

25A
2ej.



UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO

Facultad de Ciencias

SOBRE LOS ELEMENTOS DE LA PROBABILIDAD
SUBJETIVA.

T E S I S
Que para obtener el título de
M A T E M Á T I C O
p r e s e n t a
Jorge Francisco de la Vega Gongora

Diciembre 1992

FALLA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INTRODUCCION. Tres enfoques de probabilidad	1
---	---

CAPITULO I. Conceptos básicos del enfoque subjetivo

1.1. El concepto de fenómeno aleatorio	8
1.2. Cantidades aleatorias	9
1.3. Aseveraciones y particiones	11
1.4. Dependencia lógica entre eventos	12
1.5. Representación en forma lineal	16

CAPITULO II. Previsión

2.1. Introducción	20
2.2. Fuentes subjetivas de la previsión	20
2.3. Criterios de Coherencia	21
2.4. Probabilidades de eventos (caso finito)	29
2.5. Probabilidades de eventos (caso infinito)	34
2.6. Convergencia de cantidades aleatorias	44

CAPITULO III. Probabilidad Condicional

3.1. Introducción	47
3.2. Previsión y probabilidad condicional	47
3.3. Verosimilitud	49
3.4. Propiedad Conglomerativa	50
3.5. Dependencia e independencia estocástica	51

CAPITULO IV. Intercambiabilidad

4.1. Introducción	52
4.2. El problema de la inducción	52
4.3. Justificación del punto de vista frecuentista	54
4.4. Intercambiabilidad	56
4.5. Intercambiabilidad condicional	66
4.6. Interpretación de la intercambiabilidad	66

Conclusiones
Referencias

68
69

INTRODUCCION. TRES ENFOQUES DE LA PROBABILIDAD

Introducción.

La Teoría de la Probabilidad es una herramienta muy importante para las ciencias empíricas como la física, la economía, la ingeniería, la biología; además juega un papel especial en una multitud de situaciones de la vida cotidiana. En general, se puede decir que el objeto de la probabilidad es la apreciación numérica de las oportunidades: provee una herramienta para apoyar la toma de decisiones en condiciones de incertidumbre. Debido a la amplitud de aplicaciones de la probabilidad, las interpretaciones comunes no han sido suficientes para resolver diversos problemas, que surgen cuando un individuo tiene que tomar una decisión. El cómo asignar un valor numérico a un elemento tan complejo como es la oportunidad se puede resolver de distintas formas, dependiendo del punto de vista en que uno se ubique; como en el siguiente problema, por ejemplo. Un inversionista desea distribuir su capital entre n instrumentos de inversión que dan rendimientos aleatorios. Considérese la forma de resolver este problema desde dos puntos de vista: el de un estadístico y el de un jugador. El estadístico se preocupará por la frecuencia con la que un rendimiento ha caído en ciertos intervalos a lo largo del tiempo. De manera general, el estadístico comprueba que para algunos eventos cuya realización se debe al azar, existe una proporción casi constante entre el número de observaciones en que el evento se realiza y el número total de observaciones. Desde este punto de vista, la probabilidad se considera como una *característica global* de un conjunto de realizaciones, y no una característica individual. El estadístico sólo puede calcular la probabilidad de un evento después de una experiencia, e incluso, después de una larga serie de pruebas y observaciones. Es así que para el estadístico, la probabilidad de un evento es una especie de constante física relativa a cierta categoría de fenómenos colectivos. Así, observando el comportamiento histórico de los rendimientos, el estadístico llega a establecer, mediante un razonamiento inductivo, cuál es la distribución de probabilidad de los rendimientos.

Desde el punto de vista de un jugador, resulta más importante incorporar sus propias expectativas basadas en su conocimiento del mercado financiero que observar el comportamiento histórico. De manera general, un jugador declara, *antes de que se realice*

el evento, cuáles son las propiedades que él le atribuye. Es suficiente que el jugador conozca bien cuáles son las condiciones del "juego", para definir las probabilidades sin tener necesidad de realizar la experiencia. Para el jugador la probabilidad se obtiene a priori. Así, el jugador estará interesado en diseñar una estrategia de inversión de acuerdo a la forma en que él cree que sus expectativas se realizarán en el mercado.

La definición de probabilidad ha dado lugar a interminables discusiones. Sin embargo, el cálculo de probabilidades se ha desarrollado constantemente. La causa de esto se aclara si se distinguen los elementos que conforman las ciencias: un esquema de la realidad, por una parte, y un sistema de correspondencia entre los elementos de este esquema y los elementos correspondientes a la realidad. La mayoría de las dificultades radican en la elección y determinación de esta correspondencia. Las dificultades propias del esquema, por el contrario, van siendo superadas por el desarrollo de la ciencia. Esencialmente se han propuesto tres correspondencias distintas: *frecuentista*, *lógica* y *subjetiva*, y hay básicamente un esquema, resumido en los axiomas de Kolmogorov. Sin embargo, tal esquema introduce una estructura matemática muy fuerte y sofisticada basándose completamente en el análisis matemático y la Teoría de la medida. Sobre ese sistema de axiomas se encuentra soportado todo el edificio de la teoría, universalmente aceptada por los matemáticos, los cuales sólo difieren en torno a su interpretación concreta. Es importante haberse asegurado que se puede axiomatizar completamente y sin discusión posible, toda la teoría matemática de las probabilidades, pero también es necesario reconocer que si uno se coloca de esta manera en el medio de incertidumbres filosóficas, dicha teoría matemática no sólo se aleja de las aplicaciones, sino que asume un aspecto extraño y queda privada de la fuerza de invención y de la facilidad de comprensión que suscita la interpretación concreta de los problemas matemáticos que surgen.

Los axiomas de Kolmogorov son los siguientes: Sea Ω una colección de objetos $\alpha, \beta, \gamma, \dots$ llamados *eventos elementales* y sea \mathcal{F} una σ -álgebra de conjuntos de Ω . Entonces, a cada $A \in \mathcal{F}$ se le asocia un número real positivo $P(A)$ de tal manera que:

- i. $P(\Omega) = 1$.
- ii. Si $\{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F}$ tal que $A_i \cap A_j = \emptyset \quad \forall i \neq j$, entonces $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Sin embargo, una deficiencia importante de tal conjunto de axiomas es la fuerte

restricción provocada por el requerimiento de la σ -aditividad. Esta restricción no permite definir la probabilidad para todos los eventos. A pesar de esto, el cálculo de probabilidades se ha seguido desarrollando a partir de tales axiomas, puesto que en las aplicaciones concretas de la teoría de la probabilidad uno no está interesado en calcular probabilidades sino para ciertos tipos de eventos. Sin embargo, desde una perspectiva filosófica, no existe una razón lógica que otorgue diferentes status a los eventos.

Breve reseña histórica.

La teoría moderna de la probabilidad surgió como consecuencia de ciertos problemas planteados por jugadores a matemáticos, y se confinó especialmente al cálculo de probabilidades asociados a la ocurrencia de resultados específicos en juegos de azar, como obtener una tercia o un pókar al jugar con una baraja. Este tipo de problemas fueron resueltos por matemáticos italianos (Cardano, Pacioli, Tartaglia), aunque no de forma sistemática; fue hasta el siglo *XVII* con Pascal (1622-1662) y Fermat (1601-1655) que se dieron los primeros métodos generales para resolver tales problemas. De este modo se estableció un intercambio fructífero entre los juegos de azar y las matemáticas. Por otra parte, el primer libro publicado de probabilidad fue el de Huygens en 1657, *De Ratiociniis in ludo aleae*.

Como puede observarse, el origen de la probabilidad fue totalmente pragmático, ligado a la *medición* de la incertidumbre asociada a la ocurrencia de fenómenos concretos. La probabilidad era una guía para tomar una decisión. Los primeros conceptos desarrollados fueron los de *esperanza*, *evento* (ocurrencia de un fenómeno) y *espacio muestral* (todos los posibles resultados de un experimento).

Laplace publicó en 1812 el libro *Théorie Analytique des Probabilités* en el que se resume todo el conocimiento en dicha materia hasta ese año. Laplace y Bayes trabajaron con lo que se conoce "el enfoque clásico": bajo supuestos de simetría, cada evento elemental tiene la misma probabilidad, así que la probabilidad de un evento A se define como

$$P(A) = \frac{|A|}{|\Omega|},$$

es decir, es la razón entre el número de eventos que favorecen la ocurrencia de A y la totalidad de los eventos. De este modo, la probabilidad de cada evento simple es $P(\{\omega\}) = \frac{1}{n} \quad \forall \omega \in \Omega$, donde $n = |\Omega|$. Sin embargo, una de las objeciones de esta

definición de probabilidad se que se cae en un círculo vicioso en la definición, puesto que en la definición misma de la probabilidad se supone que se sabe lo que son dos probabilidades iguales.

El enfoque frecuentista.

En este punto de vista, un fenómeno aleatorio se define como un fenómeno que, observado bajo un determinado conjunto de condiciones, no siempre produce el mismo resultado, sin embargo, cada uno de sus posibles resultados es conocido y se presenta con cierta regularidad, en el sentido de que la frecuencia tiende a un valor límite a medida que se aumenta el número de observaciones. La probabilidad entonces se interpreta como ese hipotético valor límite. Esto implica que los fenómenos aleatorios que pueden ser estudiados con este enfoque son únicamente aquellos que son repetibles, al menos conceptualmente, un número infinito de veces, bajo las mismas condiciones.

La concepción empírica o frecuentista de la probabilidad fue formulada por John Venn en 1886 y después por Reichenbach en 1949 y por Von Mises en 1957 ([12]). Sin embargo, Bernoulli fue posiblemente el primero en resaltar la importancia de considerar y hacer ver la relación entre la probabilidad de un evento y la frecuencia de su realización. Se puede probar que si $\frac{S_n}{n}$ es la proporción de ocurrencias en una sucesión de n ensayos Bernoulli con probabilidad de éxito p , entonces se cumple que $\forall \epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \leq \epsilon\right) \rightarrow 1 \text{ cuando } n \rightarrow \infty.$$

Este resultado establece una relación matemática entre la frecuencia de ocurrencia de un evento y su probabilidad p . En un lenguaje informal se tiene que

$$\text{"} \lim_{n \rightarrow \infty} \text{" } \frac{S_n}{n} = p$$

El límite aparece entre comillas porque no se refiere al límite usual del análisis matemático; más bien tiene la siguiente interpretación: para cada $\epsilon > 0$ no necesariamente es posible encontrar un valor N tal que $\left|\frac{S_n}{n} - p\right| < \epsilon$, con seguridad para toda $n \geq N$. Lo único que puede decirse es que es *verosímil* que $\frac{S_n}{n}$ difiera de p a lo más por ϵ tomando n suficientemente grande. La palabra *verosímil* involucra a su vez ideas de probabilidad, así que el intento de una definición de límite utilizando el límite matemático resulta circular, y por lo tanto no es satisfactoria.

El enfoque lógico.

Esta interpretación sostiene que las afirmaciones de probabilidad no son empíricas. La probabilidad es un objeto indefinible que se incluye en argumentos y juicios que no tienen una relación directa con entidades empíricas o físicas. La probabilidad representa una relación lógica entre una proposición S y un cuerpo de conocimientos E que representan evidencia. De esta manera la probabilidad no depende de los hechos y de sus propiedades reales, sino exclusivamente de las proposiciones que las describen. En este caso sólo hay un número real p que puede ser llamado la probabilidad de S relativa a E . Sin embargo, en este enfoque no ha sido propuesto ningún procedimiento operacional satisfactorio para medir el valor de p .

Tal visión fue formulada explícitamente por John M. Keynes (*A treatise on probability*, 1921), y es defendida por Carnap y Harold Jeffrey entre otros. Carnap sugirió denotar como teoría de probabilidad₁ al problema de la inferencia inductiva, a la naturaleza de la demostración científica, a la credibilidad de proposiciones a partir de una evidencia empírica y en general a métodos de razonamiento con base en datos empíricos, dirigidos a conclusiones de experiencias futuras. Por otra parte, la teoría de la probabilidad₂ se relaciona con el estudio de eventos repetitivos que parecen tener la propiedad de que su frecuencia relativa de ocurrencia en un gran número de ensayos tiene un valor límite estable.

El enfoque subjetivo.

El enfoque subjetivo considera a la probabilidad como una medida del grado de creencia que tiene un sujeto concreto en un momento dado y con un conjunto de información específico, acerca de la ocurrencia de un evento. Precisamente esta es la característica por la que este enfoque es llamado subjetivo: la probabilidad es considerada una función de dos argumentos: $P(A|B)$ donde A es el evento que se está considerando, y B describe el conjunto de condiciones o el conjunto de información bajo la que un sujeto particular considera la ocurrencia de A .

Por este motivo, no hay sólo una forma de asignar una distribución de probabilidad para un conjunto de eventos. Cualquier grado de creencia se acepta en cualquier evento siempre y cuando sea *permisible* esta asignación, de acuerdo a ciertos criterios llamados *de coherencia*. Se verá que lo único que puede decir la teoría en relación a una asignación

específica, es si tal es o no es *coherente*. En caso de que ésta no lo sea, la teoría garantiza que ésta puede modificarse para que resulte coherente, pero no especifica cómo debe llevarse a cabo tal modificación.

El enfoque subjetivo es más general que los enfoques mencionados anteriormente, en el sentido de que es posible asignar probabilidades no sólo basadas en el conocimiento de la frecuencia de realización de un evento, o en supuestos de simetría. Sin embargo, tales asignaciones no provienen de propiedades objetivas de los fenómenos, sino de consideraciones subjetivas del sujeto que realiza la asignación.

Este enfoque no es nuevo, ya que fue compartido por Bernoulli y Laplace, aunque de forma intuitiva e informal. Pero desde principios del siglo XX comenzó a tener un fuerte desarrollo debido a los trabajos de F. P. Ramsey y Bruno de Finetti en 1926 y de J. L. Savage en 1954.

Dentro de la corriente subjetiva, Bruno de Finetti (1906-1981) ocupa un lugar preponderante, debido principalmente a que fue el primer subjetivista que expone su teoría en forma sistemática, y por haber introducido en 1931 el concepto de *intercambabilidad*: una sucesión finita o numerable de variables aleatorias es intercambiable si para $k \geq 1$, la distribución del vector $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ no depende de los índices i_1, i_2, \dots, i_k sino sólo de k . En cierto sentido, las variables intercambiables generalizan el concepto de variables aleatorias independientes e idénticamente distribuidas.

En el capítulo 1 de este trabajo se establecerán las nociones básicas de probabilidad desde el punto de vista subjetivo, resaltando las diferencias con el tratamiento axiomático de Kolmogorov. Se introducirá una representación lineal para los eventos, que será fundamental en este enfoque.

En el siguiente capítulo se introducirá el concepto de "previsión" que coincide en ciertos casos con la definición usual de esperanza; sin embargo, en este tratamiento se considera como un concepto primario, sobre el cual se define probabilidad, y no viceversa. Se introducen en el mismo capítulo los criterios que definen una previsión coherente y se demuestra el resultado fundamental que establece que siempre es posible extender una probabilidad coherente definida sobre una clase de eventos \mathcal{E} a otra clase cualquiera \mathcal{E}' que contenga a \mathcal{E} . Este resultado le da mucha importancia al enfoque de De Finetti, ya que no existe un resultado similar en otros enfoques. Por otra parte, se estudia a grandes rasgos las implicaciones que tiene el suponer sólo aditividad finita al

construir las funciones de distribución.

El capítulo 3 trata la probabilidad condicional y se comenta la noción de dependencia e independencia estocástica. Ligada a esta discusión, se introduce en el capítulo 4 la noción de intercambiabilidad y se establece cuál es la relación de este concepto con el problema del razonamiento inductivo.

CAPITULO I CONCEPTOS BASICOS DEL ENFOQUE SUBJETIVO

1.1 El concepto de fenómeno aleatorio.

En el enfoque subjetivo de la probabilidad, la propiedad más importante que caracteriza a todos los fenómenos que estudia la probabilidad es la *incertidumbre* que se manifiesta en un observador, en relación a tal o cual fenómeno. La incertidumbre no es una característica exclusiva del objeto, sino también del sujeto que lo estudia. Los fenómenos aleatorios son *inciertos* en el sentido de que se desconoce cuál de todos los posibles resultados ocurrirá. De esta forma, la corriente subjetiva considera que todos los fenómenos que son inciertos son fenómenos aleatorios. Es frecuente encontrar proposiciones lógicas, variables cuantitativas y cualitativas o acontecimientos que son inciertos en relación a un sujeto, pero que no lo son en el sentido objetivo. Por ejemplo, la veracidad de la proposición: "la ecuación $x^2 - 2x + 1 = 0$ tiene como única solución real $x = 1$ " es incierta para algunas personas pero objetivamente verdadera. De la misma manera, el resultado de las últimas elecciones en Estados Unidos es incierto para muchos, pero se trata de un hecho consumado que tuvo un resultado preciso. Aún en la verificación de verdades tautológicas, como en el caso de la determinación de el dígito un billón en la expansión decimal de π , uno puede encontrarse en un estado de incertidumbre.

En resumen, el enfoque subjetivo define un **fenómeno aleatorio** como un fenómeno sobre el que un sujeto particular tiene incertidumbre. Un evento es una afirmación acerca de un fenómeno, y un **evento aleatorio** E se define como una proposición relativa a los posibles resultados de un fenómeno aleatorio, la cual únicamente puede calificarse de falsa o verdadera.

Si Ω representa el conjunto de posibles resultados del fenómeno en consideración, entonces un evento E tiene asociado un subconjunto E_Ω formado por todos aquellos resultados que hacen verdadero a E . Este conjunto tiene a su vez asociada una función indicadora, que se denotará con la misma letra E , y que permite identificar verdadero con 1 y falso con 0.

Es importante resaltar que Ω da lugar a una colección de subconjuntos \mathcal{E} . En el tratamiento matemático de la teoría de la probabilidad, generalmente uno está intere-

sado en asignar probabilidades sobre una clase restringida \mathcal{F} de subconjuntos de Ω , con la propiedad de ser una σ -álgebra. Sin embargo, para el enfoque subjetivo, tal restricción no tiene fundamento lógico, sino exclusivamente práctico y cómodo. Por esta razón, en este enfoque, el campo sobre el cual es posible definir probabilidades puede ser cualquier clase \mathcal{E} de subconjuntos de Ω , sin importar que tal clase sea o no una σ -álgebra. Esto tendrá consecuencias importantes cuando se consideren funciones de distribución.

1.2 Cantidades aleatorias.

Con frecuencia ocurre que los fenómenos aleatorios tienen naturaleza numérica, e incluso en ocasiones los resultados pueden ser codificados con números. Considérense los siguientes ejemplos:

- i. Al lanzar una moneda, el resultado de obtener águila puede representarse con el número 0 y el resultado de obtener sol con el número 1.
- ii. Sea X el año de la muerte de Friederich Gauss. El valor verdadero de X es 1855; en 1834, la fecha de su muerte era un valor desconocido para todos, y cualquier año a partir de ese era posible para X . Después de su muerte, el valor sigue siendo desconocido para algunas personas, y por eso para esas personas, es una cantidad aleatoria. Eventos relacionados con X pueden ser:

E_1 : Gauss murió en un año par.

E_2 : Gauss murió antes de 1900.

E_3 : Gauss murió el mismo año que murió Chopin.

- iii. Sea X = tipo de cambio dólar por marco alemán el día 13 de octubre de 1991. Algunos eventos que podrían ser de interés son:

E_1 : $X \in (0, 3.5)$

E_2 : $X = 0$

E_3 : $X = 0.234$

En este ejemplo es importante observar que el evento E_2 , desde el punto de vista lógico, es imposible. Sin embargo, en la esfera de lo lógicamente posible en la práctica uno agrega un margen que no se define fácilmente, de lo

que es *personalmente posible*, asique un sujeto que sea ignorante sobre lo que pudiera ser un tipo de cambio puede considerar tal evento como posible, aunque lógicamente es imposible. Esto es así porque la lógica no siempre es suficiente para separar con claridad lo que está determinado (es verdadero o falso) de lo que no, como en el caso del evento E_3 .

- iv. Sea X el verdadero valor del parámetro σ^2 de una distribución normal con media $\mu = 3$.

En todos los casos anteriores, se dice que X es una *cantidad aleatoria*, esto es, un resultado numérico ligado a un fenómeno aleatorio. En la teoría subjetiva se utiliza el nombre de *cantidad aleatoria* en lugar del de *variable aleatoria*, para hacer notar que X sólo puede tomar un valor, el cual es desconocido. El nombre de *variable aleatoria*, para De Finetti, proviene de una consideración estadística: X puede tomar cualquier valor en un conjunto dado en cada repetición de un experimento. Para seguir con la exposición de la teoría se adoptará la nomenclatura de De Finetti. El conjunto de posibles valores de X se denotará por $I(X)$, y se dirá que X es acotada si $I(X)$ es un conjunto acotado.

Una cantidad aleatoria da lugar a una colección de eventos. Si X es una cantidad aleatoria que puede tomar valores en $I(X)$, entonces todos los eventos de la forma $E_A = \{X \in A\}$ pueden ser de interés para describir a X .

Una cantidad aleatoria se llama *discreta* si puede representarse como una combinación lineal a lo más numerable de funciones indicadoras de eventos. Se le llamará simple cuando la representación pueda hacerse con un número finito de eventos. Sea X una cantidad aleatoria discreta y $E_x = \{X = x\}$. Entonces $X = \sum_{i=1}^{\infty} x_i E_{x_i}$. Esta combinación lineal puede pensarse como una *ganancia aleatoria* asociada a un cierto juego de azar, donde se gana x_1 si ocurre E_{x_1} , x_2 si ocurre E_{x_2} , y así sucesivamente.

Cada evento E da lugar, por medio de su función indicadora, a una cantidad aleatoria simple. Debido a esta relación, se puede operar con cada E ya sea como proposición, como conjunto o como número, y para tal fin se extenderán las operaciones entre proposiciones a las operaciones aritméticas mediante los siguientes operadores:

Definición 1.2.1

Sean $x, y \in \{0, 1\}$. Se definen las operaciones binarias, máximo (\vee), mínimo (\wedge), y la operación complemento (\sim) de la siguiente manera:

$$x \vee y = \max\{x, y\}$$

$$x \wedge y = \min\{x, y\}$$

$$\bar{x} = \sim x = 1 - x$$

Entonces, si A y B son dos eventos, se tienen las siguientes relaciones considerando los eventos como conjunto, como proposición y como número, respectivamente:

- i. $A \cap B = A \wedge B = AB$
- ii. $A^c = \bar{A} = 1 - A$
- iii. $A \cup B = A \vee B = A + B - AB$

Como puede observarse, las operaciones aritméticas \sim y \wedge son equivalentes a las operaciones lógicas de negación y conjunción. Sin embargo, en general, la suma lógica de dos eventos no corresponde a la suma aritmética de sus funciones indicadoras; la relación válida es

$$(\text{suma lógica}) = 1 \wedge \text{suma aritmética},$$

o en términos de eventos,

$$E_1 \vee E_2 \vee \dots \vee E_n = 1 \wedge (E_1 + E_2 + \dots + E_n).$$

La suma aritmética de los eventos E_1, \dots, E_n puede interpretarse como una cantidad aleatoria $Y_n = \sum_{i=1}^n E_i$ que expresa el número de éxitos que ocurren entre los n eventos. La cantidad aleatoria $\frac{Y_n}{n}$ representa la frecuencia de realizaciones de los n eventos y será una cantidad aleatoria importante que se utilizará más adelante.

1.3 Aseveraciones y particiones.

Un evento aleatorio E , como proposición lógica, sólo puede ser falso o verdadero. Pero este evento, en relación a un sujeto particular (que en adelante se denotará por Z) puede recibir tres connotaciones: *cierto*, *imposible* o *posible*. Si Z tiene certeza en la

veracidad de E , se dice que E es seguro (para \mathcal{Z}); si \mathcal{Z} tiene certeza en la falsedad de E , se dice que el evento es imposible (para \mathcal{Z}), y en el caso de que \mathcal{Z} tiene incertidumbre sobre el valor de verdad de E , se dice que E es posible (para \mathcal{Z}).

Para establecer formalmente la noción de posibilidad de un evento, es necesario diferenciar por una parte, el contenido de la proposición, y por otra, la afirmación que hace \mathcal{Z} respecto al contenido de la proposición. Por ejemplo, en la afirmación "es seguro que mañana lloverá", si E es el evento "mañana lloverá", entonces la afirmación toma la forma "para \mathcal{Z} , E es verdadero", la cual es una aseveración sobre E . Las aseveraciones sobre un evento E se denotarán con el símbolo $\vdash E$, significando con esto que E es cierta para alguien. No se hará mucho uso de ésta notación posteriormente, porque se hará clara en el contexto; aquí sólo se utilizará para remarcar la diferencia.

Las siguientes expresiones son consideradas aseveraciones:

1. Si A y B son dos eventos aleatorios (para \mathcal{Z}), entonces la afirmación de que A implica B es la aseveración de que siempre que ocurra A , ocurrirá B . Como A y B son aleatorios, la implicación en realidad es una hipótesis para \mathcal{Z} . Para denotar esta aseveración se utilizará el símbolo $A \subseteq B$. Obsérvese que la afirmación A implica B puede escribirse también como $A \leq B$, ya que la desigualdad sólo es falsa si $1 \leq 0$, es decir, si $A = 1$ y $B = 0$. Si además $B \subseteq A$, se obtendrá la aseveración de que A es equivalente a B .
2. Dos eventos A y B son *incompatibles* si se asevera que es imposible (para \mathcal{Z}) que ambos ocurran, esto es: $\vdash \overline{AB}$. En general, dada una colección arbitraria de eventos $\{E_i\}_{i \in I}$, los eventos son incompatibles si lo son dos a dos.
3. Dos eventos A y B son *exhaustivos* si se asevera que es imposible que no ocurra ninguno de los dos. En símbolos: $\vdash A + B \geq 1$. Una colección de eventos es exhaustiva si $\vdash \sum_{i \in I} E_i \geq 1$.
4. Una *partición* es la aseveración de que una familia de eventos $\{E_i\}_{i \in I}$ es exhaustiva e incompatible, esto es $\vdash \sum_{i \in I} E_i = 1$.

1.4 Dependencia lógica entre eventos.

Las particiones tienen una importancia fundamental en la teoría de probabilidad, porque a partir de ellas se puede establecer cuál es la *dependencia lógica* entre eventos,

esto es, las relaciones de implicación que existen entre una clase de eventos. Aquí se considerarán sólo familias finitas de eventos.

Dada cualquier colección finita de eventos $\{E_i\}_{i \in I}^n$, siempre es posible construir una partición de la siguiente forma:

- Si los eventos son incompatibles, pero no exhaustivos, se agrega el evento $E_0 = 1 - \sum_{i=1}^n E_i$.
- Si los eventos no son incompatibles, se escribe a la suma como

$$(E_1 + \bar{E}_1)(E_2 + \bar{E}_2) \dots (E_n + \bar{E}_n) = \sum E'_{i_1} E'_{i_2} \dots E'_{i_n} = 1$$

Si se desarrolla el producto, se tendrán 2^n factores de la forma $E'_{i_1} E'_{i_2} \dots E'_{i_n}$, donde E'_{i_j} representa a E_{i_j} o a su complemento. Cada uno de tales factores, que son nuevos eventos, se denotará por C_i y reciben el nombre de *constituyentes*. Puede ser que algunos de tales constituyentes sean imposibles así que quizás algunos de los constituyentes se descarten. Si se eliminan los constituyentes imposibles, quedarán en general $s \leq 2^n$ constituyentes que forman una partición. La partición obtenida, $\{C_i\}_{i=1}^s$, recibe el nombre de *partición generada por $\{E_i\}$* .

Ejemplo:

- i. Si se lanzan dos dados, donde uno es rojo y el otro azul, y se consideran los eventos:

E_1 :la suma de las caras es par.

E_2 :el resultado del dado azul es 2.

E_3 :el resultado del dado rojo es un número impar.

Nótese que el evento E_3 está determinado si se conoce el resultado de E_1 y E_2 , de este modo E_3 es lógicamente dependiente de E_2 y E_1 . Los constituyentes son: $C_1 = E_1 E_2 \bar{E}_3$, $C_2 = \bar{E}_1 E_2 E_3$, $C_3 = E_1 \bar{E}_2 E_3$, $C_4 = E_1 E_2 E_3$, $C_5 = E_1 \bar{E}_2 \bar{E}_3$, $C_6 = \bar{E}_1 E_2 \bar{E}_3$, $C_7 = \bar{E}_1 \bar{E}_2 E_3$, $C_8 = \bar{E}_1 \bar{E}_2 \bar{E}_3$. El único constituyente imposible en este caso es C_4 .

- ii. Sean E_1, E_2, E_3 eventos exhaustivos y compatibles. Estos dan lugar a 7 constituyentes: $C_1 = E_1 E_2 E_3$, $C_2 = \bar{E}_1 E_2 E_3$, $C_3 = E_1 \bar{E}_2 E_3$, $C_4 = E_1 E_2 \bar{E}_3$, $C_5 =$

$$\bar{E}_1 \bar{E}_2 E_3, C_6 = \bar{E}_1 E_2 \bar{E}_3, C_7 = E_1 \bar{E}_2 \bar{E}_3.$$

- iii. Supóngase que se realiza un examen a un grupo de n individuos. Sea E_i el evento de que el i -ésimo candidato apruebe el examen. La partición generada por los E_i tiene 2^n constituyentes, cada uno de los cuales representa el evento de que un subconjunto de candidatos pase el examen. Esto es equivalente a afirmar que los E_i son lógicamente independientes. La cantidad aleatoria $Y = \sum_{i=1}^n E_i$ cuenta el número de candidatos aprobados, y puede tomar cualquier valor en el conjunto $I(Y) = \{0, 1, \dots, n\}$.

Los ejemplos anteriores dan lugar a la siguiente

Definición 1.4.1

n eventos son *lógicamente independientes* si dan lugar a 2^n constituyentes.

Obsérvese que de esta definición se concluye que existe dependencia lógica entre eventos cuando hay una relación de implicación entre algunos de ellos.

Inversamente, si se tienen n eventos E_1, \dots, E_n , y algún constituyente es imposible, eso significa que los eventos son lógicamente dependientes.

Ahora se considerará la dependencia lógica entre un evento E y un conjunto de eventos $\{E_i\}$. Sea $C_1 + C_2 + \dots + C_s = 1$ la partición generada por $\{E_i\}$. Entonces el evento E puede escribirse como:

$$E = EC_1 + EC_2 + \dots + EC_s$$

Cada término EC_i satisface una de las tres relaciones:

- i. $EC_i = C_i$ (C_i está contenido en E),
- ii. $EC_i = 0$ (C_i está contenido en \bar{E}), o bien,
- iii. $0 \subset EC_i \subset C_i$

Si todos los C_i son del primer tipo, E es seguro, ya que la ocurrencia de algún constituyente es segura; si todos los C_i son del segundo tipo, E es imposible, y si los C_i son del tercer tipo, entonces E es posible. Esto es, si los constituyentes del tipo (iii) no existen, significa que E está completamente determinado por los C_i , y por lo tanto, E es lógicamente dependiente de los E_i .

Si todos los constituyentes son del tercer tipo, significa que la ocurrencia de uno de ellos puede o no implicar a E , por lo que E permanece incierto y por lo tanto E es independiente de los E_i .

Por último, en el caso de que existan algunos términos del tipo (iii) pero no todos, el evento E está determinado si ocurre un constituyente del tipo (i) o (ii), y no si ocurre uno del tipo (iii). En este caso se dice que E es lógicamente semindependiente.

Obsérvese que un evento es lógicamente dependiente de los E_i si E se puede expresar como una suma de los constituyentes, ya que no existirán constituyentes del tipo (iii).

Sean $E' = \sum_{i \in I} C_i$ y $E'' = \sum_{i \in J} C_i$ donde I es el conjunto de todos los índices de los constituyentes de los tipos (i) y (iii), y J es el conjunto de todos los índices de los constituyentes del primer y tercer tipos. Es claro, de lo dicho anteriormente que $E' \subset E \subset E''$. El conjunto E' representa el evento maximal contenido en E y E'' es el evento minimal que contiene a E . Estos dos conjuntos serán importantes más adelante, cuando se introduzca la probabilidad y los criterios de coherencia.

También puede establecerse la dependencia e independencia lógica de cantidades aleatorias mediante la siguiente

Definición 1.4.2

Sean X_1, \dots, X_n cantidades aleatorias. Se dice X_1, \dots, X_n son lógicamente independientes si el conjunto de posibles valores del vector (X_1, \dots, X_n) (que se denotará por \mathcal{A}) es el producto cartesiano de los conjuntos $I(X_i)$ de puntos posibles para X_i .

La dependencia lógica de una cantidad aleatoria X de otras tiene el significado de dependencia funcional.

Ejemplo:

- i. Supóngase que un individuo queda descrito por las siguientes cantidades: su peso (X), su edad (Y) y su altura (Z). El conjunto de posibles valores \mathcal{A} del vector (X, Y, Z) no corresponde al producto cartesiano $I(X)I(Y)I(Z)$, ya que un recién nacido, por ejemplo, no puede pesar 40kg. Las cantidades aleatorias

no son, entonces, lógicamente independientes.

- ii. Tres monedas distinguibles se lanzan n veces. Sea X_i = número de soles en la i -ésima moneda. Es claro que el vector (X_1, X_2, X_3) tiene sus puntos posibles en $\{0, 1, \dots, n\}^3$, por lo que las X_i son lógicamente independientes.

1.5 Representación en forma lineal.

Sea \mathcal{X} una colección de cantidades aleatorias acotadas, y sea

$$\mathcal{L} = \left\{ X = u_1 X_1 + u_2 X_2 + \dots + u_n X_n \mid X_i \in \mathcal{X}, u_i \in \mathbb{R}^n \right\}$$

el conjunto de todas las combinaciones lineales finitas de cantidades aleatorias en \mathcal{X} . Este conjunto es el espacio vectorial constituido por las transformaciones lineales de \mathcal{X} en \mathbb{R} .

Por otra parte, el conjunto de posibles puntos de un subconjunto $\{X_1, \dots, X_n\}$ de \mathcal{X} puede pensarse como un subconjunto del espacio vectorial \mathbb{R}^n , al que en este contexto se le llama *ámbito lineal* y se le denota por \mathcal{A}^L . Se sabe que \mathcal{L} y \mathcal{A}^L son espacios isomorfos, y podría pensarse a ambos espacios como superpuestos en \mathbb{R}^n . Sin embargo, tal superposición no es útil ya que se verá más adelante que sólo las nociones afines (es decir, las propiedades no de un espacio vectorial, sino la de un espacio afín, que es como un espacio vectorial trasladado y que no contiene al cero) tendrán significado. Sin embargo, tal superposición será útil en el caso de eventos. En este caso se obtiene una representación común para \mathbb{R}^n y \mathcal{L} . Si E_1, E_2, \dots, E_n son eventos, a los constituyentes generados por éstos se les pone en correspondencia los vectores unitarios de la base canónica $\{e_1, \dots, e_n\}$; cada punto con coordenadas (x_1, x_2, \dots, x_n) , $x_i = 0, 1$, representa un constituyente, y estos a su vez son el conjunto de puntos posibles para \mathcal{E} . Cada constituyente es un vértice de un hipercubo en \mathbb{R}^n . Las cantidades aleatorias $X \in \mathcal{L}$ tienen a lo más s posibles valores distintos, donde s es el número de constituyentes; puede ser que X tome en más de un punto posible el mismo valor, esto ocurre cuando los constituyentes que dan un mismo valor c a X están sobre el plano $\sum c_i x_i = c$.

Ejemplo:

La cantidad aleatoria Y que representa el número de éxitos (es decir, el número de

eventos E_i que ocurren) está en \mathcal{L} , tomando los coeficientes x_i como 1. En este caso, si todos los vértices del hipercubo son posibles, entonces el conjunto $I(Y)$ está distribuido sobre los $n + 1$ hiperplanos $Y = 0, 1, 2, \dots, n$ de acuerdo a los coeficientes binomiales $(1, \binom{n}{1}, \binom{n}{2}, \dots, 1)$ ya que $\binom{n}{k}$ es el número de posibles formas en que pueden darse k sucesos de n eventos. \square

La suma $\sum c_i x_i$ es una función tanto de la cantidad aleatoria $X \in \mathcal{L}$ (i.e. de sus componentes c_i), como de los puntos del ámbito lineal (i.e. de sus coordenadas x_i).

Es conveniente considerar combinaciones lineales afines, i.e. combinaciones $X = \sum c_i x_i$ más una constante adicional, u_0 , suponiendo siempre la existencia de una cantidad aleatoria X_0 que toma el valor 1 con certeza.

Bajo esta representación lineal de los eventos como vectores, tiene sentido hablar de la dependencia lineal entre eventos, en el sentido algebraico usual.

La dependencia lineal es un caso especial de la dependencia lógica, en el sentido de que si $x = \sum u_i x_i$, conociendo los valores x_i , se tiene determinado el valor de x ; por esta razón, la dependencia lineal es una condición más restrictiva que la dependencia lógica. Por ejemplo, la negación \bar{E} depende linealmente de E , ya que $\bar{E} = 1 - E$. La suma de dos eventos $A + B$ depende linealmente de A , de B y de AB : $A \vee B = A + B - AB$.

La independencia lineal se corrobora simplemente calculando el determinante de la matriz que se forma con los coeficientes de las combinaciones lineales que existan entre las cantidades aleatorias.

Ejemplo:

Supongase que A, B, C, D, F, G son participantes en una competencia, y que otros seis individuos eligen de entre los participantes sus tres favoritos para ganar. (Se ofrece un premio a la persona que tenga un ganador entre sus favoritos). Supóngase que también se conoce las elecciones, que son: C, D, G para el primer individuo; B, C, G para el segundo; A, D, F para el tercero; B, F, G para el cuarto; A, C, D para el quinto; D, F, G para el sexto. Supóngase que un séptimo individuo Z participa en las apuestas, y su elección es A, B, C . En este caso, se tiene el siguiente sistema de ecuaciones:

$$\begin{array}{rcl}
1 & = & A + B + C + D + F + G \\
E_1 & = & + + C + D + + G \\
E_2 & = & + B + C + + + G \\
E_3 & = & A + + + D + F + \\
E_4 & = & + B + + + F + G \\
E_5 & = & A + + C + D + + \\
E_6 & = & + + + D + F + G \\
E & = & A + B + C
\end{array}$$

De este sistema se obtiene la relación

$$2E_1 - E_2 + E_3 - 3E_4 - 5E_5 + 5E_6 + 7E = 3(A + B + C + D + E + F + G) = 3$$

de lo cual

$$E = \frac{1}{7}(3 - 2E_1 + E_2 - E_3 + 3E_4 + 5E_5 - 5E_6)$$

□

La importancia de la dependencia lineal en \mathcal{A}^L se reflejará cuando se consideren en la siguiente sección n puntos posibles Q_1, Q_2, \dots, Q_n , y las combinaciones lineales $\sum_{i=1}^n q_i Q_i = P \in \mathcal{A}^L$ convexas, ya que se puede pensar que cada q_i representa una masa asociada a Q_i , y P será entonces el baricentro de los puntos Q_i . Entonces, el casco convexo de los puntos $\{Q_1, \dots, Q_n\}$ contiene todos los baricentros. Este conjunto jugará un papel fundamental en la teoría subjetiva de la probabilidad, porque a partir de este conjunto se establecerá el concepto de esperanza (o como se le llama en la teoría subjetiva, previsión) de una cantidad aleatoria.

El casco convexo variará dependiendo del conjunto de puntos posibles, o bien, de los constituyentes que lo integran. Por ejemplo, sean E_1, E_2, E_3 eventos. a estos corresponden 8 constituyentes. Si todos son posibles, entonces $\mathcal{A} = \{C_1, \dots, C_8\}$ y el casco convexo es el cubo completo. Por otra parte, si los eventos forman una partición, entonces $\mathcal{A} = \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$ y el casco convexo es un simplex con vértices e_1, e_2 y e_3 .

Los constituyentes no pueden representarse en el espacio \mathcal{L} porque cada constituyente es un producto de eventos que puede no ser linealmente dependiente del conjunto de eventos. Si fuera necesario estudiar en \mathcal{L} alguno de tales constituyentes, entonces sería necesario agregar a \mathcal{A}^L una dimensión o las necesarias para tal o tales constituyentes, y por lo tanto también cambiará el casco convexo. La conclusión importante de esto

es que todo puede representarse linealmente si se toma el número suficiente de dimensiones. Por ejemplo, \mathcal{A}^L puede ser \mathbf{R}^3 , y si las cantidades aleatorias están relacionadas por la ecuación $X^2 + Y^2 + Z^2 = r^2$, entonces el espacio de alternativas será una esfera de radio r . A veces, será necesario incluir la relación $X^2 + Y^2 + Z^2$ como una variable más, en cuyo caso se pensará que \mathcal{A}^L es \mathbf{R}^4 , con los puntos (x, y, z, r^2) .

Ejemplo:

- i. Sea Y el número de personas que pasarán un examen determinado. Entonces $Y = E_1 + \dots + E_n$ donde E_i es el evento "el i -ésimo candidato pase el examen". Y puede tomar los valores $0, 1, 2, \dots, n$ solamente si los eventos E_i son lógicamente independientes. Esto significa que el conjunto de todos aquellos que pasan el examen puede ser cualquiera de los 2^n subconjuntos de candidatos.
- ii. (*continuación*). Un ejemplo donde los $n + 1$ valores son posibles para Y a pesar de que los eventos no son lógicamente independientes es el siguiente. Si E_i es el evento "la persona que tuvo el lugar i -ésimo de una competencia alcanzó un puntaje p ". Es posible que cada competidor ocupe un lugar i , o bien, que no ocurra ninguno, o bien, algunos; si éste es el caso, entonces los primeros i son éxitos y los otros son fracasos. No hay independencia lógica ya que si E_i es cierto, son ciertos los precedentes, y si es falso, los siguientes son falsos.
- iii. Supóngase que habrá elecciones en Estados Unidos y que las cantidades aleatorias X, Y, Z representan, respectivamente, el número de votos a favor, en contra y abstenciones para el candidato George Bush, con $X + Y + Z = n$. En este caso, las cantidades aleatorias son linealmente (y por lo tanto lógicamente) dependientes.

2.1 Introducción.

Una vez determinadas las relaciones lógicas entre los eventos, el siguiente paso es asignar probabilidades, que desde el punto de vista bajo consideración, corresponde a un ingrediente extralógico y subjetivo, es decir, depende exclusivamente del juicio particular de un sujeto que actúa coherentemente, esto es, que no está dispuesto a sufrir las consecuencias de una mala decisión.

A continuación se definirá, en términos de dos criterios operativos equivalentes, el concepto de *previsión*. La *previsión* puede considerarse como *esperanza*, pero definida directamente como una cantidad primitiva, más que en términos de una distribución de probabilidad preestablecida. Desde el punto de vista subjetivo no hay distinción entre *probabilidad* y *esperanza*: la *probabilidad* de un evento E es la *previsión* de su función indicadora.

El dominio de definición de la *previsión* será cualquier conjunto o clase \mathcal{X} de cantidades aleatorias (y en particular eventos). Así, la *previsión* puede pensarse como una función lineal y aditiva $P : \mathcal{X} \rightarrow \mathbb{R}$ (en particular, si \mathcal{X} consta de eventos se le denota por \mathcal{E} y entonces $P : \mathcal{E} \rightarrow \mathbb{R}$ es una función de probabilidad). Un punto importante es que, en el caso de una probabilidad, no hay restricción sobre el dominio \mathcal{E} de la función, pudiendo tener la estructura y el tamaño conveniente para cada aplicación.

Primero se definirá en términos generales el concepto de *previsión* para una clase \mathcal{X} de cantidades aleatorias acotadas. Se darán dos criterios operacionales equivalentes, lo que se demostrará por medio de un argumento geométrico. A continuación se particulariza al caso de una clase finita de eventos para definir probabilidad obteniendo el esquema clásico de probabilidad. Posteriormente se considerarán clases infinitas de eventos y cantidades aleatorias con un número numerable de posibles valores, eliminando la restricción de acotamiento. Después se considerará el concepto de distribución para tratar el caso más general de cantidades aleatorias. Aquí se verán cuáles son las diferencias conceptuales con la definición usual.

En esta parte también se considera el siguiente problema, que resulta básico en cualquier enfoque de la teoría de la probabilidad. Dada una familia \mathcal{X} de cantidades aleatorias sobre la que ya se ha definido una función de *previsión* P , ¿es posible extender la función P coherentemente a un conjunto \mathcal{X}' mayor que \mathcal{X} ? Si tal extensión existe, ¿bajo que condiciones está unívocamente determinada?

2.2 Fuentes subjetivas de la *previsión*.

En términos intuitivos, la *previsión* $P(X)$ de una cantidad aleatoria X corresponde

al precio "justo" que un sujeto \mathcal{Z} asigna por la obtención de una ganancia aleatoria X . Tal precio tiene ciertas propiedades que lo hacen "justo": el precio no puede ser menor al valor más pequeño que puede tomar X , ni mayor que el valor más grande de X , esto es,

$$\inf X \leq P(X) \leq \sup X$$

Por otra parte, cuando se supone indiferencia entre los siguientes esquemas:

1. Pagar un precio $P(X)$ por una ganancia incierta X y un precio $P(Y)$ por una ganancia incierta Y , y
2. pagar un precio $P(X + Y)$ por la ganancia incierta $X + Y$,

uno está suponiendo implícitamente que $P(X) + P(Y) = P(X + Y)$. Esta suposición de linealidad, implícita en el hecho de considerar equivalentes dos riesgos distintos, supone una actitud de rigidez al enfrentar el riesgo. Esta noción no es realista desde el punto de vista económico: si un sujeto compra un bien A a un precio $P(A)$ y un bien B a un precio $P(B)$, puede no estar dispuesto a comprar ambos bienes al precio $P(A) + P(B)$, porque la utilidad asociada a un bien puede reducirse con la adquisición del otro bien, como en el caso de bienes sustitutos, por ejemplo. Sin embargo, si se introduce la escala de utilidad, tal rigidez desaparecerá. Este desarrollo alternativo no se seguirá aquí; sólo se supondrá que la función P es una función lineal.

El precio podría asignarse arbitrariamente, pero si se desea actuar con cuidado para que tal asignación no lleve a consecuencias indeseables, es conveniente basar la asignación en criterios coherentes, esto es, criterios que impidan asignaciones que lleven a resultados indeseables. Las propiedades mencionadas anteriormente son necesarias para que una asignación sea coherente. La idea fundamental de los criterios de coherencia es castigar proporcionalmente a la diferencia $X - P(X)$, esto es, la diferencia que puede haber entre el valor que toma la ganancia aleatoria X y el valor $P(X)$ asignado por \mathcal{Z} . Pareciera que la intención del criterio de coherencia fuera obligar a \mathcal{Z} a "adivinar" el valor de X ; sin embargo, quedará claro más adelante que no se trata de esto, porque una predicción obligaría a elegir a $P(X)$ como un valor posible de X , y una previsión no necesariamente es uno de tales valores posibles.

2.3 Criterios de coherencia.

Sea \mathcal{X} un conjunto arbitrario de cantidades aleatorias acotadas.

1er. criterio.

Sea $\{X_1, X_2, \dots, X_n\} \subset \mathcal{X}$. Un conjunto de previsiones $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ es coherente si no existe combinación lineal

$$Y = \sum_{i=1}^n c_i (X_i - \bar{x}_i)$$

de tal forma que $\sup Y < 0$ (esto implica que $\inf Y$ no puede ser positivo, ya que $\sup(-Y) = -\inf Y$ sería negativo).

En términos intuitivos, el primer criterio permite a \mathcal{Z} asignar cantidades ciertas $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ a las ganancias inciertas $\{X_1, X_2, \dots, X_n\}$ de tal forma que una vez hecha la asignación, \mathcal{Z} estará dispuesto a aceptar cualquier apuesta o combinación de apuestas de monto $Y = \sum_{i=1}^n c_i(X_i - \bar{x}_i)$, y tal apuesta no resultará *siempre* en una pérdida (i.e. no será pérdida para todo posible valor de X_i).

Se han hecho diversas críticas a este esquema de coherencia. Otras posturas señalan que un criterio de coherencia razonable debería excluir a todas las apuestas en las que no se gana, esto es, aquellas en donde $\sup Y \leq 0$ (o equivalentemente, aceptar sólo las apuestas con $\sup Y > 0$). Esta posición se conoce como *coherencia estricta*. Sin embargo, con esta posición, ciertos problemas no tendrían solución, como el que se verá a continuación. En este problema se hablará de probabilidad a la "manera" clásica, aunque no es posible, en ese enfoque, decir que el extraer un natural al azar constituye un experimento aleatorio con eventos equiprobables. Sin embargo, en el enfoque subjetivo se desea incluir este tipo de problemas que el modelo clásico no incluye.

Ejemplo:

Sea X una ganancia aleatoria con valores en el conjunto $\{-1, -\frac{1}{2}, -\frac{1}{3}, \dots\}$ (i.e. X representa una pérdida). Por ejemplo, puede ser la pérdida asociada al siguiente juego: se extrae un entero positivo al azar: si sale el 1, pierdo 1, si sale 2, pierdo $\frac{1}{2}$, y así sucesivamente. ¿Cuál será una previsión coherente para este problema? Una asignación coherente, según el criterio de De Finetti, es aquél valor de \bar{x} tal que $\sup\{c(X - \bar{x})\} \geq 0 \quad \forall c \in \mathbb{R}$.

Si $c \geq 0$, entonces $\sup\{c(X - \bar{x})\} = c \sup\{X - \bar{x}\}$. Puesto que $\sup\{X - \bar{x}\} = -\bar{x}$, se tiene que $\sup Y \geq 0 \iff -c\bar{x} \geq 0 \iff \bar{x} \leq 0$.

Por otra parte, si $c < 0$, entonces $\sup\{c(X - \bar{x})\} = \inf\{X - \bar{x}\} = -c(1 + \bar{x})$, ya que el ínfimo de $X - \bar{x}$ es $-1 - \bar{x}$. Por lo tanto, la condición se cumple en este caso si $\bar{x} \geq -1$.

De lo anterior, una asignación \bar{x} es coherente de acuerdo al criterio de De Finetti si $\bar{x} \in [-1, 0]$. Esto significa que para que el juego le sea justo a \mathcal{Z} , se le debe pagar una cantidad en el intervalo $[0, 1]$, que compense la pérdida segura que puede tener. Con cálculos similares, bajo el supuesto de coherencia estricta, las asignaciones coherentes son aquellas que están contenidas en el intervalo $(0, 1)$. Por otro lado, sin embargo, obsérvese que $P(X \geq -\frac{1}{n}) = 1$, esto es, la pérdida con probabilidad 1 es tan pequeña como se quiera. De este modo, la única asignación "coherente" (no en el sentido de De Finetti, sino en el sentido de "razonable") es $X = 0$, ya que implica que el jugador siempre gana, lo cuál no es coherente bajo el supuesto de coherencia estricta. Este

problema, bajo tal supuesto, no tendría solución. \square

El criterio de coherencia de De Finetti es lo más débil posible para garantizar que en cualquier problema siempre existirá una asignación coherente; dicho de una manera más formal, dada una clase arbitraria de cantidades aleatorias acotadas siempre exista una previsión coherente.

Otras equivalencias inmediatas del primer criterio son las siguientes: un conjunto $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ es coherente si se cumple cualquiera de las siguientes condiciones:

1. $\exists (c_1, c_2, \dots, c_n) \in \mathbf{R}^n$ tal que $\sup\{\sum_{i=1}^n c_i(X_i - \bar{x}_i)\} < 0$;
2. $\forall (c_1, c_2, \dots, c_n) \in \mathbf{R}^n, \exists \gamma > 0$ tal que $\sup\{\sum_{i=1}^n c_i(X_i - \bar{x}_i)\} < -\gamma$;
3. $\forall (c_1, c_2, \dots, c_n), \sup\{\sum_{i=1}^n c_i(X_i - \bar{x}_i)\} \geq 0$;
4. $\forall (c_1, c_2, \dots, c_n), \inf\{\sum_{i=1}^n c_i(X_i - \bar{x}_i)\} \leq 0 \leq \sup\{\sum_{i=1}^n c_i(X_i - \bar{x}_i)\}$.

Otra equivalencia, no tan directa, se establece en el siguiente:

Teorema 2.3.1

Una asignación de previsiones es coherente si y sólo si cada relación lineal

$$\sum_{i=1}^n c_i X_i \geq c \quad \text{con certeza,}$$

se satisface por las respectivas previsiones:

$$\sum_{i=1}^n c_i \bar{x}_i \geq c.$$

Además, si la asignación es coherente, entonces $\sum_{i=1}^n c_i X_i = c$ con certeza implica que $\sum_{i=1}^n c_i \bar{x}_i = c$.

Demostración.

\Rightarrow) Supóngase que $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ es coherente, que $\sum c_i X_i \geq c$ con certeza y que $\sum c_i \bar{x}_i < c$. Entonces, como $\inf\{\sum c_i X_i\} \geq c$, se tiene

$$\inf Y = \inf\{\sum c_i X_i - \sum c_i \bar{x}_i\} = \inf\{\sum c_i X_i\} - \sum c_i \bar{x}_i > 0$$

lo cual contradice que la asignación sea coherente.

\Leftarrow) Ahora supóngase que se cumple la condición, y que la asignación no es coherente. Entonces $\exists (c_1, \dots, c_n)$ tal que $\inf\{\sum c_i(X_i - \bar{x}_i)\} > 0$. Como $\inf\{\sum c_i(X_i - \bar{x}_i)\} = \inf\{\sum c_i X_i\} - \sum c_i \bar{x}_i$, entonces $\inf\{\sum c_i X_i\} > \sum c_i \bar{x}_i$. Sea c tal que $\inf\{\sum c_i X_i\} > c > \sum c_i \bar{x}_i$; entonces

$$\sum c_i X_i \geq c \quad \text{con certeza,}$$

pero $\sum c_i \bar{x}_i < c$, lo cual contradice el supuesto inicial.

Por último, supóngase que $\{\bar{x}_i\}$ constituye una asignación coherente y que $\sum c_i X_i = c$ y que $\sum c_i \bar{x}_i > c$. Entonces $Y = c - \sum c_i \bar{x}_i < 0$, lo que contradice que la asignación sea coherente. De manera similar, si $\sum c_i \bar{x}_i < c$, se tendría que $Y > 0$, que también contradice el supuesto de coherencia. Por lo tanto, $\sum c_i \bar{x}_i = c$. ■

2o. criterio.

Sea $\{X_1, X_2, \dots, X_n\} \subset \mathcal{X}$. Un conjunto de previsiones $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ es coherente si no existen valores x_i^* tales que

$$L^* = \sum_{i=1}^n (X_i - x_i^*)^2$$

sea menor que

$$L = \sum_{i=1}^n (X_i - \bar{x}_i)^2$$

para cualquier punto posible (X_1, X_2, \dots, X_n) , es decir, con certeza.

Equivalencia de los criterios.

Para demostrar que ambos criterios son equivalentes, se considerará la siguiente interpretación geométrica. Cualquier previsión en el ámbito lineal \mathcal{A}^L correspondiente a n cantidades aleatorias X_1, X_2, \dots, X_n consiste en un punto $P = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$.

Con el primer criterio, una condición necesaria y suficiente para que una asignación sea coherente es que cada relación lineal entre las cantidades aleatorias:

$$\sum_{i=1}^n c_i X_i = c \quad (\geq c) \text{ con certeza}$$

debe satisfacerse por las respectivas previsiones:

$$\sum_{i=1}^n c_i \bar{x}_i = c \quad (\geq c).$$

Este resultado interpretado geoméricamente implica que un punto P es una previsión coherente si y sólo si no existe ningún hiperplano que lo separe del conjunto de puntos posibles (esta propiedad caracteriza el casco convexo de un conjunto: ver apéndice). Para el segundo criterio, el punto P es un punto tal que la distancia de él a cualquier punto posible en \mathcal{A}^L es mínima. De ambas caracterizaciones, el conjunto de previsiones coherentes, que se denotará con \mathcal{P} , resulta ser el casco convexo del conjunto

de puntos posibles, esto es, el conjunto de combinaciones lineales $\sum_{i=1}^l p_i x_i$, donde cada x_i es un posible valor de X_i y $\sum_{i=1}^l p_i = 1, p_i \geq 0$.

Ejemplo:

Supongase un concurso con n participantes en donde se quiere medir la popularidad de un sujeto A . Sean

- a. $X =$ votos a favor de A
- b. $Y =$ votos en contra de A
- c. $Z =$ abstenciones
- d. $U = X - Y$
- e. $V = X/Y$

Debido a la linealidad de P , se debe cumplir, para las respectivas previsiones de X, Y, Z, U que $\bar{x} + \bar{y} + \bar{z} = n, \bar{u} = \bar{x} - \bar{y}$. Esto es así porque el conjunto de puntos posibles para (X, Y, Z) — respectivamente (X, Y, U) — es el conjunto $\{(x, y, z) | x + y + z = n, x, y, z \in \mathbb{N}\}$ — respectivamente $\{(x, y, u) | x + y \leq n, u = x - y, x, y \in \mathbb{N}\}$ — cuyo casco convexo es el conjunto $\{(x, y, z) | x + y + z = n\}$ — respectivamente $\{(x, y, u) | x + y \leq n, u = x - y\}$ —. Sin embargo, la previsión de V no necesariamente coincide con $\frac{\bar{x}}{\bar{y}}$, ya que el conjunto de puntos posibles para (X, Y, V) es un paraboloides hiperbólico, y el casco convexo de este conjunto contiene más puntos que los que se encuentran sobre el paraboloides. \square

El ejemplo anterior ilustra cómo las previsiones coherentes preservan la dependencia lineal, y de hecho esta es la única situación en que necesariamente se preserva por ser P aditiva. En otro caso esto puede no ocurrir porque el baricentro de masas de una superficie no necesariamente pertenece a la superficie.

El concepto de coherencia puede extenderse a toda la clase de cantidades aleatorias acotadas, en términos de cualquiera de los criterios; aquí se hará la extensión con respecto al primer criterio.

Definición 2.3.1

Una función $P: \mathcal{X} \rightarrow \mathbb{R}$ es una previsión coherente sobre \mathcal{X} , si para cualquier subconjunto $\{X_{i_1}, X_{i_2}, \dots, X_{i_n}\}$ de \mathcal{X} y $\forall (c_1, c_2, \dots, c_n) \in \mathbb{R}^n, n \in \mathbb{N}$, la ganancia

$$Y = \sum_{k=1}^n c_k \{P(X_{i_k}) - X_{i_k}\}$$

es tal que $\inf Y \leq 0$.

En particular, si \mathcal{X} es una familia de eventos, a la función P se le llama *función de probabilidad*.

El problema que se tratará ahora consiste en asegurar la existencia de una previsión en cualquier clase de cantidades aleatorias acotadas. Se necesitará la siguiente

Definición 2.3.2

Sea \mathcal{X} un conjunto de cantidades aleatorias y P una previsión sobre ésta. Si existe una previsión P' sobre $\mathcal{X}' \supset \mathcal{X}$ tal que $P'(X) = P(X) \quad \forall X \in \mathcal{X}$, entonces se dice que P' es una extensión de P a \mathcal{X}' .

Lema 2.3.1

Una función de previsión coherente P , definida sobre un conjunto arbitrario dado de cantidades aleatorias \mathcal{X} puede extenderse, preservando coherencia, a cualquier otra cantidad aleatoria X_0 .

Demostración.

Sean $a_1, a_2, \dots, a_k \in \mathbb{R}$, $b_1, b_2, \dots, b_m \in \mathbb{R}$ arbitrarios y sean $X_1, X_2, \dots, X_k \in \mathcal{X}$, $Y_1, Y_2, \dots, Y_m \in \mathcal{X}$ arbitrarios, entonces

$$\inf\{X_0 + \sum_{j=1}^k a_j(X_j - P(X_j))\} \leq \sup\{X_0 - \sum_{j=1}^m b_j(Y_j - P(Y_j))\}$$

en donde el inf y el sup son tomados sobre todos los posibles valores de las X_j y Y_j respectivamente. En efecto, supóngase que

$$\inf\{X_0 + \sum_{j=1}^k a_j(X_j - P(X_j))\} > \sup\{X_0 - \sum_{j=1}^m b_j(Y_j - P(Y_j))\},$$

entonces, tomando la diferencia, se tendría

$$\inf\{\sum_{j=1}^k a_j(X_j - P(X_j)) + \sum_{j=1}^m b_j(Y_j - P(Y_j))\} > 0$$

pero esto no es posible ya que P es una previsión coherente. Se definen

$$x' = \sup\{\inf\{X_0 + \sum_{j=1}^k a_j(X_j - P(X_j))\} \mid k \in \mathbb{N}; a_1, a_2, \dots, a_k \in \mathbb{R}; X_1, X_2, \dots, X_k \in \mathcal{X}\}$$

$$x'' = \inf\{\sup\{X_0 - \sum_{j=1}^m b_j(Y_j - P(Y_j))\} | m \in \mathbb{N}; b_1, b_2, \dots, b_m \in \mathbb{R}; Y_1, Y_2, \dots, Y_m \in \mathcal{X}\}$$

Por la desigualdad anterior se tiene que $x' \leq x''$. Se asigna entonces a X_0 cualquier valor $P(X_0)$ entre x' y x'' . Esta extensión de P resulta ser una asignación coherente. En efecto, supóngase que no lo es; entonces existen $X_1, X_2, \dots, X_n \in \mathcal{X}$ y $c_0, c_1, \dots, c_n \in \mathbb{R}$ con $c_0 \neq 0$ tales que

$$\inf\{c_0(X_0 - P(X_0)) + \sum_{j=1}^n c_j(X_j - P(X_j))\} > 0$$

Sin pérdida de generalidad se puede suponer que $c_0 = \pm 1$. Si $c_0 = 1$ se tiene

$$P(X_0) < \inf\{X_0 + \sum_{j=1}^n c_j(X_j - P(X_j))\} \leq x'$$

Si $c_0 = -1$ se tiene

$$\begin{aligned} P(X_0) &> -\inf\{\sum_{j=1}^n c_j(X_j - P(X_j)) - X_0\} \\ &= \sup\{X_0 - \sum_{j=1}^n c_j(X_j - P(X_j))\} \geq x'' \end{aligned}$$

Lo cual contradice el que $P(X_0)$ se eligió en el intervalo $[x', x'']$ ■

Teorema 2.3.2

Sea \mathcal{X} una familia arbitraria de cantidades aleatorias acotadas, entonces existe una previsión coherente sobre \mathcal{X} .

Demostración.

Para probar este resultado se introducirá un orden parcial sobre cierto conjunto. se mostrará entonces que el conjunto es inductivamente ordenado y se aplicará el lema de Zorn.

Sea \mathcal{X} una colección arbitraria de cantidades aleatorias acotadas y sea

$$\mathcal{N} = \{P_j : \mathcal{X}_j \rightarrow \mathbb{R} | P_j \text{ es una previsión coherente y } \mathcal{X}_j \subset \mathcal{X}\}$$

El conjunto \mathcal{N} es no vacío, pues para cualquier familia \mathcal{X}_j con un sólo elemento X se puede definir una previsión coherente $P(X)$, a saber cualquier valor $P(X)$ que satisfaga $\inf X \leq P(X) \leq \sup X$.

Se define la siguiente relación sobre \mathcal{N} : $P_i \leq P_j \iff \mathcal{X}_i \subseteq \mathcal{X}_j$ y P_j es una extensión de P_i . Esta relación es un orden parcial sobre \mathcal{N} pues la relación \subseteq es un orden parcial en la clase de conjuntos y la relación de extensión también constituye un orden parcial. Ahora se demostrará que con este orden parcial \mathcal{N} es un conjunto inductivamente ordenado.

Sea \mathcal{H}_0 una cadena en \mathcal{N} . Lo que hay que demostrar es que esta cadena está acotada superiormente. Sea

$$\mathcal{X}^* = \bigcup_{\mathcal{X}_\alpha \in \mathcal{H}_0} \mathcal{X}_\alpha$$

y sea $P^* : \rightarrow \mathbb{R}$ definida del siguiente modo: si $X \in \mathcal{X}^*$, entonces existe al menos un α tal que $X \in \mathcal{X}_\alpha$; se define entonces $P^*(X) = P_\alpha(X)$. Para ver que P^* está bien definida, supóngase que $X \in \mathcal{X}_\alpha$ y $X \in \mathcal{X}_\beta$. Como \mathcal{H}_0 es una cadena, resulta que P_α es una extensión de P_β o viceversa, así que en cualquiera de los dos casos $P_\alpha = P_\beta$ sobre $\mathcal{X}_\alpha \cap \mathcal{X}_\beta$, por lo que $P^*(X)$ tiene un único valor. Por otra parte P^* es coherente. En efecto, para cualquier conjunto finito $\mathcal{X}' = \{X_1, X_2, \dots, X_n\} \subset \mathcal{X}^*$ existe una previsión coherente $P = \max\{P_j | \mathcal{X}' \cap \mathcal{X}_j \neq \emptyset\}$ y $P^*(X_i) = P(X_i) \quad \forall X_i \in \mathcal{X}'$. Por lo tanto, $P^* \in \mathcal{N}$ y es claro que P^* , por la forma que se definió, es una cota superior de la cadena.

Por el lema de Zorn, \mathcal{N} tiene un elemento maximal P . Ahora se demostrará que el dominio de P que se denotará por D_P tiene que ser \mathcal{X} . Supóngase que no es así, es decir existe $X_0 \in \mathcal{X} - D_P$. Pero el lema demuestra que existe $P' : D_P \cup \{X_0\} \rightarrow \mathbb{R}$ que es una extensión coherente de P . Así que $P' \in \mathcal{N}$ y esto contradice la maximalidad de P . Por lo tanto, no puede existir X_0 y $D_P = \mathcal{X}$. ■

Siguiendo el método utilizado para demostrar el resultado anterior, se puede demostrar el siguiente

Teorema 2.3.3

Sea \mathcal{X}_0 una clase arbitraria sobre la que está definida una previsión coherente P_0 y sea $\mathcal{X}'_1 \supset \mathcal{X}_0$. Entonces P_0 puede extenderse coherentemente a \mathcal{X}'_1 .

Si en los resultados anteriores se agrega la restricción de que las cantidades aleatorias sean eventos, se obtiene el siguiente

Teorema 2.3.4 Teorema fundamental de la probabilidad (subjativa).

Dada una clase arbitraria de eventos \mathcal{E} y una probabilidad definida sobre esa clase, siempre es posible extender coherentemente las probabilidades a una clase mayor \mathcal{E}' que contenga a \mathcal{E} . En particular, Dadas las probabilidades $P(E_i) \quad (i = 1, \dots, n)$ de un número finito de eventos, la probabilidad de un evento adicional E satisface una de las siguientes afirmaciones:

- a. Está determinada completamente por los E_i si E depende linealmente de ellos,
o
- b. Puede ser asignado coherentemente cualquier valor en un intervalo cerrado $[p', p'']$

Los resultados anteriores marcan una diferencia fundamental con el desarrollo axiomático de Kolmogorov. Las probabilidades coherentes no sufren de las restricciones que afectan el establecimiento de una medida del probabilidad como, por ejemplo, la imposibilidad de definir una medida de probabilidad sin átomos (es decir, que todos los puntos tengan probabilidad cero) sobre el conjunto potencia de un conjunto infinito no numerable (un resultado obtenido por S. Ulam). Esta característica del enfoque subjetivo de De Finetti da fuerza a su punto de vista.

2.4 Probabilidades de eventos (Caso finito).

Las propiedades de las probabilidades de eventos son simplemente casos particulares de las propiedades de las previsiones de cantidades aleatorias. Se establecerán las propiedades y se ilustrará su significado en la representación que se ha introducido. Se demostrarán los resultados para el caso de una colección finita de eventos.

Obsérvese que si E es un evento y $P(E)$ es una previsión coherente de E , se debe tener $\inf E \leq P(E) \leq \sup E$, así que $0 \leq P(E) \leq 1$

Teorema 2.4.1 (probabilidad total)

En una clase finita exhaustiva de eventos incompatibles, la suma de las probabilidades debe ser igual a 1.

Demostración.

Sea $\{E_i\}_{i=1}^n$ la clase de eventos incompatibles y sean $\{p_i\}_{i=1}^n$ el conjunto de probabilidades coherentes evaluadas por un individuo Z .

Sean $S_1, S_2, \dots, S_n \in \mathbf{R}$ y consideremos la cantidad aleatoria $Y = \sum_{i=1}^n S_i(E_i - p_i)$. Los posibles valores de Y son de la forma

$$Y_j = S_j - \sum_{i=1}^n p_i S_i \quad j = 1, \dots, n,$$

donde cada Y_j representa la ganancia que se obtendría si ocurre el evento E_j .

Si se consideran las S_j como desconocidas, se obtiene el sistema de ecuaciones:

$$\begin{aligned}
 Y_1 &= (1 - p_1)S_1 - p_2S_2 - \dots - p_nS_n \\
 Y_2 &= -p_1S_1 + (1 - p_2)S_2 - \dots - p_nS_n \\
 &\vdots \\
 Y_n &= -p_1S_1 - p_2S_2 - \dots + (1 - p_n)S_n
 \end{aligned}$$

que tiene como discriminante

$$\Delta = \det \begin{pmatrix} 1 - p_1 & -p_2 & \dots & -p_n \\ -p_1 & 1 - p_2 & \dots & -p_n \\ \vdots & \vdots & \ddots & \vdots \\ -p_1 & -p_2 & \dots & (1 - p_n) \end{pmatrix} = 1 - (p_1 + p_2 + \dots + p_n)$$

Si $\Delta \neq 0$, las apuestas S_n pueden fijarse de tal forma que las Y_n tengan valores arbitrarios, en particular, todas negativas, lo cual resultaría en que la asignación sería incoherente. Por lo tanto, se debe tener que $\Delta = 0$, esto es,

$$p_1 + p_2 + \dots + p_n = 1$$

Teorema 2.4.2 (caso de eventos incompatibles).

Si se tiene una colección finita de eventos incompatibles, la probabilidad de la suma de los eventos es igual a la suma de las probabilidades.

Demostración.

Esto es consecuencia directa de la linealidad de la previsión. ■

Teorema 2.4.3 (eventos complementarios).

Las probabilidades de eventos complementarios deben ser a su vez complementarias.

Demostración.

Se puede escribir $1 = E + \bar{E}$ y como $P(1) = P(E + \bar{E}) = P(E) + P(\bar{E})$, se obtiene el resultado. ■

Las afirmaciones anteriores dicen cómo debe evaluar \mathcal{Z} sus probabilidades en el caso de eventos incompatibles y excluyentes; estas afirmaciones imponen necesariamente condiciones de coherencia. Si a estas afirmaciones se agrega la restricción de no negatividad, se demostrará a continuación que las afirmaciones serán también *suficientes*

para la coherencia, es decir, cualquier asignación que las satisfaga será coherente, sin importar cuál sea. Es decir, se tiene el siguiente

Teorema 2.4.4

Un conjunto de probabilidades $P(E_i) = p_i$ son coherentes si y sólo si $\sum_{i=1}^n p_i = 1$ y $p_i \geq 0 \quad i = 1, 2, \dots, n$

Demostración.

Sólo es necesario demostrar la suficiencia ya que los resultados anteriores muestran la necesidad. Sea $\{E_i\}_{i=1}^n$ una colección de eventos tales que $P(E_i) = p_i > 0$ y $\sum_{i=1}^n p_i = 1$. Supóngase ahora que tal asignación es incoherente. Entonces es posible forzar a \mathcal{Z} a apostar con una pérdida segura, es decir existen valores c_i de tal forma que

$$X = \sum_{i=1}^n c_i(E_i - p_i) < 0 \quad \text{con certeza,}$$

esto es

$$\sum_{i=1}^n c_i E_i < \sum_{i=1}^n c_i p_i \quad \text{con certeza.}$$

Esto implica $c_i < \sum_{i=1}^n c_i p_i \quad \forall i$. Pero si $c = \max_i \{c_i\}$ entonces $c_i \leq c \quad \forall i$, así que $\sum c_i p_i \leq c \sum p_i = c$, lo cual es una contradicción. ■

Con los resultados obtenidos hasta este punto, se ha reconstruido, desde el punto de vista subjetivo, el enfoque clásico de la probabilidad.

Teorema 2.4.5 (caso de eventos compatibles).

La probabilidad de la suma de los eventos debe ser menor o igual a la suma de las probabilidades.

Demostración.

Sea $E = E_1 \vee E_2 \vee \dots \vee E_n$, entonces $E = 1 \wedge (E_1 + E_2 + \dots + E_n) \leq E_1 + E_2 + \dots + E_n$, y por lo tanto, $P(E) \leq P(E_1) + P(E_2) + \dots + P(E_n)$. ■

En términos de constituyentes, si $C = \bar{E}_1 \bar{E}_2 \dots \bar{E}_n$, entonces $E = \bar{C}$, y $P(E) = 1 - P(C)$. Si se desarrolla C en productos de los E_i , se obtiene que

$$E = \sum_i E_i - \sum_{ij} E_i E_j + \sum_{ijk} E_i E_j E_k - \dots \pm E_1 E_2 \dots E_n$$

y por lo tanto se tiene el siguiente resultado:

Teorema 2.4.6

La probabilidad de la suma de los eventos es

$$P(E) = \sum_i P(E_i) - \sum_{ij} P(E_i E_j) + \sum_{ijk} P(E_i E_j E_k) - \dots \pm P(E_1 E_2 \dots E_n).$$

2.4.1 Representaciones en forma lineal.

A continuación se introducirá la representación por medio de los espacios \mathcal{A} y \mathcal{L} . Para facilitar la exposición, se considerará el espacio de dimensión 3; los ejemplos en dimensiones superiores son análogos.

Sean E_1, E_2, E_3 eventos. Los constituyentes que forman el conjunto de puntos posibles pueden ser como puntos Q_i de $\mathbb{R}^3 = \mathcal{A}^L$:

$$Q_0 = (0, 0, 0); Q_1 = (1, 0, 0); Q_2 = (0, 1, 0); Q_3 = (0, 0, 1)$$

$$Q'_0 = (1, 1, 1); Q'_1 = (0, 1, 1); Q'_2 = (1, 0, 1); Q'_3 = (1, 1, 0)$$

Los mismos puntos, pero considerados como puntos o vectores de \mathcal{L} son:

$$0, E_1, E_2, E_3$$

$$E_1 + E_2 + E_3, E_2 + E_3, E_1 + E_3, E_1 + E_2$$

Un punto $(x, y, z) \in \mathcal{A}^L$ puede ser pensado como un punto previsión, tal que $P(E_1) = x, P(E_2) = y, P(E_3) = z$, esto es, (x, y, z) representa una previsión para $\{E_1, E_2, E_3\}$ y también se puede expresar como un baricentro de los puntos posibles Q_i con pesos o masas apropiadas q_i . Este mismo punto en \mathcal{L} , representa una cantidad aleatoria $X = uE_1 + vE_2 + wE_3$ con coeficientes (u, v, w) . Como $P(X) = ux + vy + wz$, $P(X)$ puede interpretarse como el producto interior de los vectores duales $P \in \mathcal{A}$ y $X \in \mathcal{L}$. Bajo este esquema, se considerarán los siguientes casos:

2.4.1.1 El caso de particiones.

Si los E_i forman una partición, hay tres constituyentes: $Q_1 = (1, 0, 0), Q_2 = (0, 1, 0)$ y $Q_3 = (0, 0, 1)$; las previsiones admisibles $P = (x, y, z)$ pertenecen a la región triangular $x + y + z = 1, x, y, z \geq 0$. Como baricentros de los puntos Q_i , $P = q_1 Q_1 + q_2 Q_2 + q_3 Q_3$ con $q_1 = x, q_2 = y, q_3 = z$.

2.4.1.2 El caso de incompatibilidad.

Si los E_i son incompatibles, pero no exhaustivos, hay 4 constituyentes: los Q_i anteriores y $Q_4 = (0, 0, 0)$. El casco convexo de los posibles valores será ahora el tetraedro con vértices Q_1, Q_2, Q_3, Q_4 , ya que ahora se tiene la relación $x + y + z \leq 1$. Una previsión admisible se puede expresar como un baricentro de los puntos Q_i en forma única con pesos $q_1 = x, q_2 = y, q_3 = z, q_4 = 1 - x - y - z$ (esto último porque $Q_4 = 1 - E_1 - E_2 - E_3$).

2.4.1.3 El caso del producto.

Sean E_1, E_2 lógicamente independientes y sea $E_3 = E_1 E_2$. Los constituyentes son ahora $Q_0 = (0, 0, 0), Q_1 = (0, 1, 0), Q_2 = (1, 0, 0), Q_0' = (1, 1, 1)$. Los primeros tres puntos están en el plano $z = 0$, y los últimos tres en el plano $x + y - 1 = z$. El casco convexo de este conjunto de vértices está dado por el tetraedro $z \geq 0, z \geq x + y - 1, z \leq x, z \leq y$. Estas son las restricciones bajo las cuales uno puede elegir arbitrariamente las probabilidades de dos eventos lógicamente independientes y su producto. Expresando una previsión admisible como un baricentro de los Q_i , se tiene una expresión única con los pesos $q_0 = 1 - x - y + z, q_1 = x - z, q_2 = y - z, q_0' = z$.

2.4.1.4 El caso de exhaustividad.

Supóngase que los tres eventos son exhaustivos pero compatibles; entonces se tienen 7 constituyentes:

$$Q_1 = (1, 0, 0), Q_2 = (0, 1, 0), Q_3 = (0, 0, 1), Q_3' = (1, 1, 0), \\ Q_2' = (1, 0, 1), Q_1' = (0, 1, 1), Q_0' = (1, 1, 1)$$

El casco convexo \mathcal{P} es el cubo menos el tetraedro definido por el vértice $(0, 0, 0)$ y los tres adyacentes, es decir, \mathcal{P} es la parte del cubo con $0 \leq x, y, z \leq 1$ que satisfacen la condición $x + y + z \geq 1$. Cada uno de sus puntos P puede ser expresado en una infinidad de formas, como un baricentro de puntos Q (a menos que el punto P coincida con un vértice, o pertenezca a un arco, o a una cara).

En los casos precedentes, P se obtuvo como un baricentro de los puntos posibles Q con pesos q unívocamente determinados, salvo en el caso de exhaustividad. La unicidad de la representación es un caso excepcional que sucede cuando y sólo cuando los Q son (afínmente) linealmente independientes — en los casos anteriores, se tienen 3 puntos no colineales o 4 puntos no coplanares — o cuando son (como eventos) expresables como una combinación de los eventos dados, es decir, cuando los Q pertenecían a \mathcal{L} .

2.5 Probabilidades de eventos (caso infinito).

Lo que se ha dicho hasta aquí se refiere a una colección finita de eventos. Sea ahora $\{E_i\}_{i \in I}$ una partición infinita de Ω . ¿Cómo se asignan probabilidades coherentes a los eventos E_i ? En los enfoques usuales es en este punto donde se introduce el axioma de aditividad numerable, porque ya no tiene sentido calcular la probabilidad de los eventos sino agregados de ellos. Por ejemplo, en el experimento de lanzar una moneda un número infinito de veces, ya no interesa el resultado de una sucesión de lanzamientos específica, sino de una colección de sucesiones que satisfagan alguna propiedad particular, como la de todas las sucesiones que en los primeros n lanzamientos sean soles. Para estudiar este tipo de fenómenos, se introduce el esquema propuesto por Kolmogorov, en donde la medida de probabilidad necesita ser σ -aditiva, y los eventos que pueden ser medidos con tal probabilidad tienen que pertenecer a una σ -álgebra.

Para los subjetivistas, el supuesto de σ -aditividad es muy restrictivo, en el sentido que, sin una justificación lógica, reduce a una clase —que bien puede ser muy grande— los conjuntos que pueden medirse: no es posible extender siempre una medida σ -aditiva al conjunto potencia de Ω . A costa de tal restricción, es posible ganar en elegancia y simplicidad en los cálculos y unicidad. Por otra parte, surgen problemas, como la imposibilidad de tener, en una partición infinita numerable, eventos equiprobables: si fuera posible, entonces tendría que existir un número p tal que $\sum_{i=1}^{\infty} p = 1$.

Bajo el supuesto de aditividad finita, tales objeciones desaparecen, a cambio de una mayor complicación analítica. Ya no es necesario restringir los eventos medibles a una σ -álgebra: puede definirse una previsión coherente sobre cualquier clase de eventos (como se demostró anteriormente); es posible que cada evento tenga probabilidad 0 y sin embargo, cualquier unión infinita puede tener probabilidad positiva: esto es, cualquier evento posible puede tener probabilidad 0.

Una medida de probabilidad finita aditiva satisface la siguiente propiedad: si $\{E_n\}$ es una sucesión de conjuntos ajenos dos a dos de Ω , y si $E \subset \Omega$ es tal que $\bigcup_{i=1}^{\infty} E_i \subset E$, entonces

$$\sum_{i=1}^{\infty} P(E_i) \leq P(E),$$

ya que $\bigcup_{i=1}^m E_i \subset E$ y $\sum_{i=1}^m P(E_i) = P(\bigcup_{i=1}^m E_i) \leq P(E)$ para toda $m \in \mathbb{N}$.

Otras propiedades que se cumplen para este tipo de medidas (que también se llaman cargas) son las mismas que se cumplen para las medidas σ -aditivas en lo que se refiere a colecciones finitas de conjuntos.

Una consecuencia de la aditividad finita es que si $\{P_n\}$ es una sucesión de previsiones coherentes sobre una familia \mathcal{E} de eventos, y si $E_0 \subseteq \mathcal{E}$ es el subconjunto de los elementos de \mathcal{E} para los cuales la sucesión converge a un valor P , entonces P es coherente en E_0 . Esto se debe a que las condiciones de coherencia están expresadas por medio de ecuaciones (o desigualdades) lineales que involucran un número finito de términos, y

éstas se preservan al pasar al límite. En el caso de una sucesión de probabilidades $\{P_n\}$ σ -aditivas, la probabilidad límite, cuando existe, no necesariamente es σ -aditiva.

En las secciones restantes se verán las consecuencias del supuesto de aditividad finita en el proceso de asignar previsiones a cantidades aleatorias.

2.5.1 Previsiones de cantidades aleatorias de rango numerable.

Sea X una cantidad aleatoria con rango numerable $\{x_h | h \in \mathbb{N}\} = I(X)$, y los eventos $E_h = \{X = x_h\}$ para los que se tiene asignada una probabilidad $p_h = P(E_h)$. Necesariamente se tiene que $\sum_h p_h \leq 1$. Sea $p^* = 1 - \sum_h p_h$. El problema es determinar la previsión de X a partir de las previsiones p_h .

Para cualquier intervalo o conjunto I que contenga un número finito de puntos x_h , puede decirse que

$$P(X \in I) = \sum_{\{h|x_h \in I\}} p_h,$$

pero si I contiene una infinidad de tales puntos, sólo puede decirse que

$$\sum_{\{h|x_h \in I\}} p_h \leq P(X \in I) \leq \sum_{\{h|x_h \in I\}} p_h + p^*.$$

La primera desigualdad es consecuencia de la aditividad finita, y la segunda se obtiene del siguiente modo:

$$\begin{aligned} P(X \in I) &= 1 - P(X \in I^c) = \sum_h p_h + p^* - P(X \in I^c) \\ &= \sum_{\{h|x_h \in I^c\}} p_h + \sum_{\{h|x_h \in I\}} p_h + p^* - P(X \in I^c) \\ &\leq \sum_{\{h|x_h \in I\}} p_h + p^* \end{aligned}$$

En el contexto de la aditividad finita, surgen *probabilidades adherentes* a los puntos de acumulación de las x_h . Estas probabilidades son las masas que se pierden en puntos distintos a los puntos posibles.

Definición 2.5.1.1

Sea x un punto de acumulación de $I(X)$. Una probabilidad adherente a x es cualquiera de los límites $q_x^- = \lim_{\epsilon \rightarrow 0} P(x - \epsilon < X < x)$ o $q_x^+ = \lim_{\epsilon \rightarrow 0} P(x < X < x + \epsilon)$.

Para denotar en forma genérica a una probabilidad adherente se utilizará la notación q_x .

Lema 2.5.1.1

Sea Ad el conjunto de puntos con probabilidades adherentes positivas. Entonces Ad es a lo más numerable.

Demostración.

Sea $A_n = \{x \in Ad | q_x > \frac{1}{n}\}$. A_n puede contener sólo un número finito de puntos. En efecto, supóngase que A_n contiene m puntos distintos con $m > n$, entonces se pueden encontrar intervalos ajenos $B_1 = (b_1, c_1), B_2 = (b_2, c_2), \dots, B_m = (b_m, c_m)$ tales que $P(B_m) > \frac{1}{n} \quad \forall m$, pero entonces

$$P(\bigcup_m B_m) = \sum_m P(B_m) > \frac{m}{n} > 1$$

lo cual es una contradicción. Finalmente $Ad = \bigcup_{n \in \mathbb{N}} A_n$, así que el resultado queda probado. ■

Lema 2.5.1.2

Si q_x es una probabilidad adherente a x , entonces $q_x \leq p^*$.

Demostración.

Se considerará el caso $q_x = q_x^-$; el otro caso es similar. Como $P(x - \epsilon < X < x) \leq \sum_{\{h | x_h \in (x - \epsilon, x)\}} p_h + p^*$, basta demostrar que

$$\sum_{\{h | x_h \in (x - \epsilon, x)\}} p_h \rightarrow 0 \quad \text{cuando } \epsilon \rightarrow 0.$$

Como la serie $\sum_{h=1}^{\infty} p_h$ es convergente, para toda $\delta > 0$ existe $N \in \mathbb{N}$ tal que $\sum_{h=N}^{\infty} p_h < \delta$, así que tomando

$$\epsilon \leq \min_{x_h \neq x, h < N} \{d(x_h, x)\},$$

se tiene

$$\sum_{\{h | x_h \in (x - \epsilon, x)\}} p_h \leq \sum_{h=N}^{\infty} p_h < \delta.$$

Teorema 2.5.1.1

Si q_1, q_2, \dots son probabilidades adherentes a los puntos y_1, y_2, \dots , entonces

$$\sum_i q_i \leq p^*.$$

Demostración.

Nuevamente se considera el caso de probabilidades adherentes q_i^- . El caso general se puede tratar de manera similar. Sea $n \in \mathbb{N}$ y sea $\epsilon > 0$ tal que los intervalos $\{(y_i - \epsilon, y_i) | i = 1, 2, \dots, n\}$ sean ajenos. Como

$$\sum_{i=1}^n q_i \leq \sum_{i=1}^n P(y_i - \epsilon < X < y_i) \leq \sum_{\{x_h \in \bigcup_{i=1}^n (y_i - \epsilon, y_i)\}} p_h + p^*$$

nuevamente basta con demostrar que la suma del lado derecho tiende a 0 cuando ϵ tiende a 0. Como en la demostración del caso anterior, dada $\delta > 0$, si

$$\epsilon \leq \min_{h < N, x_h \neq y_i, i=1, 2, \dots, n} \{d(x_h, y_i)\}$$

entonces $\sum_{\{x_h \in \bigcup_{i=1}^n (y_i - \epsilon, y_i)\}} p_h < \delta$. ■

Ejemplo:

Considérese la cantidad aleatoria que representa una elección al azar de un número racional en el intervalo $[0, 1]$, y asociando a la probabilidad de cada intervalo su longitud. En este caso, $p_r = P(X = r) = 0 \quad \forall r \in \mathbb{Q}$, así que $p^* = 1$. Por otra parte las probabilidades adherentes son los límites cuando $\epsilon \rightarrow 0$ de $P(x - \epsilon < X < x) = \epsilon$ (o el otro caso), por lo que también valen 0. Este ejemplo muestra que el conocimiento de la distribución aún en este caso puede no dar ninguna información para las cantidades aleatorias, como se verá más adelante con mayor precisión. ■

Ahora se probará que si $p^* = 0$ (i.e. si $\sum p_h = 1$, como pasa si se cumple la aditividad numerable), se debe tener el resultado $P(X) = \sum_{h=1}^{\infty} p_h x_h$, como en el caso finito; sin embargo, fuera de este caso especial, sólo puede decirse que

$$\sum_{h=1}^{\infty} p_h x_h + p^* x' \leq P(X) \leq \sum_{h=1}^{\infty} p_h x_h + p^* x'',$$

donde x' es el mínimo de los puntos de acumulación de $I(X)$ y x'' es el máximo de los puntos de acumulación de $I(X)$, los cuales satisfacen la relación:

$$-\infty < \inf X \leq x' \leq x'' \leq \sup X < \infty.$$

Entonces, bajo este caso, $P(X)$ resulta estar unívocamente determinada si y sólo si $x' = x''$, esto es, si las x_h tienen un único punto de acumulación, y por lo tanto un límite al cual ellas convergen.

Teorema 2.5.1.2

Si $\sum_{h=1}^{\infty} p_h = 1 - p^*$, entonces

$$\sum_{h=1}^{\infty} p_h x_h + p^* x' \leq P(X) \leq \sum_{h=1}^{\infty} p_h x_h + p^* x''$$

Demostración.

Sea $\epsilon > 0$. Como la serie $\sum_{h=1}^{\infty} p_h$ es convergente, existe $N(\epsilon) = N$ tal que

$$\sum_{h \geq N} p_h < \epsilon$$

Además se puede tomar N de tal forma que $N \rightarrow \infty$ cuando $\epsilon \rightarrow 0$. Si se escribe a X como la suma $X_1 + X_2 + X_3$, donde

$$\begin{aligned} X_1 &= \begin{cases} X & \text{si } X = x_h \text{ con } h < N; \\ 0 & \text{en otro caso.} \end{cases} \\ X_2 &= \begin{cases} X & \text{si } X = x_h \text{ con } h \geq N \text{ y } (x_h < x' - \epsilon \text{ o } x_h > x'' + \epsilon); \\ 0 & \text{en otro caso.} \end{cases} \\ X_3 &= \begin{cases} X & \text{si } X = x_h \text{ con } h \geq N \text{ y } x' - \epsilon \leq x_h \leq x'' + \epsilon; \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

entonces se sigue que si $\epsilon \rightarrow 0, N \rightarrow \infty$ y

$$P(X_1) = \sum_{h=1}^{n-1} p_h x_h \rightarrow \sum_{h=1}^{\infty} p_h x_h$$

Por otra parte, como hay un número finito de puntos de X menores que $x' - \epsilon$, así como un número finito de puntos de X mayores que $x'' + \epsilon$, entonces X_2 toma únicamente un número finito de valores distintos de 0, y_1, y_2, \dots, y_k , así que $P(X_2) = \sum y_j p_j \leq \sup X \sum p_j \leq \sup X \sum_{h \geq N} p_h \leq \epsilon \sup X$, y lo mismo para el otro lado; por lo tanto

$$\epsilon(\inf X) \leq P(X_2) \leq \epsilon(\sup X)$$

De este modo $P(X_2) \rightarrow 0$ cuando $\epsilon \rightarrow 0$.

Para X_3 se requieren varios casos.

Caso 1. Supóngase que $x' - \epsilon < 0 < x'' + \epsilon$. Entonces como $P(X_3) \leq (x'' + \epsilon)P(X_3 \in [x' - \epsilon, x'' + \epsilon]) \leq (x'' + \epsilon)(\delta + p^*)$, donde δ es una cantidad que depende de ϵ y que tiende a 0 cuando $\epsilon \rightarrow 0$. Ahora, como $x' - \epsilon < 0$ entonces

$$\begin{aligned} P(-X_3) &= -P(X_3) \leq -(x' - \epsilon)P(X_3 \in [x' - \epsilon, x'' + \epsilon]) \\ &\leq -(x' - \epsilon)(\delta + p^*) \end{aligned}$$

; por lo tanto $P(X_3) \geq (x' - \epsilon)(\delta + p^*)$; de las dos desigualdades se obtiene, haciendo tender ϵ a 0, que

$$x' p^* \leq P(X_3) \leq x'' p^*$$

Caso 2. Supóngase que $0 < x' - \epsilon$. Se obtiene una cota superior para $P(X_3)$ como en el caso anterior. Por otro lado, como $\sum_{h=1}^{\infty} p_h$ es convergente y como en el complemento

de $J = [x' - \epsilon, x'' + \epsilon]$ hay sólo un número finito de puntos x_h , se tiene que

$$\begin{aligned} P(X_3) &\geq (x' - \epsilon)P(X_3 \in J) \\ &= (x' - \epsilon)(1 - P(X_3 \in J^c)) \\ &= (x' - \epsilon)\left(1 - \sum_{h=1}^k p_{jh}\right) \\ &= (x' - \epsilon)\left(1 - (1 - p^*) - \sum_{h=k+1}^{\infty} p_{jh}\right) \\ &\geq (x' - \epsilon)p^*, \end{aligned}$$

obteniéndose así el mismo resultado que en el caso 1.

Caso 3. Supóngase que $x'' + \epsilon < 0$. La cota inferior es como en el caso 1 y como $(x'' + \epsilon)(p^* + \sum_{k+1}^{\infty} p_{jh}) \leq (x'' + \epsilon)p^*$, siguiendo un razonamiento similar al caso 2 se llega nuevamente al resultado del caso 1.

De lo anterior, como $X = X_1 + X_2 + X_3$, se obtiene el resultado deseado. ■

2.5.2 Caso no numerable: Funciones de distribución.

En teoría de la probabilidad, una función de distribución es una función $F: \mathbb{R} \rightarrow \mathbb{R}$ tal que:

- i. F es no decreciente,
- ii. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ y $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$, y
- iii. F es continua por la derecha y existe $F(x-)$.

F tiene una discontinuidad en x si $F(x+) - F(x-) > 0$. Por ser F una función no decreciente, tiene a lo más un número numerable de puntos de discontinuidad, ya que si x y x' son dos puntos de salto tales que $x < x'$ y si para $y \in \mathbb{R}$, I_y denota al intervalo $(F(y-), F(y+))$, como existe un punto x^* tal que $x < x^* < x'$, entonces por la monotonía de F se tiene que los intervalos I_x y $I_{x'}$ son disjuntos. A cada uno de estos intervalos disjuntos se les puede asociar un racional, y por lo tanto la colección de estos intervalos disjuntos es numerable.

Una función de distribución puede tener las siguientes interpretaciones:

1. $F(x)$ puede pensarse como una distribución de masa sobre los reales (con masa total de tamaño 1). $F(x)$ es la masa a la izquierda del punto x , $1 - F(x)$ es la masa a la derecha de x ; el incremento $F(x'') - F(x')$ es la masa en el intervalo $(x', x'']$. Una masa p_h concentrada en un punto x_h corresponde a una discontinuidad de F en x_h y $p_h = F(x_h+) - F(x_h-)$ es el tamaño del salto. En este mismo contexto, una función de distribución puede dar lugar a una medida σ -aditiva positiva sobre los intervalos, μ_F . haciendo $\mu_F(I) = F(b) - F(a)$, para cualquier intervalo $I = (a, b]$. $\mu_F(I)$ en este

contexto puede interpretarse como la medida de Lebesgue del conjunto imagen de I sobre el eje y .

2. F también da lugar al concepto de integral, a través de μ_F . Se define

$$\varphi_F(f) = \int f dF = \int f d\mu_F,$$

donde $\int f d\mu_F$ es la integral de Lebesgue con respecto a μ_F .

La aplicación de las funciones de distribución a la teoría de la probabilidad, toma la forma siguiente: para cualquier cantidad aleatoria X , uno puede definir una distribución de probabilidad sobre X asignando a la distribución la interpretación $F(x) = P(X \leq x)$. Entonces, $P(X \in I) = \mu_F(I)$ y $\varphi_F(f) = P(f(X))$ para los conjuntos I y las funciones f para los que la notación es aplicable. De esta manera F determina las probabilidades $P(X \in I)$ para todo I elemento de una familia \mathcal{F} la cual constituye un σ -álgebra de conjuntos sobre \mathbb{R} que contiene a los Borelianos (que es ya una clase bastante extensa de conjuntos que puede ser suficiente en las aplicaciones). Es decir, F determina toda la distribución de X en \mathcal{F} .

En el caso considerado aquí, en donde la previsión P no necesariamente es σ -aditiva, también se puede definir una función $F(x) = P(X \leq x)$ a la que se seguirá llamando función de distribución de X , aunque ahora F sólo tiene las siguientes propiedades:

- i. F es no decreciente,
- ii. $F(-\infty) = 0$ y $F(+\infty) \leq 1$, y
- iii. $F(x+)$ y $F(x-)$ existen.

Dentro de este contexto, la integral $\int f dF$ se define como la integral de Riemann-Stieltjes de f con respecto a F , esto es, se define la integral de una función escalonada (constante por intervalos) $\gamma = \sum_{i=1}^{\infty} c_i E_i$ como $\int \gamma dF = \sum_{i=1}^{\infty} c_i \mu_F(E_i)$. Para una función f acotada que pueda aproximarse por arriba y por abajo por medio de funciones escalonadas γ'' y γ' , se definen los números:

$$\varphi_F^-(f) = \int_a^b f dF = \sup_{\{\gamma''(x) \leq f(x) \forall x \in [a,b]\}} \left\{ \int_a^b \gamma'' dF \right\},$$

$$\varphi_F^+(f) = \int_a^b f dF = \inf_{\{\gamma'(x) \geq f(x) \forall x \in [a,b]\}} \left\{ \int_a^b \gamma' dF \right\},$$

como las integrales inferior y superior de f (estas son integrales del tipo de Riemann-Stieltjes ya que μ_F es sólo aditiva finita). Ambas integrales siempre existen ya que f es acotada por hipótesis. Si ambas integrales coinciden, a f se le llama S -integrable y al número común se le denota por $\int f dF$. Se puede considerar una funcional φ_F del conjunto de funciones f que son S -integrables a \mathbb{R} , generada por F . El siguiente teorema resume la relación entre las funciones de distribución y las previsiones:

Teorema 2.5.2.1

Sea γ una función acotada y X una cantidad aleatoria acotada. Los valores admisibles para $P(\gamma(X))$ son aquellos que satisfacen la desigualdad

$$\varphi_F^-(\gamma) \leq P(\gamma(X)) \leq \varphi_F^+(\gamma).$$

En particular, si γ es F -integrable, se tiene que

$$P(\gamma(X)) = \varphi_F(\gamma) = \int \gamma(x) dF(x)$$

Demostración.

Sea $s = \sum c_i I_i$ una función escalonada tal que $s(x) \geq \gamma(x) \quad \forall x$. Por definición, $P(s(X)) = \sum c_i P(I_i(X)) = \sum c_i \mu_F(I_i(X)) = \varphi_F(s)$, lo cual es válido $\forall s \geq \gamma$. Entonces

$$P(\gamma(X)) \leq \inf_{s \geq \gamma} \{\varphi_F(s)\} = \varphi_F^+(\gamma).$$

Análogamente,

$$P(\gamma(X)) \geq \sup_{t \leq \gamma} \{\varphi_F(t)\} = \varphi_F^-(\gamma).$$

En particular, si $\varphi_F^-(\gamma) = \varphi_F^+(\gamma) = \varphi_F(\gamma)$, se tiene que

$$P(\gamma(X)) = \varphi_F(\gamma) = \int \gamma(x) dF(x)$$

en donde la integral es una integral del tipo Riemman-Stieltjes. ■

Obsérvese que el resultado anterior implica que $P(X) = \int x dF(x)$, y está determinada en forma única.

Para el caso de más de una cantidad aleatoria, las integrales se tomarán sobre rectángulos para 2, sobre cubos para 3, etc. En general, como las integrales siempre se refieren a la medida de Jordan, se tomarán hiperplanos de dimensión n para n cantidades aleatorias, y se tendrá que

$$P(\gamma(X_1, X_2, \dots, X_n)) = \int \gamma(x_1, x_2, \dots, x_n) dF(x_1, x_2, \dots, x_n)$$

siempre que γ sea F -integrable.

Para analizar las diferencias que existen entre el enfoque subjetivo y los que se basan en los axiomas de Kolmogorov, se introduce la siguiente nomenclatura. Al supuesto de aditividad numerable se le llamará posición fuerte, y al supuesto de aditividad finita se le llamará posición débil.

2.5.2.1 La posición fuerte.

Esta posición afirma que los conjuntos I y funciones f para los que es posible evaluar en forma única una medida de probabilidad P son los importantes; fuera de éstos, nada tiene sentido.

Existe una correspondencia uno a uno entre las medidas de probabilidad y las funciones de distribución del espacio $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, donde $\mathcal{B}(\mathbf{R})$ es la σ -álgebra de Borel. Por esta razón una función de distribución tiene asociada una cantidad aleatoria X , en el sentido de que $P(X \in I) = \mu_F(I)$ y $P(f(X)) = \varphi_F(X)$ (siempre que I y f sean F -medible en el sentido de Lebesgue-Stieltjes).

A partir de una medida definida sobre los intervalos, es posible extender en forma única esta medida a una clase mucho más amplia de conjuntos, pero tal extensión no siempre es posible a todos los subconjuntos de Ω . Este problema es importante en el sentido de que se puede calcular la probabilidad de eventos en cierto sentido "simples" y calcular la probabilidad de eventos un poco "más complejos" pero no será posible si son "demasiado complejos".

Por otro lado, una vez conocida la función de distribución se tiene completamente determinada la cantidad aleatoria X , pues su rango $I(X)$ corresponde al conjunto de puntos x tales que $\forall \epsilon, F(x + \epsilon) - F(x - \epsilon) > 0$, es decir, al conjunto de puntos en donde F es estrictamente creciente. A este conjunto de puntos se le denotará con \mathbf{D} y se le llamará el *soporte distribucional* de X o bien, el soporte de la distribución F .

2.5.2.2 La posición débil.

Sea \mathcal{G} una clase dada de funciones f (que puede incluir a los eventos a través de sus funciones indicadoras); a la funcional φ_F definida sobre \mathcal{G} se le denotará por $F_{\mathcal{G}}$. Para cada función g que no pertenece a la clase dada, siempre será posible encontrar cotas de la forma $F_{\mathcal{G}}^-(g)$ y $F_{\mathcal{G}}^+(g)$ (es decir, las integrales superiores e inferiores). En particular, será de interés considerar $\mathcal{G} = \mathbf{R}$ (las funciones que son Jordan-medibles y con $F_{\mathbf{R}}$ la integral de Riemman o Riemman-Stieltjes); $\mathcal{G} = \mathcal{B}$ (las funciones Lebesgue-medibles) y con $F_{\mathcal{B}}$ la integral de Lebesgue; y $\mathcal{G} = \mathcal{C}$ (el conjunto de todas las funciones acotadas, o bien, el conjunto potencia de Ω).

Con esta notación, la posición fuerte significa que dada una cantidad aleatoria X , se elige automáticamente una $F_{\mathcal{B}}$. La posición débil establece que dada una X , $F_{\mathcal{C}}$ puede elegirse (esto es, puede elegirse una medida μ_F para cada conjunto E de Ω y una funcional φ_F para cada función acotada) o bien, restringirse a alguna $F_{\mathcal{G}}$ parcial, eligiendo frecuentemente a $\mathcal{G} = \mathbf{R}$. Esto es, cada pieza de conocimiento parcial corresponde al conocimiento de $F_{\mathcal{C}}(\gamma)$ restringida a algún subconjunto de las funciones γ . De este modo, siempre es posible definir una medida para todo subconjunto de Ω , aunque ésta medida no es única. Por otra parte, dada la medida de Jordan, puede extenderse a toda la clase de conjuntos, aunque tampoco tal extensión es única.

Debido a esta falta de unicidad, una función de distribución ya no representa el conocimiento completo para una cantidad aleatoria, sólo es un conocimiento parcial. Esto se manifiesta en que los puntos de salto de la distribución ya no corresponden a una acumulación de masa en esos puntos, ni que el soporte lógico corresponda al soporte de la distribución.

Ejemplo:

Si X es una cantidad aleatoria con posibles puntos $\{x_0 - \frac{1}{n} | n \in \mathbb{N}\}$, y todos son equiprobables, entonces $F(x)$ tiene un salto de 1 en $x = x_0$, como si $X = x_0$ con certeza. Esto es porque $F(x) = 0$ si $x < x_0$ y $F(x) = 1$ si $x \geq x_0$: a la izquierda de cualquier punto $x < x_0$ sólo hay un número finito de valores (y por ser equiprobables, todos tienen probabilidad 0), pero a la izquierda de x_0 y por lo tanto, de cualquier punto que esté a la derecha de x_0 , están todos los puntos posibles, y por lo tanto, se tiene probabilidad de 1. ◻

El soporte distribucional está contenido en la cerradura del soporte lógico, ya que si $x \in D$, entonces $\forall \epsilon > 0, F(x + \epsilon) - F(x - \epsilon) > 0$. Entonces, cada vecindad V_ϵ de x tiene probabilidad positiva y por lo tanto cada vecindad contiene puntos posibles. Por lo tanto, $\forall \epsilon > 0, V_\epsilon \cap I(X) \neq \emptyset$. De este resultado, se concluye que $\inf I(X) \leq \inf D \leq \sup D \leq \sup I(X)$. De este modo, se hace evidente la débil relación que existe entre los dos soportes. Si se tiene la distribución, sólo puede decirse que cada punto del soporte distribucional es o un punto posible, o está arbitrariamente cerca de puntos posibles; además de esto, los puntos posibles con probabilidad total 0 podrían existir en cualquier parte y aun llenar la recta. Por otro lado, dado el soporte lógico, la función de distribución puede ser cualquiera, siempre que sea constante sobre los intervalos que no contienen ningún punto posible.

Un caso especial de conocimiento parcial de la distribución completa es el que se refiere a los intervalos: en otras palabras, aquel dado por $F(x) \forall x$. A esto se le conoce como *conocimiento de la distribución a través de la función de distribución*. Este conocimiento parcial es equivalente al conocimiento de $\mu_F(\gamma)$ para toda función γ continua. Esto es, las funciones continuas son las únicas funciones que son F -integrables para cualquier F y, conversamente, el conocimiento de $\varphi_F(\gamma)$ para toda γ continua es suficiente para determinar (salvo en los puntos de discontinuidad) a $F(x) \forall x$.

En resumen, $F(x)$ sigue siendo importante y una herramienta estándar en la teoría de la probabilidad, pero no por su carácter determinativo, ni por un estatus privilegiado en el marco lógico, sino porque permite visualizar geométrica y analíticamente el comportamiento de X . Sin embargo, como se ha dicho anteriormente, $F(x)$ ya no representará conocimiento completo, ni lo que era fuera de \mathbb{R} . Fuera de éste ámbito, sólo se tienen las cotas $\{\varphi_F^-(\gamma), \varphi_F^+(\gamma)\}$; esto da los límites para la evaluación de $P(\gamma(X))$ compatible con el conocimiento de F en el sentido distribucional. Por otra parte, el

abandono del supuesto de aditividad numerable implica que un salto en x ya no corresponda a una concentración de probabilidad en el punto x (esto es porque aparecen probabilidades adherentes) y tampoco el punto necesariamente pertenece al conjunto de puntos posibles de X . Tampoco es cierto ya que $F(X)$ deba variar entre 0 y 1; sólo se requiere que $0 \leq F(-\infty) \leq F(+\infty) \leq 1$ (a tal distribución se le llama impropia).

2.6 Convergencia de cantidades aleatorias.

A continuación se enunciarán algunas definiciones y resultados que se utilizarán más adelante relativas a convergencia de cantidades aleatorias, observando las diferencias que puede haber con respecto al enfoque fuerte.

Definición 2.6.1

Se dice que $F_n \rightarrow F$ si $F_n(x) \rightarrow F(x)$ para todo x que sea punto de continuidad de F (o bien, la convergencia de $F_n(\gamma)$ a $F(\gamma)$ para toda función γ continua y acotada.

Una formulación equivalente (que no se demostrará aquí) establece que $F_n \rightarrow F$ si y sólo si $\forall \epsilon > 0 \exists N \in \mathbb{N}$ tal que $\forall n \geq N$,

$$F(x - \epsilon) - \theta \leq F_n(x) \leq F(x + \epsilon) + \theta \quad \forall x \in \mathbb{R}.$$

Definición 2.6.2

Sea $\{X_n\}$ una sucesión de cantidades aleatorias.

- i. Se dice que $\{X_n\}$ converge débilmente (o converge en probabilidad) a una cantidad aleatoria X y se denota por $X_n \xrightarrow{d} X$ si $\forall \epsilon > 0$ y $\theta > 0 \exists N \in \mathbb{N}$ tal que $\forall n \geq N$, $P(|X_n - X| > \epsilon) < \theta$.
- ii. Se dice que $\{X_n\}$ converge en media cuadrática a una cantidad aleatoria X y se denota por $X_n \xrightarrow{2} X$ si $\forall \epsilon > 0 \exists N \in \mathbb{N}$ tal que $\forall n \geq N$, $P([X_n - X]^2) < \epsilon$.
- iii. Se dice que $\{X_n\}$ converge fuertemente (o converge casi seguramente) a una cantidad aleatoria X y se denota por $X_n \xrightarrow{a.s.} X$ si $\forall \epsilon > 0$,

$$P\left(\bigcup_{m=1}^K \{|X_{n+m} - X| \geq \epsilon\}\right) \rightarrow 0$$

cuando $n \rightarrow \infty$ y donde K es arbitraria.

En la formulación fuerte, la definición establece que la condición se cumple para $K = \infty$.

Teorema 2.6.1 (Desigualdad de Tchebyshev).

Si X es una cantidad aleatoria y $P(X^2)$ existe entonces para todo número positivo c se tiene que

$$P(|X| \geq c) \leq \frac{P(X^2)}{c^2}$$

Demostración.

$$\begin{aligned} P(X^2) &= \int_{-\infty}^{\infty} X^2 dF(t) \\ &\geq \int_{|x| \geq c} x^2 dF(t) \\ &\geq c^2 \int_{|x| \geq c} dF(t) \\ &= c^2 P(|X| \geq c) \end{aligned}$$

Teorema 2.6.2

Sea $\{X_n\}$ una sucesión de cantidades aleatorias. $X_n \xrightarrow{*} X \Rightarrow X_n \xrightarrow{<} X$

Demostración.

Sean $\epsilon > 0$ y $\theta > 0$. Por hipótesis, $\exists N_\epsilon \in \mathbb{N}$ tal que $\forall n \geq N$ $P(|x_n - X|^2) < \theta \epsilon^2$. Sea $N = N_\epsilon$. Por la desigualdad de Tchebyshev,

$$P(|X_n - X| \geq \epsilon) \leq \frac{P(|X_n - X|^2)}{\epsilon^2} < \frac{\theta \epsilon^2}{\epsilon^2} < \theta$$

Teorema 2.6.3

$X_n \xrightarrow{<} X \Rightarrow F_n \rightarrow F$.

Demostración.

Sean ϵ y θ números positivos arbitrarios y sea N tal que $P(|X_n - X| \geq \epsilon) < \theta \quad \forall n \geq N$. Supóngase que $|X_n - X| < \epsilon$ y que $X > x$. Entonces $-\epsilon < X_n - X < X_n - x$, por lo que $X_n > x - \epsilon$. Entonces $\{|X_n - X| < \epsilon\} \wedge \{X > x\} \Rightarrow \{X_n > x - \epsilon\}$ o bien,

$$\{X_n \leq x - \epsilon\} \leq \{|X_n - X| \geq \epsilon\} \vee \{X \leq x\} \leq \{|X_n - X| \geq \epsilon\} + \{X \leq x\},$$

ya que $A \vee B \leq A + B$. Si P es una previsión coherente, respeta las relaciones lineales y por lo tanto,

$$F_n(x - \epsilon) - \theta \leq F(x) \quad \forall n \geq N,$$

ya que por hipótesis $P(|X_n - X| \geq \epsilon) < \theta \quad \forall n \geq N$. Si ahora se considera $|X_n - X| < \epsilon$ y $X_n > x + \epsilon$, entonces $X > x$, y con un razonamiento similar a la desigualdad anterior se obtiene que

$$F(x) \leq F_n(x + \epsilon) + \theta \quad \forall n \geq N$$

uniendo las dos desigualdades y tomando $n \rightarrow \infty$ se obtiene el resultado. ■

Corolario 2.6.1

$$X_n \xrightarrow{P} X \Rightarrow F_n \rightarrow F$$

Bajo la formulación fuerte, se sabe que si una sucesión de cantidades aleatorias $\{X_n\}$ conforma una sucesión de Cauchy, entonces se puede asegurar la existencia de una cantidad aleatoria X tal que $X_n \rightarrow X$ en algún sentido. Sin el supuesto de aditividad numerable, y sin la referencia a un espacio medible, no es posible definir X mediante el paso al límite. Con el fin de poder hablar de X , es necesario que ésta sea una cantidad bien definida. La afirmación concerniente a la convergencia sólo tiene sentido, si para tal X fuera posible demostrar que la evaluación particular de probabilidad para las X_n y X son tales que implican que $X_n \rightarrow X$ en algún sentido probabilístico.

Sin embargo, la convergencia según Cauchy sí determina una función de distribución límite para la sucesión de funciones de distribución $\{F_n\}$ de las cantidades aleatorias X_n . La demostración es idéntica a la demostración de que la convergencia en probabilidad implica la convergencia en distribución, que en palabras informales, dice que si X_n y X_m son dos variables "cercanas", entonces sus respectivas funciones de distribución son también "cercanas".

CAPITULO III

PROBABILIDAD CONDICIONAL

3.1 Introducción.

El papel de la probabilidad condicional es fundamental en el enfoque subjetivo en el sentido de que todas las probabilidades consideradas son condicionales al estado de información del sujeto que realiza la asignación.

En ciertas ocasiones es necesario combinar previsiones que son relativas a diferentes estados de información, y en tales casos se escribe $P(E|H)$ para la probabilidad del evento E condicional sobre el evento H , entendiéndose por esta la probabilidad que \mathcal{Z} atribuye a E si, adicional a su estado inicial de información, el sabe que el evento H es cierto.

Las definiciones de previsión y probabilidad condicional se basan en los criterios dados en el capítulo anterior, pero con la suposición adicional de que cualquier apuesta o castigo pactado quedará sin efecto si H no ocurre; si H ocurre, se llevará a cabo la apuesta y la ganancia dependerá del resultado de E . Obsérvese que la probabilidad condicional no depende directamente del evento E , sino de los eventos EH , $\bar{E}H$ y \bar{H} . Es posible entonces pensar en un *evento condicional* $E|H$ con tres posibles valores de verdad:

1. *verdadero* si E y H son verdaderos.
2. *falso* si H es verdadero y E es falso, y
3. *vacío* si H es falso.

3.2 Previsión y probabilidad condicional.

1er. criterio.

Sea X una cantidad aleatoria y H un evento posible. La previsión del evento condicional $X|H$ se define como un número $P(X|H)$ tal que no existe $c \in \mathbb{R}$ tal que

$$\sup_X \{Hc(X - P(X|H))\} < 0.$$

2o. criterio

Dada una cantidad aleatoria X y un evento posible H , se dice que $P(X|H)$ es coherente si no existe x^* tal que

$$L^* = H(X - x^*)^2$$

sea uniformemente menor que

$$L = H(X - P(X|H))^2.$$

En ambos casos si X es un evento, la previsión condicional recibe el nombre de probabilidad condicional.

Teorema 3.2.1 (probabilidades compuestas).

Una condición necesaria y suficiente para que una evaluación sea coherente es que se satisfagan las relaciones

$$P(EH) = P(H)P(E|H)$$

$$0 \leq P(E|H) \leq 1$$

Demostración.

Supóngase que x, y, z denotan las probabilidades asignadas a $P(E|H), P(H), (PEH)$ respectivamente. En base al primer criterio, la ganancia

$$Y = c_1H(E - x) + c_2(H - y) + c_3(EH - z)$$

tiene como posibles valores:

$$Y_1 = c_1(1 - x) + c_2(1 - y) + c_3(1 - z)$$

$$Y_2 = c_1x + c_2(1 - y) - c_3z$$

$$Y_3 = -c_2y - c_3z$$

para los valores $EH, \bar{E}H, \bar{H}$, respectivamente.

Si el determinante

$$\det \begin{pmatrix} 1-x & 1-y & 1-z \\ -x & 1-y & -z \\ 0 & -y & -z \end{pmatrix} = xy - z$$

no es 0, se pueden elegir c_1, c_2, c_3 de tal forma que todas las Y_i sean negativas, lo que implicaría falta de coherencia. Por lo tanto,

$$P(E|H)P(H) = P(EH).$$

El evento condicional $E|H$ da lugar al evento condicional $H|E$; como $EH = HE$, se obtiene, por coherencia,

$$P(EH) = P(E|H)P(H) = P(H|E)P(E)$$

lo que implica que

$$P(E|H) = P(E) \frac{P(H|E)}{P(H)} \quad (\text{si } P(H) \neq 0).$$

Observando la ecuación anterior, puede observarse que la probabilidad condicional es una función de probabilidad que forma parte del conjunto \mathcal{P} de previsiones coherentes.

3.3 Verosimilitud.

A partir de la última ecuación en la sección anterior, es posible escribir $P(E|H) = kP(E)P(H|E)$, donde $k = \frac{1}{P(H)}$. En muchas ocasiones es conveniente hablar simplemente en términos de proporcionalidad (sobre todo desde el punto de vista bayesiano). Así, se dice que la probabilidad final $P(E|H)$ es proporcional a la probabilidad inicial $P(E)$ multiplicada por la *verosimilitud* $P(H|E)$ (como función de E). El término verosimilitud se interpreta en el sentido de que un valor mayor o menor de $P(H|E)$ corresponde al hecho de que el conocimiento de la ocurrencia de E puede hacer a H más o menos probable.

Lo anterior permite comprender cómo es posible pasar de las probabilidades iniciales a las finales a través de estados intermedios, bajo la hipótesis de que se obtienen piezas adicionales de información H_1, \dots, H_n . Esto es:

$$P(E|H) = P(E|H_1 \dots H_n) = kP(E)P(H_1|E)P(H_2|EH_1) \dots P(H_n|EH_1 \dots H_n) \quad (8)$$

3.4 Propiedad conglomerativa.

Sea $\mathcal{H} = (H_1, H_2, \dots, H_n)$ una partición finita y sean $P(E|H_j)$ las probabilidades condicionales de un evento arbitrario E a cada uno de los componentes de la partición. Como $EH_1 + EH_2 + \dots + EH_n = E$ y $P(EH_j) = P(H_j)P(E|H_j)$, se tiene que

$$P(E) = \sum_{j=1}^n P(H_j)P(E|H_j)$$

y además,

$$\min_j \{P(E|H_j)\} \leq P(E) \leq \max_j \{P(E|H_j)\}.$$

A esta propiedad de la probabilidad condicional se le llama *propiedad conglomerativa*.

3.5 Dependencia e independencia estocástica.

La probabilidad $P(E|H)$ puede ser igual a $P(E)$, o mayor, o menor. Esto significa que el conocimiento de que H es cierto o no altera la evaluación de probabilidad para E , o la incrementa o la disminuye. En el primer caso, se dice que el evento E es estocásticamente independiente de H en los otros casos se dice que E es estocásticamente dependiente de H . Esta propiedad es simétrica, y por lo tanto se puede hablar de que E y H son estocásticamente independientes o dependientes, según el caso.

Para una colección finita de eventos $\{E_1, E_2, \dots, E_n\}$ se dice que los eventos son estocásticamente independientes si se satisface la condición:

$$P(E_{i_1} E_{i_2} \dots E_{i_k}) = P(E_{i_1})P(E_{i_2}) \dots P(E_{i_k})$$

para cualquier conjunto de índices $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$

El significado de la independencia estocástica no tiene un soporte objetivo sobre el cual sostenerse; es una propiedad supuesta en base a la intuición, y por esta razón, es considerada por De Finetti como una propiedad completamente subjetiva, ya que depende de la elección particular de la evaluación de probabilidad de un sujeto particular. Es más correcto hablar de que un conjunto de eventos es estocásticamente independiente con respecto a la evaluación P dada.

Entre los factores que llevan a un individuo a juzgar si los eventos son o no estocásticamente independientes, se encuentran los siguientes:

1. Cuando la ocurrencia de un evento cambia las circunstancias que rodean a la ocurrencia del otro evento. a este caso se le llama *dependencia estocástica en sentido directo*, y es aquí donde se encuentra la mayoría de los ejemplos de dependencia estocástica, como extracciones de una urna sin reemplazo, fenómenos de contagio, descomposturas de máquinas en serie, etc.

2. Cuando la ocurrencia de un evento no tiene influencia directa sobre la ocurrencia de otro, pero hay circunstancias que influyen simultáneamente a los eventos. Por ejemplo, considerando la posibilidad de que dos barcos se hundan en el mismo mar y el mismo día, uno podría pensar que las probabilidades están influenciadas en la misma forma por circunstancias comunes, como el estado del mar, etc. A este tipo de relación se le llama *dependencia estocástica en sentido indirecto*.

3. Otra forma de dependencia, quizá la más interesante, es cuando se evalúa la probabilidad de un evento en términos de frecuencia observadas de eventos más o menos similares. Los eventos observados proveen de experiencia que es capaz de modificar las evaluaciones de probabilidad basadas en frecuencias. En este caso, se dice que hay dependencia estocástica a través de un incremento en la información. Este caso da origen al concepto de intercambiabilidad, como se verá posteriormente, y es en este caso donde se ha intentado dar justificaciones objetivas, sin poder lograrlo. Un ejemplo de este caso podría ser el de extracciones con reemplazo de una urna cuya composición es desconocida: las probabilidades de sacar bolas blancas en extracciones sucesivas están interrelacionadas porque nuevas extracciones dan más información sobre el contenido de la urna, a través de las frecuencias.

En el capítulo siguiente se estudiará este punto con más detalle.

CAPITULO IV INTERCAMBIABILIDAD

4.1 Introducción.

El concepto de intercambiabilidad fue introducido por De Finetti en 1930, y se encuentra estrechamente relacionado con el problema del razonamiento inductivo, es decir, el problema que surge cuando se obtienen resultados generales de observaciones particulares; y a la justificación, desde este punto de vista, de la evaluación de probabilidades a partir de frecuencias observadas. Para justificar el punto de vista frecuentista desde la posición subjetiva, De Finetti propone separar la justificación en dos fases y explicar sus fundamentos subjetivos. La primera fase establece cuáles son las relaciones entre las evaluaciones de probabilidad y la previsión de frecuencias futuras, y la segunda fase establece cuál es la relación entre la observación de frecuencias pasadas y la previsión de frecuencias futuras; en esta fase introduce la noción de eventos intercambiables y demuestra que en este caso es posible asignar coherentemente la frecuencia relativa de ocurrencias de eventos como probabilidad. Esta fase está a su vez íntimamente relacionada con el problema del razonamiento inductivo.

4.2 El problema de la Inducción.

Según el punto de vista subjetivo, el problema de la inducción está resuelto: todo se reduce a considerar probabilidades condicionales, y principalmente, a la noción de dependencia estocástica a través de un incremento en la información.

El razonamiento inductivo, como objeto de la teoría de la probabilidad, en términos generales, tiene las tres formas siguientes:

1. Si la ocurrencia de varios eventos similares es de la misma forma, uno espera que otros eventos similares ocurran en la misma forma, como si hubiera una ley que gobernara el fenómeno.
2. Si un evento ocurre con cierta frecuencia, uno espera que en realizaciones futuras del mismo fenómeno se comporte en función de esa frecuencia. Este es el caso típico de la inferencia estadística.
3. Del conocimiento de la conducta de algún evento en una colección o sucesión de

casos más o menos similares en el pasado, frecuentemente se está motivado a realizar algún tipo de pronóstico para el futuro. Este es el caso más general de razonamiento inductivo.

No existe una separación clara entre estas tres posibles formas. Particularmente la segunda forma resulta ser útil en la aplicaciones porque puede modelarse en términos matemáticos.

En términos de la segunda forma, el problema de la inducción se puede plantear de la siguiente manera. Si E es un evento, y en virtud de haber observado, o haber obtenido la información de que un complejo A de eventos ocurrió, ¿qué se puede decir del evento E ? Objetivamente, no se puede decir nada, porque nada justifica hacer una predicción sobre un evento futuro, a menos que se suponga que existe una ley invariable que determine completamente su comportamiento. Sin embargo, en el esquema subjetivo de probabilidad, pueden establecerse las restricciones establecidas por la coherencia al calcular su probabilidad (que en ningún sentido es objetiva).

Supóngase que $P(E)$ es la probabilidad evaluada por la información que aporta A . Si H_0 representa la información inicial con la que cuenta un individuo, y $P^0(E) = P(E|H_0)$, entonces $P(E) = P(E|H_0A)$ que es igual a la información original más la dada por A . Esto implica que

$$P(E) = \frac{P^0(E)P^0(A|E)}{P^0(A)}$$

Sobre esto se basa el razonamiento inductivo: se cambia la evaluación de P a través del conocimiento de A . Este esquema es aplicable a cada aplicación del razonamiento inductivo. Sin embargo, los problemas de carácter estadístico tienen características especiales, como es el caso en que la información disponible consiste de un determinado número de eventos análogos (como lanzamientos de monedas, etc.), lo que permite a uno hacer ciertos supuestos de simetría en las evaluaciones de probabilidad. La analogía de los eventos conduce a que las conclusiones sólo dependan (al menos principalmente), de cuántos eventos análogos ocurren, y no de cuáles son los que ocurren. La analogía de los eventos puede establecerse en distintas formas, caracterizándolos por ejemplo en su independencia, en las condiciones en las que se realiza el experimento, etc.

En el caso de independencia, el problema de inducción no existe, en el sentido de que es imposible aprender de la experiencia. Si no se supone independencia, la elección

más simple que se tiene es continuar considerando que el orden de la ocurrencia de los eventos es irrelevante. Este caso corresponde a intercambiabilidad.

4.3 Justificación del punto de vista frecuentista.

La relación entre la evaluación de probabilidades y la previsión de frecuencias futuras, está dada por el siguiente

Teorema 4.3.1

Sean E_1, E_2, \dots, E_n eventos, con probabilidades p_1, p_2, \dots, p_n respectivamente. Si $w_i = P(Y = i)$ donde Y es la cantidad aleatoria que denota el número de ocurrencias de los eventos E_i , se tiene que

$$\sum_{i=1}^n p_i = \sum_{i=0}^n i w_i$$

Demostración.

Un caso de la ocurrencia de h eventos es la siguiente:

$$E_1 E_2 \cdots E_h (1 - E_{h+1}) (1 - E_{h+2}) \cdots (1 - E_n) = \\ E_1 E_2 \cdots E_n - \sum_i E_1 \cdots E_h E_{h+i} + \sum_{ij} E_1 \cdots E_{h+i} E_{h+j} - \dots \pm E_1 \cdots E_n$$

El evento $(Y = h)$ es la suma de $\binom{n}{h}$ productos de eventos del tipo anterior. En esta suma, los términos con h factores aparecen sólo una vez, los términos con $h + 1$ factores aparecerán $h + 1$ veces, y en general, los términos con $h + k$ factores aparecen $\binom{h+k}{h}$ veces. Si $\Sigma^{(r)}$ denota la suma de términos con r factores, entonces

$$(Y = h) = \Sigma^{(h)} - \binom{h+1}{h} \Sigma^{(h+1)} + \binom{h+2}{h} \Sigma^{(h+2)} - \dots \pm \binom{n}{h} \Sigma^{(n)} \\ = \sum_{r=h}^n (-1)^{r-h} \binom{r}{h} \Sigma^{(r)}$$

Aplicando P de ambos lados de la ecuación se obtiene que

$$w_h = P(Y = h) = \sum_{r=h}^n (-1)^{r-h} \binom{r}{h} P(\Sigma^{(r)}) = \sum_{r=h}^n (-1)^{r-h} \binom{r}{h} S_r,$$

donde

$$S_r = \sum_{\{i_1, i_2, \dots, i_r\}} P(E_{i_1} E_{i_2} \cdots E_{i_r}).$$

En particular obsérvese que el coeficiente de S_k , $k = 2, 3, \dots, n$ en los términos jw_j es $S_k j \binom{k}{j} (-1)^{k-j}$, $j = 1, 2, \dots, k$. Sumando se obtiene $S_k \sum_{j=1}^k j (-1)^{k-j} \binom{k}{j}$. Como

$$\begin{aligned} \sum_{j=1}^k j (-1)^{k-j} \binom{k}{j} &= \sum_{j=1}^k (-1)^{k-j} \frac{k!}{j!(k-j)!} j \\ &= k \sum_{j=1}^k (-1)^{k-j} \frac{(k-1)!}{(j-1)!(k-j)!} \\ &= k \sum_{r=0}^{k-1} \binom{k-1}{r} (-1)^{(k-1)-r} \\ &= k(1-1)^{k-1} \\ &= 0 \end{aligned}$$

De esta manera, el único término que sobrevive en la suma es $S_0 = \sum_{i=1}^n p_i$. ■

Si se dividen ambos lados de la ecuación del teorema por n , se obtiene la relación $\bar{p} = \bar{f}$, donde \bar{p} es la media aritmética de los p_i y \bar{f} es la previsión de la cantidad aleatoria $\frac{Y}{n}$, es decir, de la frecuencia.

En ciertos casos, la ecuación $\bar{p} = \bar{f}$ puede simplificarse. Si la frecuencia es conocida, el segundo miembro simplemente es ese valor. En este caso se tiene una indicación gruesa del orden de magnitud del promedio de las probabilidades, las cuales tendrán que ajustarse (disminuyendo algunos términos y aumentando otros) de tal forma que su promedio coincida con la frecuencia. Otro caso consiste en suponer que los n eventos son equiprobables; en este caso, el lado izquierdo de la ecuación sólo involucra el valor común de las probabilidades. Si ambos casos se dan simultáneamente, la única evaluación admisible está dada por la frecuencia $\frac{m}{n}$. Cuando ninguno de los dos lados de la ecuación se conocen, ambos términos dependen de un juicio de probabilidad. Sin embargo, es posible estimar la frecuencia vía la observación de frecuencias de la ocurrencia de los eventos en el pasado. Pero ¿cuándo es posible tal estimación? Esto es posible como se verá en el caso de eventos intercambiables. Es imposible demostrar el principio según el cual la probabilidad debería ser cercana a la frecuencia observada. Este principio sólo puede ser justificado en casos particulares, debido a las características lógicas del

enfoque subjetivo: siendo la probabilidad una evaluación subjetiva, no hay nada que obligue a elegir la ley de probabilidad de un fenómeno cercana a su frecuencia, todo lo que puede mostrarse es que tal evaluación resulta coherente cuando se satisfacen ciertas propiedades.

4.4 Intercambiabilidad.

Definición 4.4.1

Una colección finita o numerable $\{E_1, E_2, \dots\}$ de eventos se llama intercambiable si para toda $k \in \mathbb{N}$ y para cada conjunto de k índices diferentes j_1, j_2, \dots, j_k , se tiene que:

$$p_k = P(E_{j_1} E_{j_2} \dots E_{j_k}) = P(E_1 E_2 \dots E_k)$$

En otras palabras, las probabilidades de k eventos intercambiables no dependen del orden de los eventos, sino del número de términos.

Ejemplo:

- i. Un agricultor está cultivando una variedad de manzanas que son resistentes a la mosca de la fruta. Tiene un lote de 50 semillas. Sea E_i el evento de que la semilla i será atacada por la mosca de la fruta. El agricultor está interesado en asignar probabilidades para los eventos E_i , y en particular, la previsión de la cantidad aleatoria X = número de plantas atacadas. Ya que no tiene información particular para cada semilla, sus probabilidades deben satisfacer

$$P(E_i) = P(E_1) \quad i = 2, 3, \dots, 50$$

Además,

$$P(E_i E_j) = P(E_1 E_2) \quad \forall i \neq j$$

Así, su probabilidad de que cualesquiera k plantas específicas sean atacadas es la misma para cualesquiera otras k , y en particular, son iguales.

- ii. Los eventos estocásticamente independientes y equiprobables son intercambiables, pero no viceversa.
- iii. Supóngase que una urna contiene un número fijo N de bolas, de las cuales M son negras y el resto son blancas. Si se seleccionan n bolas ($n \leq N$) de la urna aleatoriamente, una por una y sin reemplazo, y se denota por A_j el evento de que la j -ésima bola extraída sea negra. Entonces, para $K \geq 1$ y para $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \frac{M(M-1)\dots(M-k+1)}{N(N-1)\dots(N-k+1)}$$

- iv. Los siguientes son también ejemplos de eventos intercambiables: H_i es el evento de obtener un sol en el i -ésimo lanzamiento de una moneda (no necesariamente regular); D_i es el evento de que el i -ésimo niño en una cierta escuela morirá antes de los 40. \square

Debido a que la intercambiabilidad establece varias igualdades entre las probabilidades, sólo se necesita asignar n probabilidades para obtener las probabilidades de todas las combinaciones de un conjunto de n eventos intercambiables.

Definición 4.4.2

Se denota con p_k la probabilidad de que ocurran cualesquiera k eventos, esto es,

$$p_k = P(E_{i_1} E_{i_2} \dots E_{i_k}) \quad k = 1, 2, \dots, n$$

Cualquier probabilidad relacionada a los E_i puede expresarse en términos de las p_k . Por ejemplo: $P(E_2 \bar{E}_6) = P(E_2) - P(E_2 E_6) = p_1 - p_2$, o bien,

$$\begin{aligned} P(E_1 \bar{E}_7 E_8 \bar{E}_{10}) &= P(E_1 \bar{E}_7 E_8) - P(E_1 \bar{E}_7 E_8 E_{10}) \\ &= P(E_1 E_8) - P(E_1 E_7 E_8) - P(E_1 E_8 E_{10}) + P(E_1 E_7 E_8 E_{10}) \\ &= p_2 - 2p_3 + p_4. \end{aligned}$$

Una definición que generaliza la noción de intercambiabilidad a cantidades aleatorias es la siguiente:

Definición 4.4.3

Una sucesión finita o infinita numerable X_1, X_2, \dots de cantidades aleatorias es intercambiable si para cualquier $k \geq 1$ y para toda colección de índices $1 \leq i_1 < i_2 < \dots < i_k$, la distribución del vector $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ no depende de los índices, sino sólo de k .

Obsérvese que para las cantidades aleatorias intercambiables X_1, X_2, \dots los eventos $E_j = \{X_j \leq x\}$ son intercambiables para cada número real x , así como los eventos $E_i^c = \{X_i \in I\}$ ($I \subset \mathbb{R}$).

Uno puede tratar el estudio de eventos intercambiables como un caso especial del estudio de cantidades aleatorias intercambiables, a través de las funciones indicadoras; en este caso, se denotará $m_i = P(X_1 X_2 \dots X_i)$.

El siguiente teorema será de utilidad para dar una caracterización de una colección finita de eventos intercambiables. Como antes, a la ocurrencia de un evento se le llamará éxito y a la no ocurrencia se le llamará fracaso.

Teorema 4.4.1

Sean E_1, E_2, \dots, E_n eventos intercambiables. Sea F la ocurrencia de cualquier sucesión particular de éxitos y fracasos en la cual se tienen r fracasos y h éxitos. Entonces

$$P(F) = \sum_{j=0}^r \binom{r}{j} (-1)^j p_{h+j}$$

Demostración.

La demostración se hará por inducción sobre r , el número de fracasos. Si $r = 0$, $P(F) = \sum_{j=0}^0 \binom{0}{j} (-1)^j p_{h+j} = p_h = p_n$.

Ahora supóngase que el resultado es cierto para $r < y$ y se probará que también debe ser cierto para $r = y$. Sea E_m uno de los y eventos que no ocurre en F , y sea G el evento de h éxitos y $y-1$ fracasos. Entonces $F_r = G\bar{E}_m$ y además $G = GE_m + G\bar{E}_m = F + GE_m$; por lo tanto

$$\begin{aligned} P(F) &= P(G) - P(GE_m) = \sum_{j=0}^{y-1} \binom{y-1}{j} (-1)^j p_{h+j} \\ &\quad - \sum_{j=0}^{y-1} \binom{y-1}{j} (-1)^j p_{h+j+1} \end{aligned}$$

El coeficiente de p_{h+j} , para $j = 0, 1, \dots, y-1$ es

$$\begin{aligned} \binom{y-1}{j}(-1)^j - \binom{y-1}{j-1}(-1)^{j-1} &= (-1)^j \left[\frac{(y-1)!}{j!(y-1-j)!} + \frac{(y-1)!}{(j-1)!(y-j)!} \right] \\ &= (-1)^j \left[\frac{(y-1)!}{j!(y-j)!} (y-j+j) \right] \\ &= (-1)^j \binom{y}{j} \end{aligned}$$

de donde se concluye que el resultado también es cierto para $r = y$ y en general se cumple para toda $r \in \mathbb{N}$. ■

Como un corolario del resultado anterior, se tiene el

Teorema 4.4.2

Si E_1, E_2, \dots, E_n son eventos intercambiables, y $p_i = P(E_1 \cdots E_i)$, $i = 1, \dots, n$. Sea Y_k la suma de k términos de la sucesión, $k = 1, \dots, n$. Entonces

$$P(Y_k = h) = \binom{k}{h} \sum_{j=0}^{k-h} \binom{k-h}{j} (-1)^j p_{h+j}$$

Demostración.

Tomando en el resultado anterior $k = n$ y observando que hay $\binom{k}{h}$ formas de obtener h éxitos y $k-h = r$ fracasos en k realizaciones, se obtiene el resultado. ■

Otra posible caracterización surge del siguiente razonamiento. Supóngase que se sabe que en una sucesión de n eventos ocurren precisamente h . Si se piensa a los eventos como bolas de una urna, en donde hay h bolas blancas y $n-h$ bolas negras, y se seleccionan muestras de tamaño r sin reemplazo, entonces se tiene que la probabilidad condicional de obtener r bolas blancas (éxitos) dado que hay h de bolas blancas, está dada por

$$P(E_1 E_2 \cdots E_r | H_h) = \frac{\binom{h}{r}}{\binom{n}{r}} = \frac{h(h-1)(h-2)\cdots(h-r+1)}{n(n-1)(n-2)\cdots(n-r+1)}$$

para $r = 1, 2, \dots, h$. Si, por otra parte, la variable Y_k denota el número de bolas blancas (éxitos) que se obtienen en un total de k extracciones, entonces es claro que

$Y_k | H_h \sim \text{Hip}(n, h, k)$. Si se desconoce el número de eventos favorables (o, en términos de la analogía, el número de bolas blancas en la urna), y se denota por la cantidad aleatoria Y_n a tal número, entonces, por el teorema de probabilidad total,

$$\begin{aligned} p_r &= P(E_1 E_2 \cdots E_r) = \sum_{h=0}^n P(E_1 E_2 \cdots E_r | Y_n = h) P(Y_n = h) \\ &= \sum_{h=0}^n \frac{\binom{h}{r}}{\binom{n}{r}} P(Y_n = h) \end{aligned}$$

y

$$P(Y_k = y) = \sum_{h=0}^n \frac{\binom{h}{y} \binom{n-h}{k-y}}{\binom{n}{k}} P(Y_n = h)$$

El resultado anterior muestra que los procesos intercambiables que terminan en n pasos se representan como mezclas de procesos hipergeométricos, es decir, las probabilidades $P(Y_k = y)$ tienen la forma $P(Y_k = y) = \sum_{h=0}^n c_h P(Y_n = h)$ con $\sum c_h = 1$, $c_h \geq 0$.

Sucesiones infinitas de eventos intercambiables.

Ahora se considerará la situación para sucesiones infinitas (procesos) de eventos intercambiables.

Un ejemplo de procesos intercambiables son los que se obtienen por mezclas de procesos tipo Bernoulli; es decir, tales que

$$w_h^{(n)} = \int_0^1 \binom{n}{h} \theta^h (1 - \theta)^{n-h} dF(\theta),$$

donde $w_h^{(n)}$ denota la probabilidad de obtener h éxitos en n ensayos. Se puede probar que los procesos intercambiables se pueden obtener siempre en forma de procesos Bernoulli. La demostración se hace mediante un paso al límite, considerando un proceso intercambiable finito. Primero se demostrarán algunos resultados previos.

Teorema 4.4.3

Sea $\{X_n\}$ una sucesión de cantidades aleatorias intercambiables con varianza finita y común σ^2 y sean Y_h y Y_k los promedios de h y k de las X_i (los promedios pueden tener

o no tener términos en común). Entonces $\forall \epsilon, \delta > 0 \exists N \in \mathbb{N}$ tal que si $h, k \geq N$,

$$P(|Y_k - Y_h| > \delta) < \epsilon$$

Demostración.

Sean $m_i = P(X_{k_1}, X_{k_2}, \dots, X_{k_i})$ para $i = 1, 2, \dots$ y sea $\mu = P(X_i^2)$. Como $\sigma^2 = \text{Var}(X_i) = P((X_i - m_1)^2) = P(X_i^2) - m_1^2$, entonces $P(X_i^2) = \sigma^2 + m_1^2 = \mu \quad \forall i \in \mathbb{N}$. Por otra parte, del cálculo de $P((Y_k - Y_h)^2)$ se sigue el resultado: sea r el número de términos en común de los promedios Y_h y Y_k . Entonces

$$\begin{aligned} P((Y_k - Y_h)^2) &= P(Y_k^2 - 2Y_k Y_h + Y_h^2) \\ &= \frac{1}{k^2} P\left(\left(\sum_{j=1}^k X_{ij}\right)^2\right) - \frac{2}{hk} P\left(\left(\sum_{j=1}^k X_{ij}\right)\left(\sum_{j=1}^h X_{lj}\right)\right) \\ &\quad + \frac{1}{h^2} P\left(\left(\sum_{j=1}^h X_{lj}\right)^2\right) \\ &= \frac{k\mu + k(k-1)m_2}{k^2} + \frac{h\mu + h(h-1)m_2}{h^2} \\ &\quad - \left(\frac{2r\mu + 2(hk-r)m_2}{hk}\right) \\ &= \frac{h\mu + h(k-1)m_2 + k\mu + k(h-1)m_2 - 2r\mu + (hk-r)m_2}{hk} \\ &= \frac{\mu(h+k-2r) + (h(k-1) + k(h-1) - 2hk + 2r)m_2}{hk} \\ &= (\mu - m_2)\left(\frac{h+k-2r}{hk}\right) \\ &\leq \left(\frac{1}{h} + \frac{1}{k}\right)(\mu - m_2) \end{aligned}$$

Por lo tanto, la sucesión $\{Y_n\}$ converge en media cuadrática según Cauchy. De lo anterior, se sigue que la sucesión $\{F_n\}$ tiende a una función de distribución límite F .

Corolario 4.4.1

La sucesión $\{F_n\}$ de funciones de distribución con $F_n(x) = P(Y_n \leq x)$ tiende a una distribución límite $F_{\Theta}(\theta)$ cuando $n \rightarrow \infty$.

Demostración.

La demostración es similar a la demostración que se dió al teorema de que la convergencia débil implica la convergencia en distribución, tomando en lugar de X_n y X a X_n y X_m respectivamente. ■

La distribución de las $Y_k, k = 1, 2, \dots, n$, condicional a la ocurrencia de h éxitos en los primeros n ensayos, es, como ya se vió, una distribución $Hip(n, h, k)$. Para continuar se requiere el siguiente

Lema 4.4.1

Sea $X \sim Hip(N, k, n)$, y sea $p = \lim_{N \rightarrow \infty} \frac{k}{N}$. Entonces si $N \rightarrow \infty$,

$$Hip(N, k, n) \rightarrow Bin(n, p)$$

Demostración.

si $x \leq k$ y $n - x \leq N - k$, entonces

$$\begin{aligned} P(X = x) &= \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} \\ &= \frac{k!(N-k)!(N-n)!n!}{N!(k-x)!(n-x)!(N-k-(n-x))!x!} \\ &= \binom{n}{x} \frac{1}{N(N-1)(N-2) \cdots (N-(n-1))} k(k-1)(k-2) \cdots \\ &\quad (k-x+1)(N-k)(N-k-1)(N-k-2) \cdots \\ &\quad (N-k-(n-x)+1) \\ &= \binom{n}{x} \left(1 - \frac{k}{N}\right) \left(1 - \frac{k}{N-1}\right) \cdots \left(1 - \frac{k}{N-(n-x)+1}\right) \\ &\quad \left(\frac{k}{N-(n-1)}\right) \left(\frac{k-1}{N-n}\right) \cdots \left(\frac{k-x+1}{N-(n-x)}\right) \\ &\rightarrow \binom{n}{x} (1-p)^{n-x} p^x \end{aligned}$$

En base al resultado anterior, y como $F_N \rightarrow F$, se tiene como resultado que

$$\int_0^1 Hip(N, \theta) dF_N(\theta) \rightarrow \int_0^1 \binom{N}{H} \theta^H (1-\theta)^{N-H} dF(\theta)$$

Entonces, si se toman h y n de tal forma que cuando $n \rightarrow \infty, \frac{h}{n} \rightarrow \theta < \infty$, se tiene que la distribución condicional de Y_k dado $Y_n = h$ es $Bin(k, \theta)$. En particular, si $k = 1$, se obtiene un evento con distribución $Bin(1, \theta) = Ber(\theta)$ y por lo tanto $P(E_i | Y_n = h) = \theta$.

Sea Θ la cantidad aleatoria que tiene por función de distribución a la función F . Θ intuitivamente representa la proporción de eventos ciertos en una sucesión infinita de eventos intercambiables; de lo anterior se debe tener que $\Theta \in [0, 1]$. En lo que sigue se denotará como $\Theta = \lim \frac{Y_n}{n}$ la convergencia en distribución de la sucesión $\{Y_n\}$. Para obtener la distribución de las Y_k , se aplica el teorema de Bayes del siguiente modo:

$$\begin{aligned} P(Y_k = y) &= \int_0^1 P(Y_k | \Theta = \theta) dF_{\Theta}(\theta) \\ &= \binom{k}{y} \int_0^1 \theta^y (1 - \theta)^{k-y} dF_{\Theta}(\theta) \end{aligned}$$

Resumiendo, el resultado anterior queda expresado en el siguiente teorema, conocido como el

Teorema 4.4.4 (Teorema de representación de De Finetti.)

Sea $\{E_n\}$ una sucesión de eventos intercambiables, y sea $\Theta = \lim \frac{Y_n}{n}$. Entonces

$$p_k = P(Y_k = k) = \int_0^1 \theta^k dF_{\Theta}(\theta) = P(\Theta^k),$$

y si Y_k es el número de éxitos en una subsucesión de tamaño k de $\{E_n\}$,

$$P(Y_k = y) = \binom{k}{y} P(\Theta^y (1 - \Theta)^{k-y})$$

El resultado anterior establece cuál es la distribución de la frecuencia en el caso de una sucesión infinita de eventos intercambiables.

Intercambiabilidad condicional.

Ahora supóngase que en los primeros k ensayos de la sucesión se han observado h éxitos. ¿Cuál es la probabilidad de los eventos E_{k+1}, E_{k+2}, \dots dada la información $Y_k = h$? Para responder a esta pregunta, se demostrará que la situación no se modifica

esencialmente respecto al caso general observando que una sucesión de eventos $\{E_n\}$ es intercambiable si y sólo si cualquier subsucesión $\{E_{n_k}\}$ es intercambiable dado un evento G .

Teorema 4.4.5

Sea Y_k el número de éxitos en los primeros k eventos de una sucesión $\{E_n\}$ de eventos intercambiables. Entonces la subsucesión $\{E_{k+1}, E_{k+2}, \dots\}$ es intercambiable dado $Y_k = h$.

Demostración.

Sea $F = \{Y_k = h\}$ y G cualquier evento que afirma que una subsucesión de eventos $\{E_{k+1}, E_{k+2}, \dots\}$ tiene h' éxitos. Por el teorema de Bayes, $P(G|F) = \frac{P(FG)}{P(F)}$. El evento FG considera una sucesión con $h + h'$ éxitos y $k - h$ fracasos, así que

$$P(FG) = \binom{k+h'}{h+h'} \sum_{j=0}^{k-h} \binom{k-h}{j} (-i)^j p_{h+h'+j}.$$

Esta probabilidad no depende de cuáles son los eventos particulares que constituyen a G , y lo mismo se dice de $P(F)$, así que $P(G|F)$ sólo depende de h' . ■

Falta demostrar que si $\{E_n\}$ es una sucesión de eventos tal que $\forall k$, la sucesión $\{E_{k+1}, E_{k+2}, \dots\}$ es intercambiable dado $Y_k = h$, entonces toda la sucesión es intercambiable.

Teorema 4.4.6

Si $\{C_i\}_{i=1}^n$ denota una partición, y si $\{E_n\}$ es una sucesión de eventos intercambiables dado C_k para cada k , entonces los eventos $\{E_n\}$ son intercambiables.

Demostración.

Por el teorema de Bayes,

$$P(E_{j_1} E_{j_2} \dots E_{j_n}) = \sum_{k=1}^n P(E_{j_1} E_{j_2} \dots E_{j_n} | C_k) = \sum_{k=1}^n p_{n,k} P(C_k).$$

Entonces $P(E_{j_1} E_{j_2} \dots E_{j_n})$ sólo depende de n y no de los índices j_1, j_2, \dots, j_n . ■

De los resultados anteriores, es posible responder a la pregunta inicial a través del siguiente

Teorema

Si $\{E_n\}$ es una sucesión de eventos intercambiables, Y_k es el número de éxitos en las primeras k observaciones, y $\Theta = \lim_{n \rightarrow \infty} \frac{Y_n}{n}$, entonces

i.

$$P(\theta|Y_k = h) = \frac{\theta^h(1-\theta)^{k-h}dF_{\Theta}(\theta)}{\int_0^1 \theta^h(1-\theta)^{k-h}dF_{\Theta}(\theta)}$$

ii. Sea p_r^* el equivalente de p_r dado $Y_k = h$. Entonces

$$p_r^* = P(E_{k+1}E_{k+2} \cdots E_{k+r}|Y_k = h) = P(\Theta^k|Y_k = h)$$

iii. Si $Y_{k,r}$ es el número de éxitos en E_{k+1}, \dots, E_{k+r} , entonces

$$P(Y_{k,r} = y|Y_k = h) = \binom{r}{y} P(\Theta^y(1-\Theta)^{r-y}|Y_k = h)$$

El inciso (i) del teorema anterior establece que

$$dF(\theta|Y_k = h) = K dF(\theta)\theta^h(1-\theta)^{k-h},$$

en donde $dF(\theta)$ es la distribución inicial de Θ y la función de verosimilitud de Y_k dado que $\Theta = \theta$ tiene la forma $v(\theta) = K\theta^h(1-\theta)^{k-h}$.

A continuación se darán los detalles de un caso simple muy especial: se supondrá que la distribución inicial de Θ es uniforme ($dF(\theta) = 1(0 \leq \theta \leq 1)$). Esta es la versión clásica de Bayes-Laplace, que corresponde a la idea de que no se tiene ningún conocimiento sobre Θ . Bajo este caso, la función de verosimilitud es creciente de $[0, \frac{h}{k}]$, alcanza su máximo cuando $\theta = \frac{h}{k}$, y después decrece de $(\frac{h}{k}, 1]$. Puede demostrarse que la distribución posterior de Θ , cuando k, h crecen, es asintóticamente una $N(\bar{\theta}, \frac{\bar{\theta}(1-\bar{\theta})}{k})$, donde $\bar{\theta} = \frac{h}{k}$. De esta forma, la previsión de la cantidad aleatoria Θ corresponde a la frecuencia observada, y esto demuestra el resultado buscado.

Teorema 4.4.7

Si h y k se toman suficientemente grandes de tal forma que la proporción $\bar{\theta} = \frac{h}{k}$ se

mantenga constante, y si se supone una distribución inicial para Θ uniforme, entonces la distribución posterior de Θ es asintóticamente normal con parámetros $\bar{\theta}$ y $\frac{\bar{\theta}(1-\bar{\theta})}{k}$.

Demostración.

Sea $\sigma^2 = \frac{\bar{\theta}(1-\bar{\theta})}{k}$ y $x = \frac{\theta - \bar{\theta}}{\sigma}$. Entonces

$$\begin{aligned} v(\theta) &= K\theta^h(1-\theta)^{k-h} = \left[\theta^{\frac{h}{k}}(1-\theta)^{1-\frac{h}{k}}\right]^k \\ &= K[\theta^{\bar{\theta}}(1-\theta)^{1-\bar{\theta}}]^k \\ &= K[(x\sigma + \bar{\theta})^{\bar{\theta}}(1-x\sigma - \bar{\theta})^{1-\bar{\theta}}]^k \\ &= K' \left[\left(1 + \frac{x\sigma}{\bar{\theta}}\right)^{\bar{\theta}} \left(1 - \frac{x\sigma}{1-\bar{\theta}}\right)^{1-\bar{\theta}} \right]^k \end{aligned}$$

Obteniendo logaritmos, y después desarrollando el logaritmo en su serie de Taylor alrededor del 1, se tiene que

$$\begin{aligned} \log\left(\frac{v(\theta)}{K'}\right) &= \log\left[\left(1 + \frac{x\sigma}{\bar{\theta}}\right)^{k\bar{\theta}} \left(1 - \frac{x\sigma}{1-\bar{\theta}}\right)^{k(1-\bar{\theta})}\right] \\ &= k\bar{\theta}\log\left(1 + \frac{x\sigma}{\bar{\theta}}\right) + k(1-\bar{\theta})\log\left(1 - \frac{x\sigma}{1-\bar{\theta}}\right) \\ &= k\bar{\theta}\left(\frac{x\sigma}{\bar{\theta}} - \frac{1}{2}\frac{x^2\bar{\theta}(1-\bar{\theta})}{\bar{\theta}^2k}\right) + O(k^{-\frac{3}{2}}) \\ &\quad + k(1-\bar{\theta})\left(-\frac{x\sigma}{1-\bar{\theta}} - \frac{1}{2}\frac{x^2\bar{\theta}(1-\bar{\theta})}{(1-\bar{\theta})^2k}\right) + O(k^{-\frac{3}{2}}) \\ &= -\frac{1}{2}x^2(\bar{\theta} + 1 - \bar{\theta}) + O(k^{-\frac{1}{2}}) \end{aligned}$$

Este último valor converge asintóticamente a $-\frac{1}{2}x^2$, por lo que $v(\theta)$ se aproxima a un valor proporcional a $e^{-\frac{1}{2}x^2}$ ■

4.5 Interpretación de la Intercambiabilidad.

Hasta aquí se ha establecido la noción general de intercambiabilidad estocástica, y se ha obtenido el teorema de representación de De Finetti, que afirma que una sucesión infinita de eventos intercambiables puede caracterizarse como una combinación lineal de las distribuciones de independencia y equiprobabilidad. Para hacer claro lo anterior considérese el siguiente ejemplo. Se tiene una urna con N bolas blancas y negras en proporciones desconocidas. Se hacen extracciones de la urna reemplazando la bola

después de que es extraída. La proporción de bolas blancas en la urna puede tener varios posibles valores $\rho_i \in \{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$, a los cuales se les dan ciertas probabilidades p_i . Estas extracciones son eventos intercambiables, y la probabilidad de un resultado dado es

$$P(E) = \sum_{i=1}^{N+1} P(E|\rho_i)P(\rho_i) = \sum_{i=1}^{N+1} p_i P_i(E)$$

Entonces, los eventos intercambiables corresponden a aquellos que se conocen ordinariamente como "eventos independientes y con probabilidad constante pero desconocida p ". Por otro lado, la distribución $F_{\Theta}(\theta)$ representa la probabilidad de que la "probabilidad desconocida" p sea más pequeña que θ . De esta manera, a través de los eventos intercambiables, uno evita suponer que más allá de la distribución de probabilidad correspondiente a un juicio, exista una distribución desconocida y objetiva, y que las diferentes hipótesis acerca de esta distribución desconocida puedan constituir eventos cuya probabilidad uno puede considerar.

CONCLUSIONES.

Desde un punto de vista matemático, las diferencias básicas entre el enfoque de De Finetti y el enfoque usual de la probabilidad, consiste en los siguientes puntos:

1. Se rechaza la idea de interpretar a los eventos como conjuntos en forma sistemática; los eventos pueden ser considerados proposiciones o conjuntos o números, y no es necesario que constituyan una σ -álgebra. No necesariamente todo evento está compuesto de eventos elementales.
2. Se rechaza la idea de que exista una única función P definida sobre la clase de eventos bajo consideración. En su lugar, De Finetti caracteriza todas las P que son admisibles; éstas constituyen una clase \mathcal{P} de previsionaciones coherentes que resulta ser una clase cerrada (en el sentido topológico), así que cualquier P adherente a \mathcal{P} pertenece a la clase, lo cual no se satisface en el enfoque de Kolmogorov.
3. Se rechaza la σ -aditividad por ser demasiado restrictiva, sin tener necesidad de serlo, sobre la asignación de probabilidades.
4. Se rechaza la transformación del teorema de probabilidad total en una definición de probabilidad condicional.

La noción de intercambiabilidad ha sido y sigue siendo ampliamente estudiada, aunque desde el esquema de σ -aditividad. Bajo tal esquema, resulta mucho más claro entender los resultados, ya que en ese caso es posible garantizar la existencia de una variable aleatoria Θ , tal que $P(\Theta \in [0, 1]) = 1$ y tal que $E(\Theta^k) = p_k \quad \forall k > 1$. Esto proviene del hecho de que en este caso, la convergencia en media cuadrática según Cauchy sí garantiza la existencia de una única Θ , y no exclusivamente la convergencia en distribución. Además, es posible demostrar el teorema de representación mediante un teorema debido a Hausdorff y que reduce el teorema a un problema de momentos.

En esta tesis sólo se utilizó intercambiabilidad para justificar el punto de vista frecuentista desde un punto de vista subjetivo; por tal motivo, no se desarrolló con mayor amplitud el caso de cantidades aleatorias intercambiables en general. Por otra parte, tampoco se desarrolló una teoría de probabilidad condicional sin el supuesto de que el evento condicionante tuviera probabilidad 0, cosa que es posible en el enfoque de De Finetti debido a los puntos 3 y 4 mencionados al principio de estas conclusiones.

REFERENCIAS.

- Bartle, R.G. (1980) *Introducción al Análisis Matemático*. 1era. ed. México: Limusa.
- Bachman, G. y L. Narici. (1966) *Functional Analysis*. Academic Press, New York.
- Bhaskara Rao, K. P. S. y Bhaskara Rao, M.(1983) *Theory of Charges. A Study of Finitely Additive Measures..* Academic, London, England.
- De Finetti, B. (1930) *Sul significato soggettivo della probabilit.* *Rend. R. Acc. Naz. Lincei*, 2o. semestre.
- De Finetti, B. (1972) *Theory of Probability. A critical introductory treatment.* Vols. 1 y 2. Wiley, New York.
- Feller, W. (1975) *Introducción a la teoría de probabilidades y sus aplicaciones.* Vols. 1 y 2. México: Limusa.
- Fine, T.L. (1973) *Theories of Probability*. Academic, New York.
- Frechet, M, W. (1955) *Las matemáticas y lo concreto*. México: Plaza y Valdes - UNAM.
- Johnson, N. L. y Kotz, S. (editores) (1977) *Urn Models and their Application*. Wiley, New York.
- Koch, G. y F. Spizzichino (editores) (1982) *Exchangeability in Probability and Statistics*. North Holland.
- Kolmogorov, A.N. y S. V. Fomín (1968) *Elementos de la teoría de funciones y del análisis funcional*. Ed. Mir, Moscú.
- Kyburg, H.E., Jr. y Smokler, H.E. (editores) (1980) *Studies in Subjective Probability*. Krieger. Huntington, New York.
- Loève, M. (1963) *Probability Theory, 3rd ed.* D. Van Nostrand, New York
- Marquez Diez-Canedo, J. (1987) *Fundamentos de Teoría de Optimización*. Limusa, México.