*El significado del teorema de Bernoulli para la teoría de la inferencia estadística**

Andrés RIVADULLA
(Universidad Complutense)

1. Introducción

En 1713 apareció en Basilea con carácter póstumo el *Ars conjectandi* de Jakob Bernoulli. Lo había publicado su sobrino Niklaus ante la insistencia de los matemáticos de la época, entre ellos Pierre Remond de Montmort, quien ya había publicado anónimamente cinco años antes su *Essay d'Analyse sur les Jeux de Hazard* y estaba interesado en conocer el trabajo de Bernoulli[1].

La estadística matemática debe su origen a una extensión del cálculo en los juegos de azar al cómputo de probabilidades de vida y rentas vitalicias. En plena segunda mitad del siglo XVII los hermanos Huygens primero, y Edmund Halley después, habían analizado rigurosamente estas materias sobre base puramente empírica, mientras que Jan de Witt y Leibniz también lo habían intentado aunque sin preocuparse demasiado de si sus reflexiones estaban suficientemente apoyadas empíricamente o no[2]. No parece que Jakob

* Agradezco muy sinceramente a Miguel Ángel Gómez Villegas, catedrático del Dpto. de Estadística e I. O. de la Universidad Complutense de Madrid, su lectura y comentario de una versión anterior de este trabajo, así como su amable disposición a discutir en múltiples ocasiones sobre cuestiones relativas a la teoría de la inferencia estadística.

[1] K. Kohli (1975) ha relatado la historia de la publicación del *Ars conjectandi* en el Vol. 3 de *Die Werke von Jakob Bernoulli*, Basel 1975.

[2] Véase Rivadulla (1991*c*).

Hasta la publicación por Jakob Bernoulli de su *Ars conjectandi*, el término *probabilidad*

Bernoulli llegase a conocer nunca estos esfuerzos, aunque curiosamente su *Ars conjectandi* contenía un análisis matemático —completamente ausente entre los contemporáneos antes mencionados— que pretendía explicar la estabilidad de las frecuencias observadas y sentaba las bases para un tratamiento riguroso futuro de la inferencia estadística; se trataba de la *Ley de los grandes números*, así conocida desde que Simeon Dennis Poisson (1837) decidiese bautizar con ese nombre al *Teorema* de Bernoulli[3].

En carta a Leibniz fechada el 24 de Abril de 1704 Bernoulli declara haber presentado a su hermano un Teorema que le permitía 'calcular' empíricamente las probabilidades desconocidas de ocurrencia de los fenómenos tan bien como si fuesen conocidas *a priori*. Un teorema que, según confesión propia, ya llevaba madurando veinte años cuando escribió su *Ars conjectandi*. En efecto, el diario matemático de Bernoulli, las *Meditationes*, contiene dos artículos, el 131*a*, escrito hacia 1689, y el 151*a*, redactado entre 1688 y 1690, en los que el matemático suizo anuncia, aplica y prueba por vez primera la Ley de los grandes números. En lo que sigue de esta Introducción voy a detenerme en el primero de ellos.

El objetivo de Jakob Bernoulli en el artículo 131*a*, recogido en Schneider (1988), es mostrar que, a medida que aumenta el número de observaciones, disminuye la probabilidad de desviarse de la proporción verdadera de casos. Así, si en un juego hay tantas posibilidades de ganar como de perder, la proporción final de juegos ganados coincidirá con la de juegos perdidos. De manera que, aunque en las primeras partidas se observase que la proporción de juegos ganados no coincide con la de juegos perdidos, a medida que aumentara el número de jugadas crecería la probabilidad de que el total de partidas ganadas se sitúe dentro de unos límites previamente elegidos, que fuera de ellos.

Elijamos pues el intervalo $0{,}5 \pm 0{,}166$ como expresión de confianza en

prácticamente brilla por su ausencia en los trabajos de los autores antes mencionados. En Rivadulla (1995) §2 mencioné como excepciones a esta situación a Antoine Arnauld y Gottfried Leibniz. Glen Shafer (1993), p. 7, añade por su parte la del clérigo inglés George Hooper, quien habría hablado de "probability as a number between zero and one".

[3] "Les choses de toutes natures sont soumises à une loi universelle qu'on peut appeller *la loi des grandes nombres*. Elle consiste en ce que, si l'on observe des nombres très considérables d'événements d'une même nature,..., on trouvera, entre ces nombres, des rapports à très peu près constants. Pour chaque nature des choses, ces rapports autont une valeur spéciale dont ils s'écarteront de moins en moins, è mesure que la série des événements observés augmentera davantage, et qu'ils atteindraient rigoureusement s'il était possible de prolonger cette série à l'infini." (p. 7)

que la proporción total de jugadas ganadas no va a separarse de la igualdad más del 16,6% de los casos, y preguntémonos por la probabilidad de que el número de partidas ganadas supere los límites fijados, e.d. más de 2/3 o menos de 1/3 de las jugadas. Bernoulli limita el cálculo de esta probabilidad a 3, 6, 9 y 12 partidas. Y esto le basta para mostrar cómo se aplica el Teorema en cuestión en este caso especial.

Si jugamos tres partidas, el número de resultados lógicamente posibles en términos de 'gano' y 'pierdo' es obviamente $2^3 = 8$. Ganar más de 2/3 de las jugadas equivale a ganar las tres partidas, lo que tiene una probabilidad de ocurrencia de 1/8 -supuesta la equiprobabilidad de los 8 resultados posibles que constituyen el espacio muestral. Por otra parte, ganar menos de 1/3 de las jugadas significa perderlas todas, y la probabilidad de este suceso es igualmente 1/8. La probabilidad pues de obtener un resultado fuera del intervalo [1/3,2/3] es 2/8, menor que la de que se sitúe dentro, que es 6/8 = 0,75.

Si jugamos 6 partidas, ganar más de 2/3 de ellas equivale a ganar 5 ó 6, mientras que ganar menos de 1/3 significa ganar una o ninguna. Ahora bien, hay $2^6 = 64$ resultados lógicamente posibles en términos de 'gano' y 'pierdo'. Luego la probabilidad de ganar las seis partidas, como la de perderlas todas es 1/64. La probabilidad de ganar 5 jugadas es —aplicando la fórmula binomial[4] deducida por el propio Bernoulli en el *Ars Conjectandi*—

$$\binom{6}{5}0,5^5.0,5 = 0,09375;$$

luego la probabilidad de ganar más de 2/3 es 0,109375. Como ésta es también la probabilidad de ganar menos de 1/3, entonces resulta que la probabilidad de obtener un resultado fuera del intervalo considerado es 0,21875, mientras que la del suceso contrario, e.d. la de obtener un resultado entre los 50 restantes, es 0,78125.

Si elevamos a 12 el número de jugadas, ganar más de 2/3 equivale a ganar 9, o 10, o 11, o 12, y ganar menos de 1/3 supone ganar 3, o 2, o 1, o 0 jugadas. Como hay $2^{12} = 4096$ resultados lógicamente posibles y equiprobables, la probabilidad de ganar todas las partidas, como la de perderlas todas, es

[4] La probabilidad de obtener a 'éxitos' —p. e. a veces 'cara'— en n ensayos independientes, siendo p la probabilidad de 'éxito', viene dada por la fórmula de Bernoulli:

$$P(a \text{ 'éxitos' en } n \text{ ensayos independientes}|\, p) = \binom{n}{a}p^a(1-p)^{n-a}$$

1/4096, que, sumando con 0,002929687, que es la probabilidad de ganar 11 (ganar 1); con 0,016113281, que es la probabilidad de ganar 10 (ganar 2); y con 0,053710937, que es la probabilidad de ganar 9 (ganar 3), y multiplicando por 2, proporciona el valor 0,145996092, que es la probabilidad de ganar más de 2/3 o menos de 1/3 de las doce jugadas. Luego la probabilidad de obtener un resultado incluido en el intervalo de confianza es 0,854. Ahora bien, 0,75 < 0,78 < 0,85. Luego, a medida que aumenta el número de jugadas, resulta cada vez más probable que los resultados que se observen se incluyan dentro del intervalo 0,5 ± 0,166 que fuera de él. La conclusión que Bernoulli considera que se sigue legítimamente de este resultado ilumina el objetivo que se propone probar en el *Ars conjectandi*: inferir probabilidades a partir de frecuencias observadas, para el caso general en que la probabilidad de 'observación favorable' y 'desfavorable' no coincidan: "Con ello podría llevar a cabo tantas observaciones que sería capaz de inferir la razón [desconocida] de las posibilidades con una probabilidad próxima a la certeza, como si dispusiera de ella *a priori*."

2. El Teorema de Bernoulli

El *Teorema de Bernoulli*, la *propositio principalis* del *Ars conjectandi*, afirma lo siguiente[5]: *Teorema*. Supóngase que el número de los casos favora bles se comporta respecto al de los casos desfavorables como $\frac{r}{s}$, o sea respecto del número total de casos como $\frac{r}{r+s}=\frac{r}{t}$, estando contenida esta última razón entre los límites $\frac{r+1}{t}$ y $\frac{r-1}{t}$

Así pues hay que demostrar ahora que se pueden realizar tantas observaciones que resulte tantas veces como se quiera (p.e. c veces) más probable que la razón de las observaciones favorables respecto de todas las observaciones realizadas se sitúe dentro, que no fuera de estos límites, e.d. que no es mayor que $\frac{r+1}{t}$, ni menor que $\frac{r-1}{t}$

Hasta aquí Bernoulli, cuyo *Teorema* paso a interpretar a continuación: Sea la fracción $p=\frac{r}{t}$ la probabilidad conocida de ocurrencia de un suceso, y

[5] Bernoulli, "Teoria de probabilidades", p. 412.

f_o la variable aleatoria 'proporción de casos favorables en n observaciones'. Que sea c veces más probable que la frecuencia observable f_o se sitúe dentro a que se sitúe fuera de los límites, quiere decir que por cada caso a favor de que se sitúe fuera, hay c casos ($c \gg 1$) o posibilidades de que caiga dentro de los límites. La relación entre c y n viene dada como $n = c + 1$.

Como $\frac{r-1}{t} = p - \frac{1}{t}$ y $\frac{r+1}{t} = p + \frac{1}{t}$, entonces

$$P\left[\left(p-\frac{1}{t}\right) > f_o > \left(p+\frac{1}{t}\right)\right] = \frac{c}{c+1}$$

expresa la probabilidad de que f_o caiga fuera de los límites del intervalo; por su parte

$$P\left[\left(p-\frac{1}{t}\right) \leq f_o \leq \left(p+\frac{1}{t}\right)\right] = \frac{c}{c+1}$$

expresa la probabilidad de que f_o caiga dentro de los límites del intervalo. Y el *Teorema de Bernoulli* afirma que

$$P\left[\left(p-\frac{1}{t}\right) \leq f_o \leq \left(p+\frac{1}{t}\right)\right] = cP\left[\left(p-\frac{1}{t}\right) > f_o > \left(p+\frac{1}{t}\right)\right]$$

Ahora bien, como $\frac{1}{c+1} = \frac{1}{n}$ y $\frac{c}{c+1} = \frac{n-1}{n}$, podemos expresar el resultado de Bernoulli en notación más familiar, diciendo que

$$\lim_{n\to\infty} P\left[\left(p-\frac{1}{t}\right) \leq f_o \leq \left(p+\frac{1}{t}\right)\right] = \lim_{n\to\infty}\frac{n-1}{n}$$

con lo que

$$\lim_{n\to\infty} P\left[\left(p-\frac{1}{t}\right) \leq f_o \leq \left(p+\frac{1}{t}\right)\right] = 1.$$

Complementariamente

$$\lim_{n\to\infty} P\left[\left(p-\frac{1}{t}\right) > f_o > \left(p+\frac{1}{t}\right)\right] = 0.$$

Si llamamos $\varepsilon = \frac{1}{t}$, obtendremos una presentación del *Teorema de Bernoulli* más parecida a la usual:

$$\lim_{n \to \infty} P(p - \varepsilon \leq f_o \leq p + \varepsilon) = 1$$

Ahora bien, $p - \varepsilon \leq f_o \leq p + \varepsilon$, restando p, equivale a $-\varepsilon \leq f_o + p \leq + \varepsilon$. De ahí que la forma usual que adopta el *Teorema de Bernoulli* o *Ley* (débil) *de los grandes números* sea:

(1) $$\lim_{n \to \infty} P(| f_o - p | \leq \varepsilon) = 1$$

cuyo contenido es el siguiente: Cuando el número n de observaciones aleatorias crece indefinidamente, existe 'certeza moral' de que la frecuencia relativa observable f_0 no se desviará del valor verdadero del parámetro poblacional p más allá de un número positivo ε arbitrariamente pequeño. Luego, *cuando la probabilidad de 'éxito' existe y es conocida, si el número de observaciones crece indefinidamente, se aproxima a la certeza la esperanza de que la proporción de observaciones favorables no sobrepasará los límites fijados previamente.*

3. Un fracaso cantado: la *inversión* del Teorema de Bernoulli

¿Qué relevancia tiene el Teorema de Bernoulli, o Ley de los grandes números, para la teoría de la inferencia estadística? La intención declarada de Jakob Bernoulli en las *Meditationes* y en el *Ars conjectandi* es la de averiguar la proporción verdadera de casos —e.d. el valor verdadero *desconocido* del parámetro poblacional p— por medio de observaciones aleatorias, o sea, determinar *a posteriori* el valor de p tan bien como si nos fuera conocido *a priori*. Dicho de otro modo, el objetivo de Bernoulli es *inferir probabilidades a partir de frecuencias observadas*. En este sentido su meta es *invertir* el procedimiento usual en teoría de probabilidades, consistente en facilitar el cálculo de frecuencias observables a partir de probabilidades conocidas[6]: "Se trata pues de investigar si, por medio del incremento de las observaciones, crece constantemente la probabilidad de que el número de observaciones favorables alcance, respecto del número de observaciones desfavorables, la proporción verdadera, [...] Ahora bien, si fuera posible, y en definitiva se lograra de esta manera certeza moral (...), entonces podríamos encontrar *a*

[6] Bernoulli, "Teoría de probabilidades", pp. 401-402.

posteriori los números de los casos [e.d. las probabilidades] casi tan bien como si nos fueran conocidos *a priori*. Y esto basta,..., a fin de conducir nuestra suposición en cualquier dominio aleatorio de forma no menos científica que en los juegos de azar."

De tener éxito en su empeño, el matemático suizo habría resuelto uno de los problemas fundamentales de la metodología de la ciencia: el *problema de la inducción*; pues habría mostrado cómo inferir lo general de lo particular, o, en terminología estadística, como gustaba de decir Sir Ronald Fisher, de la muestra a la población de la que, por hipótesis, había sido extraída al azar. ¿Qué demuestra, pues, Bernoulli en su trabajo de 1713?

La Ley de los grandes números (1) equivale a

(2) $$\lim_{n \to \infty} P(f_n \in [p - \varepsilon, p + \varepsilon]) = 1.$$

¿Da el Teorema de Bernoulli ocasión para pensar que, en el caso de que los valores de p fueran desconocidos, podrían ser calculados a partir de resultados empíricamente observados tan bien como si nos fuesen conocidos *a priori*? La cuestión es, pues, si mediante una *inversión* del *Teorema de Bernoulli* es legítimo esperar

(3) $$\lim_{n \to \infty} P(p \in [f_n - \varepsilon, f_n + \varepsilon]) = 1$$

pues esta formulación parece dejar abierta la posibilidad de *estimar* el valor verdadero *desconocido* de p, con una aproximación tan grande como se quiera, a medida que crece el número de observaciones, ya que el parámetro en cuestión no se situaría más allá de $f_0 \pm \varepsilon$. De ser esto cierto, Jakob Bernoulli habría resuelto a un tiempo el problema matemático de la inferencia estadística y el problema lógico de la inferencia inductiva.

El caso es que meras transformaciones algebraicas permiten, restando p, pasar de $p - \varepsilon \leq f_0 \leq p + \varepsilon$ a $-\varepsilon \leq f_0 - p \leq +\varepsilon$. Si ahora restamos f_0, y multiplicamos por -1, con lo que las desigualdades cambian de sentido, obtendremos finalmente $f_0 - \varepsilon \leq p \leq f_0 + \varepsilon$, y (3) parecería matemáticamente justificada.

Un sencillo argumento probabilístico invalida empero el intento de *inversión* del Teorema de Bernoulli, a saber: supongamos un n fijo; entonces p, tanto si es conocida como si es desconocida, es una constante, no una variable aleatoria; luego $P(f_0 - \varepsilon \leq p \leq f_0 + \varepsilon)$ sólo puede asumir dos valores: 1 y 0, dependiendo respectivamente de que p esté o no dentro del intervalo cerrado

$[f_0 - \varepsilon, f_0 + \varepsilon]$. En conclusión, el Teorema de Bernoulli no da respuesta al problema que el matemático suizo parecía que se había propuesto resolver.

Carece de sentido pues pretender que el Teorema de Bernoulli sirve para *estimar* el valor verdadero de p, con lo que naturalmente no resuelve ni el problema de la inferencia estadística ni el de la inferencia inductiva. El Teorema de Bernoulli permite el cálculo de manera *deductiva* o *directa*, de la frecuencia con que se espera que ocurran los resultados observacionales.

4. El juicio de la historia

Muchos historiadores de la teoría de probabilidades coinciden en destacar la *naturalidad matemática* de la Ley de los grandes números. Así, para Ian Hacking (1971, p. 222, y 1975, p. 154) el Teorema de Bernoulli es un teorema de pura teoría de probabilidades, y para Anders Hald (1990, pp. 258-259) la convergencia en probabilidad de f_0 hacia p, cuando n tiende a infinito, es la propiedad puramente algebraica de la distribución binomial consistente en que la razón del cuerpo de la distribución respecto de sus extremos tiende a infinito. De ahí que no exista hoy la menor duda acerca de qué es lo que Bernoulli probó[7].

Ya antes de la publicación del *Ars conjectandi* Jakob Bernoulli había tropezado con el escepticismo de Leibniz acerca de la posibilidad de *invertir* la Ley de los grandes números a fin de obtener empíricamente, y con una aproximación arbitrariamente grande, las probabilidades de ocurrencia de los fenómenos. La correspondencia entre ambos de los años 1703 y 1704 pone de manifiesto sus divergencias al respecto. Pero también otros autores del XVIII se percataron de que la solución que ofrecía el Teorema de Bernoulli no se correspondía con su planteamiento originario del problema. Citados por

[7] Keynes (1921, p. 344) afirma al respecto: "Bernoulli's Theorem supplies the simplest formula by which we can attempt to pass from the *a priori* probabilities of each series of events to a prediction of the statistical frequency of their occurence over the whole series. ...Bernoulli's Theorem involves two assumptions... It is assumed,..., that the probability of the event's occurrence at the *r*th trial is equal *a priori* to its probability at the *n*th trial, and, further, that it is unaffected by a knowledge of what may actually have occurred at the *n*th trial."

Karl Pearson (1925, p. 205) asevera que lo que Bernoulli probó fue que si $p_n = \frac{r}{t}$ es la razón de los casos favorables, y $\frac{1}{t}$ es arbitrariamente pequeño, entonces "we are certain

Corrado Gini (1946, pp. 411-412.), P. Prevost y S. A. Lhuilier (1796) afirman

> Et d'abord Jac. Bernoulli et tous ceux qui ont suivi sa marche, n'ont fait, on doit le dire, que de vaines tentatives pour arriver à cette estimation [des causes]. Leurs méthodes, ..., ne donnoient finalement que l'estimation des effets par la cause. C'est ce qu'on peut reconnoître en jetant les yeux sur le grand problème, en apparance expérimental, qui est résolu à la fin de l'*Ars conjectandi*, ... *Dans ce problème, il s'agit de déterminer d'après la connoissance de la nature d'un dé, la probabilité qu'en jouant un très-grand nombre de coups, on obtiendra des résultats contenus entre certaines limites, voisines d'un rapport qu'indiquent les faces du dé. Ainsi on conclut dans ce problème de la cause aux effets, et non des effets à la cause.*

Y aunque Isaac Todhunter (1865, §125) parece mostrarse de acuerdo con el planteamiento general de Bernoulli[8], Keynes (1921, p. 336)[9] y Corrado Gini (1946, pp. 408-409, y 1949, p. 126)[10] apuntan acertadamente a la fala-

that in *nt* trials ... the odds are *c* to 1 that the ratio of successes to trials will fall within the limits $p_n \pm \frac{1}{t}$." Finalmente Ivo Schneider (1980, pp. 18-19) concluye: "What [Bernoulli] establishes in this theorem is that as the number of observations of a repeatable event increases, so too does the probability that the relative frequency of occurrence of a possible outcome will lie in the vicinity of the probability of this outcome. Only much later did it become clear that he did and could prove this theorem only for the relative frequency of events of known *a priori* probability, but not for those of unknown probability."

[8] "*Suppose that ... we do not know anything beforehand of the ratio [in a urn] of the white* balls to the black; but that we have made a large number of drawings, and have obtained a white ball *R* times, and a black ball *S* times: then according to James Bernoulli we are to infer that the ratio of the white balls to the black balls in the urn is approximately *R/S*. To determine the precise numerical estimate of the probability of this inference requires further investigation: ...this has been done in two ways, by an inversion of James Bernoulli's theorem, or by the aid of another theorem called Bayes's theorem; the results approximately agree."

[9] Para Keynes la expresión *Law of Great Numbers* "suggests, as perhaps Poisson intended to suggest, *but what is certainly false*, that every class of event shows statistical regularity of occurrence if only one takes a sufficient number of instances of it. It also encourages the method of procedure, by which it is thought legitimate to take any observed degree of frequency or association, which is shown in a fairly numerous set of statistics, and to assume with insufficient investigation that, because the statistics are *numerous*, the observed degree of frequency is therefore *stable*."

[10] "Auf Grund der Lösung eines Problems direkter Wahrscheinlichkeit hält man sich deshalb *ungerechtfertigterweise* zu einer Schlußfolgerung ermächtigt, welche die Lösung eines *Problems inverser Wahrscheinlichkeit voraussetzt*, zu der aber die Elemente fehlen."

"Più grave e fondamentale obiezione, che il Leibniz non aveva fatto, nè il Bernoulli si era prospettato, è che risalire dalla frequenza osservata alla probabilità ignota costituisce un problema di probabilità inversa, mentre il teorema die Bernoulli risolve un problema di probabilità diretta, insegnando a desumere dalla probabilità, nota a priori, la frequenza osservata."

cia lógica en que el matemático suizo incurre. En consonancia con este enfoque crítico Hacking (1971, p. 225) considera que no hay razón para pensar que Bernoulli estuviese interesado en calcular $P(p$ está en $f_0 \pm \varepsilon | f_0)$, e.d. en la evaluación *a posteriori* de las estimaciones de p; una tarea ésta cuya solución habría de esperar aún cincuenta años a que Thomas Bayes ofreciese en su *Essay* de 1763[11] la primera aproximación en la historia de la probabilidad al problema de la *probabilidad inversa*.

Aunque el juicio de la historia no ha podido ser más desfavorable para la intención declarada de Bernoulli, no faltaron empero voces a favor suyo. Nadie puede negar autoridad a Antoine Augustin Cournot, corresponsable con Robert Leslie Ellis y John Stuart Mill del concepto frecuencial de probabilidad. Pues bien, Cournot (1843 p. 55 y pp. 437-438) defiende así el uso del teorema de Bernoulli para la determinación experimental de la probabilidad matemática:

> le principe de Jacques Bernoulli conduit à cette détermination expérimentale [de la probabilidad matemática]: car si en désignant par x la chance inconnue de la production d'un événement, par n le nombre de fois que cet événement est arrivé en m épreuves, on peut toujours obtenir une probabilité P que l'écart fortuit $x - \frac{n}{m}$ tombe entre les limites $\pm l$ (le nombre l et la différence $1\text{-}P$ tombant au-dessous de toute grandeur assignable, pourvu que les nombres m, n soient suffisamment grands), il est clair que, si rien ne limite le nombre des épreuves, la probabilité x peut être déterminée avec une précision indéfinie;

Y aunque el representante tal vez más destacado del frecuentismo moderno, Richard von Mises, rechace la existencia de cualquier vínculo lógico entre el teorema matemático de Bernoulli y el concepto *frecuencial* de probabilidad, fundamenta no obstante su definición de probabilidad[12] en una *ley*

[11] Esta insuficiencia del Teorema de Bernoulli, e.d. su incapacidad para permitir el cálculo, a partir de la frecuencia observada, de un intervalo para el parámetro poblacional desconocido p, ha sido considerada por Anders Hald (1990), p. 263 una de las razones posibles por las que Bernoulli no completó su trabajo. Para Karl Pearson (1925), p. 202, "All Bernoulli achieved was to show that by increasing the number of observations the results would undoubtedly fall within certain limits, but he failed entirely to determine what the *adequate* number of observations were for such limits. That was entirely De Moivre's discovery". Y en el mismo sentido Stephen Stigler (1986), p. 77 asevera que el número de observaciones que Bernoulli precisa — 25.550 — para estimar la probabilidad de 'éxito' es tan grande que adolece de falta de aplicabilidad práctica.

[12] Von Mises (1919, §1): "Sea A una propiedad de los elementos de [un colectivo] K, N_A el número observado, entre los primeros N elementos, que son A; entonces existe para A el

empírica de los grandes números, es decir en el hecho empírico de que las frecuencias relativas observadas en series indefinidamente prolongadas convergen hacia valores límites constantes[13].

5. El problema de la inferencia estadística

Supongamos que llevamos a cabo el siguiente experimento aleatorio: lanzamos uniformemente n monedas supuestamente homogéneas y observamos el número a de veces que aparece 'cara'. Ignoramos cuál es la probabilidad —inicial o *a priori* — de 'cara', y desearíamos diseñar un procedimiento para su estimación. Si la intención originaria de Bernoulli en el *Ars Conjectandi* hubiese sido viable, podríamos decir que, con n tendiendo a infinito, la diferencia entre la probabilidad desconocida p de 'cara' y la frecuencia observada a/n de 'cara' sería despreciable. Pero como vimos, la teoría de probabilidades no justifica esta idea.

Sí sabemos, en cambio, que, cualquiera que sea p, la probabilidad matemática del resultado observado: a veces 'cara' en n ensayos independientes, viene dada por

$$\binom{n}{a} p^a (1-p)^{n-a}.$$

Ahora bien, como el coeficiente $\binom{n}{a}$ indica simplemente las diferentes formas —el número de combinaciones lógicamente posibles— en que las a 'caras' pueden aparecer en los n ensayos, y además es independiente de p, podemos prescindir de él, con lo que nos queda la función

$$f(p) = p^a (1-p)^{n-a}.$$

Ésta está definida sobre el intervalo compacto $[0,1]$, en el que p toma valores. $f(p)$ tiene pues un máximo, que se obtiene de

$$\frac{d}{dp}[p^a (1-p)^{n-a}] = 0$$

valor límite $\lim_{N} \frac{N_A}{N} = W_A$, o *probabilidad en el colectivo K* del atributo *A*."

[13] Von Mises (1928), pp. 115-116.

y que resulta ser $p = \frac{a}{n}$.

Este cálculo subyace a un procedimiento ideado por Sir Ronald A. Fisher a partir de 1922 con objeto de estimar, a partir de los datos proporcionados por la experiencia, los valores desconocidos de los parámetros poblacionales. El objetivo de Fisher es determinar qué valor del desconocido p es el más *verosímil*. Para ello Fisher (1922, p. 310 *et passim*) define la *verosimilitud matemática* de un valor de un parámetro como una función del mismo, proporcional a la probabilidad de que, si éste fuese verdadero, las observaciones serían las que realmente han tenido lugar.

Desconocemos la probabilidad p de 'cara', pero disponemos de un número infinito de candidatos — todos los números entre 0 y 1 — y la pregunta es: ¿Cómo explicar lo observado, en el sentido de que, si el valor paramétrico elegido para p fuera verdadero, lo observable sería lo observado con más alta probabilidad? Toda hipótesis acerca del valor verdadero de p dará cuenta *tentativamente* de las observaciones. Pero evidentemente aquel valor de p que atribuya a lo observado la máxima probabilidad matemática será el más verosímil. El procedimiento de Fisher que permite el cálculo del valor paramétrico de p máximamente verosímil recibe el nombre de *método de máxima verosimilitud*[14].

Pues bien, de alguna manera este procedimiento recupera la idea bernoulliana de inferir probabilidades a partir de observaciones, ya que, gracias a él, podemos calcular o estimar a posteriori — en sentido de Bernoulli — el valor paramétrico buscado.

Seguramente ni Karl Pearson (1925, p. 205) ni Ivo Schneider (1979, pp. 106-107) piensan concretamente en el método fisheriano de *máxima verosimilitud* cuando afirman respectivamente

> Bernoulli...turns the problem round and says that if the observed value in nt trials is p, then the true value p_0 will lie between $p \pm \frac{1}{t}$ with the given probability. This is rather stated than proved, *but it is of course the kernel of much later developments of importance.*
>
> *Für die weitere Entwicklung der Wahrscheinlichkeitstheorie war dieser Irrtum von Jakob Bernoulli sicherlich fruchtbar.* Nur auf der Basis dieses Irrtums konnte Jakob Bernoulli ein Programm formulieren, das den gesamten Bereich menschlicher, sittli-

[14] Sobre el origen e implicaciones filosóficas de este método puede consultarse Rivadulla (1991*a*), Cap. IV,y (1991*b*) § 2 y 4.

cher und wirtschaftlicher Entscheidungen für die Zukunft zum Gegenstand der von ihm geschaffenen neuen Disziplin machte.

Además, el método de Fisher está lejos de ser ese procedimiento matemático para aprender inductivamente de la experiencia, que había supuesto su autor. Pero sí es indiscutible que el empeño de Jakob Bernoulli de estimación de probabilidades a partir de resultados de experimentos aleatorios alumbra la base de la teoría de la inferencia estadística.

Bibliografía

Bernoulli, J.: 1713, "Teoría de probabilidades". *Llull*, vol. 16 (1993), nº 30, 389-418.

Cournot, A. A.: 1843, *Exposition de la théorie des chances et des probabilités*. Hachette, Paris.

Fisher, R. A.: 1922, "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society of London*, series A, vol. 222, 309-368.

Gini, C.: 1946, "Gedanken zum Theorem von Bernoulli". *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 82, 401-413.

Gini, C.: 1949, "Rileggendo Bernoulli". *Metron* XV, 117-132.

Hacking, I.: 1971, "Jacques Bernoulli's *Art of Conjecturing*". *Brit. J. Phil. Sci.* 22, 209-229.

Hacking, I.: 1975, *The Emergence of Probability*. Cambridge Uni. Pr.

Hald, A.: 1990, *A History of Probability & Statistics and Their Applications Before 1750*. New York, John Wiley & Sons.

Keynes, J. M.: 1921, *A Treatise on Probability*. London, Macmillan & Co. LTD.

Kohli, K.: 1975, "Zur Publikationsgeschichte der *Ars conjectandi*". In *Die Werke von Jakob Bernoulli*, Vol. III, Basel, Birkhäuser Verlag.

Pearson, K.: 1925, "James Bernoulli's Theorem". *Biometrika* 17, 201-210.

Poisson, S. D.: 1837, *Recherches sur la Probabilité des Jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités*. Paris, Bachelier.

Rivadulla, A.: 1991*a*, *Probabilidad e Inferencia Científica*. Barcelona, Editorial Anthropos.

Rivadulla, A.: 1991*b*, "Mathematical Statistics and Metastatistical Analysis". *Erkenntnis* 34, 211-236.

Rivadulla, A.: 1991*c*, "Apriorismo y Base Empírica en los Orígenes de la Estadística Matemática". *Llull* 14, 187-219.

Rivadulla, A.: 1995, "Historia y epistemología de los cambios de significado de *probabilidad*". *Ágora* 14/1, 53-75

Shafer, G.: 1993, "The Significance of Jacob Bernoulli's *Ars Conjectandi* for the Philosophy of Probability Today". Versión en preimpresión

Schneider, I.: 1979, "Die Mathematisierung der Vorhersage künftiger Ereignisse in der Warhscheinlichkeitstheorie vom 17. bis zum 19. Jahrhundert". *Berichte zur Wissenschaftsgeschichte* 2, 101-112.

Schneider, I.: 1980, "Why Do We Find the Origin of a Calculus of Probabilities in the Seventeenth Century?". In J. Hintikka et al. (eds.), *Pisa Conference Proceedings*, Vol. II, Dordrecht, D. Reidel.

Schneider, I.: 1988, *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Darmstadt, Wissenschaftliche Buchgesellschaft

Stigler, S. M.: 1986, *The History of Statistics. The Measurement of Uncertainty Before 1900*. Cambridge, Mass., The Belknap Press of Harvard Uni. Press.

Todhunter, I.: 1865, *A History of the Mathematical Theory of Probability*. Cambridge, Macmillan.

Von Mises, R.: 1919, "Grundlagen der Wahrscheinlichkeitsrechnung". *Mathematische Zeitschrift* 5, 52-99

Von Mises, R.: 1928, *Probability, Statistics and Truth*. G. Allen & Unwin LTD, 2ª ed. rev., Londres 1957 (1ª ed. alemana, 1928)