

## INFERENCIA ESTADÍSTICA: LÓGICA DE ACEPTACIÓN PROVISORIA DE PROPOSICIONES

por *Rolando Chuaqui*

Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile

Me siento extraordinariamente honrado y feliz de ser incorporado como miembro correspondiente de esta importante Academia. Mi alegría es mayor, ya que se trata de la Academia oficial de Ciencias de la República hermana de la Argentina, donde tengo numerosos amigos. Agradezco mucho esta designación.

La tesis fundamental que presentaré en esta conferencia es una de las tesis de mi libro "Truth, Possibility and Probability", [4]: la inferencia estadística es una forma de inferencia. Esta tesis parece tautológica, pero sin embargo, es rechazada por muchos autores. La razón de esto, es que al decir que la inferencia estadística es una forma de inferencia, yo y los otros autores que están de acuerdo con la tesis afirmamos que la inferencia estadística es similar a la inferencia lógica. La inferencia lógica se basa en reglas que permiten pasar de proposiciones verdaderas a otras proposiciones también verdaderas. Así, si aceptamos ciertas proposiciones como verdaderas y alguna proposición  $\phi$  es consecuencia lógica de ellas, entonces debemos también aceptar  $\phi$  como verdadera. La inferencia estadística, a mi juicio, tiene el mismo carácter: también incluye técnicas (en este caso probabilistas) para, partiendo de ciertas proposiciones aceptadas como verdaderas, determinar la aceptación de otras proposiciones. Sin embargo, en el caso de la estadística, la aceptación de la

conclusión es sólo provisoria. Esta concepción de la inferencia estadística, la distingue claramente de la teoría de decisión, que contiene técnicas probabilistas para decidir sobre el curso de nuestras acciones.

### 1. Inferencia estadística y teoría de las decisiones

Aunque la mayor parte de esta conferencia estará dedicada a la inferencia estadística, debo compararla con la teoría de decisión, ya que hay muchos autores que piensan que la inferencia estadística es parte de la teoría de la decisión. Esta última teoría nace al considerar la probabilidad como una "guía de la vida". La necesidad de una teoría de decisiones se basa en nuestra incerteza acerca de cómo debemos comportarnos en circunstancias en que somos ignorantes del estado del mundo. Como dice Kyburg en [12, p. 26]:

tratamos de desarrollar reglas de conducta que podamos seguir, cualquiera que sea el estado del mundo, con la esperanza que estamos siguiendo reglas cuyas características sean, en general, (1) deseables, y (2) alcanzables.

La regla más deseable es la que nos dice cómo descubrir el verdadero estado del mundo; la más alcanzable y simple es olvidarnos de la aritmética y actuar como nos parece. Un compromiso entre estos dos extremos, es asignar grados de creencia a los distintos estados del mundo, basados en lo que sabemos de sus probabilidades y grados de "deseabilidad" a los posibles resultados y combinar estos grados de una manera adecuada.

La investigación para este artículo ha sido financiada parcialmente por el proyecto DIUC 89041.

Conferencia pronunciada durante su incorporación como Académico Correspondiente en Santiago, Chile, el día 23 de julio de 1991.

La inferencia estadística se usa, por su parte, principalmente en la evaluación de hipótesis científicas. Como en estos casos tenemos que *decidir* si aceptar o rechazar una hipótesis científica, la inferencia estadística es considerada por algunos como parte de la teoría de las decisiones. Sin embargo, creo con Fisher, [7, p. 5], que:

... tales procesos (Fisher se refiere aquí a decisiones) tienen una base lógica muy diferente de los de un científico tratando de ganar de sus observaciones una comprensión mejor de la realidad.

Creo que Fisher hubiera estado de acuerdo con mi opinión que la inferencia estadística, como se usa en la evaluación y estimación de hipótesis, tiene el mismo objetivo general de la inferencia lógica. Da reglas de racionalidad para que partiendo de algunas proposiciones que son aceptadas como verdaderas, se llegue a proposiciones que deben aceptarse como verdaderas. Un caso típico de la aplicación de la inferencia estadística es en la aceptación de la hipótesis que el fumar es un factor causal para el cáncer al pulmón. Se acepta, en la actualidad, que el cigarrillo hace daño a la salud y no solamente que esta proposición es altamente probable. Así, hemos llegado a la aceptación de una proposición como verdadera basada en datos estadísticos y usando inferencia estadística.

La teoría de decisión<sup>1</sup>, por su parte, estudia reglas de racionalidad para guiar nuestras acciones frente a la incerteza. Hay, por lo menos, dos diferencias fundamentales entre la evaluación de hipótesis y la teoría de decisiones. En primer lugar, al evaluar hipótesis uno no debe considerar las "deseabilidades" (o "bondad") de las consecuencias de aceptar o rechazar las hipótesis, mientras que al decidir sobre una acción, las "deseabilidades" de las consecuencias de la acción son de importancia primaria. Esto es, al decidir sobre una acción, debemos tomar en consideración las ventajas y desventajas de las consecuencias de nuestra acción, lo que se llama, las *utilidades* de las conse-

cuencias. Por otra parte, erraríamos al considerar utilidades cuando decidimos si aceptar o no una hipótesis.

En segundo lugar, las decisiones sobre las acciones pueden basarse en las probabilidades de los estados posibles del mundo (esto es, las 'hipótesis' sobre las cuales se basa la acción), mientras que aceptar o rechazar hipótesis no puede basarse en las probabilidades de las hipótesis mismas. Tengo dos razones principales para esta posición: La primera razón se basa en el hecho de que hay una diferencia importante entre aceptar una proposición y asignarle una probabilidad alta (o, incluso, muy alta): si uno acepta un conjunto de proposiciones, uno se compromete a aceptar todas sus consecuencias lógicas. Por ejemplo, si yo acepto  $p_1, p_2, \dots, p_n$ , entonces también debería aceptar su conjunción. Por otra parte, se puede asignar a cada una de  $p_1, \dots, p_n$  probabilidad alta y su conjunción puede tener probabilidad baja. Por ejemplo, si uno juzga que  $p_1, \dots, p_n$  son independientes, la probabilidad de la conjunción es el producto de las probabilidades, y, por tanto, con  $n$  suficientemente grande, la probabilidad de la conjunción podría ser muy baja.

Cuando aceptamos una proposición, actuamos como si fuera verdadera, y, por tanto, la usamos como premisa en aplicaciones de todas las reglas normales de la lógica para obtener consecuencias, las que también deben ser aceptadas. Cuando tomamos decisiones, por otra parte, no necesitamos usar las hipótesis en inferencias lógicas. Si hemos aceptado una proposición, podemos usarla, obviamente, para nuestras decisiones, pero no es necesario aceptar una hipótesis para usarla para decidir sobre una acción. Es suficiente para tomar decisiones asignar probabilidades.

El otro argumento fuerte a favor de la imposibilidad de aceptar o rechazar hipótesis a base de sus probabilidades es la llamada paradoja de Kyburg:

**La paradoja de la lotería de Kyburg.** Creo que la paradoja de la lotería de Kyburg, [11], muestra que es imposible dar una regla de aceptación (o rechazo) de hipótesis, basada sólo en sus probabilidades altas (o bajas).

<sup>1</sup> Hacking tiene una excelente discusión de las diferencias entre aceptar o rechazar hipótesis y la teoría de las decisiones en [9, págs. 27-32].

Las siguientes condiciones de Hempel, [10], que el cuerpo  $K$  de proposiciones aceptadas debe satisfacer me parecen obvias:

(1) Cualquier consecuencia lógica de un conjunto de aserciones aceptadas, también debe ser aceptada; o, lo que es lo mismo,  $K$  contiene todas las consecuencias lógicas de cualquiera de sus subclases.

(2) La clase  $K$  de proposiciones aceptadas es lógicamente consistente.

La paradoja de la lotería de Kyburg muestra que cualquier regla de aceptación de proposiciones basada en sus probabilidades altas viola las condiciones de Hempel. Supongamos que decidimos aceptar  $h$  con evidencia  $e$  cuando la probabilidad de  $h$  dada  $e$  es mayor que 0,999, y, por tanto, rechazar  $h$  cuando esta probabilidad es menor que 0,001. Considérese una lotería con más de 1.000 números. Supóngase que un número se escoge al azar. Entonces debemos aceptar, de acuerdo a nuestra regla, las proposiciones:

(a) " $n$  no será elegido", para todo  $n$ .

También, el dispositivo de la lotería nos fuerza a aceptar:

(b) "un número será elegido".

(a) y (b) son contradictorias, y, por tanto, nos hemos visto forzados a aceptar dos aserciones contradictorias, lo que no es posible por la cláusula (2) de Hempel.

Muchos autores sobre estadística, probabilidad, o lógica inductiva, evitan este problema diciendo que nunca debemos aceptar hipótesis, ni siquiera provisoriamente. Por ejemplo, de acuerdo a Carnap, una hipótesis científica  $h$  se hace probable hasta un cierto grado por la evidencia  $e$ . Deberíamos tener un grado de creencia en  $h$  correspondiente a su probabilidad, pero nunca debemos aceptar  $h$ , ni siquiera provisoriamente.

Un subjetivista puede hablar acerca de la aceptación de hipótesis, pero sólo en los casos extremos. Para ellos, aceptar una hipótesis  $h$  sería asignarle una probabilidad subjetiva de uno. Pero la única manera de asignar una probabilidad condicional de uno a  $h$  bajo la evidencia  $e$  es asignar una probabilidad a priori de uno a  $h$ .

Muchos estadísticos no-bayesianos también tienen una tendencia conductista. For-

mulan el problema de la inferencia estadística como un problema de elección entre cursos de acción alternativos. Neyman, por ejemplo, habla sólo de comportamiento inductivo, no de inferencia inductiva.

La objeción principal a este punto de vista conductista sobre la ciencia, es que no parece estar de acuerdo ni con la práctica, ni con el punto de vista teórico de los científicos: los científicos simplemente *aceptan* (y *rechazan*) hipótesis, cuando la evidencia a su favor (o en contra) es muy grande. Parece difícil, dada la importancia enorme que las teorías científicas han adquirido en los tiempos modernos, admitir que las leyes científicas sirven sólo como resúmenes de una gran cantidad de datos, y que su único fin es para simplificar nuestros cálculos o guiar nuestras acciones. Así, creo que la inferencia estadística es inferencia, esto es, que contiene reglas para pasar de proposiciones aceptadas como verdaderas a otras proposiciones que se deben aceptar (provisoriamente, al menos) como verdaderas.

## 2. Probabilidad

Es claro que aunque no podemos aceptar o rechazar hipótesis por sus probabilidades, la inferencia estadística se basa en las probabilidades de ciertas proposiciones. Discutiremos brevemente la interpretación de probabilidad que sirve de soporte a nuestra concepción de inferencia estadística. Esta interpretación se basa en tres elementos:

(1) *La probabilidad se basa en una noción objetiva.* Aunque creo que hay espacio para una noción subjetiva de probabilidad, especialmente en la teoría de la decisión, pienso que la noción básica es la de azar o probabilidad fáctica. La probabilidad fáctica está determinada por las posibilidades reales, las que acepto como propiedades físicas de los objetos y las condiciones experimentales. El conjunto de posibilidades para una situación dada está representado, desde el punto de vista matemático, por un modelo  $K$ , que es un conjunto con cierta estructura.

(2) *La probabilidad es la medida de la posibilidad de verdad.* Además de la noción objetiva de azar ya explicada, considero dos nociones de tipo espistémico: grado de apo-

yo y grado de creencia. La probabilidad de una proposición  $\phi$  es considerada fundamentalmente como el grado de apoyo a  $\phi$ , dado que aceptamos un modelo  $\mathbf{K}$  para el fenómeno involucrado, y sólo secundariamente, como el grado de creencia. El grado de apoyo a una proposición  $\phi$  dado que  $\mathbf{K}$  es un modelo adecuado de la situación es la medida del conjunto de posibilidades (esto es, miembros de  $\mathbf{K}$ ) donde la proposición es verdadera.

(3) *La definición clásica de probabilidad es esencialmente correcta.* La medida de los diferentes conjuntos de posibilidades (esto es, subconjuntos de  $\mathbf{K}$ ) se basa en una noción de equiprobabilidad, tal como la definición clásica de de Moivre y Laplace. Creo que todas las dificultades que presenta la definición clásica son resueltas por mi interpretación.

El azar es una propiedad de *dispositivos aleatorios*. La única característica esencial para que un dispositivo esté sometido al azar es que debe tener un resultado único que es un elemento de un conjunto fijo de resultados realmente posibles. En esencia, un dispositivo aleatorio consiste de un conjunto de objetos más un conjunto de condiciones bajo las cuales estos objetos tienen ciertas posibilidades reales (una discusión de posibilidad en general y de la noción de posibilidad real en particular, aparece en [3] y en [4, Capítulo II]).

En el sentido técnico de mi teoría, un *resultado* de un dispositivo aleatorio incluye además del resultado particular mismo, una descripción del dispositivo. Así, un resultado codifica el mecanismo del dispositivo. Un dispositivo aleatorio puede, entonces, ser identificado con su conjunto de resultados posibles.

Los sucesos son conjuntos de resultados. Un suceso ocurre, si uno de los resultados que son sus elementos ocurre. Es natural pensar que un suceso tiene más probabilidad de ocurrir, si contiene 'más' resultados. La probabilidad (fáctica) de un suceso es, entonces, la medida del conjunto de resultados posibles que conforman el suceso. Cada resultado posible perteneciente a un suceso representa una posibilidad real invisible para que el suceso ocurra. Cada po-

sibilidad real tiene, por así decirlo, una cierta propensión a ocurrir y suponemos que las posibilidades últimas representadas por los resultados tienen la misma fuerza. En otras palabras, cada resultado tiene la misma tendencia a ocurrir. Así, creo que mi análisis coincide con las ideas de Leibniz<sup>2</sup> expresadas en la siguiente afirmación:

La probabilidad es el grado de la posibilidad.

Una característica de mi punto de vista es que la probabilidad se aplica a sucesos singulares, no necesariamente como parte de una sucesión de experimentos similares. Es posible mirar a esta definición de probabilidad fáctica como una explicación de propensiones singulares. La propensión se piensa habitualmente como una forma de poder, fuerza o tendencia. Cada posibilidad de que un suceso ocurra representa poder adicional para su obtención. Así, si hay más posibilidades, hay más poder; y posibilidades iguales determinan igual poder. Como J. Bernoulli dice en [1, p. 219], refiriéndose a sucesos equiprobables:

Todos los casos son igualmente posibles, esto es, cada uno puede ocurrir tan fácilmente como cualquier otro.

Por la incorporación del mecanismo del dispositivo en cada resultado, el conjunto de los resultados posibles es todo lo que necesitamos saber acerca de un dispositivo aleatorio para decidir cuál es el espacio de probabilidad adecuado para él. Así, el conjunto de los resultados posibles para un dispositivo particular determina, en mi teoría, la familia de sucesos y la medida de probabilidad que son adecuados para el dispositivo.

El primer paso para la determinación de las probabilidades es un análisis detallado de lo que es un resultado, y la delimitación de las propiedades y relaciones que son relevantes para el dispositivo. Los resultados pueden tener muchas propiedades diferentes. Mirando al ejemplo de tirar un dado, estamos normalmente interesados en el número que aparece en la cara superior, pero no en la distancia hacia cierta pared o en su color. Al formalizar el conjunto de resulta-

<sup>2</sup> Citado en [9].

dos posibles, incluimos sólo las propiedades relevantes. Algunas propiedades, sin embargo, pueden resultar ser relevantes en mi teoría, aunque en las formulaciones probabilistas usuales no se consideran explícitamente. Por ejemplo, al tirar un dado cargado, la distribución de pesos en el dado es relevante. Es posible incluir las propiedades relevantes representando un resultado por un modelo conjuntista, en el sentido de la teoría lógica de modelos.

Otro elemento que debe ser incluido en la descripción del dispositivo, y, por tanto, en la estructura que representa los resultados, son las relaciones de dependencia e independencia estocástica entre los diferentes componentes de los resultados, cuando estos resultados estén formados de otros más simples. Supóngase, por ejemplo, que el dispositivo consiste en la elección de una urna al azar, y, después, en la elección de una bola de esta urna. La elección de la urna determina el campo de posibilidades para la elección de las bolas. Así vemos que dependencia es dependencia del campo de posibilidades. Por otra parte, cuando tiramos una moneda dos veces, el campo de posibilidades de la segunda tirada no depende del resultado de la primera tirada; en este caso tenemos independencia. Estas nociones de dependencia e independencia están representadas en mis modelos por órdenes parciales.

La descripción de un dispositivo aleatorio está dado por un conjunto de sistemas, cada uno representando un resultado posible. La medida de probabilidad se define como la medida (si existe y es única) que satisface las siguientes dos condiciones. En primer lugar, debe ser invariante bajo las transformaciones que preservan el dispositivo. Estas transformaciones son las permutaciones de los objetos involucrados en el dispositivo que transforman un resultado posible en otro resultado posible. La invarianza bajo estas transformaciones es una versión precisa de un principio objetivo de simetría. Este principio es objetivo, porque el grupo de transformaciones que determina las simetrías se obtiene del conjunto de resultados posibles, el que, a su vez, depende sólo del dispositivo aleatorio, el que es objetivo.

En segundo lugar, la medida debe preservar las relaciones de dependencia e indepen-

dencia incluídas en los resultados. Esta preservación de la dependencia es, de hecho, otro principio de invarianza: Dos dispositivos aleatorios que tienen la misma *forma* deben asignar las mismas probabilidades a sucesos correspondientes. Dos modelos de dispositivos aleatorios  $K_1$  y  $K_2$  se llaman *homomorfos*, esto es, tienen la misma "forma", si hay una función (el *homomorfismo*) que transforma  $K_1$  en  $K_2$  preservando la estructura de dependencia e invarianza interna de los dispositivos. Si un suceso  $A$  de  $K_1$  contiene todos los resultados cuyas imágenes por el homomorfismo conforman un suceso  $B$  de  $K_2$  (esto es,  $A$  es la imagen inversa por el homomorfismo de  $B$ ), entonces la probabilidad de  $B$  en  $K_2$  debe ser la misma que la probabilidad de  $A$  en  $K_1$ .

La familia de sucesos también está determinada por el conjunto de los sistemas que representan los resultados posibles del dispositivo. Los sucesos son aquellos conjuntos de resultados posibles cuya medida está determinada por el grupo de simetrías y la relación de homomorfismo.

Como muestro en mi libro [4], todos los dispositivos aleatorios para los cuales se ha definido una medida de probabilidad en el pasado pueden ser modelados por mis métodos.

Otro aspecto de la probabilidad es el grado de apoyo que tiene una proposición  $\phi$  relativo a un cierto modelo, dado por un conjunto de resultados posibles  $K$ . Esta probabilidad la notaré  $Pr_K(\phi)$ . Este grado de apoyo,  $Pr_K(\phi)$ , se entiende como el *grado de posibilidad de verdad* (o, más informalmente, de verdad parcial) de  $\phi$  en  $K$ .  $Pr_K(\phi)$  se define como la medida de probabilidad (invariante y preservada bajo homomorfismos) del conjunto de resultados en  $K$  que hacen a  $\phi$  verdadera; esto es, el grado de apoyo a  $\phi$ , dado  $K$ , es lo mismo que la medida de la probabilidad fáctica del suceso que consiste de los resultados en  $K$  donde  $\phi$  es verdadera. Si  $\phi$  es verdadera siempre que ocurre un resultado en  $K$ , decimos que  $\phi$  es verdadera en  $K$ ; en este caso, el conjunto de los resultados que hacen a  $\phi$  verdadera es  $K$  mismo, y, luego,  $Pr_K(\phi) = 1$ . Cuando  $\phi$  es falsa siempre que un resultado en  $K$  ocurre, decimos que  $\phi$  es falsa en  $K$ ; en este caso, el conjunto de resultados que hacen a  $\phi$  verdadera es vacío, y, luego,  $Pr_K(\phi) = 0$ . Para los casos in-

termedios,  $\text{Pr}_{\mathbf{K}}(\varphi)$  es un número real entre 0 y 1. De esta manera, el grado de apoyo a  $\varphi$ , dado  $\mathbf{K}$ , se puede entender como el grado de "posibilidad de verdad" (o "verdad parcial") de  $\varphi$  en  $\mathbf{K}$ . Luego, una proposición  $\varphi$  es "más posible que sea verdadera" en  $\mathbf{K}$  que otra proposición  $\psi$ , si el conjunto de las posibilidades donde  $\varphi$  es verdadera es mayor (en el sentido de la medida) que el conjunto de las posibilidades donde  $\psi$  es verdadera.

Mi teoría también da sentido al siguiente principio, que llamo, el *Principio de Inferencia Directa*:

**Principio de Inferencia Directa.** *Sea A un suceso. Entonces el grado de apoyo para la proposición que A ocurra, dado que la probabilidad fáctica de A es r, es también r. Luego, el grado de creencia de una persona que cree que la probabilidad de A es r, si es racional, debe ser también r.*

Este principio puede ser justificado fácilmente en mi teoría. El grado de posibilidad de verdad  $\varphi$  en  $\mathbf{K}$  es lo mismo que la probabilidad fáctica del suceso A que consiste de los resultados en  $\mathbf{K}$  donde  $\varphi$  es verdadera. Ya que en este caso podemos tomar  $\varphi$  como la proposición que A ocurre, tenemos que el grado de apoyo para la proposición que A ocurra (esto es, de  $\varphi$ ) dado que la probabilidad de A es r (esta probabilidad se obtiene de  $\mathbf{K}$ ), es r. Uno también puede justificar la segunda parte del principio: si uno es racional, uno debe creer en  $\varphi$  en la misma medida que uno tiene apoyo para que  $\varphi$  sea verdadera.

### 3. La inferencia estadística como lógica

La teoría de la probabilidad, desde mi punto de vista, estudia la teoría de los modelos probabilistas  $\mathbf{K}$  mencionados en la sección anterior. En cambio, la inferencia estadística estudia la relación entre estos modelos probabilistas y proposiciones. Esta es la tesis que desarrollaremos en el resto de esta charla.

Modelos se usan tanto en lógica como en las ciencias naturales y sociales<sup>3</sup>. En las

ciencias naturales o sociales, se supone que un modelo es una representación matemática del fenómeno que se está investigando. Como la matemática puede ser formalizada en la teoría de conjuntos, esta representación matemática puede ser considerada como un sistema dentro de esta teoría. Por otra parte, en lógica matemática, un modelo de una oración de un lenguaje formal también es un sistema conjuntista, en el cual se interpreta el lenguaje de la oración, y donde la oración es verdadera o falsa de acuerdo a la interpretación. En estos sistemas, que representan mundos posibles completos, la verdad o falsedad de todas las oraciones del lenguaje está determinada. Creo que los modelos de ambas aplicaciones pueden representarse por el mismo tipo de sistemas conjuntistas.

No discutiré aquí la forma precisa de los modelos usados tanto en ciencias como en lógica, sino que supondré que tienen ciertas propiedades mínimas que son necesarias para discutir la inferencia estadística. Una explicación más completa aparece en [4, Capítulo III, 5].

Los modelos en ciencias naturales y sociales son estructuras matemáticas que reflejan ciertos aspectos de la situación real. En nuestro caso probabilista, son del tipo  $\mathbf{K}$  discutido anteriormente, donde  $\mathbf{K}$  es el conjunto de resultados realmente posibles. Se supone definida una medida de probabilidad sobre subconjuntos de  $\mathbf{K}$ , la que está determinada por la situación física de la manera explicada someramente en la sección anterior. Esta es la *probabilidad fáctica o real*.

Hay otra manera de mirar nuestra relación con la naturaleza. Usamos un lenguaje para describir la realidad. En este lenguaje, mencionamos objetos y hablamos acerca de relaciones entre ellos. Cuando nuestras aserciones corresponden con la realidad, decimos que son verdaderas. Las palabras en un lenguaje natural como el castellano pueden referirse a objetos del mundo. También comprendemos lo que significa que una proposición expresada en castellano sea verdadera o falsa: hablando informalmente, una proposición es verdadera cuando 'corresponde' a la realidad. O, como lo dice mejor Aristóteles<sup>4</sup>:

<sup>3</sup> Mi visión en esta sección está influenciada por [17]. Véase también [5].

<sup>4</sup> Aristóteles, *Metafísica*, Libro  $\Gamma$  1011b26. Para una discusión completa de la definición de verdad, véase [19].

Decir de lo que es que no es, o de lo que no es que es, es falso, mientras que decir de lo que es que es, y de lo que no es que no es, es verdadero.

Esta concepción de verdad como correspondencia con la realidad, ha sido representada matemáticamente por Tarski, [18]. Para esto, se necesita matematizar el lenguaje. Así, un lenguaje natural se aproxima por ciertas estructuras matemáticas llamadas lenguajes formales. Estos lenguajes formales son aproximaciones a los lenguajes naturales.

Tal como la interpretación de los lenguajes naturales es en la realidad, la de los lenguajes formales es en la representación matemática de la realidad. Una definición matemática de la verdad es, entonces, posible. Tarski definió una relación entre una oración  $\varphi$  y una estructura matemática  $\omega$ : la relación de  $\varphi$  ser verdadera en  $\omega$  (o, de acuerdo a la interpretación  $\omega$ ). Esta relación habitualmente se expresa simbólicamente por  $\omega \models \varphi$ , y también se lee  $\omega$  es un modelo de  $\varphi$ <sup>5</sup>.

Al estudiar lógica pura, qué estructura corresponde a la realidad no es relevante: para la lógica, todo lo que importa es la relación abstracta de verdad entre una oración de un lenguaje formal y un mundo posible representado matemáticamente.

**Las variables aleatorias como parte de un lenguaje.** Los probabilistas y estadísticos han usado tradicionalmente, para describir matemáticamente propiedades de una situación, lo que se llaman variables o cantidades aleatorias. Como se usan en teoría de probabilidad, las variables aleatorias son funciones de un espacio muestral a los números reales. Como nuestro espacio muestral consiste de un conjunto  $\mathbf{K}$  de resultados posibles, nuestras variables aleatorias son funciones,  $X$ , de  $\mathbf{K}$  en los números reales.

Los estadísticos, por su parte, usan variables aleatorias, como los lógicos usan un lenguaje artificial. Cuando realizan un test es-

<sup>5</sup> Para una explicación elemental de esta definición véase [14]. Una exposición menos elemental aparece en [6]. Una exposición informal excelente aparece en [19].

tadístico, por ejemplo, una variable aleatoria  $X$  puede tener distribuciones de probabilidad diferentes. De hecho, esto significa que puede tener interpretaciones diferentes. Supongamos que estamos midiendo una cantidad  $X$  y que sabemos que está normalmente distribuída, pero que no conocemos la media o la varianza. Así, para cada media  $\mu$  y desviación standard  $\sigma$ ,  $X$  puede ser normalmente distribuída con estos parámetros. De hecho, estas diferentes distribuciones son interpretaciones posibles de  $X$ . La variable  $X$  adquiere, así, las características de los símbolos de un lenguaje.

Podemos expresar propiedades de los resultados por medio de variables aleatorias. Por ejemplo, si queremos expresar que el resultado de la tirada de un dado es 6, podemos hacerlo por la oración  $[X = 6]$ . Otros tipos de oraciones posibles son  $[X \leq r]$  y  $[X \geq r]$ , donde  $r$  es un número real. Es claro lo que significa que una oración de una de estas formas sea verdadera en un elemento  $\omega \in \mathbf{K}$ . A saber

$[X \leq r]$  es verdadera en  $\omega$  si y sólo si  $X(\omega) \leq r$ .

Simbolizamos  $\varphi$  es verdadera  $\omega$  con  $X$  como variable aleatoria por

$$\langle \omega, X \rangle \models \varphi.$$

Podemos tener más de una variable aleatoria en nuestro lenguaje. Por ejemplo, podríamos tener  $X_1, X_2, \dots, X_n$ . Es fácil de ver cómo construir un lenguaje con estos símbolos: incluimos, además, funciones continuas de estas variables. Más precisamente, si  $g$  es una función continua de  $n$  variables reales, entonces tenemos la oración  $[g(X_1, \dots, X_n) \leq r]$  en nuestro lenguaje. Por ejemplo,  $g(X_1, \dots, X_n)$  podría ser

$$g(X_1, \dots, X_n) = \frac{1}{n} \sum_{k=1}^n X_k$$

Así, tenemos

$$\langle \omega, X_1, \dots, X_n \rangle \models [g(X_1, \dots, X_n) \leq r]$$

si y sólo si  $g(X_1(\omega), \dots, X_n(\omega)) \leq r$ .

La relación entre oraciones, modelos, y realidad, en el caso de modelos determi-

nistas que asignan verdad o falsedad a las oraciones, puede resumirse como sigue: Una oración se interpreta en un modelo, y puede ser verdadera o falsa en el modelo. El modelo, a su vez, se interpreta en la realidad, y puede ser fiel o no a la realidad. Pero las oraciones también se interpretan en la realidad, y pueden ser verdaderas o falsas en ella. La conexión entre el modelo y la realidad puede analizarse a través de esta interpretación doble de las oraciones: si una oración verdadera en el modelo es falsa en la realidad, entonces el modelo es rechazado como modelo de la realidad.

Los modelos probabilistas discutidos anteriormente también juegan este doble papel: son posibles modelos de la realidad, cuando la realidad es un dispositivo aleatorio. Por otra parte, como veremos a continuación, las estructuras probabilistas sirven como interpretaciones de un lenguaje, el que será, en nuestro caso, un lenguaje de variables aleatorias. Estas interpretaciones no determinan la verdad o falsedad de las oraciones, sino sólo sus probabilidades. Este uso de variables aleatorias es similar a su uso en inferencia estadística. En estadística, las variables aleatorias se interpretan en espacios muestrales. Habitualmente hay en consideración varios espacios muestrales, esto es, distribuciones de la variable aleatoria.

### Modelos probabilistas y lenguajes.

Hemos visto informalmente como se puede obtener un espacio de probabilidad  $\langle \mathbf{K}, \mathcal{F}_{\mathbf{K}}, \text{Pr}_{\mathbf{K}} \rangle$ , donde  $\mathbf{K}$  es un conjunto de resultados,  $\mathcal{F}_{\mathbf{K}}$  es un álgebra de subconjuntos de  $\mathbf{K}$ , y  $\text{Pr}_{\mathbf{K}}$  es una medida de probabilidad sobre  $\mathcal{F}_{\mathbf{K}}$ .

Como vimos anteriormente, estas estructuras probabilistas se consideran como posibles modelos, en el sentido de las ciencias naturales o sociales, de dispositivos aleatorios. Podemos realizar muchas mediciones diversas sobre  $\mathbf{K}$ . Una medición  $X$  para  $\mathbf{K}$  puede ser simplemente una variable aleatoria sobre  $\mathbf{K}$ . Cuando  $\mathbf{K}$  es un modelo de un dispositivo compuesto, esto es, incluye un orden causal  $T$ , entonces una medición  $Y$  para  $\mathbf{K}$  puede ser una función de  ${}^n T \times \mathbf{K}$  en los reales, para algún número natural  $n$ . Esto es, para cada  $t_1, \dots, t_n \in T$  y  $\omega \in \mathbf{K}$ ,  $Y(t_1, \dots, t_n, \omega)$  es un número real. Por sim-

plicidad, tomamos aquí  $n = 1$ . Así,  $Y(t, \omega)$  da la medición del resultado compuesto  $\omega$  en el momento causal  $t$ . En vez de escribir  $Y(t, \omega)$  escribiremos, como es habitual,  $Y_t(\omega)$ . Así,  $Y$  es un proceso estocástico sobre  $T$ , con lo que  $Y_t$  es una variable aleatoria e  $Y(\cdot, \omega)$  es una trayectoria del proceso.

Por ejemplo, si nuestra estructura compuesta representa  $n$  repeticiones independientes del mismo experimento con modelo  $\mathbf{K}$ , lo que se denota por  ${}^n \mathbf{K}$ , y  $X$  es una variable aleatoria en  $\mathbf{K}$ , entonces podemos definir la función  $X$  en  ${}^n \mathbf{K}$  por  $X(m, \omega) = X(\omega(m))$ , para  $1 \leq m \leq n$  y  $\omega \in {}^n \mathbf{K}$ . También escribimos  $X(m, \omega) = X_m(\omega)$ . Aquí,  $X_m(\omega)$  es el valor de  $\omega$  en la repetición  $m$ -ésima.

Como otro ejemplo, tomemos la estructura compuesta para el modelo de las dos urnas, donde primero se elige una urna al azar y, luego, una bola de la urna elegida. En este caso, tenemos dos momentos causales, digamos  $t$  y  $s$ , con  $t$  antes de  $s$ . En  $t$  se elige una urna y en  $s$ , una bola de esta urna. Los sucesos compuestos (que son, en realidad trayectorias) son funciones  $\omega$  sobre  $\{t, s\}$ , tales que  $\omega(t)$  es una de las urnas y  $\omega(s)$ , una bola de la urna  $\omega(t)$ . Si numeramos las urnas, podemos considerar el proceso  $Z$ , tal que  $Z(t, \omega) =$  el número de la urna elegida en  $\omega(t)$ , y  $Z(s, \omega) =$  el número de, digamos, bolas rojas elegidas en  $\omega(s)$ .

Luego, nuestros modelos probabilistas son pares  $\langle \mathbf{K}, Y \rangle$ , donde  $\mathbf{K}$  es una estructura probabilista e  $Y$  es una variable aleatoria o proceso apropiada para  $\mathbf{K}$ . La misma estructura probabilista puede servir de base a variables o procesos diferentes, y así obtenemos modelos probabilistas diferentes. Permitted, también, más de una variable o proceso. Por simplicidad, sólo describiré aquí modelos con una variable o proceso.

El lenguaje para expresar propiedades de un modelo de la forma  $\langle \mathbf{K}, Y \rangle$ , donde  $\mathbf{K}$  es una estructura compuesta con orden causal  $T$ , consiste de expresiones de la forma  $[Y_t \leq r]$  o  $[Y_t = r]$ , donde  $t \in T$  y  $r$  es un número real, y está cerrado bajo negaciones, conjunciones y disyunciones.

También incluimos funciones continuas de las variables  $Y_t$ . Así, si  $g$  es una función continua de  $n$  variables reales, entonces  $[g(Y_{t_1}, \dots, Y_{t_n}) \leq r]$  y  $[g(Y_{t_1}, \dots, Y_{t_n}) = r]$  son oraciones. La negación de una oración  $\phi$  se

denota  $\neg\phi$ , la conjunción del conjunto de oraciones  $\Phi$ , por  $\wedge\Phi$ , y la disyunción del conjunto  $\phi$ , por  $\vee\Phi$ . La conjunción de dos oraciones  $\phi$  y  $\psi$ , por  $\phi \wedge \psi$ , y su disyunción por  $\phi \vee \psi$ . Una conjunción es verdadera si todos sus términos lo son; una disyunción, si por lo menos uno de sus términos lo es, y una negación, si la oración que es negada no es verdadera.

Combinando estas expresiones, podemos decir que  $Y$  está en cierto conjunto de números reales  $B$ , para algunos conjuntos de reales  $B$ . Así, podemos tener expresiones que pueden abreviarse, con el significado obvio, por

$$[Y_1 \in B_1] \wedge [Y_2 \in B_2] \wedge \dots \wedge [Y_n \in B_n],$$

y, también

$$[g(Y_{t_1}, \dots, Y_{t_m}) \in B],$$

donde  $B, B_1, \dots, B_n$  son conjuntos de reales.

Es claro lo que significa que  $[g(Y_{t_1}, \dots, Y_{t_m}) \in B]$  sea verdadera en un sistema  $\omega$  con el proceso  $Y$ ; a saber

$\langle \omega, Y \rangle \models [g(Y_{t_1}, \dots, Y_{t_m}) \in B]$  si y sólo si

$$g(Y_{t_1}(\omega), \dots, Y_{t_m}(\omega)) \in B.$$

Por ejemplo, supongamos que el modelo  $\mathbf{K}$  es un posible modelo para la tirada de una moneda, y  $X(\omega) = 1$ , si  $\omega$  es una tirada con resultado de cara, y  $X(\omega) = 0$ , si es sello. Entonces  $\mathbf{K}$  representa  $n$  tiradas independientes de la moneda. Sea  $X$  definida como antes por  $X(m, \omega) = X(\omega(m))$  para  $1 \leq m \leq n$  y  $\omega \in \mathbf{K}$ . Sea también  $X_m(\omega) = X(m, \omega)$  e  $Y_m(\omega) = \sum_{k=1}^m X_k(\omega)$ . Entonces  $Y_m(\omega)$  es el número de caras en las  $m$  primeras tiradas, para  $1 \leq m \leq n$  y  $\omega \in \mathbf{K}$ . Tenemos que  $\langle \omega, Y \rangle \models [Y_m \leq 6]$  si y sólo si el número de caras en la sucesión  $\omega$  restringida a  $m$  es menor o igual a 6.

La variable  $\text{Fr}_m = (1/m) Y_m$  representa la frecuencia relativa de caras en las primeras  $m$  tiradas. Por ejemplo

$$\langle \omega, Y \rangle \models [|\text{Fr}_m - \frac{1}{2}| \geq \delta]$$

$$\text{si y sólo si } |\text{Fr}_m(\omega) - \frac{1}{2}| > \delta.$$

Esta definición de verdad fija una interpretación de las oraciones en los resultados  $\omega$ . Pero necesitamos una interpretación en los modelos probabilistas  $\langle \mathbf{K}, Y \rangle$ . Cada oración  $\phi$  se interpreta  $\langle \mathbf{K}, Y \rangle$  por el conjunto de resultados en  $\mathbf{K}$  en los cuales  $\phi$  es verdadera. Esto es, definimos

$\text{Mod}_{\langle \mathbf{K}, Y \rangle}(\phi) =$  el conjunto de los  $\omega \in \mathbf{K}$  tales que  $\langle \omega, Y \rangle \models \phi$ .

Ahora bien,  $\langle \mathbf{K}, Y \rangle$  no asocia, en general, verdad o falsedad a las oraciones, sino sólo probabilidad. Exigimos que  $\text{Mod}_{\langle \mathbf{K}, Y \rangle}(\phi) \in \mathcal{F}_{\mathbf{K}}$  y definimos

$$\text{Pr}_{\langle \mathbf{K}, Y \rangle}(\phi) = \text{Pr}_{\mathbf{K}}(\text{Mod}_{\langle \mathbf{K}, Y \rangle}(\phi)).$$

Así, a una oración se le asigna como probabilidad la medida del conjunto de resultados donde es verdadera, esto es, la probabilidad de una proposición es la medida del conjunto de posibilidades que la hacen verdadera.

Como ejemplo, discutiremos el dispositivo de una ruleta que se hace girar aplicando una fuerza constante, pero con posiciones iniciales escogidas al azar. Para tratar de ruletas con diferentes cargas, introducimos un lenguaje con una variable aleatoria, digamos  $X$ . Sean  $u$  y  $v$  dos números reales entre 0 y  $2\pi$ . La oración  $[u \leq X \leq v]$  expresa el hecho que la posición final de la ruleta está en el arco entre  $e^{iu}$  y  $e^{iv}$ . La probabilidad asignada a esta oración depende de la estructura. Supongamos que  $\langle \mathbf{C}, X_1 \rangle$  es cualquiera de estas estructuras. Primero recordamos la definición de verdad. Decimos que

$$\langle \omega_c, X_1 \rangle \models [u \leq X \leq v], \text{ si } u \leq X_1(\omega_c) \leq v,$$

donde  $\omega_c$  es el resultado con posición inicial  $c$ .

Tenemos, de acuerdo a nuestra definición

$$\text{Pr}_{\langle \mathbf{C}, X_1 \rangle} [u \leq X \leq v] =$$

$$= \text{Pr}_{\mathbf{C}}(\text{Mod}_{\langle \mathbf{C}, X_1 \rangle}([u \leq X \leq v])),$$

donde  $\text{Mod}_{\langle \mathbf{C}, X_1 \rangle}([u \leq X \leq v])$  es el conjunto de  $\omega_c$  tales que  $u \leq X(\omega_c) \leq v$  (esto es, tales que  $\langle \omega_c, X_1 \rangle \models [u \leq X \leq v]$ ).

Así

$$\Pr_{(C, X_1)} [u \leq X \leq v] = \Pr_C (\{c: u \leq X_1(\omega_c) \leq v\}),$$

donde,  $\Pr_C$  es la medida de Lebesgue dividida por  $2\pi$ .

Luego, si  $X_1$  es la variable para una ruleta simétrica, la probabilidad de esta oración es  $(v - u)/2\pi$ , porque el conjunto de los  $c$  tal que  $u \leq X_1(\omega_c) \leq v$  tiene longitud  $v - u$ .

Por otra parte, si  $(C, X_2)$  representa una ruleta asimétrica, entonces la longitud del intervalo que consiste de los  $c$  tal que  $u \leq X_2(\omega_c) \leq v$  puede ser diferente de la longitud del intervalo  $[u, v]$ , y, luego,

$$\Pr_{(C, X_2)} [u \leq X \leq v]$$

podría no ser  $(v - u)/2\pi$ .

Por esto, a la misma oración se le puede asignar probabilidades muy diferentes por modelos distintos. Creo que este procedimiento es similar al usado en estadística, donde las distintas distribuciones posibles le asignan diferentes probabilidades a una variable aleatoria.

Como otro ejemplo, supóngase que  $\mathbf{K}$  es un modelo para un número infinito de tiradas de una moneda simétrica, y sea  $\text{Fr}_m(\omega)$ , como antes, la frecuencia relativa de caras en las primeras  $m$  tiradas. Entonces, la ley débil de los grandes números demuestra que

$$\lim_{n \rightarrow \infty} \Pr_{(\mathbf{K}, Y)} [|\text{Fr}_n - \frac{1}{2}| \geq \alpha] = 0,$$

para cualquier  $\alpha > 0$ .

Vemos que dos modelos probabilistas distintos, que pueden ser interpretaciones de la misma oración  $\varphi$ , podrían asignar probabilidades muy diferentes a  $\varphi$ . Por ejemplo, si  $\mathbf{J}$  es un modelo para un número infinito de tiradas de una moneda no simétrica el

$$\lim_{n \rightarrow \infty} \Pr_{(\mathbf{J}, Y)} [|\text{Fr}_n - \frac{1}{2}| \geq \alpha] \neq 0,$$

para algunos  $\alpha > 0$ .

Esta es la situación habitual en la inferencia estadística clásica. Tenemos varios modelos alternativos que asignan probabilidades diferentes a las mismas oraciones. Como discutiremos más adelante, estas probabilidades distintas se usan para decidir

entre los diferentes posibles modelos. Por ejemplo, supóngase que estamos tirando una moneda y que no sabemos si es simétrica o no. Entonces los modelos posibles representan monedas con cargas diferentes. Estos modelos asignan probabilidades diferentes a las oraciones  $[a \leq |\text{Fr}_n - p|]$ , donde  $p$  es la probabilidad de cara de cierta moneda. Las probabilidades de estas oraciones nos ayudan a decidir cuál es la carga de la moneda.

Como último ejemplo, volvamos al modelo de las dos urnas,  $\mathbf{U}$ , y describamos un lenguaje que podríamos interpretar en él. Podríamos tener una variable aleatoria simple  $X$ , tal que, por ejemplo,  $X(\xi) = 1$ , si la bola elegida en el resultado  $\xi$  es blanca, y  $X(\xi) = 0$ , si la bola es negra. En este lenguaje, tenemos, entonces, oraciones de la forma  $[X = 1]$  y  $[X = 0]$ . Estructuras probabilistas distintas  $\langle \mathbf{U}_1, X_1 \rangle$  y  $\langle \mathbf{U}_2, X_2 \rangle$  pueden asignar probabilidades diferentes a estas oraciones.

Pero supongamos que nos gustaría hablar de la urna elegida y la bola elegida. Necesitamos, entonces, un proceso, digamos  $Y$ , que es una función de dos variables. Si  $\xi \in \mathbf{U}$ , entonces podríamos tener  $Y(t, \xi) = 1$  si la urna  $u_1$  es elegida en  $\xi$ , e  $Y(t, \xi) = 0$ , si la urna  $u_2$  es elegida. Para las bolas, podríamos tener  $Y(s, \xi) = 1$  para una bola blanca e  $Y(s, \xi) = 0$ , si es negra. El lenguaje, ahora, además de  $Y$  incluye las constantes  $t$  y  $s$ , para los momentos causales. Las oraciones básicas son de las formas  $[Y_t = i]$  e  $[Y_s = i]$  para  $i = 1$  ó  $0$ . Las probabilidades se asignan como antes. Por ejemplo

$$\Pr_{(\mathbf{U}, Y)} [Y_t = 1] = \Pr_{\mathbf{U}} (\{\xi \in \mathbf{U}: Y(t, \xi) = 1\}).$$

Como en los ejemplos anteriores, puede haber varias estructuras que asignan probabilidades diferentes a cada oración.

#### 4. Reglas de aceptación y rechazo

Al estudiar probabilidad pura (lo que se puede llamar lógica probabilista), relativizamos la probabilidad a conjuntos diferentes  $\mathbf{K}$  de resultados posibles. Cada  $\mathbf{K}$  es un modelo, que, en vez de determinar el valor veritativo de una proposición, abre sus posibilidades, y, al hacerlo, define su probabilidad. Para aplicar la teoría, debe suponerse que  $\mathbf{K}$  es una descripción fiel de las posibi-

lidades fácticas en el mundo real. Esto es, debemos aceptar a **K** como verdadero. Tal como para los modelos que determinan la verdad o falsedad de las proposiciones relevantes, hay reglas para el rechazo o la aceptación de **K** como una representación del mundo real. Discutiremos estas reglas a continuación.

En esta discusión, escribiré **K** para los modelos probabilistas, aunque, si fuéramos más precisos deberíamos escribir  $\langle \mathbf{K}, Y \rangle$ , donde  $Y$  es una variable aleatoria o un proceso estocástico. A menos que haya peligro de confusión, omitiré la  $Y$ , la que será obvia del contexto.

En la inferencia estadística, el principal problema es la aceptación de hipótesis como verdaderas o su rechazo como falsas sin tomar en cuenta la deseabilidad de las consecuencias de la aceptación o el rechazo. La inferencia estadística clásica da reglas para el rechazo o la aceptación de hipótesis, pero estas reglas no se basan en las probabilidades de las hipótesis. Creo que el principio que se formulará aquí es la base para estas reglas. Al igual que en la inferencia estadística clásica, este principio basa la aceptación y el rechazo de una hipótesis, no en su probabilidad, sino en la probabilidad asignada de acuerdo a la hipótesis a otras proposiciones.

Comenzaré analizando reglas para el rechazo. La regla para el rechazo de una hipótesis determinista  $h$  (esto es, una que determina un resultado único) es muy simple. Supóngase que se realiza un experimento. Si se obtiene un resultado que sería falso si la hipótesis fuera verdadera (esto es, que es falso bajo la hipótesis), entonces debemos rechazar  $h$ <sup>6</sup>. Desde el punto de vista de Popper, [15], lo que caracteriza a las aserciones de la ciencia es su posible falsificación.

En vez de considerar las hipótesis deterministas, discutimos ahora las probabilistas.

<sup>6</sup> Como es bien sabido, estoy simplificando demás el rechazo de hipótesis. Es posible salvar una hipótesis, a pesar de la falsedad de una de sus consecuencias  $e$ , cambiando otros aspectos de la relación de la hipótesis con la experiencia, y, así,  $e$  puede dejar de ser una consecuencia. Aquí, y en el resto de la conferencia, continuaré haciendo esta simplificación.

La falsificación de hipótesis probabilistas ha sido un problema para el punto de vista de Popper mencionado antes. Creo que aquí estoy presentando una solución a este problema, que es consistente con las ideas de Popper. El problema para la falsificación estricta de hipótesis probabilistas es que de una tal hipótesis no podemos deducir proposiciones interesantes que pueden ser verificadas empíricamente, sino sólo asignarles probabilidades. Por ejemplo, de la suposición que la probabilidad de cara en la tirada de una moneda es  $1/2$ , no podemos deducir ninguna aserción sobre frecuencias relativas, sino sólo asignarles probabilidades. Así, por ejemplo, no se puede deducir que la frecuencia relativa de caras en 100 tiradas es  $1/2$ , sino sólo que esta frecuencia tiene una cierta probabilidad alta. Por esto, si la frecuencia relativa obtenida realmente no es  $1/2$ , no podemos concluir que la hipótesis no es válida. Como hemos dicho antes, una aserción que tiene probabilidad baja significa que su posibilidad de verdad es baja. Así, si ocurre en la realidad una aserción con probabilidad baja de acuerdo a la hipótesis, entonces hay evidencias que la hipótesis es falsa<sup>7</sup>. La interpretación de la probabilidad de Popper como propensión no incluye una conexión con verdad, y, luego, no tiene disponible esta justificación de la falsificación de hipótesis probabilistas.

Una hipótesis probabilista es, en mi versión, la proposición que el conjunto de resultados posibles, digamos **K**, es un modelo adecuado de la realidad, esto es, del dispositivo aleatorio en cuestión. Brevemente, llamaré a esta hipótesis simplemente **K**. Esta hipótesis **K** determina una distribución de probabilidad sobre los sucesos de un experimento. Esta distribución asigna el grado de apoyo a las diferentes proposiciones, dado **K**.

Es claro que si una proposición falsa en **K** fuera verdadera en la realidad, deberíamos rechazar **K**. Esto es, si el resultado del experimento fuera imposible según **K**, deberíamos rechazar **K**. Pero es muy raro que los experimentos considerados admitan resulta-

<sup>7</sup> La situación es más complicada que esto, como se verá más adelante.

dos que son imposibles de acuerdo a una hipótesis probabilista  $\mathbf{K}$ . Así, un procedimiento natural sería introducir la noción de aproximadamente falso y, entonces, rechazar  $\mathbf{K}$ , si determináramos que una proposición aproximadamente falsa en  $\mathbf{K}$  fuera verdadera en la realidad. Por conveniencia de expresión, llamaré a un suceso *inverosímil*<sup>8</sup>, si la proposición que expresa su obtención en la realidad es aproximadamente falsa en este sentido. Luego, podemos parafrasear el principio "rechace  $\mathbf{K}$ , si se obtiene una proposición aproximadamente falsa" por "sucesos inverosímiles no ocurren"<sup>9</sup>. Pero, ¿qué es un suceso inverosímil? Como considero a la probabilidad de verdad, es tentador decir que una proposición  $\phi$  es aproximadamente falsa en  $\mathbf{K}$ , si  $\phi$  tiene probabilidad baja en  $\mathbf{K}$ , esto es, si su grado de apoyo dado  $\mathbf{K}$  es bajo. Con esto, podríamos decidir que si ocurre un resultado para el cual  $\mathbf{K}$  determina probabilidad baja, entonces deberíamos rechazar  $\mathbf{K}$ . Esta regla, sin embargo, no funciona:

Supóngase que el dispositivo aleatorio para el cual  $\mathbf{K}$  es un modelo es una lotería similar a las discutidas antes. La obtención de cualquier boleto particular tiene probabilidad muy baja de acuerdo a este modelo  $\mathbf{K}$ , pero un tal resultado no es razón en absoluto para rechazar  $\mathbf{K}$ .

Los próximos párrafos contienen un análisis de la noción de resultado inverosímil. La primera parte del análisis es el hecho que para rechazar  $\mathbf{K}$ , debemos tener algunas razones para cuestionarlo, a saber, debemos tener en mente algunos modelos alternativos para el dispositivo aleatorio. Así, aun hablar de proposiciones aproximadamente falsas según  $\mathbf{K}$  exige la consideración de alternativas a  $\mathbf{K}$ . Nos vemos forzados a tratar con clases de hipótesis alternativas  $\mathbf{K}_i$ , para cada  $i$  en un cierto conjunto de índices  $I$ . Estas son todas las hipótesis que consideramos epistémicamente posibles para la situación involucrada. Todas estas hipótesis son

compatibles con lo que aceptamos como las leyes del fenómeno en cuestión. Es claro que no es que  $\mathbf{K}_i$  pueda suceder, sino sólo que es compatible con las leyes de la naturaleza que  $\mathbf{K}_i$  es el modelo correcto para la situación que estamos considerando. Como decidir cuál es la clase adecuada de hipótesis alternativas no se hace, en general, por consideraciones estadísticas o probabilistas. Puede haber leyes naturales generales o hechos particulares que son relevantes para esta decisión. En cualquier caso, debemos aceptar como verdadero que la clase de los  $\mathbf{K}_i$  para  $i$  en  $I$ , contiene todos los modelos que son posibles en la situación dada.

Ahora podemos mejorar la definición de una proposición aproximadamente falsa. Una proposición  $\phi$  es aproximadamente falsa en  $\mathbf{K}_i$ , relativa a la clase  $\mathbf{K}_i$ , para  $i$  en  $I$ , si  $\phi$  tiene probabilidad mucho más baja en  $\mathbf{K}_j$ , que en un cierto  $\mathbf{K}_i$ , con  $i$  en  $I$  e  $i \neq j$ . Sin embargo, esta definición, como veremos más adelante, no funciona, así que sólo la aceptamos provisoriamente, como base de discusión.

No es posible, en general, tomar la familia de todas las hipótesis lógicamente posibles como nuestro conjunto de alternativas: Supóngase que tiramos una moneda diez veces. Una sucesión  $x = 0100101110$  (donde 0 es para cara, y 1, para sello), se considera a favor de la hipótesis, llamémosla  $\mathbf{K}_{1/2}$ , que la moneda es simétrica, si las alternativas son que la moneda está cargada con diferentes cargas. Sin embargo, es aproximadamente falsa en  $\mathbf{K}_{1/2}$ , si una de las posibilidades es "la sucesión está generada por una máquina que produce exactamente  $x$ ". Esta última hipótesis es ciertamente lógicamente posible, por lo que vemos, que un cambio en las alternativas produce un cambio de la propiedad de ser aproximadamente falso.

Además de ser necesaria para una regla natural de rechazo, la idea que la consideración de un conjunto de alternativas afecta la aceptabilidad de una hipótesis, es intuitivamente atractiva. Por ejemplo, antes de la formulación de la teoría de la relatividad de Einstein, era muy difícil rechazar la mecánica Newtoniana. Sin embargo, después de la formulación de su alternativa, la mecánica clásica fue casi inmediatamente rechazada.

<sup>8</sup> En vez de *inverosímil*, podríamos usar el neologismo *implausible*, pero creo que *inverosímil* es mejor, porque un resultado *inverosímil* es uno que no parece ser verdadero. En inglés se dice "unlikely".

<sup>9</sup> Véase [2].

Sin embargo, para declarar a  $K_j$  falsa, sólo la ocurrencia de un suceso inverosímil no es suficiente. No estaríamos satisfechos, en general, con sólo un tal suceso. Podemos en el caso que ocurriera un suceso inverosímil, rechazar provisoriamente  $K_j$ , pero deberíamos poder revisar este juicio. Así, debemos tener a la mano una sucesión de sucesos inverosímiles, cuyas probabilidades se aproximan a cero. Si tuviéramos una tal sucesión, entonces, en principio, podríamos obtener, con suficiente esfuerzo o tiempo, la ocurrencia de un resultado tal que su probabilidad sea tan baja como se quiera, y, así, podríamos revisar nuestro juicio. Pero, como antes, necesitamos considerar alternativas. Luego, necesitamos ser capaces de realizar<sup>10</sup> una sucesión  $E_n$  de experimentos, o simplemente observaciones, tal que para cada  $j \in I$ , si  $K_j$  es verdadera, una sucesión  $a_n$  de resultados de estos experimentos ocurre eventualmente casi seguramente, satisfaciendo que para cada  $i \in I$  con  $i \neq j$ , la probabilidad en  $K_i$  del resultado  $a_n$  tiende a cero. Tenemos que exigir que si  $K_j$  es verdadero, entonces una tal sucesión  $a_n$  ocurra eventualmente casi seguramente. Esta sucesión de experimentos, la que será definida precisamente más adelante, se llamará un *experimento discriminante* para el sistema  $K_i$  con  $i \in I$ . Abreviaremos la proposición 'el resultado de  $E_n$  es  $a_n$ ' por  $[E_n = a_n]$ .

La importancia de tratar con sucesiones de experimentos cuyas probabilidades se aproximan a cero, no es que podemos obtener de este modo proposiciones con probabilidad baja, sino con probabilidades arbitrariamente bajas. Esto es, si esperamos suficiente tiempo o gastamos suficiente esfuerzo, podemos obtener casi seguramente una proposición con probabilidad tan baja como queramos. De este modo, podemos poner en movimiento un proceso dialéctico: una persona fija la probabilidad baja con la cual estaría satisfecha, y el oponente (que puede ser la misma persona) prueba la sucesión de experimentos hasta que obtiene el nivel correcto. Para tener un proceso dialéctico real para rechazar  $K_j$ , si  $K_m$  es verdadera para

algún  $m \neq j$ , debe ocurrir casi seguramente una sucesión que determina oraciones cuya probabilidad bajo  $K_j$  tiende a cero. Esto es, el conjunto de sucesiones que rechaza a  $K_j$  debe tener, si  $K_m$  es verdadera, probabilidad infinitamente cerca de uno. En la práctica uno sólo es capaz de llegar a un cierto nivel finito, esto es, uno puede probar  $[E_n = a_n]$  sólo hasta un cierto  $n$  finito. Esto da cuenta del hecho que, aunque podemos rechazar una hipótesis y considerarla falsa, este rechazo es provisorio. Si se obtiene nuevas proposiciones con probabilidad alta, podríamos estar forzados a aceptarla de nuevo<sup>11</sup>.

Las reglas de rechazo que hemos estado considerando son asimétricas en el siguiente sentido. Para rechazar las hipótesis deterministas, no necesitamos hipótesis alternativas, mientras que para las probabilistas, sí las necesitamos. Para las reglas de aceptación de hipótesis, la simetría se restaura. No podemos aceptar una hipótesis determinista  $h$ , sólo porque una proposición  $\phi$  implicada por  $h$  ha ocurrido en realidad. Estamos obligados a considerar hipótesis alternativas  $h_i$ , para  $i$  en un conjunto de índices  $I$ . Si una proposición  $\phi$  es verdadera en la realidad, siendo también verdadera de acuerdo a  $h_j$ , pero falsa en todas las  $h_m$ , para  $m \neq j$ , entonces  $h_j$  debe ser aceptada en vez de  $h_m$ , con  $m \neq j$ . Así, aceptamos  $h_j$ , si podemos rechazar todas las  $h_m$ , con  $m \neq j$ , pero no  $h_j$ .

Una de las reglas para la aceptación de una hipótesis no-determinista  $K$  es similar. En general, no es posible, de acuerdo a esta regla, aceptar un modelo particular  $K$ , sino una clase de tales modelos, digamos los  $K_j$ , para  $j \in J$ , donde  $J$  es un subconjunto de  $I$ . Esta regla dice que aceptamos provisoriamente tal clase, si podemos rechazar provisoriamente todos los modelos que no están en la clase, esto es, todos los  $K_i$ , para  $i \in I - J$ .

Tal como en el caso del rechazo, no es suficiente para la aceptación tener una proposición aproximadamente verdadera. Lo que necesitamos es una sucesión discriminante de experimentos  $E_n$ , tal que la verdad o falsedad de  $\phi$  depende del resultado de un

<sup>10</sup> Nótese que debemos ser capaces de realizar, y no realizar.

<sup>11</sup> Cf. [13, Capítulo V].

elemento de la sucesión. Para que el proceso dialéctico sea posible, los experimentos discriminantes deben tener las dos condiciones mencionadas más arriba. Idealmente, la verdad en la realidad de una sucesión infinita de proposiciones  $[E_n = a_n]$  determinaría la verdad de una hipótesis. En la práctica, debemos conformarnos con una sucesión finita que se aproxima a la verdad, y, por tanto, nuestra aceptación es sólo provisoria.

Además de tener un experimento discriminante, para rechazar  $K_j$ , el experimento  $E_n$ , para algún  $n$ , debe tener un resultado que es inverosímil de acuerdo a  $K_j$ . Sin embargo, la noción resultado inverosímil para  $K_j$  como un resultado que tiene probabilidad baja de acuerdo a  $K_j$  y alta, de acuerdo a una alternativa no funciona: supóngase que tiramos una moneda 10 veces, y que ocurre la sucesión 0010000000. Intuitivamente, si las alternativas son los modelos  $K_p$ , donde  $p$  es la probabilidad de cara, este resultado es inverosímil de acuerdo a las hipótesis  $K_{1/2}$  que la moneda es simétrica. Sin embargo, la sucesión tiene probabilidad baja según cualquiera de las hipótesis.

Procedo, ahora, a introducir la noción de resultado inverosímil que considero correcta. Nuestra noción resultado inverosímil puede ejemplificarse como sigue: Supóngase que tiramos una moneda 10 veces y anotamos la frecuencia de caras. Dado que la moneda es simétrica, ¿qué es un "resultado inverosímil del experimento"? El resultado de exactamente 5 caras tiene probabilidad baja, pero no lo llamaríamos inverosímil. Supongamos que se obtienen 6 caras. Llamamos a este resultado inverosímil (con respecto a una moneda simétrica), si el suceso de obtener por lo menos 6 caras o a lo sumo 4, tiene probabilidad baja. Esto es, el suceso de 6 caras o peor (esto es, menos probable de acuerdo a la hipótesis que la moneda es simétrica) tiene probabilidad baja. Usaremos esta idea acerca de sucesos inverosímiles para nuestra regla de rechazo.

Pasamos ahora a delimitar más precisamente la noción de un resultado inverosímil. Para hacer esto, necesitamos introducir dos nociones: la noción de 'equivalencia evidencial', y la noción de 'resultado peor'. Suponemos que las hipótesis alternativas son  $K_i$ , para  $i \in I$ . Comenzamos con la primera de

estas nociones. Nótese el siguiente hecho, el que está detrás del principio de suficiencia de la inferencia estadística. Supóngase que haya dos resultados posibles de  $E_n$ ,  $a$  y  $b$ , que satisfacen la condición: existe un  $c > 0$  tal que para todo  $i \in I$

$$\Pr_{K_i} [E_n = a] = c \Pr_{K_i} [E_n = b]. \quad (1)$$

Esto significa que para todo  $i, j \in I$

$$\frac{\Pr_{K_i} [E_n = a]}{\Pr_{K_j} [E_n = a]} = \frac{\Pr_{K_i} [E_n = b]}{\Pr_{K_j} [E_n = b]}, \quad (2)$$

siempre que  $\Pr_{K_j} [E_n = a] \neq 0$  y  $\Pr_{K_j} [E_n = b] \neq 0$ .

Un ejemplo de esta situación es la siguiente. Supóngase que tiramos una moneda 60 veces y que las hipótesis son que la moneda tiene diferentes cargas. Dos sucesiones diferentes con el mismo número de caras, tienen la misma probabilidad bajo cualquiera de las hipótesis. Así, (1) es verdadera para este caso con  $c = 1$ .

Un modo natural de medir el grado en que una probabilidad es menor que otra es su razón. Así, que

$$r = \frac{\Pr_{K_j} (\varphi)}{\Pr_{K_i} (\varphi)}$$

sea baja para un cierto  $i$  en  $I$  da evidencia contra  $K_j$ . No es necesario exigir que  $\Pr_{K_j} (\varphi)$  sea baja y  $\Pr_{K_i} (\varphi)$  sea alta, sino sólo que su razón sea baja. Por ejemplo, si tuviéramos que  $\varphi$  fuera imposible bajo  $K_j$ , pero posible bajo  $K_i$ , rechazaríamos  $K_j$ , aunque la probabilidad bajo  $K_i$  fuera muy baja.

Por esto, en una situación como (1),  $a$  y  $b$  no discriminan entre los diferentes modelos. Esto es, la evidencia que se dan con respecto a los modelos es la misma. Decimos, en el caso en que (1) sea satisfecha, que  $a$  y  $b$  son *evidencialmente equivalentes con respecto a  $E_n$  y los sistemas  $K_i$ , para  $i \in I$* . En el ejemplo de la moneda, entonces, cualquier par de sucesiones con el mismo número de caras son evidencialmente equivalentes. Es claro que la relación de equivalencia evidencial es una relación de equivalencia entre resultados posibles. Escribimos la cla-

se de equivalencia de  $a$ ,  $[a]$ . Ya que todos los elementos de una clase de equivalencia tienen la misma importancia evidencial, tenemos que considerarlos juntos. Luego, podríamos reemplazar el experimento  $E_n$  por otro  $E'_n$  tal que  $E'_n = [a]$  en vez de  $E_n = a$ . En el caso de la moneda, el nuevo experimento, en vez de tener como resultado las sucesiones mismas, tendría el número de caras en la sucesión. Nótese que  $[E'_n = [a]]$  puede también escribirse como  $[E_n \in [a]]$ . En la práctica estadísticas, se hace habitualmente el reemplazo de  $E$  por  $E'$ .

Definimos, ahora, a la segunda noción que anunciamos. Llamamos un resultado posible  $b$  por lo menos tan malo para  $K_j$  como un resultado posible  $a$ , en símbolos  $b \leq a$ , si  $\Pr_{K_j}[E_n \in [b]] \leq \Pr_{K_j}[E_n \in [a]]$  y existe un  $i \in I$  donde la desigualdad se invierte. Entonces, para el rechazo de  $K_j$ , un experimento parcial  $E_n$  debe tener un resultado  $a$  tal que la probabilidad bajo  $K_j$  de que  $E_n$  tenga el valor  $a$  o cualquier resultado por lo menos tan malo como  $a$  es baja, digamos menor que un cierto  $\alpha$ . Así, llegamos a la noción de "resultado inverosímil": un resultado  $a$  de un experimento  $E_n$  es *inverosímil para  $K_j$* , si la probabilidad del suceso que  $E_n$  tenga el valor  $a$  o cualquier valor por lo menos tan malo para  $K_j$  como  $a$ , es baja según  $K_j$ . Creo que ésta es una caracterización natural de resultados inverosímiles.

En el caso que para cualquier par de resultados  $a$  y  $b$ , si  $\Pr_{K_j}[E_n = b] \leq \Pr_{K_j}[E_n = a]$  haya siempre otro  $K_j$  donde se invierte la desigualdad, entonces podemos simplificar la definición de  $b \leq a$  eliminando la cláusula que la desigualdad se invierta. De hecho, como veremos en algunos ejemplos más adelante, en estos casos, es como si la hipótesis  $K_j$  se prueba aisladamente, esto es, sin considerar alternativas. Estos casos son frecuentes.

**El principio de inferencia inversa.** El principio que hemos estado discutiendo es un principio de inferencia inversa, en el sentido que argüimos desde la evidencia a la hipótesis (esto es, a la aceptación de un cierto modelo), y no desde el modelo a los hechos particulares, como en el Principio de Inferencia Directa. Llamaré a este nuevo principio *Principio de Inferencia Inversa*. De

acuerdo al Principio de Inferencia Inversa, consideramos modelos alternativos para el dispositivo aleatorio en cuestión. Cada modelo es una de las hipótesis epistémicamente posibles que pueden explicar el dispositivo. Los modelos alternativos no se consideran como posibilidades reales por lo que no se les asigna probabilidades. Una de las hipótesis, digamos  $K$ , es rechazada (esto es, considerada falsa), si hay un experimento discriminante  $E_n$  para el conjunto de hipótesis alternativas que tiene un resultado  $a_n$ , tal que la probabilidad de  $a_n$  o peor es baja según  $K$ .

Una parte esencial de las reglas que hemos discutido es el hecho que las hipótesis alternativas deben ser decidibles por una sucesión de experimentos con las propiedades indicadas arriba. Aunque podríamos no usar en el hecho esta sucesión, debe existir la posibilidad de emplearla en el caso en que nos objeten nuestro rechazo (o aceptación). Debería ser posible, entonces, pedir al que nos objeta que fije un nivel de probabilidad  $\alpha$  que sea lo suficientemente bajo para él, y luego realizar el experimento hasta encontrar un resultado  $a$  tal que el suceso que contiene los resultados por lo menos tan malos como  $a$  para la hipótesis tenga probabilidad menor o igual que  $\alpha$ . Creo que éste es el principio intuitivo detrás de la inferencia estadística clásica.

Daremos a continuación una versión más precisa del Principio de Inferencia Inversa. Haremos esto definiendo más precisamente el concepto de un experimento discriminante para un sistema de modelos probabilistas alternativos posibles  $K_j$ ,  $i \in I$ .

Los resultados de los experimentos pueden ser cualquier cosa. Por simplicidad, consideraremos que todos los resultados de los experimentos son números reales o  $n$ -tuplas de números reales. Así, las oraciones que consideraremos son de la forma  $[E_n = r]$ , donde  $r$  es un número real, o  $[X_1 = r_1] \wedge [X_2 = r_2] \wedge \dots \wedge [X_n = r_n]$ , que puede escribirse  $[E_n = r]$ . Como no todos los espacios de probabilidad son discretos, también necesitamos oraciones donde desigualdades reemplazan la relación de igualdad. Esto es, oraciones de las formas  $[E_n \leq r]$ ,  $[E_n < r]$ ,  $[r \leq E_n \leq s]$ , etc., donde  $r$  y  $s$  pueden ser números reales o  $n$ -tuplas de números reales. Estas son

precisamente las oraciones básicas o conjunciones de oraciones básicas de los lenguajes que introdujimos en la sección anterior.

Ahora discutimos en forma más precisa los modelos de repeticiones de dispositivos aleatorios. Estamos en una situación en la que necesitamos tener la posibilidad de  $n$  repeticiones del dispositivo, para un conjunto no acotado de números naturales  $n$ . Modelaremos estas repeticiones por infinitas repeticiones, una para cada número natural. Por ejemplo, si el modelo del dispositivo es  $\mathbf{K}$  y las repeticiones son independientes, el modelo para  $n$  repeticiones independientes es el conjunto de  $n$ -tuplas de elementos de  $\mathbf{K}$ , denotado por  ${}^n\mathbf{K}$ . El modelo para infinitas repeticiones es el conjunto de las sucesiones infinitas de elementos de  $\mathbf{K}$ , que denotaremos  ${}^{\infty}\mathbf{K}$ .

Para la interpretación de las oraciones  $[E_n = r]$  en un modelo probabilista  $\mathbf{K}$ , debemos tener que  $E_n$  es una variable aleatoria sobre  ${}^n\mathbf{K}$ . Habitualmente cuál es el sistema de variables aleatorias  $X$  estará claro del contexto, en particular, de la oración  $\phi$ . Cuando sea claro del contexto, escribiremos  $\text{Pr}_{\mathbf{K}_j}$  en vez de  $\text{Pr}_{{}^n\mathbf{K}_j}$  o  $\text{Pr}_{{}^{\infty}\mathbf{K}_j}$ .

Para simplificar la definición de un experimento consideraremos en esta definición oraciones de la forma  $[E_n = r]$  con una nueva interpretación: una oración  $[E_n = r]$  se interpreta, ahora, como que afirma  $[r - dr \leq E_n \leq r + dr]$ , donde  $dr$  es un infinitésimo positivo<sup>12</sup>. En el caso de distribuciones continuas, la densidad  $f(r)$  puede reemplazar la probabilidad de  $[E_n = r]$  en muchos casos. Es bien sabido que si  $f_j$  es la densidad asociada a la distribución  $\text{Pr}_{\mathbf{K}_j}$ , entonces

$$f_j(r) \cdot dr = \text{Pr}_{\mathbf{K}_j} [r - dr \leq E_n \leq r + dr].$$

Así, con la interpretación dada arriba

<sup>12</sup> Esto se puede hacer formalmente usando análisis no-standard. Seré bastante informal aquí en el uso de infinitésimos. Una versión no-standard rigurosa de todas las nociones que veremos a continuación aparece en [4, Capítulos IV y XVIII].

$$\frac{\text{Pr}_{\mathbf{K}_j} [E_n = r]}{\text{Pr}_{\mathbf{K}_j} [E_n = s]} \approx \frac{f_j(r)}{f_j(s)}$$

y

$$\text{Pr}_{\mathbf{K}_i} [E_n = r] < \text{Pr}_{\mathbf{K}_i} [E_n = s]$$

si y sólo si  $f_i(r) < f_i(s)$ .

Para espacios muestrales discretos, una oración de la forma  $[E_n = r]$  es la misma con las interpretaciones vieja y nueva.

Escribiremos, como es habitual,  $[E_n \in A]$ , donde  $A$  es un conjunto de números reales, para decir que existe un  $r \in A$  tal que  $[E_n = r]$ . También, si  $A$  es un conjunto de sucesiones infinitas de números reales,  $[E \in A]$  dice que existe una sucesión  $a_n$ ,  $n \in \mathbb{N}$ , en  $A$ , tal que  $[E_n = a_n]$  para todo  $n \in \mathbb{N}$ . Esto es,  $[E \in A]$  representa la proposición que dice que la sucesión de resultados del experimento  $E$  está en el conjunto  $A$ . Como es habitual, decimos en este caso que  $E$  está *casi seguramente* en  $A$  según un modelo  $\mathbf{K}_j$  si el complemento del suceso  $[E \in A]$  es un conjunto nulo, según  $\text{Pr}_{\mathbf{K}_j}$ .

La definición precisa de un experimento discriminante varía con el conjunto de hipótesis sobre las cuales el experimento tiene que discriminar, y con su objetivo, esto es, aceptación o rechazo. La única característica común es que en el caso que el experimento  $E_n$  fuera realizado con un número infinito de repeticiones y se obtuviera como resultados la sucesión de resultados,  $r_n$ ,  $n \in \mathbb{N}$ , entonces casi seguramente se cumpliría el propósito del experimento, esto es, una hipótesis sería aceptada o rechazada. Por supuesto, es imposible realizar un experimento un número infinito de veces, y, por lo tanto, la aceptación o el rechazo es sólo provisorio.

Seguiré esta conferencia dando la definición de un tipo de experimento discriminante, que se usará en algunos ejemplos. Otros tipos de experimentos discriminantes aparecen en [4, Capítulo XVIII].

Repetimos, primero, más formalmente, la definición de equivalencia evidencial entre resultados. En el resto de la charla, supongamos que  $E_n$ ,  $n \in \mathbb{N}$ , es una sucesión de variables aleatorias para el sistema de modelos  $\mathbf{K}_i$ ,  $i \in I$ .

**Definición 1 (Equivalencia evidencial).**

Sean  $a$  y  $b$  resultados posibles de  $E_n$ . Entonces  $a$  y  $b$  son *evidencialmente equivalente*, en símbolos

$$a \sim_n b,$$

si hay un  $c > 0$  tal que, para todo  $i \in I$

$$\Pr_{K_i} [E_n = a] = c. \Pr_{K_i} [E_n = b].$$

La clase de equivalencia correspondiente es

$$[a] = \{b \mid b \sim_n a\}.$$

Ahora, introducimos más formalmente la definición de 'por lo menos tan malo como' que ya fue mencionada.

**Definición 2 (Resultados peores).** Sean  $a$  y  $b$  resultados posibles de un experimento  $E_n$ . Entonces  $b$  es *por lo menos tan malo como  $a$  para  $K_j$*  (relativo a  $K_j$ , para  $i \in I$ , en  $E_n$ ), en símbolos

$$b \leq_{nj} a,$$

si

$$\Pr_{K_j} [E_n \in [b]] \leq \Pr_{K_j} [E_n \in [a]]$$

y existe un  $i \in I, i \neq j$ , tal que

$$\Pr_{K_i} [E_n \in [b]] \geq \Pr_{K_i} [E_n \in [a]].$$

Usamos el símbolo  $R_{nja}$  para el conjunto de los  $b$  que son por lo menos tan malos como  $a$  para  $K_j$  en  $E_n$ , llamado *el conjunto de rechazo determinado por  $a$  para  $K_j$  en  $E_n$* , esto es

$$R_{nja} = \{b \mid b \leq_{nj} a\}.$$

Para acortar algunas de las definiciones, introducimos la siguiente expresión. Sea  $E_n$ , para  $n \in N$ , una sucesión de variables aleatorias, y sea  $A$  un conjunto de sucesiones infinitas de resultados posibles de  $E_n, n \in N$ . Decimos que  $E_n$  está *casi seguramente (c.s.) eventualmente en  $A$* , si casi seguramente hay una sucesión  $r_n, n \in N$ , en  $A$ , tal que

$$\lim_{n \rightarrow \infty} E_n = \lim_{n \rightarrow \infty} r_n.$$

Esto es, el conjunto de resultados posibles donde no hay una tal sucesión  $r_n, n \in N$  en  $A$  es un conjunto nulo. Decimos también que  $E_n$  *no está casi seguramente eventualmente en  $A$* , si casi seguramente para ninguna sucesión de resultados posibles,  $r_n, n \in N$ , en  $A$ , tenemos que

$$\lim_{n \rightarrow \infty} E_n = \lim_{n \rightarrow \infty} r_n.$$

Ahora estamos listos para definir nuestros experimentos.

**Definición 3 (Experimentos discriminantes).** Supóngase que los modelos epistémicamente posibles para un dispositivo son  $K_i, i \in I$ . Una *sucesión discriminante de experimentos* o, simplemente, un *experimento discriminante*,  $E_n, n \in N$  para  $K_i, i \in I$ , es una sucesión de variables aleatorias  $E_n$  en los espacios de probabilidad de  $K_i$ , para  $i \in I$  y  $n \in N$ , que satisface las siguientes condiciones:

(1) Para cada  $j \in I$ , existe un conjunto  $A$  de sucesiones de números reales tal que  $E_n$  está casi seguramente (c.s.) eventualmente en  $A$ , de acuerdo a  $K_j$ , y  $E_n$  no está c.s. eventualmente en  $A$ , de acuerdo a  $K_i$  para  $i \neq j$ , y tal que para cualquier sucesión  $r_n, n \in N$ , en  $A$ , cualquier  $i \in I, i \neq j$ , y cualquier  $\alpha > 0$ , existe un  $n$ , tal que para cada  $m$  con  $m \geq n$

$$\Pr_{K_i} [E_m \in R_{mir_m}] \leq \alpha.$$

(2) Es posible determinar en principio si las oraciones  $[E_n = r]$  son verdaderas en la realidad o no, para todo  $n \in N$ .

La condición (2) no es de carácter matemático, por lo que no la tomaremos en cuenta para consideraciones matemáticas.

La idea detrás de un experimento discriminante, es que si fuéramos capaces de realizar una sucesión infinita de experimentos, entonces casi seguramente se decidiría cual hipótesis es la correcta. Como podemos realizar sólo una sucesión finita, somos capa-

ces sólo de aceptar o rechazar provisoriamente.

Un ejemplo simple de experimento discriminante es el siguiente: Supóngase que estamos tirando una moneda. Sea nuestro experimento  $E_n$ ,  $Fr_n$ , la frecuencia relativa de caras en  $n$  tiradas, para cada número natural  $n$ . Este experimento es discriminante para el conjunto de modelos para monedas con todas las cargas diferentes posibles. Para demostrar esto, sea  $K_p$  el modelo que asigna probabilidad  $p$  a cara. Por la ley fuerte de los grandes números, si  $A$  es el conjunto de sucesiones infinitas que tiende a  $p$ , entonces  $Fr_n$  está c.s. eventualmente en  $A$ , según  $K_p$ , y no está c.s. eventualmente en  $A$ , según  $K_q$ , para  $q \neq p$ . Este es nuestro conjunto  $A$  de la definición, llamémoslo  $A_p$ . Ahora bien, para cualquier par de reales  $r$  y  $q$  entre 0 y 1, el conjunto de rechazo para  $K_q$ ,  $R_{nqr}$ , es de la forma

$$R_{nqr} = \{ |Fr_n - q| \geq |r - q| \}.$$

Esto es así, porque, si  $|b - q| \geq |r - q|$ , entonces

$$\Pr_{K_q} [Fr_n = b] \leq \Pr_{K_q} [Fr_n = r]$$

y

$$\Pr_{K_b} [Fr_n = b] \geq \Pr_{K_b} [Fr_n = r].$$

Es claro que si la sucesión  $r_n$ ,  $n \in N$ , está en  $A_p$  y  $q \neq p$ , entonces de nuevo por la ley de los grandes números, para cada  $\alpha > 0$ , hay un  $n$  finito tal que para cada  $m \geq n$

$$\Pr_{K_q} [Fr_m \in R_{mqr}] \leq \alpha.$$

Así las condiciones para un experimento discriminante se satisfacen.

Una regla ideal de aceptación puede formularse fácilmente con nuestra noción de experimento. Esta regla es ideal en el sentido que es imposible llevarla a cabo en la práctica. Se formula sólo como modelo que otras reglas más prácticas deben aproximar.

**Regla 1 (Regla ideal de aceptación).** Sea  $K_i$ ,  $i \in I$ , un sistema de modelos probabilistas. Decimos que un modelo  $K_j$  debe ser aceptado con respecto a  $K_i$ , para  $i \in I$ , si hay un

experimento discriminante  $E_n$ ,  $n \in N$ , para  $K_i$ ,  $i \in I$ , tal que:

- (1) Se obtiene la sucesión de resultados  $[E_n = r_n]$  en la realidad.
- (2) Esta sucesión,  $r_n$ ,  $n \in N$ , tiene la propiedad que para cada  $i \in I$ ,  $i \neq j$ , y para cada  $\alpha > 0$ , existe un  $n$  tal que para cada  $m \geq n$

$$\Pr_{K_i} [E_m \in R_{mir_m}] \leq \alpha.$$

- (3) No existe ningún otro experimento tal que un cierto  $K_j$ , para  $i \neq j$ , es aceptado (de acuerdo a (1) y (2)).

La regla ideal de rechazo es muy simple:

**Regla 2 (Regla ideal de rechazo).** El modelo  $K_j$  es rechazado si existe un experimento tal que un modelo alternativo  $K_i$ , con  $i \neq j$ , es aceptado (de acuerdo a la Regla 1).

Ahora hacemos más precisa la regla dialéctica para rechazar provisoriamente una hipótesis probabilista. El carácter provisorio del rechazo es muy importante.

**Regla 3 (Regla dialéctica de rechazo).** Sea  $K_i$ ,  $i \in I$ , un sistema de modelos probabilistas. Decimos que un modelo  $K_j$  debe ser provisoriamente rechazado con respecto a  $K_i$ , para  $i \in I$ , si existe un experimento discriminante  $E_n$ ,  $n \in N$ , para  $K_i$ ,  $i \in I$ , tal que:

- (1) Se obtiene  $[E_n = a]$ , para algún  $a$  y  $n \in N$ .
- (2)  $\Pr_{K_j} [E_n \in R_{nja}]$  es baja (esto es, menor que  $\alpha$  para un cierto  $\alpha$  pequeño).

Cuan bajo tiene que ser el  $\alpha$ , se decide dialécticamente. Tiene que haber un acuerdo entre nuestro oponente y nosotros acerca del nivel de probabilidad que lo conveniría. También es necesario llegar a un acuerdo acerca del experimento a usar. Con este acuerdo, si  $K_j$  es en realidad falso, entonces, probando con números cada vez más grandes, casi seguramente se obtendría un  $n$  y un resultado  $a$  para  $E_n$ , tal que la probabilidad de  $[E_n \in R_{nja}]$  es menor que  $\alpha$ .

*Ejemplo 1.* Tomemos como primer ejemplo tirar una moneda para determinar si la moneda está cargada o no. Aquí el  $K_j$  que estamos probando es el modelo para una moneda simétrica. Esto es, con la termino-

logía introducida antes,  $K_j$  es  $K_{1/2}$ . Supóngase que los modelos alternativos son los modelos para monedas con todas las cargas posibles diferentes, esto es,  $K_p$  con  $0 \leq p \leq 1$ . Ya hicimos notar que tirar la moneda  $n$  veces y observar la frecuencia de caras es un experimento discriminante para este caso. Supóngase que tiramos la moneda 100 veces, y que observamos 60 caras. Determinamos, entonces, la probabilidad, bajo la hipótesis que la moneda es simétrica, del resultado de 60 caras o peor. En este caso, este suceso es el resultado de más de 60 caras o menos de 40. Si su probabilidad es baja (esto es, menor que  $\alpha$  para un cierto  $\alpha$  predeterminado), entonces rechazamos la hipótesis.

Cuando, como en el caso que estamos examinando, tenemos todos los sistemas  $K_p$ , para  $0 \leq p \leq 1$  como alternativas, y el experimento para tirar la moneda  $n$  veces, es  $Fr_n$ , para cada  $n \in N$ , podemos simplificar algunas de las definiciones. En primer lugar, valores diferentes para la frecuencia relativa nunca son evidencialmente equivalentes (Definición 1). En segundo lugar, para cada par de valores posibles diferentes,  $a$  y  $b$ , de  $Fr_n$  hay  $p$  y  $q$  tales que

$$\Pr_{K_p} [Fr_n = a] < \Pr_{K_q} [Fr_n = b],$$

y

$$\Pr_{K_q} [Fr_n = b] < \Pr_{K_p} [Fr_n = a].$$

Así, la definición de  $\leq$  puede simplificarse a

$a \leq_{np} b$  si y sólo si

$$\Pr_{K_p} [Fr_n = a] \leq \Pr_{K_q} [Fr_n = b].$$

Por lo tanto, la definición de  $R_{npa}$  (de acuerdo a la Definición 2) parece no depender de las alternativas, y podríamos considerar que  $K_p$  se prueba aisladamente. Pero, nótese que estamos suponiendo que todas las probabilidades alternativas son epistémicamente posibles, que estamos considerando un dispositivo con repeticiones independientes, etc. Si estos factores no estuvieran presentes, podríamos tener alternativas diferentes y obtener resultados distintos. Así, no se considera todas las alternativas lógicamente posibles.

*Ejemplo 2.* Para entender la influencia de alternativas, consideremos como modelos posibles, sólo los  $K_p$  con  $1 \geq p \geq 1/2$ . El experimento, para este caso, puede ser, de nuevo,  $Fr_n$ . También tenemos, en este caso, que ningún par de resultados diferentes son evidencialmente equivalentes. Sin embargo, si obtenemos 60 caras en 100 tiradas, el conjunto de rechazo para  $K_{1/2}$  sólo consiste de resultados mayores que  $3/5$ . Esto es así, porque un resultado,  $a$  menor que  $2/5$  no es peor que  $3/5$ , ya que, aunque

$$\Pr_{K_{1/2}} [Fr_{100} = a] < \Pr_{K_{1/2}} [Fr_{100} = \frac{3}{5}],$$

no hay alternativa donde esta desigualdad se invierta.

*Ejemplo 3.* Ahora discutimos el dispositivo de tirar la moneda con un experimento diferente. Supóngase que el experimento,  $E_n$ , tiene como resultados las sucesiones mismas que se obtienen en  $n$  tiradas independientes de la moneda. Supongamos, primero, que las hipótesis alternativas son las mismas que en el primer caso, a saber,  $K_p$ , para  $0 \leq p \leq 1$ , donde  $p$  es la probabilidad de cara. Puede demostrarse, en forma similar que para  $Fr$ , que este nuevo experimento,  $E_n$ , es discriminante.

Como fue demostrado en la página 90, dos sucesiones con el mismo número de caras son evidencialmente equivalentes. Así, la probabilidad que  $E_n$  esté en el conjunto de rechazo determinado por una sucesión,  $x$ , es la misma que la probabilidad que  $Fr_n$  esté en el conjunto de rechazo correspondiente determinado por la frecuencia de caras en  $x$ . Supóngase que tiramos la moneda diez veces y que obtenemos  $x = 0100101110$  (donde 0 es para caras, y 1, para sello). Entonces la probabilidad que  $E_{10}$  esté en el conjunto de rechazo determinado por  $x$ , de acuerdo a  $K_{1/2}$  es uno, y, por lo tanto,  $K_{1/2}$  no es rechazado en absoluto.

*Ejemplo 4.* Supóngase, ahora, que estamos en la situación explicada en la página 88 y que se tenga como una de las hipótesis alternativas (además de todos los  $K_p$  con  $0 \leq p \leq 1$ ) que las tiradas se obtienen de una máquina, la que, en las primeras 10 tiradas, produce exactamente la sucesión  $x = 0100101110$ , introducida en el ejemplo anterior. Llamemos a esta alternativa, que es

ciertamente lógicamente posible,  $H$ . Como puede demostrarse fácilmente, el experimento  $E_n$ ,  $n \in \mathbb{N}$  también es discriminante para el sistema de hipótesis alternativas,  $H$  y  $K_p$  para  $0 \leq p \leq 1$ : el conjunto  $A_H$ , para  $H$ , es simplemente el singleton que consiste de la sucesión infinita, digamos  $y$  (cuyos diez primeros términos constituyen  $x$ ), producida por la máquina. Entonces  $E_n$  está seguramente eventualmente en  $A_H$ , si  $H$  es verdadera, ya que, en este caso,  $E_n$  es  $y$  restringida a  $n$ , para  $n \in \mathbb{N}$ . Por otra parte, bajo cualquier  $K_p$ , la sucesión  $E_n$  no es c.s. esta sucesión  $y$ , y, así,  $E_n$  no está c.s. en  $A_H$ . Para las otras hipótesis, como en el caso anterior, los conjuntos  $A_{K_p}$  consisten de las sucesiones cuyo límite es  $p$ , excepto posiblemente, de  $y$ .

Volvamos al caso  $E_{10}$ . Entonces,  $x$ , esto es, la sucesión de 10 resultados producidos por la máquina, no es evidencialmente equivalente a ninguna otra sucesión, porque, tenemos bajo  $H$

$$\Pr_H [E_{10} = x] = 1 \text{ y } \Pr_H [E_{10} = z] = 0,$$

para  $z \neq x$ . Luego, no hay  $c > 0$  que satisfaga la Definición 1 de equivalencia evidencial. Por otra parte, si  $u$  y  $z$  son sucesiones diferentes de  $x$ , pero con el mismo número de caras entre ellas, entonces, como antes

$$\Pr_{K_p} [E_{10} = u] = \Pr_{K_p} [E_n = z],$$

y también

$$\Pr_H [E_{10} = u] = \Pr_H [E_{10} = z] = 0.$$

Luego,  $u$  y  $z$  son evidencialmente equivalentes.

Por esto, es fácil calcular que el conjunto de rechazo determinado por  $x$  para  $K_{1/2}$  contiene sólo a  $x$  y a las dos sucesiones con sólo unos, digamos 1, y ceros, digamos 0: tenemos, de acuerdo a la Definición 2, que

$$z \leq_{10, 1/2} x \Leftrightarrow \Pr_{K_{1/2}} [E_{10} \in [z]] \leq \Pr_{K_{1/2}} [E_{10} = x].$$

Así, es claro que sólo  $x$ ,  $0$ ,  $1 \leq_{10, 1/2} x$ . Por tanto

$$R_{10, 1/2, x} = \{x, 1, 0\}.$$

Luego

$$\Pr_{K_{1/2}} [E_{10} \in R_{10, 1/2, x}] = 3 (1/2)^{10},$$

que es muy baja. Por lo tanto, de acuerdo a nuestra regla de rechazo, estamos justificados en rechazar  $K_{1/2}$ , como es intuitivamente natural.

*Ejemplo 5.* Como otro ejemplo, considérese las mismas hipótesis alternativas de la moneda con cargas diferentes,  $K_p$ , y la máquina que produce  $x$ ,  $H$ , pero ahora con el experimento Fr. Supóngase que el límite de la frecuencia relativa de caras en la sucesión producida por la máquina es  $1/2$ . Fr no es un experimento discriminante para estas alternativas: no hay ningún conjunto  $A$  de sucesiones de resultados posibles de Fr tal que Fr esté c.s. eventualmente en  $A$ , de acuerdo a  $H$  y Fr no esté c.s. eventualmente en  $A$ , de acuerdo a las alternativas. Si tomamos  $A$  como el conjunto de sucesiones cuya frecuencia relativa converge a  $1/2$ , entonces Fr está seguramente (no sólo c.s.) eventualmente en  $A$  de acuerdo a  $H$ , pero también está c.s. eventualmente en  $A$ , de acuerdo a  $K_{1/2}$ . Considérese otro conjunto como nuestro candidato para el conjunto  $A$ : Consista  $A$  de sucesiones de posibles frecuencias en  $n$  tiradas,  $r_n$ ,  $n \in \mathbb{N}$ , tales que  $r_n$  es la frecuencia exacta en las primeras  $n$  tiradas de  $y$ , la sucesión producida por la máquina. En este caso, de nuevo Fr está c.s. eventualmente en  $A$ , bajo  $H$ , pero también está c.s. eventualmente en  $A$  bajo  $K_{1/2}$ .

Vemos, de estos ejemplos, que la consideración de hipótesis alternativas es esencial. Es posible, como lo vimos más arriba, que podamos considerar todas las alternativas "razonables", por ejemplo, todas las alternativas  $K_p$ , para  $0 \leq p \leq 1$ ,  $p$  real. Pero, en este caso, no estamos tomando en cuenta todas las alternativas lógicamente posibles, ya que, entre otras, la alternativa producida por la máquina que mencionamos estaría excluida.

Una regla simple de aceptación, que es conveniente en muchos casos, es:

**Regla 4 (Regla dialéctica de aceptación).** Sea  $K_i$ ,  $i \in I$  un sistema de modelos probabilistas. Aceptamos provisoriamente el modelo  $K_i$  relativo a  $K_p$  para  $i \in I$ , si todos

los  $K_i$ , para  $i \neq j$ , han sido rechazados provisoriamente con respecto a  $j$ , de acuerdo con la Regla 3.

Además, no hay experimento discriminante  $E'$  tal que  $K_j$  es rechazado provisoriamente con respecto a  $E'$  (de acuerdo a la Regla 3).

Esta regla es demasiado estricta para la práctica estadística, especialmente cuando hay un número infinito de hipótesis posibles. Sin embargo, una regla semejante está detrás de la estimación por regiones de confianza, pero en este caso, las hipótesis involucradas son compuestas, esto es, son clases de distribuciones de probabilidad. Se discute regiones de confianza en [4, Capítulo XVIII].

Existe la posibilidad que todo  $K_i$ , para  $i \in I$ , sea rechazado. Supóngase en el caso de la moneda, por ejemplo, que la hipótesis  $K_p$  determina que la probabilidad de cara es  $p$ , y que aceptamos como posibles todos  $K_p$  para  $0 < p < 1$ ,  $p$  real, pero no  $p = 1$  ó  $p = 0$ . El experimento  $E_n$  es la sucesión de resultados en  $n$  tiradas de la moneda. Supóngase que se obtiene una sucesión de ceros. Entonces, si la sucesión es suficientemente larga, todas las hipótesis deberían ser rechazadas provisoriamente. En verdad, si se obtuviera una sucesión infinita de ceros, entonces todas las hipótesis deberían ser absolutamente rechazadas. En este caso, llamamos al sistema *insostenible*. Entonces, debemos agregar hipótesis al sistema para transformarlo en sostenible. Por supuesto, como en general sólo obtenemos sucesiones finitas, sólo podemos determinar insostenibilidad provisoria.

Hay otro modo por el cual podemos llegar a la conclusión que un sistema de modelos probabilistas es insostenible. Este es cuando tenemos dos o más experimentos y todos los modelos que se consideran posibles son rechazados provisoriamente en por lo menos uno de los experimentos.

### 5. Comparación con la teoría de Neyman y Pearson

Veremos cómo se comparan nuestras reglas para el rechazo de hipótesis con los test de hipótesis de Neyman-Pearson, a través del análisis de algunos ejemplos. Aunque mi justificación de los test para el rechazo de hipótesis es diferente de las de Neyman-

Pearson, en la gran mayoría de los casos los test obtenidos dan casi los mismos resultados. Sin embargo, hay diferencias de énfasis y, en algunos casos, los resultados son divergentes. Los ejemplos 1, 2 y 3, formulados como un test para la hipótesis nula  $K_{1/2}$  contra la hipótesis compuesta que incluye todas las alternativas indicadas en cada ejemplo, dan tests similares a los clásicos, pero con algunas diferencias con los que se obtienen de los criterios de Neyman-Pearson.

Examinemos, a continuación el ejemplo 4, considerado como un test de la hipótesis nula  $K_{1/2}$  contra la hipótesis alternativa compuesta, llamémosla  $J$ , formada por todos los  $K_p$ , con  $p \neq 1/2$ ,  $0 \leq p \leq 1$ , más la hipótesis  $H$  que dice que la sucesión de resultados es producida por la máquina. Las diferencias entre nuestro test y el test de Neyman-Pearson que veremos en este caso, son las mismas, *mutatis mutandi*, con las que se producen para los ejemplos 1, 2 y 3.

El test, según el procedimiento explicado en la sección anterior, podría ser formulado de la siguiente manera. Supóngase que se obtenga el resultado  $z$  de  $E_n$ , para un cierto  $n$ . Para fijar ideas, tomemos  $n = 10$ . Calculamos el conjunto de rechazo determinado por  $z$  para  $K_{1/2}$ , que llamamos  $R_{10, 1/2, z}$ ; este conjunto está formado por  $u$  tales que

$$\Pr_{K_{1/2}} [E_{10} = u] \leq \Pr_{K_{1/2}} [E_{10} = z],$$

más  $x$ , la sucesión producida por la máquina. Computamos la probabilidad de este conjunto  $R_{10, 1/2, z}$  la que llamamos  $R_{10, 1/2, z}$ . Este es el valor- $p$  del test. Si este valor es menor que un  $\alpha$  pequeño predeterminado, se rechaza  $K_{1/2}$ . Nótese que si  $z = x$ ,  $p_{10, 1/2, z} = \{0, 1, x\}$  y su probabilidad es muy baja, seguramente mucho menor que  $\alpha$ . Así, el valor- $p$  del test da información adicional, la que está claramente justificada de acuerdo a nuestro procedimiento. Creo que esta formulación del test, aunque no es estrictamente la de Neyman-Pearson, es de una forma en que los tests son usados normalmente por estadísticos y, también, aparecen así formulados en algunos textos de estadística, por ejemplo en [16].

Veamos, ahora, el procedimiento de Neyman-Pearson (reconstruido por mí) para la misma situación del ejemplo 4. Se fija el

nivel  $\alpha$ , el tamaño del test. En la teoría de Neyman-Pearson se escoge el test de tamaño  $\alpha$  que es uniformemente más poderoso, UMP, o si no hay un tal test, uno que sea uniformemente más poderoso entre los insesgados, UMPU. Ya que en este caso no hay UMP, escogemos uno UMPU. Por tanto, se elige un conjunto  $S_0$  de sucesiones de 10 resultados tal que

$$\Pr_{K_{1/2}} [E_{10} \in S_0] \leq \alpha,$$

$$\Pr_K [E_{10} \in S_0] \geq \Pr_{K_{1/2}} [E_{10} \in S_0]$$

para cualquier hipótesis  $K$  en la alternativa  $J$ , y, además

$$\Pr_K [E_{10} \in S_0] \geq \Pr_K [E_{10} \in C],$$

para cualquier  $K \in J$  y cualquier  $C$  tal que

$$\Pr_{K_{1/2}} [E_{10} \in C] \leq \alpha, \text{ y}$$

$$\Pr_K [E_{10} \in C] \geq \Pr_{K_{1/2}} [E_{10} \in C].$$

para cualquier  $K \in J$ .

Si el resultado obtenido  $z$  está en  $S_0$ , se rechaza  $K_{1/2}$  y si no, se acepta. Es claro que el conjunto  $S_0$  debe contener  $x$ . Además, se calcula el mínimo número  $\delta$  tal que

$$\Pr_{K_{1/2}} \left[ \left| Fr_{10} - \frac{1}{2} \right| \geq \delta \right] \leq \alpha - \frac{1}{2^{10}}.$$

Se agregan a  $S_0$  todas las sucesiones cuya frecuencia relativa de caras es mayor o igual que  $1/2 + \delta$  o menor o igual que  $1/2 - \delta$ .

$$\text{Si la } \Pr_{K_{1/2}} \left[ \left| Fr_{10} - \frac{1}{2} \right| \geq \delta \right] \leq \alpha - \frac{1}{2^{10}},$$

podríamos tener que agregar algunas de las sucesiones con una de las frecuencias inmediatamente más cercanas a  $1/2$  (pero no podríamos agregar todas esas sucesiones, porque en ese caso nos pasaríamos de  $\alpha$ ) para aproximarse lo más posible al nivel  $\alpha$ . Me parece que ésta es una característica incorrecta del punto de vista de Neyman-Pearson<sup>13</sup>. La razón para considerarla inco-

<sup>13</sup> Una discusión de este problema desde el punto de vista de los tests "randomizados", que es similar a la situación que discutimos aquí, aparece en [4, Capítulo XIX, 3].

recta es que la elección entre las sucesiones de las frecuencias indicadas que se agregan a  $S_0$  es arbitraria. Todas estas sucesiones dan la misma evidencia a favor o en contra de  $K_{1/2}$ , por lo que o se deberían incluir todas o, ninguna.

Salvo por este último punto, el test es similar al nuestro. La única otra diferencia es que no se ve muy justificado el uso de los valores- $p$ , en este contexto.

En el caso del experimento Fr usado en el ejemplo 5 para este mismo test, parece que tampoco en el caso de Neyman-Pearson existe un test adecuado, esto es, un test UMPU.

Veremos ahora ejemplos de Gillies, [8, págs. 204, 205], discutido en [4, Capítulo XIX, 1], que creo que muestran que los criterios de Neyman-Pearson no son intuitivamente adecuados.

Supóngase que estamos haciendo un test para la hipótesis nula  $H_0$  que la variable aleatoria  $X$  tiene distribución uniforme con densidad

$$f_1(x) = \begin{cases} \frac{1}{2}, & \text{para } -1 \leq x \leq 1 \\ 0, & \text{en los otros casos,} \end{cases}$$

contra las alternativas

$$f_\epsilon(x) = \begin{cases} \frac{1}{2}\epsilon, & \text{para } -1 \leq x \leq -1 + \epsilon \text{ y } 1 - \epsilon \leq x \leq 1, \\ 0, & \text{en los otros casos,} \end{cases}$$

donde  $0 < \epsilon \leq 1$ . Aquí, el problema es de hacer un test de  $\epsilon = 1$  contra las alternativas  $0 < \epsilon < 1$ . La Fig. 1 muestra las distribuciones.

Según la teoría de Neyman-Pearson, si fijamos

$$S_0 = [-1, -1 + \alpha] \cup [1 - \alpha, 1],$$

entonces  $S_0$  es un test UMP (uniformemente más poderoso) de tamaño  $\alpha$ . Este resultado no parece intuitivamente correcto, porque, como la hipótesis nula,  $H_0$ , es de una distribución uniforme, ningún resultado en el intervalo  $[-1, 1]$  debería ir en contra de ella. Nótese que estamos considerando un experimento algo artificial con un sólo resultado. Si nuestro experimento fuera una sucesión de resultados, la situación sería muy diferente.

Analícemos, ahora, este ejemplo de acuerdo a mis principios. En primer lugar, ningún par

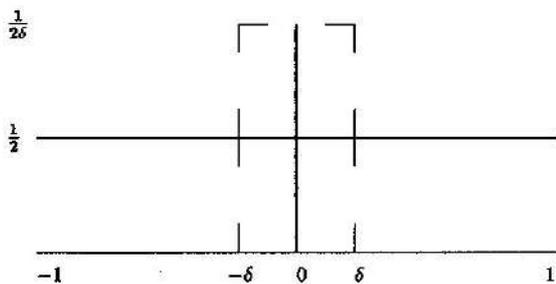


Figura 1.- El primer ejemplo de Gillies

de resultados diferentes son evidencialmente equivalentes. También, el conjunto de rechazo para  $f_1$ , determinado por cualquier resultado, es el intervalo entero  $[-1, 1]$ , porque todos los resultados son equiprobables. Así, no hay experimento discriminante. Además, el conjunto  $S_0$  no es un conjunto de rechazo, y, por lo tanto, el test no está justificado.

Un análisis similar puede darse para el segundo ejemplo de Gillies, en el que se considera como alternativas a  $f_1$  las distribuciones con densidades

$$f_2(x) = \begin{cases} \frac{1}{2}\delta, & \text{para } -\delta \leq x \leq \delta \\ 0, & \text{en los otros casos,} \end{cases}$$

donde  $0 < \delta \leq 1$ , dada en la Fig. 2.

En este caso, el conjunto

$$S_0 = [-\delta, \delta]$$

es un test UMP. Como en el otro ejemplo, el conjunto de rechazo para  $f_1$  es el intervalo entero, y, por lo tanto, éste no es un buen test según mis principios.

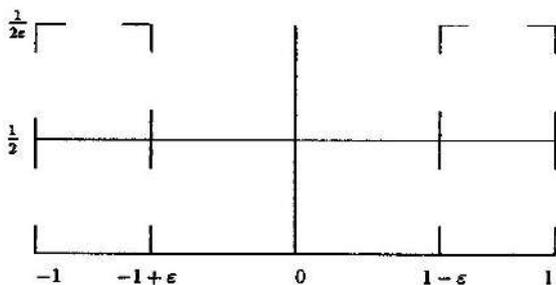


Figura 2.- El segundo ejemplo de Gillies

## BIBLIOGRAFIA

1. BERNOULLI, J., *Ars conjectandi*, 1713, hay una traducción inglesa de la Parte IV por Bing Sung, Harvard University Department of Statistics Technical Reports 2, 1966.
2. BOREL, E., *Elements of the theory of probability*, Prentice-Hall, Englewood Cliffs, NJ, 1950, traducción inglesa de Freund de la edición francesa de 1950.
3. CHUAQUI, R., *Probability based on possibility*, Actas del Congreso Internacional Extraordinario de Filosofía (Córdoba, Argentina, 1987), págs. 945-954 (traducción española, mismo volumen, págs 955-965).
4. —, *Truth, possibility, and probability: New logical foundations of probability and statistical inference*, North-Holland Pub. Co., Amsterdam, 1991, XVIII+484 pages.
5. DA COSTA, N. C. A. y CHUAQUI, R., *On Suppes's set theoretical predicates*, Erkenntnis 29 (1988), 95-112.
6. ENDERTON, H. E., *A mathematical introduction to logic*, Academic Press, New York, 1972.
7. FISHER, R. A., *Statistical methods and scientific inference*, Hafner Press, New York, tercera ed., 1973, primera ed. en 1956.
8. GILLIES, D. A., *An objective theory of probability*, Methuen & Co. Ltd., London, 1973.
9. HACKING, I., *The logic of statistical inference*, Cambridge University Press, 1965.
10. HEMPEL, C., *Deductive nomological vs. statistical explanation*, Scientific Explanation, Space, and Time (Feigl and Maxwell, eds.), University of Minnesota Press, Minneapolis, 1962, pp. 150-151.
11. KYBURG, H. Jr., *Probability, rationality and a rule of detachment*, Logic, Methodology, and Philosophy of Science (Y. Bar-Hillel, ed.), Studies in Logic and the Foundations of Mathematics, North-Holland Pub. Co., Amsterdam, 1965, pp. 301-310.
12. —, *Logical foundations of statistical inference*, Reidel, Dordrecht, 1974.
13. LUCAS, J., *The concept of probability*, Oxford University Press, Oxford, 1970.
14. MATES, B., *Elementary logic*, Oxford University Press, New York, 1972.
15. POPPER, K., *The logic of scientific discovery*, Hutchinson & Co., London, 1959, traducción inglesa del original alemán de 1934.
16. ROSS, S. M., *Introduction to probability and statistics for engineers and scientists*, John Wiley & Sons, New York, 1987.
17. SUPPES, P., *Set-theoretical structures in science*, Stanford University, 1967.
18. TARSKI, A., *Der Wahrheitsbegriff in den formalisierten Sprache*, Studia Philosophica 1 (1935), 261-405, traducción inglesa en *Logic, Semantics and Metamathematics*; 2a ed. (J. Corcoran, editor), Hackett Publishing, Indiana, 1983. 1a. ed. (ed. y traducida por J. H. Woodger), Oxford, 1956.
19. —, *Truth and proof*, Sci. Amer. 220 (1969), 63-77.