LA SOLUCIÓN DEL PROBLEMA DE ELEGIR UN PROMEDIO EN LAPLACE (MEMORIA DE 1774)

Jesús Basulto Santos
José Javier Busto Guerrero
Francisco Javier Ortega Irizo
Universidad de Sevilla

INTRODUCCIÓN

Cuando tenemos la necesidad de estimar una magnitud desconocida μ, realizamos una muestra de mediciones sobre la magnitud, $x_1, x_2, \ldots, x_n$, y procedemos a calcular un valor, $\hat{\mu}$, a partir de la mediciones, por medio de algún criterio que nos asegure una buena estimación de la magnitud.

La cuestión de estimar una magnitud a partir de un conjunto de mediciones, era denominada en el siglo XVIII el problema de elegir un promedio.

Laplace (1749-1827) abordará el problema en 1772 y lo recogerá, como problema 3,en su memoria de 1774 "Mémoire sur la probabilité des causes par les évenemens". Laplace ha encontrado una solución al problema, de elegir un promedio, en 1772 pero no lo incluirá en su memoria de 1772 "Sur les Séries récurrorécurrentes" por creer que era de poca utilidad. Sólo, después de leer en la revista de astronomía de Jean Bernoulli, que el problema es de gran importancia en astronomía y que Daniel Bernoulli y Lagrange han considerado el problema en sus trabajos de investigación, retomará la solución encontrada en 1772 y la recogerá en su memoria de 1774.

Laplace partirá de que las mediciones son realizaciones independientes de una variable aleatoria X tal que $X - \mu = Z$, donde Z es una variable aleatoria (v.a.) con una función de densidad conocida $\phi(z)$. La v.a. Z será la Ley de los Errores. El supuesto de que los errores deben seguir una Ley aleatoria ha sido aceptado por muchos investigadores, aunque cada uno de ellos ha propuesto su propia función de densidad.

Como explicaremos más adelante, Laplace desarrollará su propuesta únicamente para **tres** mediciones. Ordenando las tres mediciones x_1, x_2, x_3 en los estadísticos ordenados, $y_1 \leq y_2 \leq y_3$, Laplace calculará la densidad conjunta de observar las tres mediciones cuando provienen de una Ley $\phi(Y - \mu)$, para todo valor de la magnitud μ, es decir,

$$\phi(y_1, y_2, y_3) \propto \phi(y_1 - \mu)\, \phi(y_2 - \mu)\, \phi(y_3 - \mu)$$

que es nuestra **función de verosimilitud**.

Para Laplace, el valor a priori de la magnitud μ es una v.a. con una función de densidad $\pi(\mu)$. Con lo que la densidad de la v.a., μ, condicionada a las mediciones, vendrá dada por nuestra fórmula de Bayes

$$\pi(\mu \mid y_1, y_2, y_3) \propto \pi(\mu)\, \phi(y_1, y_2, y_3 \mid \mu)$$

Ahora bien, Laplace se apoyará en el principio de la Razón Insuficiente (“si no tenemos una causa para asignar diferentes probabilidades a priori, debemos suponer que todas las probabilidades son iguales de verosímiles”) para suponer que $\pi(\mu) = 1$, es decir, una **función a priori imparcial**. Stigler (1986) señala que el principio de razón insuficiente es más que un axioma metafísico, un método para simplificar los cálculos.

Bajo el supuesto de tomar una función a priori imparcial, Laplace obtendrá la siguiente función de densidad a posteriori, una vez observadas las tres mediciones,

$$\pi(\mu \mid y_1, y_2, y_3) \propto \phi(y_1, y_2, y_3 \mid \mu)$$

y al considerar que la integral

$$\int \phi(y_1, y_2, y_3 \mid \mu)\, d\mu$$

es finita, obtiene la densidad a posteriori siguiente

$$\pi(\mu \mid y_1, y_2, y_3) = \frac{\phi(y_1, y_2, y_3 \mid \mu)}{\int \phi(y_1, y_2, y_3 \mid \mu)\, d\mu}$$

Esta última fórmula es la que propondrá Laplace, al comienzo de su memoria, como un **Principio** para poder **invertir**, llamado **principio de la probabilidad inversa**, el proceso que va de las causas a los sucesos, se invierte para ir de los sucesos a las causas.

A partir de aquí, Laplace introducirá la siguiente **función de pérdida**

$$g(\mu, \hat{\mu}) = |\hat{\mu} - \mu|$$

donde $\hat{\mu}$ es una estimación de μ. Laplace tomará como la mejor estimación del valor de la magnitud, μ, el valor μ^* que haga mínima la pérdida esperada a posteriori. Es decir, el valor esperado de la función de pérdida cuando μ sigue la función a posteriori. Por razonamientos geométricos, Laplace probará que μ^* debe ser la mediana de la densidad a posteriori de la v.a. μ. Es decir,

$$\int_{-\infty}^{\mu^*} \pi(\mu \mid y_1, y_2, y_3)\, d\mu = \frac{1}{2}$$

Ahora bien, si queremos que esta fórmula sea útil necesitamos especificar la función de densidad $\phi(z)$, es decir debemos elegir la **Ley o Curva de los Errores**.

Laplace rechazará la hipótesis de que la curva de los errores es uniforme, $\phi(z) = k$ en un intervalo finito centrado en el cero. La razón es que son más probables errores pequeños que grandes. Igualmente rechazará la hipótesis de que la curva sea triangular, $|\phi'(z)| = k$ en un intervalo finito centrado en el cero, ya que cree que esta derivada debe disminuir cuando aumenta z.

Laplace supondrá que la curva de los errores debe verificar que sea simétrica respecto de cero y que $\phi'(z)/\phi(z) = -k$ para $z > 0$. Estas hipótesis le conducirá a que la curva de los errores sea

$$\phi(z) = \frac{k}{2} e^{-k|z|}$$

que hoy denominamos distribución de Laplace o doble exponencial.

Una vez elegida la curva de los errores, Laplace obtiene la siguiente fórmula para la mediana a posteriori,

$$\mu^* = y_1 + p + \frac{1}{k} \ln\left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right)$$

donde $p = y_2 - y_1$, $q = y_3 - y_2$, suponiendo que $p > q$. La constante k es conocida. Laplace observa que cuando k tiende a cero, entonces μ^* se aproxima a la media aritmética. Igualmente, cuando k tiende a infinito, entonces μ^* se aproxima a la mediana de las mediciones, que es el máximo de la función de verosimilitud.

Resuelto el problema, Laplace emprende estimar μ cuando el parámetro k es desconocido. De nuevo Laplace utiliza el principio de la razón insuficiente para el vector (μ, k), proponiendo como función a priori $\pi(\mu, k) = 1$. El principio de la probabilidad inversa le lleva a la siguiente distribución a posteriori,

$$\pi(\mu, k \mid y_1, y_2, y_3) \propto \phi(y_1, y_2, y_3 \mid \mu, k)$$

donde la distribución a posteriori para μ, es

$$\pi(\mu \mid y_1, y_2, y_3) = \int_0^\infty \pi(\mu, k \mid y_1, y_2, y_3)\, dk$$

y la mediana de esta distribución es el valor μ^* que verifica la siguiente ecuación,

$$\int_{-\infty}^{\mu^*} \pi(\mu \mid y_1, y_2, y_3)\, d\mu = \frac{1}{2}$$

y esta última ecuación es equivalente a la siguiente,

$$\int_0^\infty \int_{-\infty}^{\mu^*} \pi(\mu, k \mid y_1, y_2, y_3)\, d\mu\, dk = \frac{1}{2} \int_0^\infty \int_{-\infty}^{\infty} \pi(\mu, k \mid y_1, y_2, y_3)\, d\mu\, dk$$

que haciendo el cambio de variable $\theta = \mu - y_1$, y teniendo en cuenta las definiciones de p y q se transforma en

$$\int_0^\infty \int_{-\infty}^{\theta^*} \pi(\theta, k \mid p, q)\, d\theta\, dk = \frac{1}{2} \int_0^\infty \int_{-\infty}^{\infty} \pi(\theta, k \mid p, q)\, d\theta\, dk$$

que a su vez es equivalente a

$$\int_0^\infty \int_{-\infty}^{\theta^*} \pi(\theta, k \mid p, q)\, \pi(k \mid p, q)\, d\theta\, dk = \frac{1}{2} \int_0^\infty \int_{-\infty}^{\infty} \pi(\theta \mid k, p, q)\, \pi(k \mid p, q)\, d\theta\, dk$$

Ahora, teniendo en cuenta que se verifica

$$\pi(\theta, p, q \mid k) \propto \pi(\theta, k \mid p, q) \qquad (*)$$

(ver fórmulas (1) y (2) más adelante), donde la proporcionalidad actúa sobre ambas funciones y las funciones dependen de las variables θ y k, obtenemos la siguiente ecuación

$$\int_0^{\infty}\int_{-\infty}^{\theta^*} \pi(\theta,\ p,\ q \big| k)\ d\theta\ dk = \frac{1}{2}\int_0^{\infty}\int_{-\infty}^{\infty} \pi(\theta,\ p,\ q \big| k)\ d\theta\ d \Rightarrow$$

$$\Rightarrow \int_0^{\infty} k^2 e^{-k(2p+q-\theta^*)}\ dk = \int_0^{\infty} k^2 e^{-k(p+q)} \left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right) dk$$

que conduce a la ecuación de grado tres:

$$(2p + q - \theta^*)^3 = \left((p+q)^{-3} + \frac{(2p+q)^{-3}}{3} - \frac{(p+2q)^{-3}}{3} \right)^{-1}$$

donde hemos supuesto que $p > q$.

Ahora bien, según señala Stigler(1986), Laplace ha cometido un error al calcular θ*, ya que en vez de utilizar la expresión (*), utilizó la siguiente:

$$\pi(\theta \big| p,\ q,\ k) \propto \pi(\theta,\ p,\ q \big| k)$$

lo que le llevó a la ecuación

$$\int_0^{\infty}\int_{-\infty}^{\theta^*} \pi(\theta,\ p,\ q \big| k)\ \pi(k \big| p,\ q)\ d\theta\ dk = \frac{1}{2}\int_0^{\infty}\int_{-\infty}^{\infty} \pi(\theta,\ p,\ q \big| k)\ \pi(k \big| p,\ q)\ d\theta\ dk \Rightarrow$$

$$\Rightarrow \int_0^{\infty} k^2 e^{-k(2p+q-\theta^*)}\ \pi(k \big| p,\ q)\ dk = \int_0^{\infty} k^2 e^{-k(p+q)} \left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right) \pi(k \big| p,\ q)\ dk$$

que finalmente le condujo a la ecuación de grado 15:

$$(3p + 2q - \theta^*)^{-5} - \frac{(4p + 2q - \theta^*)^{-5}}{3} - \frac{(3p + 3q - \theta^*)^{-5}}{3} =$$
$$= (3p + 2q)^{-5} - \frac{2(2p+3q)^{-5}}{3} - \frac{(4p+2q)^{-5}}{9} + \frac{(2p+4q)^{-5}}{9}$$

donde hemos supuesto que $p > q$. Laplace probará que sólo existe una raíz válida (que la mediana sea una raíz de una ecuación de grado 15 explica por qué Laplace tomó sólo tres mediciones).

Stigler conjetura que este error fue debido a que, en tiempos de Laplace, no se tenía una idea clara del concepto de probabilidad condicionada. No obstante, analizando el camino seguido por Laplace, **tal vez** pueda darse otra interpretación al error cometido y que le llevó a la ecuación errónea.

En primer lugar, Laplace calcula la distribución marginal a posteriori de k, obteniendo

$$\pi(k \mid p,q) \propto k^2 e^{-k(p+q)} \left(1 - \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right)$$

donde se observa que depende sólo de p y q. A continuación, obtiene una ecuación que le permite calcular la mediana a posteriori para cada valor de k conocido, valor que Laplace llamó x y que nosotros vamos a llamar $\theta^*(k)$, ya que se trata de una función que depende del valor de k. Dicho valor ha de verificar

$$k^2 e^{-k(2p+q-\theta^*(k))} = k^2 e^{-k(p+q)} \left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right)$$

y al tomar esperanzas en ambos miembros, obtenemos

$$\int_0^\infty k^2 e^{-k(2p+q-\theta^*(k))}\, \pi(k \mid p,q)\, dk = \int_0^\infty k^2 e^{-k(p+q)} \left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right) \pi(k \mid p,q)\, dk$$

Laplace resolverá las integrales pero considerando un θ^* fijo que no depende de k, lo que le conduce a resolver la ecuación incorrecta indicada anteriormente. Aunque las soluciones que se obtienen son diferentes, cálculos realizados por Stigler y por nosotros, nos llevan a sostener que la diferencia entre ellas es muy pequeña.

A partir de aquí, nuestro trabajo consta de la sección 2, donde obtenemos, para k conocida, los resultados de Laplace. En esta sección discutimos la obtención de la función a priori imparcial $\pi(\mu) = 1$, señalamos que p y q son estadísticos conjuntamente auxiliares y demostramos que los intervalos bayesianos se comportan como intervalos de confianza, en muentras condionadas a los estadísticos auxiliares p y q. En la sección 3 consideramos que k es desconocido, discutimos la función a priori de Laplace $\pi(\mu, k) = 1$ y probamos que si eligimos una función a priori imparcial del tipo $\pi(\mu, k) = 1/k$, entonces los intervalos bayesianos se comportan como intervalos de confianza, cuando se condiciona a los estadísticos auxiliares p y q. Por último, finalizamos el trabajo con una discusión del trabajo de Laplace.

ESTIMACIÓN DE μ CUANDO k ES CONOCIDO

A partir de la curva de errores

$$f(x \mid \mu, k) = \frac{k}{2} e^{-k|x-\mu|}$$

donde $\mu \in R$, $k > 0$ y $x \in R$, y los estadísticos ordenados $y = \{y_1, y_2, y_3\}$, obtenemos la siguiente función de verosimilitud

$$L(\mu \mid k, y) \propto e^{-k\sum_{i=1}^{3}|y_i-\mu|}$$

que posee un máximo en $\hat{\mu} = y_2$ que denominamos estimador máximo verosímil.

Llamando $p = y_2 - y_1$, $q = y_3 - y_2$, podemos escribir la función de verosimilitud de la forma siguiente,

$$L(\mu \mid p, q, \hat{\mu}, k) \propto e^{-k(|\hat{\mu}-\mu-p|+|\hat{\mu}-\mu|+|q+\hat{\mu}-\mu|)}$$

que si ahora tomamos como función imparcial, $\pi(\mu) = 1$, obtenemos la siguiente distribución a posteriori

$$\pi(\mu \mid p, q, \hat{\mu}, k) = \frac{L(\mu \mid p, q, \hat{\mu}, k)}{\int_{-\infty}^{\infty} L(\mu \mid p, q, \hat{\mu}, k)\, d\mu}$$

y haciendo el cambio de variable $\theta = \mu - y_1$ obtenemos

$$\pi(\theta \mid p, q, k) = \frac{e^{-k(|\theta|+|p-\theta|+|p+q-\theta|)}}{I(p, q, k)}$$

donde

$$I(p, q, k) = \frac{2}{k} e^{-k(p+q)} \left(1 - \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right)$$

Ahora, si consideramos la función de pérdida absoluta de Laplace, podemos probar que el valor de μ que minimiza la perdida esperada a posteriori es la mediana de la distribución a posteriori. Llamando μ^* al valor de la mediana, éste debe ser solución de la ecuación siguiente,

$$\int_{-\infty}^{\mu^*} \pi(\mu \mid p, q, \hat{\mu}, k)\, d\mu = \frac{1}{2}$$

y, si suponemos que los tres estadísticos ordenados son diferentes, entonces se demuestra que para $p > q$ resulta $y_1 < \mu^* < y_2$, si $p < q$ entonces $y_2 < \mu^* < y_3$ y si $p = q$ entonces $\mu^* = y_2$.

Si ahora suponemos que $p > q$, entonces es fácil probar que

$$\mu^* = y_1 + p + \frac{1}{k} \ln\left(1 + \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3}\right)$$

Estos resultados de Laplace contienen ideas que han sido fundamentales para el desarrollo posterior de la inferencia estadística. En primer lugar vamos a obtener la distribución conjunta de los estadísticos ordenados,

$$f\,(y_1, y_2, y_3 \mid k, \mu) = 3!\left(\frac{k}{2}\right)^3 e^{-k\sum_{i=1}^{3}|y_i-\mu|}$$

que si hacemos el cambio de variables: $\{\theta = \mu - y_1,\ p = y_2 - y_1,\ q = y_3 - y_2\}$ obtenemos

$$f\,(\theta, p, q \mid k) = 3!\left(\frac{k}{2}\right)^3 e^{-k(|\theta|+|p-\theta|+|p+q-\theta|)} \qquad (1)$$

que no depende de μ. La distribución marginal de los estadísticos p y q, es

$$f\,(p, q \mid k) = 3!\left(\frac{k}{2}\right)^3 I\,(p, q, k)$$

que al no depender de μ son denominados estadísticos auxiliares. La distribución de la v.a. θ, condicionada a los estadísticos auxiliares, no depende de la cantidad desconocida μ,

$$f\,(\theta \mid p, q, k) = \frac{e^{-k(|\theta|+|p-\theta|+|p+q-\theta|)}}{I\,(p, q, k)}$$

que coincide con la función $\pi(\theta \mid p, q, k)$. Al depender θ de μ y del estadístico y_1, recibe el nombre de cantidad pivotal. Igualmente se prueba que la marginal de θ no depende de μ.

Laplace ha resuelto el problema de elegir un promedio para una curva doble exponencial. Construye una función de densidad sobre el parámetro μ a partir de la muestra observada, y esta función mide la incertidumbre que tenemos sobre dicho parámetro. En cambio, Lagrange ha tomado la media aritmética y ha calculado su distribución en el muestreo, con el objeto de apoyar su elección. Todas las muestra del mismo tamaño intervienen en la solución de Lagrange, la observada y el resto. Sabemos que la estadística que se impuso en el primer cuarto de nuestro siglo va a seguir el camino de Lagrange. Ahora bien, Fisher va a retomar los resultados de Laplace y va a proponer un camino intermedio, entre una muestra y todas las muestras. Fisher propondrá el **Principio de Condicionar a estadísticos auxiliares**. Este principio consiste en trabajar con muestras que tienen los mismos valores para los estadísticos auxiliares calculados en la muestra observada. Es decir, si $p=2$ y $q=1$, sólo consideraremos muestra de tamaño $n=3$ donde $p=2$ y $q=1$. Por lo tanto, en la construcción de un intervalo de confianza para μ, deberemos utilizar la distribución $f(\theta \mid p, q, k)$.

Fisher también retomará la idea de asignar una función de densidad al parámetro desconocido μ. Ahora bien, Fisher rechazará las funciones a priori uniformes de Laplace por cambiar frente a transformaciones de los parámetros. Es decir, si $\pi(\mu)=1$, el principio de razón insuficiente conduce a que $\pi(\mu^2)=1$, que no está de acuerdo con la regla del cambio de variable. Fisher inventará la denominada **Inferencia Fiducial**, cuyo objetivo es determinar una función de densidad para μ a partir de una muestra observada.

Una solución al problema de determinar la funciones a priori de Laplace va a surgir de Jeffreys (ver su libro de 1961). La idea de Jeffreys consiste en definir en el conjunto de los modelos de Laplace,

$$f(x \mid \mu, k) = \frac{k}{2} e^{-k|x-\mu|}$$

con conocido, la denominada distancia de Hellinger entre dos modelos dis tintos, es decir

$$H(\mu_1, \mu_2) = \int_{-\infty}^{\infty} \left[\left(f(y \mid \mu_1, k)^{1/2} - f(y \mid \mu_2, k)^{1/2} \right) \right]^2 dy$$

que llamando $h = \mu_2 - \mu_1$, obtenemos que la distancia de Hellinger vale

$$H(\mu_1, \mu_2) = \frac{1}{2} \left(4 \left(1 - e^{-\frac{kh}{2}} \right) - 2khe^{-\frac{kh}{2}} \right)$$

Esta distancia considerada por Jeffreys es invariante a cambios de parámetros. Esta propiedad es necesaria ya que la distancia no debe depender del tipo de parámetro considerado. Por ejemplo, el modelo de Laplace puede definirse en función de $\nu = \exp(\mu)$.

Jeffreys definirá una función a priori sobre μ, $\pi(\mu)$, observando (con notaciones muy diferentes) que $H(\mu, \mu + h) = I(\mu)\, h^2 + o(h^2)$ (donde $I(\mu)$ representa la Información de Fisher y se supone que el modelo es regular), con lo que la distribución a priori

$$\pi(\mu) \propto \sqrt{I(\mu)} = \left(\lim_{h \to 0} \frac{H(\mu, \mu + h)}{|h|^2} \right)^{\frac{n}{2}}$$

resulta ser invariante ante reparametrizaciones. En Ortega, J. Se generaliza esta definición, proponiendo

$$\pi(\mu) \propto \left(\lim_{h \to 0} \frac{H(\mu, \mu + h)}{|h|^z} \right)^{\frac{1}{2}}$$

donde $z = 1$ si el límite existe y es distinto de cero e ∞, en otro caso $z = 2$ si el correspondiente límite existe y es distinto de cero e ∞, etc. En los modelos regulares, resultará ser $z = 2$ y en los no regulares se obtiene $z = 1$. En nuestro caso se prueba que debemos tomar $z = 2$, donde el límite vale $k^2/4$, que al no depender de μ conduce a tomar como distribución a priori la función $\pi(\mu) = 1$, que es la elegida por Laplace. Ahora bien, si representamos el modelo de Laplace por un nuevo parámetro, $\nu = \exp(\mu)$, la aplicación de esta regla de Jeffreys, conduce a que $\pi(\nu) \propto \pi(\mu)\left| d\mu/d\nu \right|$, que está de acuerdo con la fórmula de cambio de variable, y ya no es uniforme.

Podemos relacionar la distribución $\pi(\mu) = 1$ con el concepto de distribución imparcial. En Basulto (1998) introducimos el concepto de función a priori imparcial, que está relacionado con los estadísticos auxiliares. En nuestro modelo de Laplace, la función $\pi(\mu) = 1$ es imparcial porque es solución de la siguiente ecuación integral,

$$\int_{-\infty}^{+\infty} \pi(\mu) \frac{L(\mu \mid k, y)}{L(\hat{\mu} \mid k, y)} d\mu \propto \left(1 - \frac{e^{-kp}}{3} - \frac{e^{-kq}}{3} \right)$$

es decir, la función $\pi(\mu) = 1$ hace que la integral sea constante para todas las muestras de tamaño $n = 3$ que tenga los mismos valores que p y q. La función