*South American Research Journal, 3*(1), 67-83
https://sa-rj.net/index.php/sarj/article/view/39

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

# Explorando la Reconciliación entre los Enfoques Frecuentista y Bayesiano en Estadística

## *Exploring Reconciliation between Frequentist and Bayesian Approaches to Statistics*

Juan Carlos Abril[1] y María de las Mercedes Abril[1]

[1]Universidad Nacional de Tucumán y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Av. Independencia 1900, San Miguel de Tucumán, Tucumán, Argentina.

Correspondencia: jabril@herrera.unt.edu.ar; mabrilblanco@hotmail.com

Recepción: 4 de junio de 2023 - Aceptación: 1 de agosto de 2023 - Publicación: 16 de agosto de 2023

## Resumen

En estadística, la estadística frecuentista a menudo se ha considerado como la única vía. No obstante, desde la década de 1950, la estadística bayesiana ha ido ganando progresivamente terreno en la academia. El presente estudio tiene como propósito demostrar los puntos de encuentro entre estas dos corrientes aparentemente opuestas. Para ello, los autores realizan un recorrido por varios tópicos, explicando qué es el Teorema de Bayes mediante ejemplos didácticos. En contraparte, se muestra que los frecuentistas rechazan el postulado central del enfoque Bayesiano, pero se ven obligados a reemplazarlo con soluciones alternativas, siendo la más generalizada la Máxima Verosimilitud. Frente a esta discrepancia, los autores sugieren que podría tratarse de una mala interpretación entre ambas corrientes y ofrecen ejemplos en los que el postulado de Bayes y el principio de Máxima Verosimilitud arrojan la misma respuesta numérica. Luego, se analizan las inferencias a partir de información a priori, tanto no informativa como informativa, y se exploran las propuestas inferenciales de ambas escuelas. Además, se aborda el enfoque fiducial, que trabaja con cantidades ficticias. Todos estos aspectos son discutidos desde las perspectivas matemáticas de reconocidos estadísticos como Fisher, Keynes, Carnap, Good, Durbin, Box, Giere, Neyman, Pearson, entre otros. Además, se buscan suposiciones filosóficas que filósofos como Lakatos, Popper y Kuhn, entre otros, no han logrado ofrecer para establecer una posible reconciliación entre estas corrientes en aparente conflicto.

**Palabras claves:** Enfoque clásico, Enfoque frecuencial, Enfoque basado en la verosimilitud, Enfoque fiducial, Enfoque Bayesiano objetivo, Enfoque Bayesiano subjetivo, Teoría de la decisión.

## *Abstract*

In statistics, frequentist statistics has often been considered the only way. However, since the 1950s, Bayesian statistics has been progressively gaining ground in academia. The purpose of the present study is to demonstrate the meeting points between these two apparently opposing currents. To this end, the authors review several topics, explaining what Bayes' Theorem is by means of didactic examples. On the other hand, it is shown that the frequentist reject the central postulate of the Bayesian approach, but are forced to replace it with alternative solutions, the most generalized being the Maximum Likelihood. Faced with this discrepancy, the authors suggest that it could be a misinterpretation between both currents and offer examples in which Bayes' postulate and the Maximum Likelihood principle yield the same numerical answer. Then, inferences from a priori information, both non-informative and informative, are analyzed and the inferential proposals of both schools are explored. In addition, the fiducial approach, which works with fictitious quantities, is discussed. All these aspects are discussed from the mathematical perspectives of renowned statisticians such as Fisher, Keynes, Carnap, Good, Durbin, Box, Giere, Neyman, Pearson, among others. In addition, philosophical assumptions that philosophers such as Lakatos, Popper and Kuhn, among others, have failed to offer are sought in order to establish a possible reconciliation between these currents in apparent conflict.

**Keywords:** Classical approach, Frequential approach, Likelihood-based approach, Fiducial approach, Objective Bayesian approach, Subjective Bayesian approach, Decision theory.

## 1. Introducción

La teoría de la probabilidad, nos lleva de las probabilidades dadas de eventos primarios a las probabilidades de eventos más complejos basadas en ellas. En la práctica estadística generalmente buscamos hacer inferencias en la dirección inversa; es decir, dadas las observaciones, requerimos saber algo sobre la población de donde emanaron o el mecanismo generador por el cual se produjeron.

La inferencia estadística es un proceso inductivo que va de la muestra a la población. Al pensar en una hipótesis $(H)$ y datos observacionales, o evidencia $(E)$, no hay problema en hacer enunciados probabilísticos de la forma $P(E|H)$; de hecho, estos están justificados por la lógica deductiva una vez que se especifican los axiomas de probabilidad, y tales afirmaciones se han utilizado repetidamente por muchos autores, incluso nosotros. Sin embargo, se ha cuestionado la existencia misma de enunciados inductivos de la forma $P(H|E)$ y muchos filósofos, en particular Sir Karl Popper (1968, 1969), han concluido que tales probabilidades no existen. Tales probabilidades son, por supues-

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

to, las probabilidades a posteriori del enfoque Bayesiano, por lo que el debate, que no muestra signos de disminuir, es de vital interés para los estadísticos. Para más detalles consultar Popper y Miller (1987), Good (1988), Gemes (1989) y Miller (1990).

Cualquier procedimiento inferencial debe basarse en un conjunto de reglas más o menos racional, pero la racionalidad de cualquier sistema dado y el valor aparente de las conclusiones que permite alcanzar permanecen abiertos a debate.

En nuestra vida académica y profesional hemos adoptado el paradigma *frecuentista*, a veces conocido como enfoque *clásico* o *frecuencial*, que ha sido la escuela dominante de pensamiento estadístico durante la mayor parte de los siglos XX y XXI. Sin embargo, el punto de vista *Bayesiano* ha ganado popularidad desde la década de 1950 y en los últimos años se han desarrollado varios otros enfoques de la inferencia, algunos más completos que otros.

En este trabajo intentamos esbozar tanto las áreas de acuerdo como las diferencias entre las principales escuelas; no es nuestro interés desarrollar cada enfoque en detalle.

Dado que nuestra discusión es una evaluación bastante breve de una extensa y compleja literatura, enfatizaremos solo los puntos principales en cuestión. Por lo tanto, examinamos las posiciones "estándares" dentro de cada escuela y no enfatizamos los debates dentro de una escuela (por ejemplo, la elección de axiomas para la probabilidad subjetiva). Esperamos que estos grandes trazos sirvan para producir retratos y no caricaturas.

Por lo tanto, la teoría de la probabilidad, nos lleva de las probabilidades dadas de eventos primarios a las probabilidades de eventos más complejos basadas en ellas. En la práctica estadística generalmente buscamos hacer inferencias en la dirección inversa; es decir, dadas las observaciones, requerimos saber algo sobre la población de donde emanaron o el mecanismo generador por el cual se produjeron. Más adelante iniciaremos un estudio sistemático de los diversos métodos y procesos inferenciales que se emplean en Estadística con este fin. En esta etapa nos limitaremos a dar un relato introductorio, en términos muy amplios, con el objeto de dar algún punto inicial a los temas considerados más adelante.

## 2. El Teorema de Bayes

Sea $B_1, B_2, \ldots, B_n$ un conjunto de eventos mutuamente excluyentes y exhaustivos del espacio muestral $\Omega$, sea $A$ otro evento de $\Omega$ y sea $H$ la información actualmente disponible. De

$$P(A \cap B) = P(A|B)\,P(B)$$

y

$$P(A \cap B) = P(B|A)\,P(A)$$

tenemos

$$\begin{aligned} P(B_r \cap A|H) &= P(B_r, A|H) \\ &= P(A|H)\,P(B_r|A,H) \\ &= P(B_r|H)\,P(A|B_r,H). \end{aligned} \tag{1}$$

Por lo tanto

$$P(B_r|A,H) = \frac{P(B_r|H)\,P(A|B_r,H)}{P(A|H)}. \tag{2}$$

Usando la Ley de la Probabilidad Total podemos sustituir a $P(A|H)$ en (2). Luego encontramos

$$\begin{aligned} P(B_r|A,H) &= \frac{P(B_r|H)\,P(A|B_r,H)}{\sum_r \{P(B_r|H)\,P(A|B_r,H)\}} \\ &= \frac{P(B_r,A|H)}{\sum_r P(B_r,A|H)}. \end{aligned} \tag{3}$$

Esto se conoce como Teorema de Bayes, en honor a Thomas Bayes (1764), quien lo propuso por primera vez. Establece que la probabilidad de que $B_r$ ocurra dada la ocurrencia de $A$ y la información $H$ es proporcional a la probabilidad de $B_r$ dado $H$ multiplicada por la probabilidad de $A$ dado $B_r$ y $H$.

El teorema da las probabilidades de $B_r$ *cuando se sabe que ha ocurrido* $A$. Las cantidades $P(B_r|H)$ se denominan *probabilidades a priori*, las de tipo $P(B_r|A,H)$ se denominan *probabilidades a posteriori* y $P(A|B_r,H)$ se denomina *verosimilitud*. El teorema de Bayes puede entonces replantearse en la siguiente forma: la probabilidad a posteriori varía como la probabilidad a priori multiplicada por la verosimilitud.

## 3. El postulado de Bayes

De esta forma, el teorema se ve como una simple consecuencia lógica de las reglas de probabilidad y es indiscutible. Lo que ha suscitado críticas en el pasado ha sido el uso que se le ha dado al teorema. Hay un principio implícito de que, si tenemos que elegir una de las $B_r$, tomamos la de mayor probabilidad a posteriori. Esto es equivalente a elegir la hipótesis que maximiza la *probabilidad conjunta* de $B_r$ y $A$ como se ve inmediatamente del extremo derecho de la ecuación (3). La dificultad surge del hecho de que para calcular las probabilidades a posteriori requerimos conocer las probabilidades a priori. Estas son, en general, desconocidas, y Bayes sugirió que cuando esto sea así, se debería suponer que son iguales; o más bien, que debían ser asumidas iguales donde nada se supiera en contrario. Esta suposición, conocida como el Postulado de Bayes, el Principio de Equidistribución de la Ignorancia y por uno o dos nombres más, proporcionó uno de los puntos más polémicos en la teoría de la inferencia estadística. Antes de discutir el punto, puede ser útil dar algunos ejemplos.

### 3.1. Ejemplos

1. Una urna contiene cuatro bolitas, que se sabe que son (a) todas blancas o (b) dos blancas y dos negras. Se saca una bolita y se encuentra que es blanca. ¿Cuál es la probabilidad de que todas las bolitas sean blancas?

   Tenemos aquí dos hipótesis, $B_1$ y $B_2$. En $B_1$ la probabilidad de sacar una bolita blanca es 1, en $B_2$ es $1/2$. De (3)

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

tenemos

$$\begin{aligned} P(B_1|A,H) &= \frac{P(B_1|H)}{P(B_1|H)+\frac{1}{2}P(B_2|H)} \\ P(B_2|A,H) &= \frac{\frac{1}{2}P(B_2|H)}{P(B_1|H)+\frac{1}{2}P(B_2|H)}. \end{aligned}$$

Ahora, de acuerdo con el postulado de Bayes asumimos

$$P(B_1|H) = P(B_2|H) = \frac{1}{2}$$

y encontramos

$$\begin{aligned} P(B_1|A,H) &= \frac{2}{3} \\ P(B_2|A,H) &= \frac{1}{3}. \end{aligned}$$

Si tuviéramos que elegir entre las dos posibilidades (a) y (b) deberíamos seleccionar la de mayor probabilidad a posteriori, es decir, aceptamos el supuesto de que las bolitas son todas blancas.

Ahora supongamos que reemplazamos la bolita y nuevamente sacamos una al azar. Si se encuentra que es negra, la hipótesis (a) se rechaza rotundamente. Pero si resulta ser blanca, podemos calcular nuevas probabilidades a posteriori en las que nuestras probabilidades a posteriori anteriores se vuelven a priori. Ahora tenemos $P(B_1|H) = 2/3, P(B_2|H) = 1/3$, donde $H$ incluye $A$, y una aplicación renovada de (3) nos da las probabilidades a posteriori basadas en el nuevo evento, digamos $A'$,

$$\begin{aligned} P(B_1|A',H) &= \frac{\frac{2}{3}}{\frac{2}{3}+\frac{1}{2}\frac{1}{3}} = \frac{4}{5} \\ P(B_2|A',H) &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{2}{3}+\frac{1}{2}\frac{1}{3}} = \frac{1}{5}. \end{aligned}$$

Estará claro que si repetimos el proceso y nuevamente obtenemos una bolita blanca, la nueva probabilidad a posteriori de (a) será aún mayor. Esto está de acuerdo con el requisito del sentido común; cuanto más tiempo pasemos muestreando (con reemplazo) sin obtener una bolita negra, más probable es que no haya bolitas negras presentes.■

2. Generalizando el Ejemplo anterior, supongamos que sacamos bolitas una a la vez, reemplazándolas después de cada extracción, y obtenemos $n$ bolitas blancas en sucesión. La probabilidad de este evento en la hipótesis (a) es la unidad; en la hipótesis (b) es $1/2^n$. De (3) tenemos, ($A$ se refiere a la observación de todas las $n$ bolas como blancas),

$$\begin{aligned} P(B_1|A,H) &= \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}2^{-n}} = \frac{2^n}{2^n+1}. \\ P(B_2|A,H) &= \frac{1}{2^n+1}. \end{aligned}$$

A medida que $n$ crece, $P(B_1|A,H)$ tiende a la unidad y $P(B_2|A,H)$ a cero.

Además, esta será la verdadera cualesquiera que hayan sido las probabilidades a priori originales. De hecho, si la de la hipótesis (a) es $t$ y la de (b) es $1-t$, encontramos

$$P(B_1|A,H) = \frac{2^n t}{2^n t + (1-t)},$$

que tiende a la unidad para cualquier $t$ distinto de cero. Esto también está de acuerdo con el sentido común. Cualesquiera que sean las probabilidades originales, la nueva evidencia es tan fuerte que las supera.■

3. De una urna llena de bolitas de color desconocido se extrae una bolita al azar y se reemplaza $m$ veces y se saca una bolita negra cada vez ¿Cuál es la probabilidad de que si se extrae otra bolita, ésta sea negra?

   La pregunta tal como está formulada no admite una respuesta definitiva, pues habiendo un infinito número de colores y combinaciones de colores posibles, no sabemos cuáles son las hipótesis a comparar. Supongamos que las bolitas son blancas o negras y, por lo tanto, consideremos las hipótesis (1) de que todas son negras, (2) de que todas menos una son negras, (3) de que todas menos dos son negras, y así sucesivamente. El problema aún carece de precisión, ya que no se especifica el número de bolitas. Supongamos que hay $N$ bolitas. Más adelante dejaremos que $N$ tienda a infinito para obtener el caso límite.

   Considere la hipótesis $B_R$ de que hay $R$ bolitas negras y $N-R$ blancas. La probabilidad de sacar una bolita negra es $R/N$ y la de hacerlo $m$ veces seguidas es $(R/N)^m$. Si los $B$ tienen probabilidades a priori iguales, tenemos, de (3),

$$P(B_R|A,H) = \frac{(R/N)^m}{\sum_{R=0}^{N}(R/N)^m}.$$

   Ahora la probabilidad de obtener otra bolita negra en la hipótesis $B_R$ es $R/N$. Dado que las hipótesis $B_R$ son mutuamente excluyentes, la probabilidad de obtener otra bolita negra es

$$\sum_{R=0}^{N}\frac{R}{N}P(B_R|A,H) = \frac{\sum_{R=0}^{N}(R/N)^{m+1}}{\sum_{R=0}^{N}(R/N)^m}. \qquad (4)$$

   Esta es la respuesta a la forma limitada de la pregunta. Como $N \to \infty$ esto tiende al cociente de integrales definidas

$$\frac{\int_0^1 x^{m+1}dx}{\int_0^1 x^m dx} = \frac{m+1}{m+2}. \qquad (5)$$

   Este es un caso particular de la llamada Regla o Ley de Sucesión de Laplace. Los entusiastas la han aplicado indiscriminadamente en alguna forma incondicional como la afirmación de que si se observa que un evento sucede $m$ veces en sucesión, las posibilidades son $m+1$ a 1 de que vuelva a suceder. Esto es claramente injustificado.■

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

Las principales dificultades que surgen del postulado de Bayes aparecen desde el punto de vista de la teoría frecuencial o frecuentista de la probabilidad, que requeriría que los estados correspondientes a los diversos $B$ se distribuyeran con igual frecuencia en alguna población de la que haya emanado el $B$ real, si debe aplicarse el postulado de Bayes. A algunos estadísticos, aunque no a todos, esto les ha parecido pedir demasiado del universo. Sin embargo, si adoptamos el punto de vista "lógico" de la probabilidad, es razonable considerar que las probabilidades a priori son iguales cuando no se sabe nada en contrario. Asimismo, para los seguidores de la escuela subjetiva, todo lo que se requiere es que no se debe privilegiar ninguna hipótesis sobre cualquier otra al contemplar una serie de apuestas. Así, la mayoría de los que ven la probabilidad como un grado de creencia aceptan el postulado de Bayes, al igual que muchos frecuentistas lo rechazan explícitamente.

J. L. Savage (1954, 1961) defiende el uso puramente subjetivo de las distribuciones a priori de Bayes. Para una exposición y discusión penetrante de esos puntos de vista, ver Savage (1962).

Las distribuciones a priori que reflejan la ausencia de información a priori se conocen como distribuciones a priori *vagas* o *no informativas*. Una dificultad que surge en el caso continuo es que tales a priori pueden ser impropias. Cuando la información a priori está disponible, podemos utilizar a priori informativas.

Todavía hay tanto desacuerdo sobre este tema que uno no puede presentar ningún conjunto de puntos de vista como ortodoxo. Sin embargo, una cosa está clara: cualquiera que rechace el postulado de Bayes debe poner algo en su lugar. El problema que Bayes intentó resolver es sumamente importante en la inferencia científica y apenas parece posible tener ningún pensamiento científico sin alguna solución, por muy intuitiva y empírica que sea. Nos vemos constantemente obligados a evaluar el grado de credibilidad que se concede a las hipótesis sobre dados datos; la lucha por la existencia, en frase de Thiele (1903), nos obliga a consultar los oráculos. Pero, añadió, los oráculos no nos eximen del pensamiento y de la responsabilidad.

# 4. Máxima verosimilitud

Se han propuesto varios sustitutos del postulado de Bayes. Algunos de ellos se plantean como soluciones a problemas específicos; tales son los principios de Mínimos Cuadrados y Chicuadrado Mínimo. Hay un principio, sin embargo, de aplicación general, el de Máxima Verosimilitud.

Volviendo a (3) podemos escribir el teorema de Bayes en la forma

$$P\left(B_r \left| A, H\right.\right) \propto P\left(B_r \left| H\right.\right) L\left(A \left| B_r, H\right.\right), \qquad (6)$$

donde ahora escribimos $L\left(A \left| B_r, H\right.\right)$ para la verosimilitud. El principio de Máxima Verosimilitud establece que, cuando nos enfrentamos a una elección de hipótesis $B_r$, elegimos aquella (si la hay) que maximiza $L$. En otras palabras, debemos elegir la hipótesis que da la mayor probabilidad al evento observado. Mientras que el teorema de Bayes impone la maximización de la probabilidad conjunta de $B_r$ y $A$, la Máxima Verosimilitud exige la maximización de la probabilidad condicional de $A$ dado $B_r$.

Es de notar particularmente que esto no es lo mismo que elegir la hipótesis con la mayor probabilidad. Algunos defensores del principio de Máxima Verosimilitud niegan explícitamente cualquier significado a expresiones como "la probabilidad de una hipótesis". Veremos más adelante que en la práctica las diferencias entre los resultados obtenidos con Máxima Verosimilitud y el postulado de Bayes no son tan grandes como cabría esperar. Hay, sin embargo, una importante diferencia conceptual involucrada.

De hecho, hay un cambio de énfasis en la forma en que consideramos la función de verosimilitud, reflejada en que la escribimos con una $L$ en lugar de una $P$. La función de probabilidad ordinaria da la probabilidad de $A$ sobre los datos $B_r$ y $H$; $A$ varía, $B_r$ y $H$ están dados. Desde el punto de vista de la verosimilitud, consideramos varios valores de $B_r$ para $A$ observado y $H$ dado; $B_r$ varía, $A$ y $H$ están dados. Es esta variación de la función para diferentes valores de los $B$ lo que tenemos en mente al hablar de verosimilitud.

Supongamos (como suele ser el caso en el trabajo estadístico) que las hipótesis que nos ocupan afirman algo sobre el valor numérico de un parámetro $\theta$. Por ejemplo, las hipótesis podrían ser $B_1 \equiv \theta < 0$, $B_2 \equiv \theta \geq 0$, en cuyo caso hay dos alternativas. O podríamos tener $B_1 \equiv \theta = 1$, $B_2 \equiv \theta = 2$, etc., en cuyo caso hay una infinidad denumerable de hipótesis.

Si ahora $\theta$ puede tener solo valores discretos, podemos, frente a un evento observado $A$, requerir estimar $\theta$, o preguntar cuál es el "mejor" valor de $\theta$ para tomar, dada la evidencia $A$. El método de Bayes sería que en (3) deberíamos buscar que $B_r$ haga de $P\left(B_r \left| A, H\right.\right)$ un máximo. Si no sabemos nada de las probabilidades a priori $P\left(B_r \left| H\right.\right)$, deberíamos, de acuerdo con el postulado de Bayes, suponer que todas esas probabilidades son iguales. Entonces simplemente tenemos que encontrar el $B_r$ que maximiza $L\left(A \left| B_r, H\right.\right)$. En otras palabras, el postulado de Bayes y el principio de Máxima Verosimilitud dan como resultado la misma respuesta numérica.

## 4.1. Ejemplo

Consideremos tomar una muestra con reemplazo de una urna que se sabe que contiene $N$ bolitas, un número desconocido $R$ de las cuales son rojas. Supondremos que se sabe que $R$ es uno de los números enteros $R_1$, ..., $R_k$; este conjunto puede corresponder a todos los números enteros $0 \leq R_j \leq N$.

Si hacemos $n$ selecciones, la probabilidad de $r$ bolitas rojas viene dada por la fórmula binomial

$$p\left(r \left| n, R, N\right.\right) = \binom{n}{r} \pi^r (1-\pi)^{n-r}, \quad r = 0, 1, \ldots, n,$$

donde $R = \pi N$. Una vez realizado el experimento, la verosimilitud de $R$, dado $r = t$ digamos, es

$$L\left(R \left| n, t, N\right.\right) = \binom{n}{t} \frac{R^t (N-R)^{n-t}}{N^n}, \qquad (7)$$

para $R = R_1, \ldots, R_k$. La probabilidad puede evaluarse para cada $R_j$ a su vez, y el valor de $R_j$ que da la $L$ más grande es la estimación por Máxima Verosimilitud. Claramente, multiplicar (7) por una probabilidad a priori constante no afecta el resultado.■

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

Esta proposición no se cumple necesariamente si los valores permisibles de $\theta$ son continuos. Ahora debemos reemplazar expresiones como

$$P\left(B_r\,|H\right)$$

por una función de densidad a priori $f(\theta\,|H)$ y en lugar de (6) tenemos la función de densidad a posteriori

$$g\left(\theta\,|A,H\right) \propto f\left(\theta\,|H\right) L\left(A\,|\theta,H\right). \tag{8}$$

Si ahora requerimos el "mejor" valor de $\theta$, deberíamos, de acuerdo con el postulado de Bayes, tomar la densidad a priori como una constante y, una vez más, deberíamos maximizar $L$ para las variaciones de $\theta$.

Sin embargo, podríamos haber optado por representar nuestras hipótesis, no por $\theta$, sino por alguna cantidad $\phi$ que sea una función de $\theta$, por ejemplo la desviación estándar en lugar de la varianza. En este caso deberíamos haber llegado a la ecuación (8) con $\phi$ escrito en todas partes en lugar de $\theta$; deberíamos haber tomado la probabilidad a priori como constante; y deberíamos haber llegado a la conclusión de que debemos maximizar $L$ para variaciones de $\phi$.

Pero, ¿estamos siendo coherentes al hacerlo? Por el habitual argumento de cambio variable, la densidad a priori para $\phi$ es

$$f_\phi\left(\phi\,|H\right) = f\left(\theta\,|H\right)\frac{d\theta}{d\phi},$$

de modo que si $f_\theta$ es constante, $f_\phi$ no puede serlo siempre que $\phi$ sea una función no lineal de $\theta$. Así, el uso del postulado de Bayes parece implicar autocontradicciones. Sin embargo, el principio de Máxima Verosimilitud está libre de esta dificultad, porque si $L(\theta)$ se maximiza en $\widehat{\theta}$, y $\phi(\theta)$ es una función de $\theta$, $L(\phi)$ se maximiza en $\widehat{\phi} = \phi(\widehat{\theta})$. Por lo tanto, no importa cuál sea el resultado de la parametrización utilizada.

Esta es una de las razones por las que los seguidores de la escuela frecuencial han rechazado el postulado de Bayes en favor del principio de Máxima Verosimilitud; pero en nuestra opinión, el asunto ha sido malinterpretado. Parece que el postulado de Bayes y el principio dan la misma respuesta tanto en el caso continuo como en el caso discreto. cuando se tiene debidamente en cuenta los procesos límites implicados. Vimos que al hablar de probabilidad en un continuo era esencial especificar la naturaleza del proceso hasta el límite. Si consideramos que $\theta$ (desde el punto de vista frecuencial) ha emanado de una población especificada por una densidad rectangular para $\theta$, entonces el postulado de Bayes aplicado a este proceso claramente dará una respuesta diferente de la que se obtiene al suponer que $\theta$ emana de una población cuya función de densidad es rectangular para $\phi$. Así, la inconsistencia aparente no es una inconsistencia en absoluto, sino una dificultad introducida al ignorar el proceso límite en poblaciones continuas.

Sigue siendo cierto, por supuesto, que para muchos propósitos prácticos no sabemos cómo surgió el valor real de $\theta$. Si requerimos una teoría de la inferencia que no se vea afectada por nuestra ignorancia sobre tales puntos, la objeción al postulado de Bayes permanece y no se aplica al principio de Máxima Verosimilitud. Por otro lado, todavía no hemos aducido razones convincentes por las que deberíamos adoptar el principio de Máxima Verosimilitud como principio de inferencia estadística, pero ya conocemos sus excelentes propiedades.

Ahora ilustramos la discusión anterior comparando los resultados obtenidos de los argumentos Bayesiano y de Máxima Verosimilitud para problemas relacionados con la distribución normal.

## 4.2. Ejemplo

Considere una muestra independiente de tamaño $n$ de la distribución normal

$$dF = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx.$$

Si las observaciones son $x_1, x_2, \ldots, x_n$, la función de verosimilitud puede escribirse

$$L = \frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma}\right)^2\right\}.$$

Consideramos un rango de posibles valores de $\mu$ que podrían haber generado estas observaciones. Para estimar $\mu$, tomamos el valor que maximiza $L$. Dado que $L$ es una función regular de $\mu$, y $L \to 0$ cuando $\mu \to \pm\infty$, requerimos que $\mu$ satisfaga

$$\frac{\partial L}{\partial \mu} = 0, \qquad \frac{\partial^2 L}{\partial \mu^2} < 0.$$

Dado que $L$ es positivo, obtenemos el mismo resultado al maximizar $\log L$, a veces (como aquí) un procedimiento más conveniente. Entonces tenemos

$$\frac{\partial \log L}{\partial \mu} = +\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma^2}\right) = 0, \tag{9}$$

y así el estimador de $\mu$, digamos $\widehat{\mu}$, viene dado por

$$\sum_{j=1}^{n} x_j = n\widehat{\mu}$$

o

$$\widehat{\mu} = \overline{x}, \tag{10}$$

la media de las $x$. Ya que

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0,$$

este es un máximo único y, por lo tanto, es la solución de Máxima Verosimilitud.

Si quisiéramos estimar tanto $\mu$ como $\sigma$, deberíamos encontrar, además de (9),

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{j=1}^{n}\frac{(x_j-\mu)^2}{\sigma^3} = 0, \tag{11}$$

dando

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2. \tag{12}$$

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

Mientras que $\widehat{\mu}$ no depende de $\sigma$, $\widehat{\sigma}$ sí depende de $\mu$. Elegimos aquellos estimadores que maximizan la verosimilitud para variaciones simultáneas en $\mu$ y $\sigma$, es decir, resolvemos (9) y (11) simultáneamente. esto nos da

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2, \tag{13}$$

y (10) y (13) conjuntamente maximizan la verosimilitud.■

### 4.3. El principio de verosimilitud

Frecuentemente consideramos la recomendación de Fisher de que se utilice la función de verosimilitud (FV) como resumen de información. Sin embargo, es posible llevar más lejos esta línea de razonamiento y argumentar que todo procedimiento inferencial debe basarse únicamente en la FV. Este punto de vista puede expresarse formalmente como el *principio de verosimilitud* (PV), que también se presenta en formas débiles y fuertes. El principio débil (PVD) establece que toda la información sobre $\theta$ obtenida del experimento estadístico, $E$, está contenida en la FV, $L\,(x\,|\theta\,)$. Si dos repeticiones, que arrojan observaciones $x_1$ y $x_2$, conducen a probabilidades proporcionales:

$$L\,(x_1\,|\theta\,) = c(x_1, x_2)L\,(x_2\,|\theta\,),$$

donde la función $c$ es independiente de $\theta$, $x_1$ y $x_2$ proporcionan la misma información sobre $\theta$, o

$$Ev(E, x_1) = Ev(E, x_2), \tag{14}$$

donde la igualdad anterior significa que la evidencia obtenida de $x_1$ es exactamente igual a la obtenida de $x_2$. La forma fuerte (PVF) amplía el principio para incluir dos experimentos diferentes, $E_1$ y $E_2$, de modo que

$$Ev(E_1, x_1) = Ev(E_2, x_2).$$

Edwards (1974) rastrea la historia del PV.

## 5. A priori no informativas

La ignorancia a priori sobre $\mu$ puede ser expresada por la distribución rectangular a priori no informativa

$$f(\mu)d\mu \propto d\mu, \qquad -\infty < \mu < \infty. \tag{15}$$

Al combinar esto con la función de verosimilitud del Ejemplo de §4.2, vemos que la probabilidad a posteriori se maximiza en $\widehat{\mu} = \overline{x}$, como antes. Sin embargo, debe notarse que (15) es una a priori impropia en el sentido de que $\int f(\mu)d\mu$ no existe.

Las a priori impropias pueden conducir a paradojas en problemas multiparamétricos, y recientemente el énfasis se ha desplazado a favor del uso de la idea de *intercambiabilidad* de De Finetti para representar la ignorancia a priori.

Cuando se desconoce $\sigma$, Jeffreys (1961) recomendó el uso de una a priori rectangular en la línea real para $\log\sigma$, o

$$f(\sigma)d\sigma \propto \frac{d\sigma}{\sigma}, \qquad 0 < \sigma < \infty. \tag{16}$$

Multiplicando (16) por la verosimilitud, obtenemos la densidad a posteriori (8), y vemos de inmediato que el valor de maximización viene dado por (11) con $(n+1)$ en lugar de $n$, de modo que

$$\widehat{\sigma}^2 = \frac{1}{n+1}\sum_{j=1}^{n}(x_j - \mu)^2. \tag{17}$$

Finalmente, si tanto $\mu$ como $\sigma$ son desconocidos, combinamos las a priori (15) y (16) y llegamos a $\widehat{\mu} = \overline{x}$ y (17) con $\mu$ reemplazado por $\widehat{\mu}$.

## 6. A priori informativas

Cuando se dispone de información a priori y se puede incorporar a la función de probabilidad a priori, la probabilidad a posteriori puede determinarse a partir de (8). El estimador del parámetro desconocido $\theta$ aún se obtendrá maximizando la probabilidad a posteriori pero, en general, diferirá del estimador por Máxima Verosimilitud.

### 6.1. Ejemplo

Suponga que la información a priori sobre la media normal $\mu$ puede representarse por

$$f\,(\mu\,|\lambda, \omega\,) \propto \exp\left\{-\frac{1}{2}\frac{(\mu-\lambda)^2}{\omega^2}\right\}, \qquad -\infty < \mu < \infty. \tag{18}$$

Como en el Ejemplo de §4.2, suponga que tenemos una muestra independiente de tamaño $n$ de la distribución normal. Entonces, de (8), la densidad a posteriori es

$$f\,(\mu\,|\lambda, \omega, \sigma, x\,) \propto \exp\left\{-\frac{1}{2}\left[\frac{\mu-\lambda}{\omega}\right]^2 - \frac{1}{2}\sum_{j=1}^{n}\left[\frac{x_j-\mu}{\sigma}\right]^2\right\}. \tag{19}$$

Derivando con respecto a $\mu$, obtenemos

$$\frac{\partial f}{\partial \mu} = \sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma^2}\right) + \frac{(\lambda-\mu)}{\omega^2},$$

que da un máximo en

$$\widehat{\mu} = \frac{\overline{x}n\omega^2 + \lambda\sigma^2}{n\omega^2 + \sigma^2}. \tag{20}$$

Cuando $n \to \infty$, $\widehat{\mu} \to \overline{x}$ independientemente de la información a priori contenida en $(\lambda, \omega^2)$. Esto refuerza el punto mencionado anteriormente, que la información de una muestra suficientemente sólida eventualmente abrumará las opiniones a priori. Además, notamos que $\widehat{\mu} \to \overline{x}$ como $\omega^2 \to \infty$. Hacer que $\omega^2 \to \infty$ es una forma de expresar la ignorancia a priori, ya que (18) muestra que $\omega$ representa la dispersión de $\mu$ alrededor de $x$. De hecho, una forma de superar el problema de las a priori impropias es seleccionar una a priori informativa y luego evocar un argumento límite apropiado. Como en ocasiones anteriores, la elección adecuada del argumento límite es fundamental.■

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

Cuadro 1: Formas comunes de distribuciones a priori conjugadas

| Verosimilitud | Parámetro | A priori/A posteriori |
|---|---|---|
| Normal | $\mu$ | Normal |
| Normal | $\sigma^2$ | Gamma (para $\sigma^{-2}$) |
| Binomial | $\pi$ | Beta |
| Poisson | $\lambda$ | Gamma |

La elección de una densidad a priori normal en el Ejemplo de §6.1 es ciertamente conveniente, pero ¿es apropiada? Recordando que tales a priori expresan grados de creencia, la respuesta final para el subjetivista debe ser individual, aunque la teoría lógica puede esperar una respuesta más definitiva.

En general, las a priori arbitrarias hacen que las matemáticas sean intratables. Dado que el conocimiento de la forma funcional de la a priori es a menudo vago, esto ha llevado al desarrollo de una clase de distribuciones a priori *conjugadas*, para las cuales la a priori y la a posteriori tienen la misma forma funcional. Algunas de las formas comunes se resumen en el Cuadro 1.

La introducción de a priori conjugadas abre un camino por el cual la información a priori puede introducirse en un análisis frecuentista. Dado que la verosimilitud y la a priori son compatibles en forma, el frecuentista puede especificar una probabilidad a priori que se considera "equivalente" a $n_0$ observaciones. Cuando $n_0 = 0$, la verosimilitud a priori sería plana pero aún adecuada. Cuando $n_0 > 0$, el estimador por Máxima Verosimilitud se modifica de la misma manera que el estimador de probabilidad a posteriori.

### 6.2. Ejemplo

Sea la probabilidad a priori para la media normal

$$L_p(\mu\,|\lambda, n_0\,) \propto \exp\left\{-\frac{n_0(\mu-\lambda)^2}{2\sigma^2}\right\}, \qquad -\infty < \mu < \infty,$$

por lo que en (18) ponemos $\omega^2 = \sigma^2/n_0$. Siguiendo el funcionamiento del Ejemplo de §6.1, (20) es ahora

$$\widehat{\mu} = \frac{n\overline{x} + n_0\lambda}{n + n_0}. \tag{21}$$

Comparando (20) y (21), vemos la diferencia de énfasis en que la verosimilitud a priori requiere la especificación de $\lambda$ y $\omega^2$. Una vez más, las ideas utilizadas conducen a formulaciones diferentes, aunque los resultados finales pueden ser muy similares.■

## 7. Inferencia Estadística Comparada

La inferencia estadística es un proceso inductivo que va de la muestra a la población. Al pensar en una hipótesis $(H)$ y datos observacionales, o evidencia $(E)$, no hay problema en hacer enunciados probabilísticos de la forma $P\,(E\,|H\,)$; de hecho, estos están justificados por la lógica deductiva una vez que se especifican los axiomas de probabilidad, y tales afirmaciones se han utilizado repetidamente por muchos autores, incluso nosotros. Sin embargo, se ha cuestionado la existencia misma de enunciados inductivos de la forma $P\,(H\,|E\,)$ y muchos filósofos, en particular Sir Karl Popper (1968, 1969), han concluido que tales probabilidades no existen. Tales probabilidades son, por supuesto, las probabilidades a posteriori del enfoque Bayesiano, por lo que el debate es de vital interés para los estadísticos. De hecho, no muestra signos de disminuir dicho debate, y el lector interesado debe consultar Popper y Miller (1987), Good (1988), Gemes (1989) y Miller (1990) para conocer más desarrollos.

Cualquier procedimiento inferencial debe basarse en un conjunto de reglas más o menos racional, pero la racionalidad de cualquier sistema dado y el valor aparente de las conclusiones que permite alcanzar permanecen abiertos a debate.

En nuestra vida académica y profesional hemos adoptado el paradigma *frecuentista*, a veces conocido como enfoque *clásico* o *frecuencial*, que ha sido la escuela dominante de pensamiento estadístico durante la mayor parte de los siglos XX y XXI. Sin embargo, el punto de vista *Bayesiano* ha ganado popularidad desde la década de 1950 y en los últimos años se han desarrollado varios otros enfoques de la inferencia, algunos más completos que otros.

En este trabajo, intentamos esbozar tanto las áreas de acuerdo como las diferencias entre las principales escuelas; no es nuestro interés desarrollar cada enfoque en detalle. Barnet (1982), Dawid (1984) y el volumen editado por Godambe y Sprott (1971) ofrecen discusiones generales sobre la inferencia comparativa. En Howson y Urbach (1989) aparece una discusión más filosófica que respalda el enfoque Bayesiano subjetivo.

Dado que nuestra discusión es una evaluación bastante breve de una extensa y compleja literatura, enfatizaremos solo los puntos principales en cuestión. Por lo tanto, examinamos las posiciones "estándares" dentro de cada escuela y no enfatizamos los debates dentro de una escuela (por ejemplo, la elección de axiomas para la probabilidad subjetiva). Esperamos que estos grandes trazos sirvan para producir retratos y no caricaturas.

## 8. Un marco para la inferencia

En términos generales, el proceso inferencial contiene los siguientes ingredientes:

- Una *variable aleatoria medible* (*vectorial*) $X$, que toma valores en el espacio muestral $\mathfrak{X}$.
- El o los *parámetro*(*s*) *desconocido* $\theta$, que se puede dividir en parámetros de interés directo y parámetros no deseados (en inglés "nuisance parameter"), que luego se denotan por $\theta$ y $\phi$, respectivamente. El conjunto de valores posibles de $\theta$ está definido en el espacio paramétrico $\Omega$.
- La *población de interés* que tomamos es representable en términos de una familia de distribuciones de probabilidad $\{F(x,\theta)\}$, indexada por $\theta$. Usamos

  $$F \equiv F(\theta) \equiv F(x,\theta) = P(X \le x\,|\theta\,)$$

  indistintamente cuando no surja ambigüedad. La forma funcional de $F$ puede estar completamente especificada o ser miembro de alguna clase de distribuciones, $\mathfrak{F}$.

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

- Un *experimento estadístico* que produce un conjunto de observaciones, descrito por el vector aleatorio $X = (X_1, X_2, \ldots, X_n)'$, con una realización particular, los *datos de la muestra*, denotados por $x = (x_1, x_2, \ldots, x_n)'$. El procedimiento experimental especifica el modo de muestreo y la forma de la regla de muestreo, $S$, se requiera o no dicha información.

Nuestra notación no distinguirá entre vectores y escalares, a menos que la discusión explísitamente requiera que se haga la distinción.

Además, puede haber información histórica (o previa) con respecto a $\theta$ de carácter personal u objetiva que resumimos en alguna función $p(\theta)$. Dado que la especificación, el uso e incluso la existencia de dicha información es un tema de considerable debate, aplazamos la discusión adicional de este tema. La forma general del problema de inferencia es usar la información disponible

$$I = \{\mathfrak{X}, \Omega, \mathfrak{F}, x, S, p\} \tag{22}$$

para hacer declaraciones inductivas sobre $\theta$. Ahora examinamos los diversos enfoques de este problema, comenzando con una descripción general del enfoque frecuentista o frecuencial que hemos adoptado hasta ahora. Luego dirigimos nuestra atención a la inferencia Bayesiana. Se concluye con una evaluación de los diferentes enfoques y una discusión de los intentos de reconciliación entre estas escuelas de pensamiento.

# 9. El enfoque frecuencial

La teoría frecuencial de probabilidad supone que es posible considerar una sucesión infinita de réplicas independientes del mismo experimento estadístico.

Ahora limitamos la atención principalmente a la estimación puntual. Podemos considerar un estadístico o estimador, $T(X)$, como un resumen de la información sobre $\theta$; por simplicidad, a menudo restringiremos la atención a un solo parámetro. En el estudio de la estimación identificamos ciertas propiedades deseables para $T$, como la *consistencia* y la *falta de sesgo*. Dado que a menudo hay una multiplicidad de estimadores que satisfacen estos requisitos, buscamos medidas de *eficiencia* e identificamos estimadores deseables como IVM, insesgados de varianza mínima. El criterio más amplio de MECM o menor error cuadrático medio a veces se considera más apropiado y aplicable.

Aunque estos criterios pueden considerarse deseables en sí mismos, carecen de un método para construir estadísticos adecuados, $T$. Dentro de la familia exponencial, se puede identificar el conjunto de estadísticos suficientes que conducen al estimador IVM de $\theta$, si existe. De manera más general, establecimos que el estimador por máxima verosimilitud (EMV), obtenido como

$$\widehat{\theta} = \max_{\theta \in \Omega} L(\theta \,|\mathbf{x}\,), \tag{23}$$

donde

$$L(\theta \,|\mathbf{x}\,) = \prod_{i=1}^{n} f(x_i \,|\theta\,), \tag{24}$$

es consistente y asintóticamente insesgado bajo condiciones de regularidad moderada cuando las observaciones son independientes y de la misma distribución. Además, el EMV es una función de los estadísticos suficientes y es asintóticamente IVM.

Incluso en esta etapa, encontramos alguna separación de caminos, que las propiedades para muestras grandes del EMV tienden a oscurecer. Si $T$ es un estimador insesgado para $\theta$, entonces $g(T)$ no es insesgado para $\phi = g(\theta)$, mientras que el EMV es funcionalmente invariante, de modo que

$$\widehat{\theta} = T \Longleftrightarrow \widehat{\phi} = g(T). \tag{25}$$

## 9.1. Ejemplo

Dada una muestra aleatoria de $n$ observaciones, $X$, de una población normal con media $\theta$ y varianza 1, tenemos que

$$\widehat{\theta} = \overline{X} \quad \text{y} \quad \widehat{\phi} = (\overline{X})^2,$$

cuando $\phi = \theta^2$. Sin embargo, el IVM para $\phi$ es

$$T = (\overline{X})^2 - \frac{1}{n}. \tag{26}$$

Aunque $E(T) = \phi$ y $\phi \geq 0$, puede suceder que el valor observado de $T$ sea negativo. El sentido común sugiere reemplazar los valores negativos de $T$ por cero, aunque esto viola la propiedad de imparcialidad. En general, tales ajustes producen estimadores con un error cuadrático medio más pequeño, por lo que diferentes criterios pueden conducir a diferentes estimadores.■

Los estimadores *ad hoc* obtenidos al resolver $T = g(\widehat{\theta})$, donde $E(T) = g(\theta)$, se usan ampliamente y se justifican apelando a la falta de sesgo para $g(\theta)$, aunque estos estimadores están sesgados para $\theta$ a menos que $g(\theta)$ es una función lineal.

# 10. Inferencia Bayesiana

El enfoque Bayesiano del problema de la inducción es suponer que se puede especificar una distribución a priori para el parámetro $\theta$, $p(\theta)$, por ejemplo, definida en el espacio de parámetros $\theta \in \Omega$. Dada la función de verosimilitud, $L(\mathbf{x}\,|\theta\,)$, se deduce de una aplicación del teorema de Bayes que la distribución a posteriori es

$$P(\theta \,|\mathbf{x}\,) \propto p(\theta) L(\mathbf{x}\,|\theta\,). \tag{27}$$

Debe notarse que la versosimilitud, $L$, dada en (24) difiere en la forma de escribir su argumento de aquella dada en (27). En un caso (el primero) esa función está escrita como función de $\theta$ para una muestra $\mathbf{x}$ dada, según lo considerado por Fisher, y en el otro caso (la segunda) se la considera como función de $\mathbf{x}$ para $\theta$ dado. Pero ambas formas son matemáticamente equivalentes.

Una vez que se acepta la noción de especificar una distribución a priori para $\theta$, el marco de la inferencia Bayesiana puede desarrollarse deductivamente a partir de uno de varios sistemas de axiomas (por ejemplo, Ramsey, 1926; Good, 1950; Savage, 1962; De Groot, 1970); para una evaluación detallada, véase Fishburn (1986).

Por lo tanto, la pregunta clave es cómo especificar la distribución a priori. Se pueden considerar tres enfoques posibles:

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

(i) como una distribución de frecuencias, basada en la experiencia pasada;

(ii) como una representación objetiva de creencias iniciales racionales sobre el parámetro;

(iii) como un enunciado subjetivo sobre lo que Usted (una persona específica) cree antes de que se recopilen los datos.

Consideraremos la alternativa (i) sólo brevemente. De acuerdo con el enfoque frecuencial, necesitaríamos tener un proceso subyacente que genera los valores del parámetro que es estable, o al menos predecible. Los ejemplos incluyen esquemas de producción industrial donde una distribución a priori para la proporción de defectuosos, por ejemplo, puede evaluarse a partir de registros anteriores. De manera más general, los modelos de espacio de estado en series de tiempo suponen que los parámetros (estado) se desarrollan en el tiempo de acuerdo con una ecuación de estado como

$$\theta_t = \theta_{t-1} + \delta_t, \tag{28}$$

donde $\delta_t$ representa un disturbio aleatorio en el tiempo $t$. Véase Abril (1999 y 2004), Abril y Abril (2018) para una discusión más detallada.

De alguna manera, esto puede verse como una mezcla de aceite y agua y podría hacerse la reconvención de que la información a priori no está permitida en el esquema frecuentista. Por cierto, tal afirmación la hacen los críticos del enfoque frecuencial, pero parece representar una interpretación demasiado literal de ese punto de vista. De hecho, se debe notar que, aunque lo anterior se especifica en términos frecuentistas, (28) todavía requiere que estemos dispuestos a considerar la distribución a posteriori para $\theta$.

## 10.1. Probabilidad objetiva

La probabilidad objetiva o lógica fue desarrollada, en particular por Jeffreys (1961, versión revisada de su libro de 1939) y otros, para proporcionar una medida sustancial del peso de la evidencia que favorece una hipótesis dada a la luz de los datos. Es decir, se buscó una distribución a priori acordada que permitiera hacer afirmaciones de probabilidad a posteriori sobre la base de un ensayo particular.

Gran parte del trabajo de Jeffreys se centró en las especificaciones de una distribución a priori en situaciones en las que no se sabe nada acerca de los parámetros antes de que se lleve a cabo el experimento estadístico. Curiosamente, la mayoría de los Bayesianos subjetivos, como Lindley (1971), argumentarían ahora que siempre hay *alguna* información disponible y que la especificación de la ignorancia a priori no es un problema. Cuando el número de valores de $\theta$ en $\Omega$ es finito, es factible hacer uso del *postulado* de Bayes (también conocido como *principio de razonamiento insuficiente* o *principio de indiferencia*) y asignar probabilidades a priori iguales a cada valor posible. Esto requiere que se pueda establecer una base satisfactoria de posibles valores de parámetros, lo que no siempre es una tarea trivial.

### 10.1.1. Ejemplos

1. Una urna contiene un número desconocido de bolitas de igual tamaño y peso que están hechas del mismo material. ¿Cuál es la probabilidad a priori de que se seleccione una bolita blanca en la primera extracción cuando se le dice que la urna contiene bolitas que son:

   (a) blanca o no blanca,

   (b) blanca, roja o azul?

   El principio de razonamiento insuficiente nos lleva a concluir que $p = 1/2$ en el caso (a), pero $p = 1/3$ en el caso (b).■

   A pesar de este ejemplo, el principio a menudo puede servir como un punto de partida razonable. Una implicación de ese principio es la *Ley de Sucesión de Laplace* que muestra que si se parte de

   $$\Omega = \{0, 1/N, 2/N, \ldots, (N-1)/N, 1\}, \tag{29}$$

   y asigna probabilidades a priori iguales $1/(N+1)$ a cada estado, con

   $$\begin{aligned} a_{m+1} &= \{\text{la prueba } (m+1)\text{-ésima es un éxito}\}, \\ b_m &= \{\text{las primeras } m \text{ pruebas son éxito}\}, \end{aligned}$$

   entonces

   $$P\{a_{m+1} \,|b_m\} = \frac{m+1}{m+2}, \tag{30}$$

   para cualquier $m$ y $N \geq 1$.■

2. Si se lanza una moneda $m$ veces y sale cara cada vez, ¿aceptaríamos que la probabilidad de que en el próximo lanzamiento salga cara viene dada por (30)?

   La respuesta probablemente sea no, porque nos basamos en mucha experiencia pasada que dice que la moneda tiene cara y cruz y que cualquier lado tiene "igual probabilidad" de caer boca arriba. Sin embargo, esto no viola el principio, sino que nos dice que asignar probabilidades iguales a los valores en (29) no fue una declaración precisa de creencia a priori. Por el contrario, si hay tres monedas: una con dos caras, una estándar y otra con dos cruces, especificar probabilidades iguales en (29) con $N = 2$ sería muy plausible. Tenga en cuenta que no requerimos que se seleccione una moneda al azar, sino que ignoramos el proceso de selección.■

Ahora supongamos que $\Omega$ es continuo; incluso si la a priori para $\theta$ es rectangular en un intervalo finito, eso para cualquier transformada no lineal de $g(\theta)$ no lo será. Esto llevó a Jeffreys a proponer el uso de la a priori

$$p(\theta) \propto \{I(\theta)\}^{1/2}, \tag{31}$$

dónde $I(\theta) = -E(\partial^2 \log L/\partial\theta^2)$. Él llegó a (31) seleccionando la forma de $g(\theta)$ para la cual $p\{g(\theta)\}$ es rectangular, incluso si es impropia en algunos casos; la función de $g(\theta)$ corresponde entonces a un parámetro de posición para la distribución, al

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

menos localmente. Jeffreys llamó a las a prioris dadas por (31) *invariantes*.

Aunque el concepto que Jeffreys estaba tratando de hacer operativo es atractivo, no parece posible desarrollarlo de manera consistente; véanse las críticas en Barnett (1982, Capítulo 6) y Howson y Urbach (1989, Capítulo 9). Es interesante especular si Jeffreys habría adoptado (31) si sus resultados no hubieran coincidido con los existentes.

## 10.2. Probabilidades subjetivas

Dejamos ahora el punto de vista objetivista y aceptamos que las probabilidades a priori son necesariamente personales y se basan en nuestra propia experiencia. Para que un esquema de este tipo sea operativo, es necesario que

(a) Uno tenga creencias sobre los parámetros de interés, que se pueden expresar en forma de probabilidades;

(b) Sus probabilidades pueden compararse entre sí (aunque no es necesario que sean comparables con las de nadie más);

(c) Sus probabilidades pueden evaluarse mediante algún esquema de apuestas hipotéticas.

Si Su[1] comportamiento de apuestas es internamente consistente, se deduce que Sus probabilidades satisfacen las reglas estándar de probabilidad y se dice que Usted es *coherente*; de lo contrario, eres *incoherente* y un Bayesiano podría hacer apuestas Contigo de tal manera que perderías dinero. Este es el *principio de coherencia*, que establece que su sistema de apuestas debe ser internamente consistente. Presumiblemente, se usó la coherencia para evitar confusiones con el uso de la consistencia de Fisher en la estimación y los tests de hipótesis. ¡Claramente, los no Bayesianos no tienen el monopolio de las palabras clave virtuosas!

El requisito clave ahora es la evaluación de la distribución a priori. La mayoría de los subjetivistas (p. ej., Ramsey, 1931; Savage, 1954) utilizan algún método para evaluar apuestas justas, ya sea directamente para el fenómeno en estudio o en comparación con algunos experimentos estandarizados (p. ej., un esquema de urna). Se supone que dichas evaluaciones pueden hacerse directamente para las probabilidades, sin estar contaminadas por utilidades relativas de diferentes resultados.

Una vez que Usted haya establecido Su distribución a priori, el análisis Bayesiano subjetivo procede directamente, aunque a menudo será deseable usar conjugadas a priori como se las definió anteriormente para simplificar el álgebra. Si el conjunto de parámetros es $(\theta, \phi)$, donde $\phi$ denota parámetro(s) no deseados (nuisance), el enfoque estándar es examinar la *distribución marginal a posteriori*

$$\begin{aligned} P(\theta\,|\mathbf{x}) &= \int P(\theta,\phi\,|\mathbf{x})d\phi \\ &= \int L(\mathbf{x}\,|\theta,\phi)p(\theta,\phi)d\phi. \end{aligned} \tag{32}$$

[1]A partir de aquí se invita al lector, es decir a Usted, a involucrarse en este juego y se usa mayúscula en Sus pronombres porque suponemos que es Usted quien realiza la acción

La evaluación explícita de (32) puede resultar muy difícil para problemas de dimensiones altas. Sin embargo, los innovadores procedimientos de integración numérica desarrollados por Naylor y Smith (1988), entre otros, han contribuido en gran medida a la viabilidad de este enfoque.

Para reglas de actualización más generales, ver Diaconis y Zabell (1982).

## 10.3. Estimación Bayesiana

La estimación puntual generalmente se basa en el modo o en la media de la distribución a posteriori. El *modo a posteriori* viene dado por $\widetilde{\theta}$, donde

$$P(\widetilde{\theta}\,|\mathbf{x}) = \max_{\theta} P(\theta\,|\mathbf{x}); \tag{33}$$

cuando la distribución a priori es rectangular, $\widetilde{\theta}$ será equivalente al estimador por MV $(\widehat{\theta})$.

La *media a posteriori*, dada por

$$\overline{\theta} = E(\theta\,|\mathbf{x}), \tag{34}$$

será igual al estimador por MV solo para elecciones específicas de la distribución a priori.

### 10.3.1. Ejemplo

Sea $\pi$ la probabilidad de éxito en un ensayo Bernoulli con función de frecuencia a priori

$$p(\pi) \propto \pi^{a-1}(1-\pi)^{b-1}.$$

Dados $n$ ensayos con $x$ éxitos, la a posteriori es

$$P(\pi\,|x) \propto \pi^{a+x-1}(1-\pi)^{b+n-x-1},$$

de donde obtenemos

$$\widetilde{\theta} = \frac{a+x-1}{n+a+b-2} \qquad \text{y} \qquad \overline{\theta} = \frac{a+x}{n+a+b},$$

comparado con $\widehat{\theta} = x/n$. Tras la inspección, $\widetilde{\theta} = \widehat{\theta}$ para la a posteriori rectangular ($a = b = 1$), mientras que $\overline{\theta} = \widehat{\theta}$ cuando $a = b = 0$, una elección degenerada que no es factible.■

Las estimaciones por intervalo se pueden obtener directamente de la distribución a posteriori; la inferencia básica permite el enunciado "con probabilidad $1-\alpha$, $\theta$ se encuentra entre los valores $\theta_1$ y $\theta_2$" o

$$P(\theta_1 \le \theta \le \theta_2) = P(t_2\,|\mathbf{x}) - P(t_1\,|\mathbf{x}) = 1-\alpha. \tag{35}$$

El intervalo $[\theta_1, \theta_2]$ se conoce como una *región creíble* del $100(1-\alpha)$ por ciento. Paralelamente a la noción de un intervalo físicamente más corto, podemos elegir el conjunto $\Omega_1$ de valores $\theta$, tal que (35) se satisface y

$$\left\{\theta \in \Omega_1 : \frac{\partial P(\theta)}{\partial \theta} \ge c\right\}. \tag{36}$$

Tal intervalo (o región) se conoce como la región creíble de *mayor densidad a posteriori* (MDP).

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

### 10.3.2. Ejemplo

Para una muestra aleatoria de tamaño $n$ de una población normal con varianza conocida, digamos $N(\mu, \sigma^2)$, considere la distribución a priori $N(\phi, \tau^2)$. De §6, la distribución a posteriori para $\mu$ es $N(\mu_p, \sigma_p^2)$, donde

$$\mu_p = \frac{\phi\sigma^2 + n\overline{x}\tau^2}{\sigma^2 + n\tau^2} \qquad \text{y} \qquad \sigma_p^2 = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

La región creíble de MDP para $\mu$ es

$$\mu_p \pm z_{1-\alpha/2}\sigma_p,$$

donde $z$ representa los puntos porcentuales de $N(0, 1)$. En este ejemplo, $\widetilde{\theta} = \overline{\theta}$ y estos serán iguales a $\widehat{\theta}$ para la a priori rectangular impropia dado al hacer $\tau \to \infty$; los intervalos creíble y de confianza serán idénticos (¡numéricamente hablando!).■

## 10.4. Tests Bayesianos

Las dos hipótesis unilaterales

$$H_0 : \theta \leq \theta_0 \qquad \text{y} \qquad H_1 : \theta > \theta_0$$

se comparan fácilmente calculando sus probabilidades a posteriori

$$P(H_0) = P(\theta_0 \,|\mathbf{x}), \qquad P(H_1) = 1 - P(\theta_0 \,|\mathbf{x}). \tag{37}$$

Sin embargo, la comparación de

$$H_0 : \theta = \theta_0 \qquad \text{y} \qquad H_1 : \theta \neq \theta_0$$

plantea algunas dificultades. Jeffreys (1961, Capítulo 5) argumenta que el valor de $\theta_0$ se distingue de todos los demás valores de $\theta$ y, por lo tanto, se puede asignar una probabilidad a priori al punto:

$$p_0 = p(\theta_0) > 0.$$

Las probabilidades a posteriori a favor de $H_0$ son entonces

$$\frac{P(\theta_0 \,|\mathbf{x})}{\int\limits_{\Omega-\theta_0} dP(\theta_0 \,|\mathbf{x})}.$$

Tal suposición es claramente plausible en algunos casos, como probar si un coeficiente de regresión es cero, pero depende en gran medida del valor de $p_0$ seleccionado. El punto de vista frecuentista sería que la hipótesis nula a menudo merece una atención especial, pero que no hay una forma razonable de llegar a un valor apropiado de $p_0$.

Bernardo (1980) examinó la estructura de los tests Bayesianos y concluyó que no hay problemas cuando $H_0$ y $H_1$ tienen la misma dimensionalidad. En otros casos, parece que las conclusiones que se extraen de tales tests son claramente interpretables solo cuando $p(\theta_0)$ depende de la a priori general $p(\theta)$, $\theta \neq \theta_0$.

En general, los tests de hipótesis ahora reciben menos atención por parte de los Bayesianos, quienes tienden a favorecer el uso de la teoría de la decisión.

## 10.5. La relación entre los enfoques Bayesianos y fiduciales

Como se sabe, si $t$ es el estadístico suficiente (mínimo) para el el único parámetro $\theta$, con función de distribución $F(t \,|\theta)$, la distribución fiducial de $\theta$ dado $t$ tiene función de densidad (de probabilidad)

$$g(\theta \,|t) = \frac{\partial G(\theta \,|t)}{\partial\theta} = -\frac{\partial F(t \,|\theta)}{\partial\theta}, \tag{38}$$

siempre que $F$ sea monótona decreciente en $\theta$. Algunas de las dificultades de este enfoque son cómo proceder en ausencia de un estadístico suficiente, la falta de unicidad (en casos multiparamétricos) y la falta de una interpretación frecuencial.

Los trabajos de Fisher sobre inferencia fiducial fueron evidentemente influenciados por Keynes (1921), Carnap (1962), y otros, que buscaron desarrollar una visión epistémica de la probabilidad que mediría el "grado de credibilidad racional" de una hipótesis $H$ en relación con los datos o evidencia $E$. Por lo tanto, aunque el desarrollo inicial de la probabilidad fiducial fue confuso, el objetivo era claro: hacer enunciados de probabilidad de la forma $P(H \,|E\,)$ o, en nuestro contexto actual, desarrollar una función de distribución $G(\theta \,|t\,)$. Por construcción e intención, $G$ está diseñada para proporcionar información sobre $\theta$ para *un solo ensayo*, por lo que la ausencia de una interpretación frecuencial no es sorprendente. Está claro que el enfoque fiducial busca establecer un enunciado inductivo completamente diferente al que está disponible desde el punto de vista frecuencial.

Lindley (1958) obtuvo un resultado simple pero de gran alcance que no solo ilumina la relación entre los argumentos fiduciales y Bayesianos, sino que también limita las afirmaciones de la teoría fiducial para proporcionar un método general de inferencia, consistente y combinable con los métodos Bayesianos. De hecho, Lindley muestra que el argumento fiducial es consistente con los métodos Bayesianos si y solo si se aplica a una variable aleatoria $x$ y un parámetro $\theta$ que pueden transformarse (por separado) en $u$ y $\tau$ respectivamente, de modo que $\tau$ es un parámetro de locación de $u$; y en este caso, es equivalente a un argumento Bayesiano con una distribución a priori rectangular para $\tau$. Esta crítica se aplica igualmente a las "distribuciones de confianza" definidas en la teoría general de la estimación por intervalos, en la medida en que coincidan con distribuciones fiduciales.

## 10.6. Métodos empíricos de Bayes

Una variación interesante del enfoque Bayesiano es el esquema empírico de Bayes desarrollado por Robbins (1956, 1964); ver Maritz y Lwin (1989) para una exposición detallada. Suponga que se dispone de una muestra de $n$ observaciones con función de frecuencia $f(\mathbf{x} \,|\theta_i\,)$, donde $\theta_i$ representa una extracción aleatoria de una distribución a priori $p(\theta \,|\phi\,)$ y $\phi$ representa los parámetros de la distribución a priori. Entonces podemos considerar la distribución marginal

$$f(\mathbf{x} \,|\phi\,) = \int f(\mathbf{x} \,|\theta\,)p(\theta \,|\phi\,)d\theta \tag{39}$$

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

y utilizar métodos de MV para estimar $\phi$. La distribución a posteriori de $\theta_i$ se aproxima por

$$P(\theta_i \,|x_i\,) \propto f(x_i \,|\theta_i\,)p(\theta_i \,\big|\widehat{\phi}\,). \tag{40}$$

En casos particulares (por ejemplo, con conjugadas a priori), puede ser posible la determinación explícita de (39), de lo contrario, se deben usar procedimientos numéricos.

Este enfoque es algo así como una amalgama de ideas Bayesianas y frecuentistas y tuvo una recepción mixta. Por ejemplo, Neyman (1962) lo aclamó como un gran avance, mientras que Lindley (1971) considera que no involucra ningún nuevo punto de principio.

### 10.7. Teoría de la decisión

El trabajo de Abraham Wald sobre el análisis secuencial condujo también al desarrollo de una teoría general de la toma de decisiones. Considere una situación donde, dados los datos, es necesario tomar una decisión; además, suponga que se conocen las consecuencias de estas decisiones y que pueden evaluarse numéricamente. Estas no son suposiciones triviales; por ejemplo, en su desarrollo de tests de hipótesis, Nayman y Pearson concluyen que es *poco probable* que tal información esté disponible. Dados los antecedentes necesarios, el problema es decidir sobre reglas de decisión óptimas con referencia a alguna medida de desempeño. Ahora procedemos a esbozar los fundamentos de dicha teoría; para exposiciones más detalladas, véase Wald (1950), Blackwell y Girshick (1954), Ferguson (1967) y De Groot (1970), entre otros.

Supongamos que podemos especificar un conjunto de acciones posibles $A = \{a\}$ y una *regla de decisión* $d(x)$ que especifica la acción a realizar cuando se observa $x$. La consecuencia de tomar esa acción es incurrir en una pérdida $L[d(x), \theta]$ cuando el valor del parámetro es $\theta$. Algunos autores utilizan una función de utilidad en lugar de una función de pérdida; para la mayoría de los propósitos, la pérdida se puede considerar como una utilidad negativa, aunque se puede considerar que la utilidad está acotada, mientras que a menudo se permite que las funciones de pérdida no sean acotadas.

La pérdida esperada se conoce como la *función de riesgo*:

$$R(d, \theta) = \int L[d(\mathbf{x}), \theta] f(\mathbf{x}\,|\theta\,) dx. \tag{41}$$

Una regla de decisión, $d$, es *admisible* si no hay una regla $d'$ tal que

$$R(d', \theta) \leq R(d, \theta) \quad \text{para todo } \theta \tag{42}$$

con desigualdad estricta para al menos algunos $\theta$. En general, no se pierde nada restringiendo la atención a la clase de reglas de decisión admisibles, aunque esta clase puede ser grande.

Para seleccionar una regla de decisión particular, podemos usar un criterio como *minimax*; es decir, elegimos la regla $d(x)$ que minimiza el riesgo asumido sobre todo $\theta$:

$$\min_{d} \max_{\theta} R(d, \theta). \tag{43}$$

## 11. Discusión

Ha habido tanta controversia acerca de los diversos métodos de estimación que hemos descrito que a partir de aquí dejaremos nuestro enfoque objetivo habitual. El resto de este trabajo es una expresión de puntos de vista personales. Pensamos que esta es la posición correcta; y representa el resultado de muchos años de reflexión sobre los temas en cuestión, un serio intento de comprender lo que dicen los protagonistas y de adivinar lo que quieren decir.

Tenemos, entonces, que examinar seis enfoques principales, aunque algunos están más estrechamente relacionados que otros: frecuencia, verosimilitud, fiducial, Bayesiano objetivo, Bayesiano subjetivo y teoría de la decisión. No debemos dejarnos engañar por la similitud de los resultados a los que conducen en ciertos casos simples, aunque podemos obtener algún consuelo de ello. Sin embargo, desarrollaremos la tesis de que, cuando difieren, la razón básica no es que uno o más estén equivocados, sino que, consciente o inconscientemente, responden a diferentes preguntas o se basen en diferentes postulados.

Al establecer las diferencias, es útil adoptar el concepto de *programas de investigación en competencia* de Lakatos (1974) y establecer el *núcleo duro* de los supuestos que subyacen a cada teoría. Cada teoría está respaldada por una *capa protectora* de supuestos auxiliares, por lo que las conclusiones que se pueden extraer se derivan de manera deductiva de estos fundamentos. No nos interesa debatir extensamente qué supuestos son principales y cuáles auxiliares, sino más bien utilizar esto como marco para nuestras discusiones.

El núcleo duro que subyace a la teoría frecuencial se puede resumir de la siguiente manera:

(a) los axiomas de Kolmogorov;

(b) procedimientos de muestreo aleatorio bien definido, que incluyan la especificación del espacio muestral y la regla de parada;

(c) la interpretación frecuencial de la probabilidad;

(d) una versión del *principio de muestreo repetido* (Cox y Hinkley, 1974, p. 45) que establece que los procedimientos estadísticos deben evaluarse por su comportamiento en repeticiones hipotéticas bajo las mismas condiciones. Esta es la versión *fuerte* de Cox y Hinkley; la versión *débil* requiere únicamente que no sigamos procedimientos que induzcan a error para alguna combinación de parámetros (la mayoría de las veces, en repeticiones hipotéticas). Este principio es esencialmente el mismo que el *principio de confianza* de Birbaum (1977). Como se señaló anteriormente, es el conflicto entre este principio y el principio de verosimilitud lo que está en la raíz del debate entre frecuentistas y Bayesianos.

El cinturón protector incluye conceptos tales como consistencia, insesgamiento, eficiencia, suficiencia, poder, etc. En su análisis de máxima verosimilitud y la teoría de la decisión, Efron (1982, p. 343) se refiere a estos conceptos como "evasivas ingeniosas" utilizadas por Fisher para evitar un enfoque basado en la teoría de la decisión. Sin embargo, cabe señalar que Fisher, al

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

igual que Neyman y Pearson, se esforzó por evitar fuertes suposiciones sobre la existencia y la forma de las funciones de pérdida y las distribuciones a priori; las nociones son ciertamente ingeniosas pero forman parte de un paradigma alternativo, no de una evasión.

El enfoque frecuencial conduce entonces a estimaciones puntuales y por intervalos y tests de hipótesis que son claves para una interpretación del desempeño *a largo plazo*. Los otros enfoques que hemos descrito constituyen varios intentos de desarrollar una noción adicional, o alternativa, de probabilidad que permita al investigador hacer enunciados inferenciales condicionados a los datos registrados en un experimento estadístico *particular*.

El cinturón protector de cualquier teoría evoluciona con el tiempo, por ejemplo, cuando el enfoque de Neyman-Pearson para testar hipótesis suplantó los tests puros de significancción de Fisher; tengase en cuenta que no afirmamos que tales cambios sean instantáneos o estén libres de controversia. Otra posible modificación sería el uso de ecuaciones de estimación insesgadas como las propuestas por Godambe (1960, 1976) en lugar de insesgamiento.

Dos de las dificultades que enfrenta el enfoque frecuencial en la práctica son la especificación del espacio muestral y la necesidad de garantizar un muestreo aleatorio. Johnstone (1989) argumenta que no es necesario que sepamos que la muestra se extrajo al azar; "todo lo que es necesario lógicamente es que *no tengamos conocimiento de lo contrario*". Siguiendo a Fisher, Johnstone llama a esto un *postulado de ignorancia* que es distinto del postulado de Bayes en que se aplica al espacio muestral en lugar del espacio de parámetros.

Las ideas de Johnstone están claramente abiertas al abuso pero, usadas con cuidado, tienen un mérito considerable. Por ejemplo, la distribución del término de error en una ecuación de regresión aplicada a algún agregado macroeconómico tiene una interpretación mucho más plausible cuando se usa el postulado de Johnstone.

El enfoque frecuencial es bastante general en el sentido de que puede aplicarse a cualquier situación de muestreo una vez que el proceso de muestreo esté completamente especificado. Sin embargo, puede haber dificultades en la ejecución. Por ejemplo, cuando no existe un único estadístico suficiente, los intervalos de confianza pueden no ser reales o ser nulos. Así, la suficiencia es deseable, aunque no se requiera. Quizás, sería mejor decir que pueden existir problemas de interpretación cuando no se pueden obtener intervalos anidados y conexos simples.

El principal argumento a favor de la teoría frecuencial de la probabilidad es que no presupone ninguna distribución a priori, como las que son esenciales para el enfoque Bayesiano. Esto, en nuestra opinión, es innegable. Pero es justo preguntarse si logra esta economía de supuestos básicos sin perder algo que posee la teoría Bayesiana. Nuestra opinión es que en ocasiones pierde algo, y que ese algo puede ser importante a los efectos de la estimación.

## 11.1. Información a priori

Considere el caso en el que estamos estimando la media $\mu$ de una población normal con varianza conocida (sin pérdida de generalidad igual a 1), y suponga que sabemos que $\mu$ se encuentra entre 0 y 1. De acuerdo con el postulado de Bayes, deberíamos tener

$$P(\mu\,|\overline{x}\,) = \frac{\exp\left\{-\frac{n}{2}(\mu-\overline{x})^2\right\}}{\int_0^1 \exp\left\{-\frac{n}{2}(\mu-\overline{x})^2\right\}\,d\mu}, \tag{44}$$

y el problema de poner límites a $\mu$, aunque no exento de complejidad matemática, es determinante. ¿Qué tiene que decir la teoría de los intervalos de confianza sobre este punto? No puede hacer más que reiterar enunciados como

$$P\left\{\overline{x} - \frac{1,96}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{1,96}{\sqrt{n}}\right\} = 0,95.$$

Estos siguen siendo ciertos en la proporción requerida de casos, pero el enunciado no tiene en cuenta nuestro conocimiento a priori sobre el rango de $\mu$ y ocasionalmente puede ser inútil. Puede ser cierto, pero es absurdo afirmar $1 \leq \mu \leq 2$ si ya sabemos que $0 \leq \mu \leq 1$. Por supuesto, podemos truncar nuestro intervalo de acuerdo con la información a priori. En nuestro ejemplo, solo podríamos afirmar que $0 \leq \mu \leq 1$: las observaciones no habrían agregado nada a nuestro conocimiento.

Así, parece que la teoría frecuencial tiene el defecto de su principal virtud: alcanza su generalidad al precio de no poder incorporar conocimientos a priori a sus enunciados. Cuando hacemos nuestro juicio final sobre $\mu$, tenemos que sintetizar la información obtenida de las observaciones con nuestro conocimiento a priori. El teorema de Bayes intenta esta síntesis desde el principio. La teoría frecuencial lo deja para el final (y, nos sentimos obligados a señalar, en la mayoría de las exposiciones actuales ignora el punto por completo).

La teoría fiducial, como hemos señalado, ha sido confinada por Fisher al caso en el que se utilizan estadísticos suficientes o, en general, a casos en los que se puede utilizar toda la información de la función de verosimilitud. No se ha dado una exposición sistemática del procedimiento a seguir cuando se dispone de información a priori, pero no parece haber motivo para no utilizar un método similar al explicado por la ecuación (44). Es decir, si derivamos la distribución fiducial $f(\mu)$ sobre un rango general pero tenemos la información adicional de que el parámetro debe estar en el rango $\mu_0$ a $\mu_1$ (dentro de ese rango general), modificamos la distribución fiducial por truncamiento a

$$\frac{f(\mu)}{\int_{\mu_0}^{\mu_1} f(\mu)d\mu}.$$

## 11.2. Falsacionismo

Una observación final es relevante con respecto al enfoque frecuentista. Su desarrollo fue paralelo al desarrollo del *falsacionismo* en la filosofía de la ciencia, encabezado por Sir Karl Popper (cf. Popper, 1968). La base del esquema de Popper es que la evidencia puede o no refutar una teoría pero no la sostiene; es decir, la ciencia progresa realizando experimentos que desafían las teorías. Esta visión de la ciencia no permite que los resultados de un experimento proporcionen una corroboración explícita de la teoría; está simbolizado por la restricción de que hablamos de "no rechazar $H_0$" en lugar de "aceptar $H_0$". Más

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

fundamentalmente, el enfoque frecuentista no busca proporcionar medidas de corroboración, y es la búsqueda de estas medidas lo que, en parte, ha impulsado el desarrollo de paradigmas alternativos para la inferencia estadística. De hecho, todos los demás enfoques descritos en este trabajo permiten hacer afirmaciones corroborativas sobre la base del experimento recién realizado y condicionalmente a las observaciones.

### 11.3. Inferencia basada en la verosimilitud

Todas las escuelas de pensamiento reconocen que la función de verosimilitud (FV) es un resumen completo de los datos. De hecho, Fisher (1956) sugirió graficar la FV contra $\theta$; otros (por ejemplo, Efron, 1982) también apoyan firmemente el uso de la FV como un *resumen* eficaz. Edwards (1972), argumentando que la FV describe el *soporte* relativo para diferentes valores de $\theta$, fue más allá y sugirió que la inferencia se hiciera sobre la base de estos valores de soporte. Claramente, tal enfoque es consistente con el principio de verosimilitud fuerte (PVF), aunque es incompleto a menos que se complemente con algún procedimiento para manejar parámetros no deseados (nuisance), como el uso de la verosimilitud parcial o el uso de la razón de verosimilitud como medida de credibilidad. Dichos métodos tienen la ventaja de que se puede incorporar información a priori a través de una función de verosimilitud a priori.

Los procedimientos Bayesianos siempre son consistentes con el PVF, aunque los métodos empíricos de Bayes no necesitan serlo. La inferencia fiducial puede violar el PVF, aunque tales violaciones tienden a ser poco comunes.

### 11.4. La probabilidad como un grado de creencia

En los argumentos Bayesianos y fiduciales, primero debemos asumir la existencia de un concepto diferente de probabilidad que mide el grado de creencia o credibilidad en una hipótesis o teoría. Carnap (1962) denominó a esta probabilidad$_1$, a diferencia del concepto frecuencial, probabilidad$_2$. Visto desde este punto de vista, el fracaso (?) del enfoque frecuencial para ofrecer aseveraciones sobre la credibilidad de una hipótesis es casi axiomático, ya que los frecuentistas no están dispuestos a aceptar ningún concepto de probabilidad$_1$ que no tenga una interpretación frecuencial.

El argumento fiducial se basa en el supuesto de que la probabilidad$_2$ se puede convertir en probabilidad$_1$ mediante una operación de pivote. Sabemos que el proceso es posible; la pregunta clave es si la medida de probabilidad resultante tiene sentido.

El núcleo duro de la inferencia Bayesiana es un desarrollo axiomático que proporciona el marco para especificar probabilidades a priori y actualizar dichas probabilidades mediante el teorema de Bayes. Para el objetivista, esto significa que debe haber un proceso acordado mediante el cual se pueda generar una a priori que sea aceptable para todos. Tal regla es necesariamente mecanicista, ya que la interpretación subjetiva no es admisible; sin embargo, si el cumplimiento de la regla no puede juzgarse ni por la frecuencia ni por criterios subjetivos, su significado sigue siendo bastante oscuro. En efecto, se nos pide que aceptemos que la a priori sea plana en $(-\infty, \infty)$ pero inversamente proporcional a $\theta$ en $(0, \infty)$. Los argumentos sofisticados relacionados con la distinción de alguna manera no logran impresionarnos como tocando la raíz del problema. Además, se encuentra que trabajar con a priori no informativas puede conducir a algunas dificultades teóricas; ver Stone (1976).

El núcleo duro de los subjetivistas exige que el individuo esté dispuesto a apostar por cualquier cosa, pero de manera lógica, como se señaló en §10.2. Dado este marco, ciertamente Uno puede comenzar con Su enunciado de probabilidad a priori y derivar Su enunciado de probabilidad a posteriori con respecto a la plausibilidad de una hipótesis. Comencemos por considerar el proceso de especificación de la distribución a priori.

Si la distribución a priori se especifica en forma conjugada, como la media normal siendo $N(\phi, \tau^2)$, entonces nos enfrentamos a una posible regresión infinita al especificar la a priori para $(\phi, \tau^2)$ y así sucesivamente. Esto se resuelve sólo alegando el conocimiento de los (hiper) parámetros en algún momento (cf. Lindley y Smith, 1972). Si la distribución a priori se determina dentro de un marco de apuestas, Usted debe poder especificar Su función de utilidad. Una vez que esto está disponible, un desarrollo axiomático como el de Savage (1954) muestra que el comportamiento coherente conduce a grados de creencia que satisfacen los axiomas de probabilidad.

Una vez que esté disponible la información a priori, Usted puede proceder a hacer inferencias de una manera que sea consistente con el PVF (y, por lo tanto, posiblemente inconsistente con el principio de confianza). Si esto es una fuente de fortaleza o debilidad depende del ojo del espectador, pero el hecho es que todas las inferencias hechas son subjetivas, Sus propias evaluaciones.

Si tales declaraciones individuales son aceptables es problemático. Al tomar una decisión en un contexto que carece de oportunidades de replicación, el uso de Tus probabilidades parece razonable cuando Tú eres el responsable de la decisión. Sin embargo, creemos que muchos análisis estadísticos, si no la mayoría, no pueden encajar razonablemente en un marco de teoría de decisiones. Además, la expresión de creencias personales no ha resultado aceptable como forma de informar sobre los resultados de una investigación.

## 12. ¿Reconciliación?

Como era de esperar, ha habido varios intentos de reconciliar los diferentes enfoques de la inferencia estadística; Revisaremos algunos de estos brevemente. Comenzamos observando que, en muestras grandes, todos los métodos son consistentes con el principio de verosimilitud fuerte.

Es posible ver que el uso del teorema de Bayes con una distribución a priori rectangular da un modo a posteriori que es igual al estimador por MV. Incluso si se utiliza una distribución a priori no rectangular, los métodos son *asintóticamente* equivalentes. La ecuación (6) puede escribirse en nuestra notación actual como

$$P(\theta|\mathbf{x}) \propto p(\theta)L(\mathbf{x}|\theta). \tag{45}$$

Maximizar esto con respecto a $\theta$ es equivalente a maximizar su

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

logaritmo,

$$\log p(\theta) + \log L(\mathbf{x}\,|\theta) = \sum_{i=1}^{n} \left\{ \log f(x_i\,|\theta) + \frac{1}{n} \log p(\theta) \right\}. \tag{46}$$

Cuando $n \to \infty$, el segundo término entre llaves del segundo miembro es insignificante y estamos maximizando efectivamente $\log L(x\,|\theta)$ para obtener el estimador por MV. Podemos expresar esto diciendo que, dadas suficientes observaciones, la distribución a priori se vuelve irrelevante; esto se conoce como el principio de *estimación estable*. Sin embargo, para $n$ pequeño, puede haber grandes diferencias entre las estimaciones por MV y las Bayesianas.

Diaconis y Freedman (1986a, b) muestran que cuando el espacio paramétrico es de alta dimensión (o infinitamente dimensional como en algunos problemas no paramétricos), la distribución a priori puede empantanar los datos sin importar cuántas observaciones estén disponibles. En este sentido, los estimadores Bayesianos pueden carecer de consistencia; también debe consultarse la discusión que siguió a su artículo de 1986a.

Un aspecto del enfoque Bayesiano es, como hemos sugerido en alguna ocasión, que exige demasiado. Por ejemplo, necesitamos poder especificar la forma funcional de la FV y enumerar todas las variables de interés. Sin embargo, gran parte del atractivo del procedimiento, como la validación cruzada y el botstrapping, se deriva de sus aplicaciones en circunstancias en las que puede que no sea posible especificar la FV con precisión. Asimismo, la aleatorización en el diseño experimental protege contra factores que pueden no haber sido reconocidos.

Siguiendo este tema, Durbin (1988) señala que la complejidad general de muchos modelos hace que las especificaciones de la FV y, por lo tanto, la aplicación del principio de verosimilitud, sean poco prácticas. Sin embargo, los tests de diagnóstico simples a menudo guían bien al constructor del modelo, y Durbin sugiere que los efectos prácticos de las diferencias filosóficas suelen ser pequeños en comparación con la necesidad de un modelo estadístico efectivo.

Box (1980) identifica dos componentes en el modelado estadístico: *crítica* y *estimación*. A partir de (27), Box usaría la distribución a posteriori de $\theta$ para la estimación, pero la distribución predictiva

$$f(\mathbf{x}) = \int p(\theta) L(\mathbf{x}\,|\theta)\, d\theta \tag{47}$$

para la crítica de modelos. Aunque $f(x)$ se deriva bajo el supuesto de que la distribución a priori $p(\theta)$ está disponible, Box recomienda procedimientos frecuentistas para la parte crítica del proceso de modelado. Esto es similar en espíritu a los comentarios de Durbin dados anteriormente.

Giere (1977) distinguió entre *test* e *información* en la inferencia estadística, sugiriendo que el criterio de información permite una medida directa de evidencia para una hipótesis para que se pueda invocar el enfoque Bayesiano. En el marco de los tests, no existe tal medida, como señalaron muchos escritores frecuentistas desde Neyman y Pearson en adelante. Gieres continúa argumentando a favor de la probabilidad como una medida de propensión que permitiría hacer afirmaciones para experimentos individuales.

I. J. Good (cf. 1976, 1983, 1988) pide un compromiso Bayesiano-no-Bayesiano desde un punto de vista diferente. Para Good, los métodos frecuentistas a menudo representan una colección de procedimientos *ad hoc*, y aceptaría los procedimientos frecuentistas siempre que coincidan lo suficientemente bien con la solución Bayesiana. Si bien tal enfoque puede servir para reducir la controversia, "compromiso" es quizás una descripción inapropiada.

Juntando estas diversas consideraciones, vemos que la buena práctica estadística a menudo puede surgir de diferentes paradigmas y que, de hecho, diferentes nociones de probabilidad pueden ser apropiadas en diferentes circunstancias. Sin embargo, el enfoque frecuentista permanece firmemente arraigado en la tradición Popperiana del falsacionismo, y cualquier intento de ir más allá requiere el reconocimiento de algún otro concepto de probabilidad.

Puede ser tentador pensar en términos de la noción de Kuhn (1970) de una revolución científica en la que el paradigma actual (frecuentista) es desafiado por el recién llegado (Bayesiano), de la cual surgirá una nueva ortodoxia. Sin embargo, este punto de vista es algo inapropiado; más bien deberíamos reconocer que el enfoque Bayesiano busca entregar más pero, para hacerlo, requiere suposiciones más fuertes.

Para concluir, conviene citar algunas palabras escritas hace mucho tiempo (Kendall, 1949):

> El frecuentista busca objetividad al definir sus probabilidades por referencia a frecuencias; pero tiene que usar una idea primitiva de aleatoriedad o equiprobabilidad para calcular la probabilidad en cualquier caso práctico dado. El no-fecuentista comienza tomando las probabilidades como una idea primitiva, pero tiene que suponer que los valores que sus cálculos dan a la probabilidad reflejan, de alguna manera, el comportamiento de los eventos... *Ninguna de las partes puede evitar usar las ideas del otro para establecer y justificar una teoría integral y profunda*.

## Referencias

Abril, J. C. (1999), *Análisis de Series de Tiempo Basado en Modelos de Espacio de Estado*, EUDEBA: Buenos Aires.

Abril, J. C. (2004). *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas: Buenos Aires.

Abril, J. C. y Abril, M. de las M. (2018). *Métodos Modernos de Series de Tiempo y sus Aplicaciones*. Editorial Académica Española: Saarbrücken (Alemania).

Barnett, V. D. (1982). *Comparative Statistical Inference*, 2nd edition. Wiley: Chichester.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Phil. Trans.*, **53**, 370. (Reproducido en *Biometrika*, **45**, 293 (1958), editado e introducido por G. A. Barnard).

Bernardo, J. M. (1980). A Bayesian analysis of classical hypothesis testing. In *Bayesian Statistics: Proceedings of the*

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

*first International Meeting*. Valencia Univ. Press: Valencia, Spain.

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. *Synthèse*, **36**, 19.

Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley: New York.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modeling and robustness (with discussion). *J. R. Statist. Soc.*, **A**, **143**, 383.

Carnap, R. (1962). *Logical Foundations of Probability*. 2nd. edition. Univ. of Chicago Press: Chicago.

Cox, D. R. y Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall: London.

Dawid, A. P. (1984). Present position and potential developments: some personal views. Statistical theory, the frequential approach (with discussion). *J. R. Statist. Soc.*, **B, 49**, 1.

De Groot, M. H. (1970). *Optimal Statistical Decissions*. McGraw-Hill: New York.

Diaconis, P. y Freedman, D. A. (1986a). On the consistency of Bayes estimates (with discussion). *Ann. Statist.*, **14**, 1.

Diaconis, P. y Freedman, D. A. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.*, **14**, 68.

Diaconis, P. y Zabell, S. L. (1982). Updating subjective probability. *J. Amer. Statist. Ass.*, **77**, 822.

Durbin, J. (1988). Is a philosophical consensus for statistics attainable? *J. Econometrics*, **37**, 51.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge Univ. Press: Cambridge.

Edwards, A. W. F. (1974). The history of likelihood. *Int. Statist. Rev.*, **42**, 9.

Efron, B. (1982). Maximum likelihood and decision theory. *Ann. Statist.*, **10**, 341.

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press: New York.

Fishburn, P. C. (1986). The axioms of subjective probability. *Statist. Sci.*, **1**, 335.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd: Edinburgh.

Gemes, K. (1984). A refutation of Popperian inductive scepticism. *Brit. J. Phil. Sci.*, **40**, 183.

Giere, R. N. (1977). Allan Birnbaum's conception of statistical evidence. *Synthèse*, **36**, 5.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208.

Godambe, V. P. (1976). Conditional likelihood and unconditional estimating equations. *Biometrika*, **63**, 277.

Godambe, V. P. and Sprott, D. A., eds. (1971). *Foundations of Statistical Inference*. Holt, Rinehart and Winston: Toronto, Canada.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. Griffin: London.

Good, I. J. (1976). The Bayesian influence, or how to sweep subjectivism under the carpet. In *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. 2. C. A. Hooker and W. Harper (eds.). Reidel: Dordrecht, Holland, 125.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and its Applications.* U. Minnesota Press: Minneapolis.

Good, I. J. (1988). The interface between statistics and philosophy of science. *Statist. Sci.*, **3**, 386.

Howson, C. and Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach*. Open Court: La Salle, Illinois.

Jeffreys, H. (1961). *Theory of Probability*, 3rd. edn. Oxford University Press: Oxford.

Johnstone, D. J. (1989). On the necessity for random sampling. *Brit. J. Phil. Sci.*, **40**, 443.

Kendall, M. G. (1949). On the reconciliation of theories of probability. *Biometrika*, **36**, 101.

Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan: London.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. 2nd. edition. Univ. of Chicago Press: Chicago.

Lakatos, I. (1974). Falsification and the methodology of scientific research programs. In *Criticism and the Growth of Knowledge*, I. Lakatos and A. E. Musgrave (eds.). Cambridge Univ. Press: Cambridge, 91.

Lindley, D. V. (1958). Fiducial distributions and Bayes' Theorem. *J. R. Statist. Soc.*, **B**, **20**, 102.

Lindley, D. V. (1971). *Bayesian Statistics Review*. S.I.A.M.: Philadelphia.

Lindley, D. V. y Smith, A. F. M. (1972). Bayesian estimates for the linear model. *J. R. Statist. Soc.*, **B**, **34**, 1.

Maritz, J. S. y Lwin, T. (1989). *Empirical Bayes Methods*. 2nd edition. Chapman and Hall: London.

Miller, D. (1990). A restoration of Popperian inductive scepticism. *Brit. J. Phil. Sci.*, **A**, **147**, 389.

Naylor, J. C. y Smith, A. F. M. (1988). Econometric illustrations of novel numerical integration strategies for Bayesian inference. *J. Econometrics*, **38**, 103.

Neyman, J. (1962). Two breakthroughs in the theory of statistical decision making. *Rev. Int. Statist. Inst.*, **30**, 11.

Popper, K. R. (1968). *The Logic of Scientific Discovery*. Hutchinson: London.

Popper, K. R. (1969). *Conjeturas y Refutaciones*. Routledge and Kegan Paul: London.

Popper, K. R. y Miller, D. (1987). A proof of the impossibility of inductive probability. *Nature*, **302**, 687.

Ramsey, F. P. (1926, 1931). Truth and probability. In *The Foundations of Mathematics and Other Essays*. Kegan, Paul Trench, Tubner. Reimpreso en H. E. Kyburg, Jr. y H. E. Smokler (eds. 1964). *Studies in Subjective Probability*. Wiley: New York, 61.

https://doi.org/10.5281/zenodo.8205883

SOUTHAMERICAN RESEARCHJOURNAL
ISSN 2806-5638

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. 3rd Berkeley Symp. Math. Statist. and Prob.*, **1**, 157. U. California Press: Berkeley.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.*, **35**, 1.

Savage, L. J. (1954). *The Foundation of Statistics*. Wiley: New York.

Savage, L. J. (1961). The foundation of statistics reconsidered. *Proc. 4th Berkeley Symp. Math. Statist and Prob.*, **1**, 575.

Savage, L. J. (1962). *The Foundation of Statistical Inference: a Discussion... at a Meeting of the Joint Statistical Seminar, Birkbeck and Imperial Colleges, in the University of London*. Methuen: London.

Stone, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Ass.*, **71**, 114.

Thiele, T. N. (1903). *Theory of Observations*. Reimpreso (1931) en *Ann. Math. Statist.*, **2**, 165, de la versión en inglés publicada en 1903; el original (danés) apareció en 1889 y 1897.

Wald, A. (1950). *Statistical Decision Functions*. Wiley: New York.