

ahora, usando que la sucesión $\{1/m, m \in N\}$ converge a 0, se tiene que el conjunto donde hay convergencia puede reescribirse como:

$$\bigcap_{m\in N} \bigcup_{n\in N} \bigcap_{k\geq n} \{\omega \in \Omega / |X_k(\omega) - X(\omega)| < 1/m\}.$$

Se puede ver que la diferencia de variables aleatorias es también una variable aleatoria, por lo tanto los conjuntos

$$\{\omega \in \Omega / |X_k(\omega) - X(\omega)| < 1/m\} = \{\omega \in \Omega / |(X_k - X)(\omega)| \in (-1/m, 1/m)\}$$

deben estar en β (por definición de variable aleatoria). Así, para que el conjunto $\{\omega \in \Omega \ / \ X_n(\omega) \to X(w)\}$ esté en β es suficiente que la familia β sea cerrada para intersecciones y uniones *enumerables*. Sabemos que las álgebras son apenas cerradas para intersecciones y uniones finitas, por este motivo en la definición de espacio de probabilidad se substituye el álgebra por una σ-álgebra (cuya definición veremos en seguida). Dado que toda σ-álgebra es álgebra pero no vale la recíproca, esta decisión torna la definición más restrictiva, pero por otro lado permite abordar problemas como la convergencia de variables aleatorias y afines, de suma importancia en la teoria de probabilidades.

La caracterización del conjunto de convergencia de una sucesión de variables aleatorias que dimos arriba sólo tiene por objetivo ilustrar la necesidad de introducir la σ-álgebra. Por lo tanto el lector puede omitir la demostración, de la cual dimos aquí apenas un esbozo.

En lo que sigue nos limitaremos apenas a presentar:
a) la definición de σ-álgebra y algunas propiedades;
b) la definición definitiva de probabilidad y algunas propiedades.

5.4.2. Definición y propiedades de una σ-álgebra

Definición
Sea Ω un conjunto no vacío. Una σ-álgebra en Ω es una familia β de subconjuntos de Ω que satisface:

i) $\Omega \in \beta$;
ii) si $A \in \beta$ entonces $A^c \in \beta$;
iii) si $A_n \in \beta$ para todo $n \in N$ entonces $\bigcup_{n \in N} A_n \in \beta$.

Es equivalente expresar las condiciones precedentes de la siguiente forma: Ω está en β y la familia β es cerrada para complementos y uniones numerables.

Propiedades
Sea β una σ-álgebra en Ω. Afirmamos que:
iv) si $A \in \beta$ y $B \in \beta$ entonces $A \bigcup B \in \beta$, portanto β es álgebra con todas sus propiedades;
v) si $A_n \in \beta$ para todo $n \in N$ entonces $\bigcap_{n \in N} A_n \in \beta$.

Demostración:
iv) sean $A \in \beta$ y $B \in \beta$, consideremos la sucesión de conjuntos siguiente: $A_1 = A$, $A_2 = B$ y $A_n = \emptyset$ para $n \geq 3$; entonces, dado que $\emptyset \in \beta$ por ser β álgebra, se tiene que $A \bigcup B = \bigcup_{n \in N} A_n \in \beta$;
v) sea si $A_n \in \beta$ para todo $n \in N$; entonces $\bigcap_{n \in N} A_n = (\bigcup_{n \in N} (A_n)^c)^c \in \beta$, en consecuencia de ii) y iii).

5.4.3. Definición Definitiva de Probabilidad

Un espacio de probabilidad (Ω, β, P) consiste en un conjunto no vacío Ω, una σ-álgebra β de subconjuntos de Ω y una función de conjunto P definida en β que satisface:
I) $P(A) \geq 0$ para todo $A \in \beta$;
II) $P(\Omega) = 1$;
III) si $A_n \in \beta$ para todo $n \in N$ y además $A_i \bigcap A_j = \emptyset$ para todo $i \neq j$, entonces $P(\bigcup_{n \in N} A_n) = \sum_{n \in N} P(A_n)$ (σ-aditividad).

Equivalentemente se dice que un espacio de probabilidad (Ω, β, P) se compone de un conjunto no vacío Ω, una σ-álgebra β de subconjuntos de Ω y una función de conjunto P no negativa y σ-aditiva definida en β tal que $P(\Omega) = 1$. P es también llamada una *medida de probabilidad*.

5.4.4. Observaciones

i) En este nuevo contexto, un suconjunto A de Ω será llamado evento si y solo si $A \in \beta$;
ii) Siendo P σ-aditiva necesariamente será $P(\emptyset) = 0$: dado que $\emptyset = \bigcup_{n \in N} E_n$ con $E_n = \emptyset$ para todo n natural y P es *sigma*-aditiva entonces se tiene que $P(\emptyset) = \sum_{n \in N} P(E_n)$. Si fuese $P(\emptyset) = r$ con $r > 0$ la igualdad anterior sería imposible. Luego debe ser $P(\emptyset) = 0$.
iii) Ahora es fácil verificar que si P es σ-aditiva entonces P es aditiva.
En efecto, sean $A \in \beta$, $B \in \beta$ y $A \bigcap B = \emptyset$. Llamemos $A_1 = A$, $A_2 = B$ y $A_n = \emptyset$ para $n \geq 3$; entonces $P(\bigcup_{n \in N} A_n) = \sum_{n \in N} P(A_n) = P(A) + P(B)$.

5.4.5. Otras propiedades

En los espacios de probabilidad, al ser P aditiva, valen las propiedades IV), V), VI), VII) y VIII) relativas a la Primera Definición Provisoria dadas en el Teorema 4.5.2, desde que todos los conjuntos mencionados en tales propiedades pertenezcan a la σ-álgebra β. Con precisión, las nuevas versiones de IV) a VIII) son:
IV) $\emptyset \in \beta$ y $P(\emptyset) = 0$;
V) si $A \in \beta$ entonces $A^c \in \beta$ y $P(A^c) = 1 - P(A)$;
VI) si $A \in \beta$, $B \in \beta$ y $A \subset B$ entonces $B - A \in \beta$, $P(A) \leq P(B)$ y $P(B - A) = P(B) - P(A)$;
VII) si $A \in \beta$ y $B \in \beta$ vale que $B - A \in \beta$ y $P(B - A) = P(B) - P(B \bigcap A)$;
VIII) si $A \in \beta$ y $B \in \beta$ entonces $A \bigcup B \in \beta$, $A \bigcap B \in \beta$ y $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$.

A seguir, un par de nuevas propiedades de los espacios de probabilidades, no necesariamente satisfechas por las funciones de conjunto apenas aditivas. Cada una de ellas es equivalente a la σ-aditividad si P es una función de conjunto no negativa y aditiva definida en una σ-álgebra. Enunciaremos una versión más débil de esta afirmación, tal cual consta en el siguiente lema (sin demostración).

5.4.6. Lema

En todo espacio de probabilidad (Ω, β, P) valen las siguientes afirmaciones:
i) si $\{A_n, n \in N\} \subset \beta$, $A_n \subset A_{n+1}$ para todo $n \in N$ y $A = \bigcup_{n \in N} A_n$, entonces $P(A_n) \to P(A)$;
ii) si $\{A_n, n \in N\} \subset \beta$, $A_n \supset A_{n+1}$ para todo $n \in N$ y $A = \bigcap_{n \in N} A_n$, entonces $P(A_n) \to P(A)$.

5.4.7. Aplicaciones: propiedades de la función de distribución.

Definición
Sea (R, β, P) un espacio de probabilidad, donde R denota el conjunto de los números reales y β es una σ-álgebra que continene la familia de todos los intervalos. $F : R \to [0, 1]$ definida por $F(t) = P((-\infty, t])$ es denominada función de distribución de P.

5.4.8. Proposición

F tiene las siguientes propiedades:
i) si $a < b$ $F(a) \leq F(b)$ y $P((a, b]) = F(b) - F(a)$;
ii) $F(\infty) = limite_{\ n\to\infty} F(n) = 1$;
iii) $F(-\infty) = limite_{\ n\to\infty} F(-n) = 0$;
iv) si la sucesión $a_n, n \in N$ satisface $a_n \leq a_{n+1}$ para todo $n \in N$ y $a_n \to a$ entonces $F(a_n) \to F(a)$.

Demostración:
i) si $a < b$ vale que $(-\infty, a] \subset (-\infty, b]$ y entonces $F(a) \leq F(b)$;
además $P((a, b]) = P((-\infty, b] - (-\infty, a]) = P((-\infty, b]) - P(-\infty, a]) = F(b) - F(a)$;
ii) sean $A_n = (-\infty, n]$ para todo $n \in N$ y $A = R$, aplicando la parte i) del lema anterior a $A_n, n \in N$ y A se tiene el resultado enunciado;
iii) aplique i) del Lema anterior a $A_n = (-\infty, -n]$ y $A = \emptyset$;
iv) aplique ii) del Lema anterior a $A_n = (-\infty, a_n]$ y $A = (-\infty, a]$.

6. Apéndices

6.1. El Teorema de Bernoulli

Daremos sólo el enunciado de este importantísimo resultado demostrado por Bernoulli.

Supongamos que repetimos veces n (n natural), siempre en las mismas condiciones, un experimento aleatorio cuyo espacio de resultados sea el conjunto Ω. Sea A um subconjunto de Ω. Como antes, $fr(A, n)$ denota la frecuencia relativa de A en n repeticiones independientes del experimento y $P(A)$ la probabilidad de A. Entonces:

$$\textit{para todo } \epsilon > 0 \qquad \textit{limite}\ _{n\to\infty}\ \textit{Probabilidad}\ \{|fr(A,n) - P(A)| \geq \epsilon\} = 0$$

Es en este sentido que decimos que la frecuencia relativa converge a la probabilidad. Este tipo de convergencia se conoce como convergencia en probabilidad, caso particular de la convergencia en medida.

El Teorema de Bernoulli fue posteriormente fue generalizado, entre otros, por Tchebycheff (1821-1894). Esta generalización se llamó Ley Débil de los Grandes Números. Una desigualdad desarrollada por Tchebycheff (que lleva su nombre y que no explicitaremos aquí) permite demostrar el resultado siguiente:

$$\textit{para todo } \epsilon > 0 \qquad \textit{Probabilidad}\ \{|fr(A,n) - P(A)| \geq \epsilon\} \leq \frac{1}{4\ n\ \epsilon^2}$$

Se puede verificar que a partir de esta afirmación puede demostrarse el Teorema de Bernoulli (la demostración que dio Bernoulli es otra, porque Tchebycheff es posterior a Bernoulli). Además de eso, se puede acotar la probabilidad del error cuando se estima $P(A)$ mediante la frecuencia relativa de A (observe a presencia de n en el denominador de $\frac{1}{4\ n\ \epsilon^2}$). Por lo tanto, a través de esta afirmación se llega a una respuesta a la pregunta " cuando habría que parar? " al trabajar con frecuencias relativas. Hay respuestas mejores usando el Teorema Central del Límite, no veremos esto aquí.

Por ejemplo, supongamos que se desea estimar la probabilidad de cierto evento A (denotada por $P(A)$) de tal forma que la probabilidad de errar en más de 0,1 en la estimación no supere 0,02. Esto significa que pretenderemos que $\quad Probabilidad \ \{|fr(A,n) - P(A)| \geq 0,1\} \leq 0,002.$

Para tal es suficiente elegir un entero n tal que $\frac{1}{4\ n\ \epsilon^2} = \frac{1}{4\ n\ 0{,}1^2} \leq 0,02$ o equivalentemente un entero n tal que $n \geq \frac{1}{4\ 0{,}1^2\ 0{,}02}$.

6.2. Conjuntos borelianos y variables aleatorias

Como definimos en 5.4.1, dado un espacio de probabilidad (Ω, β, P), una variable aleatoria es una función $X : \Omega \longrightarrow R$ tal que para todo intervalo $I \subset R$ sea verdad que $X^{-1}(I) \in \beta$.

Daremos una definición equivalente, donde se note la verdadera extensión de la condición impuesta en esta definición.

En primer lugar, podemos observar que la condición puede ser extendida a uniones finitas o enumerables de intervalos. Este hecho es consecuencia de ser β una σ-álgebra y de la siguiente propiedad de la función *imagen inversa*:

$$X^{-1}(\textstyle\bigcup_{n \in N} I_n) = \bigcup_{n \in N} X^{-1}(I_n).$$

También, si ahora cada conjunto $A_n, n \in N$ es unión finita o numerable de intervalos, la propiedad de la definición vale para cualquier intersección finita o enumerable de los conjuntos B_n, por ser β una σ-álgebra y por ser válido que $X^{-1}(\bigcap_{n \in N} A_n) = \bigcap_{n \in N} X^{-1}(A_n)$.

Es claro que podríamos seguir indefinidamente con este proceso, pero seguiremos otro camino para encontrar todos los subconjuntos de A de R que satisfazen la condición $X^{-1}(A) \in \beta$.

Admitiremos sin prueba que la intersección de un número arbitrario de σ-álgebras es también una σ-álgebra. Consideremos la σ-álgebra denominada $Partes \ \ de \ \ R = \{B/B \subset R\}$. Es claro que $\eta = \{intervalos\ en\ R\} \subset Partes\ de\ R$.

Está bien definida entonces $\quad \beta^* = \bigcap_{\gamma\ s.a.,\ \eta \subset \gamma} \gamma, \quad$ donde "s.a." significa σ-álgebra. Por lo dicho en el parágrafo anterior β^* es una σ-álgebra. También es fácil ver que $\eta \subset \beta^*$. Entonces podríamos decir que β^* es la "menor" σ-álgebra en R que contiene al conjunto η de todos los intervalos de R. β^* se denomina la σ-álgebra de los borelianos de R. Se puede probar (no lo haremos) que las tres siguientes afirmaciones son equivalentes y por lo tanto constituyen una definición de variable aleatoria.

6.2.1. Proposición

Sea (Ω, β, P) un espacio de probabilidad y una función $X : \Omega \to R$.
Son equivalentes:
i) para todo intervalo $I \in \eta$ vale que $X^{-1}(I) \in \beta$;
ii) para todo $B \in \beta^*$ vale que $X^{-1}(B) \in \beta$;
iii) para todo abierto A de R vale que $X^{-1}(A) \in \beta$.

6.3. Probabilidad inducida por una variable aleatoria

6.3.1. Definición

Sea (Ω, β, P) un espacio de probabilidad y $X : \Omega \to R$ una variable aleatoria. Está claro que el dominio de la función de conjunto P es la σ-álgebra β en Ω. Vamos a definir ahora en la σ-álgebra β^* de los borelianos de R una función de probabilidad P_X de la forma siguiente:

$$\text{para todo } B \in \beta^* \qquad P_X(B) = P(X^{-1}(B)).$$

Se puede probar que P_X es una probabilidad definida en β^*, o sea que (R, β^*, P_X) es un espacio de probabilidad. P_X se denomina probabilidad inducida en los borelianos de R por la variable aleatoria X.

6.3.2. Definición

Sea (Ω, β, P) un espacio de probabilidad y $X : \Omega \to R$ una variable aleatoria. Por definición, la función de distribución F_X de la variable aletoria X es la función de distribución de P_X, o sea:

$$F_X(t) = P_X((-\infty, t]) = P(X^{-1}(-\infty, t]) = P(\{\omega \in \Omega \ / \ X(\omega) \leq t\}).$$

6.3.3. Proposición

Sean (Ω, β, P) un espacio de probabilidad, $X : \Omega \to R$ una variable aleatoria, P_X la probabilidad inducida em R por X y F_X la función de distribución de X. Entonces F_X tiene las siguientes propiedades:

i) si $a < b \quad F_X(a) \leq F_X(b)$ y $P_X((a,b])) = P(\{\omega \in \Omega / a < X(\omega) \leq b\}) = F_X(b) - F_X(a)$;
ii) $F_X(\infty) = limite_{\ n\to\infty}\ F_X(n) = 1$;
iii) $F(-\infty) = limite_{\ n\to\infty}\ F_X(-n) = 0$;
iv) si la sucesión $a_n, n \in N$ satisface $a_n \leq a_{n+1}$ para todo $n \in N$ y $a_n \to a$ entonces $F_X(a_n) \to F_X(a)$;
v) Por definición, P_X determina F_X. También vale la recíproca, o sea que F_X determina P_X.

La demostración de las 4 primeras propiedades es muy parecida con las 4 propiedades de la función F de la Proposición 5.4.8. Ya la propiedad v) no tiene equivalente en dicha proposición. El motivo es el siguiente: en la Proposición 5.4.8 la σ-álgebra β contiene la familia de los intervalos y por lo tanto contiene la σ-álgebra de los borelianos. Ya en la situación actual, el dominio de P_X es estrictamente la σ-álgebra de los borelianos. No daremos la prueba de v) aquí, ya que para tal son necesarios conocimientos de la teoría de la medida que quedan fuera del alcance de este trabajo.

Referencias

Apostol, T. M., 1969. Calculus, Volume II, Multi-Variable Calculus and Linear Algebra, with Applications to Differential Equations and Probability, John Wiley and Sons, 2nd Edition.

Cramer, H., 1964. Elementos de la teoría de Probabilidades, Aguilar, Madrid, 5a edición.

Feller, W., 1968. An Introduction to Probability Theory and Its Applications, Volume I, Wiley, 3rd edition.

Gnedenko, B., 1978. The Theory of Probability, Mir Publishers, Moscow, 4th Printing.

Kolmogorov, A. N., 1956. Foundations of the Theory of Probability, Chelsea, N. York. Es traducción del trabajo original en alemán: Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin, 1933.

Peña, M., 1972. El paraíso terrateniente, Ediciones Fichas, 2a edición.

Todhunter, I., 1965 (la versión original es de 1865). A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace. Chelsea, N. York.