“El paradigma de la verosimilitud”

palabras clave: verosimilitud, test de hipótesis.

Dr. Jesus Luis Bescos Sinde.

Universidad de Castilla-La Mancha.

RESUMEN:

El uso de los p-valor como medida de la evidencia en favor o en contra de la hipótesis nula sometida a prueba en un contraste estadístico es, cuando menos, equívoco. El enfoque de Neyman y Pearson responde a la necesidad de seguir un curso de acción tras la realización de un contraste (aceptar o no la hipótesis), con lo que excluye la posibilidad de que el soporte de los datos para evaluar la idoneidad de hipótesis alternativas sea insuficiente.

Esta ponencia trata de remarcar que la medida de evidencia adecuada para evaluar el soporte relativo de hipótesis rivales es la función de verosimilitud (o la razón de verosimilitud), entendida en los términos originalmente acuñados por Fisher, quien definio el cociente de verosimilitud como aquella función que muestra la preferencia relativa de los datos por las hipótesis bajo estudio.

Además de consideraciones de tipo lógico, en esta ponencia se recogen los resultados de diversos experimientos que, por métodos de Montecarlo, muestran cómo bajo ciertos requisitos expresados en términos de preferencia relativa, el mecanismo decisorio de Neyman y Pearson falla al apoyar hipótesis que resultan estar escasamente sustentada por los datos.

En el corazón de lo que viene en denominarse teoría de Neyman-Pearson para la determinación de un contraste óptimo, se encuentra el concepto de evidencia estadística que se defiende en esta ponencia: la recogida en la función de verosimilitud de la muestra. Sin embargo, en las interpretaciones habituales de los contrastes, el concepto de evidencia barajado se relaciona con las denominadas probabilidades de error de tipo I y tipo II, probablemente al interpretar estos contrastes de un modo análogo a los denominados contrastes de significación.

En esta ponencia se señalan las principales inconsistencias lógicas del enfoque NP, y se pergeñan las líneas principales del paradigma de la estadística paramétrica que es una consecuencia lógica de aceptar el principio de verosimilitud. Tal aceptación se produce de manera implícita cuando los experimentos se diseñan en atención al concepto de desvianza, o cuando se contrastan hipótesis recurriendo a los test de la razón de verosimilitud o de los multiplicadores de Lagrange. Es por ello que parece adecuado revisar los fundamentos de este paradigma, tratando de contribuir a una mejor comprensión de los métodos prácticos.

A efectos de establecer el marco conceptual de referencia de esta ponencia, supondremos que enfrentamos un fenómeno que contiene una variabilidad, descrita por una familia de posibles distribuciones de probabilidad $\{P_\theta:\theta \in \Theta\}$, donde θ se denomina *parámetro*, y el conjunto Θ al que pertenece el parámetro *espacio paramétrico*. Tal espacio puede variar en diferentes situaciones concebibles, desde una familia de funciones hasta un conjunto de números reales, o vectores. Aquí consideraremos que el espacio paramétrico es un subconjunto de $\mathfrak{R}^r$. Podemos expresar la distribución de probabilidad a través de su función de densidad o cuantía que denotamos por p_θ. La probabilidad de que se produzca la observación x se denota por $p_\theta(x)$ o más brevemente p(x).

Las conjeturas (hipótesis) acerca del comportamiento del fenómeno bajo estudio, se expresan en términos estadísticos afirmando que la variabilidad en el fenómeno se describe por un subconjunto de la familia de distribuciones de probabilidad postulado o, alternativamente, que pertenece a un determinado subconjunto de Θ.

Supongamos que se formulan dos hipótesis rivales simples (aquellas que especifican completamente la distribución de probabilidades de la/s variable/s observada/s, denotadas por x, que toman valores en el espacio muestra, $\mathcal{X}$), denotadas por H_1 y H_2, y tratamos de determinar cuál de ambas es "soportada" mejor por los datos observados, x.

La forma de proceder para determinar la región crítica, C, óptima del test en el paradigma de Neyman-Pearson (en adelante NP), consiste en preguntarse por la existencia una región $C \subseteq \mathcal{X}$, tal que el suceso { $p_1(x)$ / $p(_2(x) < k\}$ bajo H_1 tenga una probabilidad menor o igual a un nivel predeterminado, α, denominada probabilidad de error de tipo I. En estas condiciones, el lema de NP garantiza que tal región será la que presente menor probabilidad de cometer error de tipo II, denotada por β. Error consistente en "elegir" H_1 bajo H_2.

Habitualmente, se entiende que si se han predeterminado probabilidades de error (α, β) suficientemente pequeñas –digamos (0.05, 0.05)-, y las observaciones ulteriores conducen al rechazo de H_1 a favor de H_2 , tal rechazo se realiza sobre una supuesta evidencia "fuerte" o "moderadamente fuerte" en tal sentido, mientras que si dichas probabilidades de error toman los valores (0.1;0.2), la "evidencia" con la que se rechazaría H_1 se considera moderada, sino débil.

Pues bien, uno de los propósitos de esta ponencia es poner de relieve que tal interpretación es lógicamente inconsistente, lo que no constituye una novedad: Pratt (1961 y 1977), Birnbaum (1962, 1970, 1977), Royall (1997).

En primer lugar, la "forma" de la región crítica viene dada por la desigualdad,

$$\{p_1(x) / p(_2(x) < k\} \quad (1)$$

[1]

es decir, por la verosimilitud relativa de la muestra bajo las hipótesis alternativas. En concreto, la regla prescribe inclinarse hacia H_2 a partir de un determinado valor "k", sobre el que más adelante volveremos, y que es el número de veces que es relativamente más verosímil H_1 que H_2. Dada una muestra donde k=2 significa que H_1 es 2 veces "más preferida"[2] por los datos –en términos relativos- que H_2, mientras

[1] Obviamente también se puede definir en términos del suceso $\{p_2(x) / p_1(x) > k\}$, sin que conlleve problema alguno de interpretación.

[2] Sobre la interpretación de la funcion de verosimilitud como la expresión del orden natural de

que k=1/2, expresaría la situación contraria. Por tanto, el lema de NP establece una división del espacio muestra en función de la preferencia relativa de los resultados, x, por las distintas hipótesis, estableciendo como región crítica el subconjunto de resultados que "soporten" en mayor medida relativa H_2. Además de esta condición, se impone el cumplimiento de

$$P_\theta [\{p_1(x) / p(_2(x) < k\} / H_1] = \alpha \quad \textbf{(2)}$$

para que el test tenga el tamaño deseado y si hay otro más potente que el obtenido como consecuencia de aplicar (1) y (2), necesariamente ha de corresponder a otra probabilidad α.

En definitiva, ¿cuál es el concepto de evidencia implícitamente manejado en el anterior razonamiento?. Es decir, ¿qué se emplea como portador de la información de los datos para dilucidar entre las diferentes hipótesis? Es claro que la función de verosimilitud. Antes de seguir avanzando, examinemos un ejemplo de función de verosimilitud.

Supongamos que se trata de determinar la probabilidad 'p' asociada al resultado 'cara' al lanzar una moneda. Imaginemos que realizamos 30 lanzamientos y obtenemos 12 caras. La función de verosimilitud de este resultado, que denotamos por l(x,p), es proporcional como sabemos a la probabilidad de observar el resultado 'x' cuando la probabilidad de éxito toma el valor 'p'

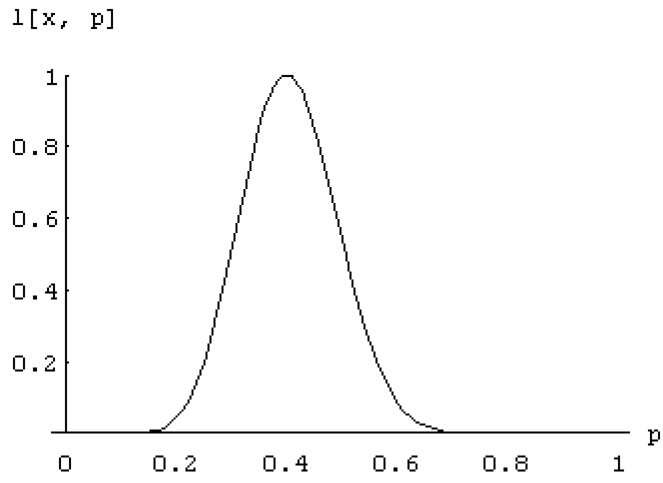
$$l(x,p) = c\ p_\theta(x)$$

donde 'c' es una constante de proporcionalidad, que podemos establecer como

$$c = \frac{1}{\underset{\{p\}}{\max}\ l(x,p)}$$

de modo que la función de verosimilitud está "normalizada" con respecto a su máximo. El gráfico siguiente recoge la forma de esta función para el ejemplo referido

preferencias de los datos por las distintas hipótesis véase Fisher (1973, pag 73)



Como se observa, la función de verosimilitud alcanza su máximo en 0.4, el valor de p más soportado por los datos (o preferido), para disminuir a medida que nos alejamos de este punto. La ordenada de la función para cada valor de 'p' es la versosimilitud de tal valor. Si la comparación se produce por cociente, es obvio que la verosimilitud relativa no queda afectada por el cambio de escala consecuencia de 'normalizar'.

Si nuestro interés estuviera en comparar el soporte relativo de las hipótesis H_1:p=0,3; frente a H_2: p=0,6, el cociente o razón de verosimilitud sería 5.78529, es decir, la hipótesis H_1 es soportada 5.7 veces más, en términos relativos, que H_2. La pregunta inmediata es:¿qué significa ser relativamente preferido 5.7 veces? y ¿cuál es la escala en la que se miden estas preferencias relativas? Dada la definición de la función de verosimilitud, 5.7 es el número de veces que la observación, x, resulta más probable bajo H_1 que bajo H_2. La respuesta en relación a la escala queda, momentáneamente, diferida.

Como puede apreciarse, en el caso de hipótesis simples, la observación y eventual tabulación de la función de verosimilitud permite reflejar la intensidad con que los datos observados soportan los diferentes valores del parámetro. Como consecuencia lógica de tal representación de la evidencia, surgen los ***intervalos de verosimilitud***, - véase por ejemplo Hudson (1971) o Edwards, (1972). En concreto, Hudson los define como:

$$I(\theta) = \left\{ \theta : \mathrm{Ln}\, l(x,\theta) \geq \mathrm{Ln}\, l\left(x,\hat{\theta}\right) - \mathrm{Lnk} \right\}$$

(3)

Es decir, el conjunto de valores del parámetro tales que el valor más preferido lo es, con repecto a ellos, k veces como máximo[3]. A efectos ilustrativos, si en el ejemplo anterior de 12 caras en 30 lanzamientos, consideramos una preferencia relativa máxima de 3 (el máximo, p=0.4, es preferido *como mucho* 3 veces más), el intervalo que obtenemos es (0.274; 0,535)[4]

La aceptación de tal función como la única portadora de 'evidencia' es un argumento extra-matemático, puesto que no es derivable a partir del cuerpo de axiomas de probabilidad. Tal argumento se formula como ***axioma de verosimilitud***: (Edwards, 1972, pag 31): "*Dentro del sistema de un modelo estadístico,* ***toda*** la *información que proveen los datos en cuanto a los méritos relativos de dos hipótesis está contenida en el cociente de verosimilitud de esas hipótesis sobre los datos, y el cociente de verosimilitud se interpretará como el grado en que los datos soporta una hipótesis frente a la otra*".[5]

Para Lindsey (1996, pag 73), al aceptar este principio "*lo que hemos hecho es utilizar la variabilidad de los posibles resultados del mecanismo generador de los datos, específicamente, la incertidumbre sobre que el resultado observado hubiera ocurrido, para proveer una medida de la incertidumbre sobre los parámetros del modelo utilizado para describir ese mecanismo. Esto debe ser distinguible de cualquier incertidumbre acerca de la elección del modelo (función) debida a la carencia de información (teórica), que no aparece en la verosimilitd (...) El principio de verosimilitud no establece que la inferencia estadística deba basarse solamente en este principio. Es más, es una afirmación relativa, comparando modelos, pero sin proveer conocimiento absoluto acerca de cualquier modelo posible. Esto es lógico si ningún modelo puede ser cierto. Lo único que queremos es el mejor modelo entre los disponibles para que nos ayude a comprender cómo pueden ser generados los*

---

[3] El logartimo de la función de verosimilitud es denominado **"función de soporte".** En caso de considerar la función de verosimilitud normalizada, lo que tenemos es la expresión de las preferencias relativas en términos aditivos. De ahí el nombre de soporte.

[4] Obsérvese que obtenemos un conjunto de valores "ordenados" en cuanto a la preferencia que la observación manifiesta por ellos, a diferencia de los tradicionales intervalos de confianza que constituyen una clase de equivalencia, en términos de "compatibilidad" de los valores del parámetro incluidos en el intervalo con la observación. Además, el intervalo de verosimilitud del ejemplo **no** es simétrico respecto a 0.4.

[5] Birnbaum (1962) demuestra que el principio de verosimilitud (L) - sostenido por la llamada estadística bayesiana - es una consecuencia de aceptar conjuntamente los principios de suficiencia (S) y condicionalidad (C) con los que los autodenominados no bayesianos están ampliamente de acuerdo En escritos posteriores (1977) concluyó que el principio de condicionalidad implica el de suficiencia, por lo

datos"

Admitido este axioma, es relativamente fácil comprender por qué las probabilidades de error no deben interpretarse como medida de la evidencia.

En primer lugar, puesto que en general, β es la probabilidad en la cola de la distribución bajo H_2, resulta que cuanto más potente sea el test (menor β) menos compatibles resultaran los datos con H_2 y, por tanto, menor evidencia relativa para H_2 en favor de H_1.. Tomemos un ejemplo de Pratt (1977, pag 65) : "si x es N(0,1) bajo H_1 y N(θ,1) bajo H_2 , y x=1.645 (α=0.05), entonces H_2 es más plausible si θ=2 (β=0.36) que si θ=4 (β=0.009).

Además, dado que la relación entre las probabilidades de error del test y la razón de verosimilitud es[6]:

$$\alpha \le \frac{\alpha}{1-\beta} \le \lambda(x) \le \frac{1-\alpha}{\beta} \le \frac{1}{\beta}$$

(donde λ(x) es la razón de verosimilitud.), es inmediato que si las proabilidades de error se fijan de antemano en, digamos, (α=0.05, β=0.05) la razón de verosimilitud puede tomar cualquier valor comprendido entre 1/19 y 19, y si tal cociente expresa la "intensidad" del soporte relativo de los datos por las dos hipótesis en juego, está claro que podemos encontrarnos con muestras que soporten H_1 , o H_2 ,o que carezcan de información para dilucidar entre ambas hipótesis.

Consideremos el siguiente ejemplo tomado de Royall (1996, pag16). Se trata de dilucidar entre dos hipóteis relativas a la probabildad de éxito, θ, en 30 repeticiones independientes de un experimento de Bernoulli, H_1 θ=1/4 y H_2: θ=3/4. Cuando el número observado de éxitos se representa por x, la razón de verosimilitud es $p_2(x)/p_1(x) = 3^{2x-30}$ (en el numerador H_2). La mejor región crítica con tamaño α=0.05 contiene todos los valores de x para los cuales la razón de verosimilitud es como mínimo $k = 3^{24-30} = 1/729$, es decir $x \ge 12$. (Bajo H_1 la probabilidad de 12 éxitos o más en 30 repeticiones es sólo 0.05).

Como se ve en este ejemplo, el "mejor" test llama a elegir H_2 cuando la evidencia favorece, de hecho, a H_1: un factor k menor que la unidad. En concreto, si

---

que (C) y (L) resultan ser equivalentes.

[6] Una demostración puede encontrarse en Silvey (1970) en el capítulo dedicado al análisis secuencial.

se observa x=12 el test determina la elección de H_2 cuando los datos soportan 729 veces más a H_1, en términos relativos. De modo análogo, los resultados x=13 y x=14 representan evidencia bastante fuerte en favor de H_1; y en cuanto a la observación x=15, que corresponde a una tasa de éxito de 1/2, igualmente distante de las hipótesis formuladas ($\theta=1/4$ y $\theta=3/4$), mientras que el método NP determina la elección de H_2, la evidencia relativa en términos de verosimilitud es neutral entre las hipótesis $p_2(x)/p_1(x)=1$.

Abordemos ahora el problema de la "escala" en la que medir la verosimilitud relativa. Lindley (1996) refiere los resultados de Kass y Raftery[7] que proporcionan una interpretación de los valores de la función de verosimilitud normalizada (con repecto a su máximo) recogida en la tabla siguiente:

Valores de la razón de verosimilitud normalizada	***Evidencia***
0.37 - 1	Pequeña
0.05 - 0.37	Positiva
0.007-0.05	Fuerte
<0.007	Muy Fuerte

Por su parte, Royall (1996) define una escala de evidencia por comparación a un experimento que denomina "canónico" y que resumimos a continuación. Supongamos que se tienen dos hipótesis sobre las proporciones de la composición de colores de las bolas que integran una urna. Las dos hipótesis son: 'todas blancas' ,y 'la mitad blancas'. Las bolas son extraidas con reposición, removiendo la urna antes de cada extracción.

Supongamos que se extrae una bola que resulta ser blanca, a continuación otra blanca. Si en la tercera extracción la bola resulta blanca, se podrá caracterizar tal situación como evidencia "bastante fuerte" en favor de "todas blancas" frente a "la mitad blancas". La razón de verosimilitud es $2^3=8$.

A partir de este ejemplo construye la tabla siguiente razonando de este modo: si observamos b blancas, la razón de verosimilitud en favor de 'todas blancas' sobre

[7] Kass y Raftery (1995) Bayes factors. Journal of the American Statistical Association, 90, 773-795

'la mitad blancas' es igual a $1/\left(1/2\right)^b$, o 2^b. Por tanto una verosimilitud relativa de 'k' corresponde a 'b' bolas blancas consecutivas, donde $k = 2^b$.

Razón verosimilitud	10	20	50	100	1000
b	3.3	4.3	5.6	6.6	10.0

Estamos ahora en condiciones de evaluar sobre una base más objetiva, cuánta evidencia representa un factor k= 5.7 o un factor k=729.

Para ilustrar convenientemente la inconsistencia del paradigma NP con la interpretación de la evidencia en términos de la razón de verosimilitud -corazón de dicho paradigma- hemos elaborado el siguiente ejemplo. Tomamos observaciones procedentes de una población $N(\mu,\sigma^2)$, donde la media es desconocida y se barajan dos hipótesis alternativas para la varianza, H_1: $\sigma^2=16$ y H_2 : $\sigma^2=49$. Se consideran muestras de tamaño 20 y la regla de decisión viene determinada por $C = \left\{ X: \sum \left(x - \bar{x} \right)^2 > 482{,}28 \right\}$ que conduce al rechazo de H_1 en favor de H_2. Esta regla se ha fijado de modo que las probabilidades de error se pueden cifrar en (0.05, 0.043), por lo que, de acuerdo con la interpretación al uso, cualquier resultado que se alcance bajo estas condicones, se considerará apoyado por evidencia "fuerte". A continuación se procedio a generar, mediante simulación, 1000 muestras de tamaño 20 a partir de x~N(20,4) -H_1- , y otras tantas a partir de x~N(20,7) -H_2-.[8]. Para "representar" la evidencia en términos de verosimilitud, el supuesto establecido de desconocimiento de μ , implica que en la función de verosimilitud obtenida a partir de la función de densidad de la distribución normal, está presente μ como parámetro de ruido, por lo que hemos elegido representar la verosimilitud de la muestra sobre σ^2, a través de la denominada *verosimilitud marginal* [9]-obtenida a partir de la distribución en el muestreo del estadístico $(n-1)s^2$. El logaritmo de tal función -normalizada con respecto al máximo- es:

[8] Tomar la media de la población como 20, se ha hecho por conveniencia, para poder generar las muestras. De hecho para realizar las "inferencias" a partir de la función de verosimilitud, y fijar las probabilidades de error de tipo I y II, se ha operado considerando μ desconocida.

[9] La idoneidad de esta elección descansa en que esta función de verosimilitud es "robusta" para distintos valores de la media . Tsou (1995)

$$L(\sigma^2) = -\frac{n-1}{2}\left\{Ln\left(\frac{\sigma^2}{s^2}\right) - 1 + \frac{s^2}{\sigma^2}\right\}$$

Se aplica esta función, particularizándola para las dos hipótesis sobre la varianza, a cada una de las muestras, y se obtiene la razón de verosimilitud $p_2(x)/p_1(1x)$, como $\lambda(x) = Exp[L(\sigma^2{}_2)-L(\sigma^2{}_1)]$.

Elegimos considerar el valor 8 para la razón de verosimilitud, como nivel de evidencia "bastante fuerte" -tres extracciones consecutivas de bolas blancas en el experimento canónico de Royall- de modo que si $\lambda>8$ se considera evidencia fuerte en favor de H_2; si $\lambda<1/8$ se considera evidencia fuerte en favor de H_1; y si toma valores intermedios, se considera evidencia débil para dilucidar entre las dos hipótesis.

Por supuesto, a cada una de las muestras se les aplicó la regla de decisión derivada del paradigma NP, tal y como se formuló con anterioridad.

La tabla siguiente resume los principales resultados de esta simulación y muestra como, bajo el concepto de evidencia subyacente en el cociente de verosimilitud, las "decisiones" derivadas de la apliación de la regla NP, son inconsistentes. En concreto, lo más relevante es que muchas decisiones-NP se toman con un nivel de evidencia "débil". Esto es especialmente importante si consideramos que a la hora de contrastar hipótesis no sabemos, a direfencia de este ejemplo, cuál de ellas es cierta, si es que alguna lo es. Y que el verdadero problema consiste en dilucidar cuál es "relativamente" más consistente con la observación.

La primera columna indica el número de la serie correspondiente a las 10 tandas de muestras de tamaño 20 generadas bajo las dos hipótesis. La segunda columna, el número de veces que la regla NP condujo al rechazo de H_1. La tercera, el número de veces que el cociente de verosimilitud favoreció H_2 (el rechazo de H_1)con un factor de al menos 8. Obviamente esta evidencia es equívoca cuando los datos se generan a partir de N(20,4)). La cuarta, el número de veces que el cociente de verosimilitud favoreció H_1 en un factor de al menos 8. La última columna el número de veces que la evidencia, en forma de cociente de verosimilitud, fue débil.

Datos generados bajo H1: σ^2=16				
Numero de la serie	Rechazo H1 según NP	Evidencia fuerte errónea en contra de H1	Evidencia fuerte a favor de H1	Evidencia Débil
1	5	1	85	14
2	2	0	93	7
3	4	0	88	12
4	3	0	84	16
5	5	0	79	21
6	4	0	90	10
7	6	0	84	16
8	7	4	81	15
9	4	1	87	12
10	5	2	87	11
Total	45	8	858	134
Porcentaje	**0,045**	**0,008**	**0,858**	**0,134**

Datos generados bajo H2: σ^2=49				
	Rechazo H1 según NP	Evidencia fuerte en contra de H1	Evidencia fuerte errónea en contra de H2	Evidencia Débil
1	96	88	2	10
2	95	90	3	7
3	94	84	3	13
4	96	87	3	10
5	98	91	1	8
6	97	88	1	11
7	96	84	1	15
8	97	91	1	8
9	95	86	2	12
10	100	93	0	7
Total	964	882	17	101
Porcentaje	**0,964**	**0.882**	**0,017**	**0,101**

Como conclusiones podemos señalar que la supuesta base evidencial resultante de trabajar con una regla cuyo comportamiento medio es óptimo según cierto criterio (paradigma NP) no es válida. Si bien se observa, en esta simulación, que al operar con probabilidades de error pequeñas, es muy infrecuente que se presenten situaciones con evidencia fuerte “equívoca" -fecuencia del 8 por mil bajo H_1 y del 1,7% bajo H_2- , la frecuencia de aparición de evidencia “débil” es muy superior: por encima del 10% bajo las dos hipótesis. Por lo tanto, de la mera aplicación de la regla de decisión dicótoma (NP) no se puede garantizar, ni siquiera en términos de “frecuencia aceptable” a largo plazo en sucesivas repeticiones del muestreo, que la observación ***soporte fuertemente*** tal decisión.

Si se quiere hacer consistente ambas visiones es necesario enfocar el problema precisamente en el orden inverso al empleado por Neyman-Pearson. Primero determínese qué nivel de "evidencia" -en términos del cociente de verosimilitud- se desea para discernir entre diferentes hipótesis y, luego podremos, eventualmente y siempre que interese, determinar las probabilidades de error.

Ahora bien, si empleamos la razón 'k' como medida de la evidencia, sería tranquilizador, a la hora de diseñar un experimento, estar en condiciones de determinar el tamaño muestral necesario para ser capaces de discernir entre hipótesis "verdaderas" e hipótesis "falsas" con cualquier nivel de exigencia (en términos de k) y con casi seguridad. En otras palabras, supongamos que X se distribuye de acuerdo con p_B, (la hipóteisis "cierta") y que p_A es otra distribución de probabilidad (la hipótesis "falsa"); dada una constante k>0 podemos afirmar:

$$\Pr\left[p_A(X) \big/ p_B(X) \geq k \right] \leq \frac{1}{k}$$

(4)

Es decir, la probabilidad de obtener evidencia equívoca en favor de la hipótesis falsa es menor, cuanto mayor sea la fuerza de la evidencia (k) requerida.

La demostración es inmediata. Si S es el conjunto de valores de x que producen un cociente de verosimilitud en favor de la hipóteisis A de al menos k unidades, cuando B es cierta:

$$\Pr(S) = \sum_S p_B(X) \leq \sum_S p_A(X) \big/ k \leq 1/k$$

(5)

Expresión que establece la práctica imposibilidad de encontrar evidencia fuerte en favor de una hipótesis falsa.

Pero además, a partir del resultado anterior es inmediato concluir que si X_i se distribuyen iid bajo H_B, y A etiqueta otra hipótesis, el cociente $\prod_{i=1}^{n} p_A(x_i) / p_B(x_i)$ converge a cero con probabilidad 1. Lo que significa que es posible especificar cualquier número k por grande que sea, tal que encontraremos un número 'n' de observaciones que con probabilidad prácticamente 1, la evidencia en favor de B será de al menos k.

Por supuesto, la mayor parte de las conclusiones son trasladables al caso de los denominados "contrastes de significación", donde la medida de la evidencia allí manejado se expresa a través de determinadas probabilidades. Por cuestiones de espacio no abordamos aquí este problema. Baste señalar que, la idea subyacente bajo el paradigma de la verosimilitud es evaluar el soporte relativo de los datos por hipótesis rivales, sin pretender aportar una media *absoluta* del soporte, como se deriva de la apliación corriente de los contrastes de significación.

REFERENCIAS

1	Birnbaum	A.	1962	"On the Foundations of Statistical Inference", Journal of the American Statistical Association, 57, 269-326
2	Birnbaum	A.	1970	"Statistical Methods in Scientific Inference" (Letter), Nature, 225,1033
3	Birnbaum	A.	1977	"The Neyman-Pearson Theory as Decision Theory, and as Inference Theory; with a Criticism of the Lindley-Savage argument for Bayesian Theory", Synthese 36, 19-49
4	Edwards	A.W.	1972	"Likelihood". Cambridge University Press.
5	Fisher	R.A	1973	"Statistical Mathods and Scientific Inference", Hafner Press, London
6	Hudson	D.J.	1971	"Interval Estimation from the Likelihood Function", Journla of the Royal Statistical Society, B33, 256-262
7	Lindsey	J.K	1996	"Parametric Statistical Inference",Clarendon Press, Oxford
8	Pratt	J.W.	1961	"Review of Testing Statistical Hypotheses by E.L Lehmann", Journal of the American Statistical Association, 56, 163-167
9	Pratt	J.W.	1977	" 'Decisions' as statistical evidence and Birnbaum's 'confidence concept". Synthese 36, 59-69
10	Royall	R.	1997	"Statistical Evidence. A Likelihood paradigm", Chapman & Hall
11	Silvey	S.D.	1970	"Statistical Inference", Chapman & Hall
12	Tsou	T.S	1995	"Robust Likelihood". Journal of the American Statistical Association", v 90, 429, 316-320