6 Inferencia basada en la verosimilitud

En este capítulo vamos a estudiar algunas de las aproximaciones más básicas a la inferencia. En esencia, deseamos que nuestras inferencias dependan solamente del modelo $\{P_\theta : \theta \in \Omega\}$ y de los datos s. Estos métodos son minimalistas, en el sentido de que requieren muy pocas suposiciones o hipótesis de trabajo, y aunque resuelven satisfactoriamente algunas situaciones, para casos algo más complejos parece necesaria una estructuración adicional que se expondrá en los Capítulos 7 y 8.

La función de verosimilitud es uno de los conceptos más básicos en inferencia estadística, habiéndose construido teorías completas de inferencia en base a dicho concepto. En los Apartados 6.1, 6.2, 6.3 y 6.5 estudiaremos los métodos de verosimilitud y en el Apartado 6.4 introduciremos algunos métodos de inferencia de distribución libre. Estos no son realmente ejemplos de métodos de verosimilitud pero siguen la misma idea básica de que las inferencias dependen de las mínimas suposiciones posibles.

6.1 La función de verosimilitud

Las inferencias por verosimilitud se basan solamente en los datos s y el modelo $\{P_\theta : \theta \in \Omega\}$, es decir, un conjunto de posibles medidas de probabilidad para el sistema que es objeto de estudio. Con estos elementos obtenemos la entidad fundamental de la inferencia basada en la verosimilitud, denominada la función de verosimilitud.

Para fundamentar la definición de la función de verosimilitud suponga que tenemos un modelo estadístico en el que cada P_θ es discreta y viene dada por una función de probabilidad f_θ. Una vez se ha observado s, considere la función $L(\,\cdot\,|s)$ definida en el espacio de parámetros Ω y con valores de R^1, dada por $L(\theta|s) = f_\theta(s)$. Nos referiremos a la función $L(\,\cdot\,|s)$ determinada por el modelo y los datos como la *función de verosimilitud*, y al valor de $L(\theta|s)$ lo denominaremos *verosimilitud* de θ. Observe que para la función de verosimilitud fijamos los datos y variamos el valor del parámetro.

Vemos que $f_\theta(s)$ no es más que la probabilidad de obtener los datos s cuando el valor verdadero del parámetro es θ. Esto implica establecer un orden de preferencia o confianza en Ω, es decir, creemos que θ_1 es el valor verdadero de θ, por encima de θ_2, siempre que $f_{\theta_1}(s) > f_{\theta_2}(s)$. Ello se debe a que la desigualdad indica que los datos son más probables si se considera θ_1 que si se considera θ_2. Por el contrario, si $f_{\theta_1} = f_{\theta_2}$ permanecemos indiferentes respecto a θ_1 y θ_2. La inferencia de θ por verosimilitud está basada en esta ordenación.

Es importante recordar la interpretación correcta de $L(\theta\,|\,s)$. El valor de $L(\theta\,|\,s)$ es la probabilidad de s dado que θ es el valor verdadero, y *no* la probabilidad de θ dado que hemos observado s. También puede ocurrir que el valor de $L(\theta\,|\,s)$ sea muy pequeño para cada valor de θ. Sin embargo, no es el valor concreto de la verosimilitud el que nos informa de la confianza para cada valor de θ, sino el valor relativo de la verosimilitud de los diferentes valores posibles del parámetro.

EJEMPLO 6.1.1 Suponga que $S = \{1, 2, \ldots\}$ y que el modelo estadístico es $\{P_\theta : \theta \in \{1, 2\}\}$, donde P_1 es la distribución uniforme para los valores enteros $\{1, \ldots, 10^3\}$ y P_2 es la distribución uniforme para los enteros $\{1, \ldots, 10^6\}$, y suponga además que el valor observado ha sido $s = 10$. En consecuencia $L(1|10) = 1/10^3$ y $L(2|10) = 1/10^6$. Ambos valores son bastante pequeños, pero tenga presente que la verosimilitud de $\theta = 1$ es mil veces mayor que la de $\theta = 2$. □

Por lo tanto, y de acuerdo con lo expuesto, nuestro interés se centra en la *relación de verosimilitud* o *ratio de verosimilitud* $L(\theta_1\,|\,s)/L(\theta_2\,|\,s)$ para θ_1, $\theta_2 \in \Omega$ cuando deseamos establecer inferencias para θ basadas en la función de verosimilitud. Esto implica que cualquier función que sea un múltiplo positivo de $L(\cdot\,|\,s)$, es decir, $L^*(\cdot\,|\,s) = cL(\cdot\,|\,s)$ para algún valor de $c > 0$, puede utilizarse también como una función de verosimilitud. Diremos que dos verosimilitudes son equivalentes si son proporcionales tal como se ha expuesto y en general nos referiremos a cualquier múltiplo positivo de $L(\cdot\,|\,s)$ como una función de verosimilitud.

EJEMPLO 6.1.2 Suponga que lanzamos una moneda $n = 10$ veces y que obtenemos $s = 4$ caras. Sin ningún conocimiento sobre la probabilidad de obtener cara en un lanzamiento, el modelo estadístico adecuado para la situación es el modelo Binomial$(10, \theta)$ con $\theta \in \Omega = [0, 1]$. La función de verosimilitud viene dada por

$$L(\theta\,|\,4) = \binom{10}{4}\theta^4(1-\theta)^6, \tag{6.1.1}$$

representada en la Figura 6.1.1.

El máximo de la función de verosimilitud es 0,2508 y corresponde a $\theta = 0{,}4$. Más adelante estudiaremos aplicaciones de la verosimilitud para estimar el valor desconocido de θ y la precisión de dicha estimación. En líneas generales, el procedimiento consistirá en determinar dónde aparece un máximo de la verosimilitud y cuál es la dispersión de la verosimilitud alrededor de dicho máximo.

Existe un abanico de aproximaciones para obtener inferencias mediante la función de verosimilitud. Un extremo es el principio de verosimilitud.

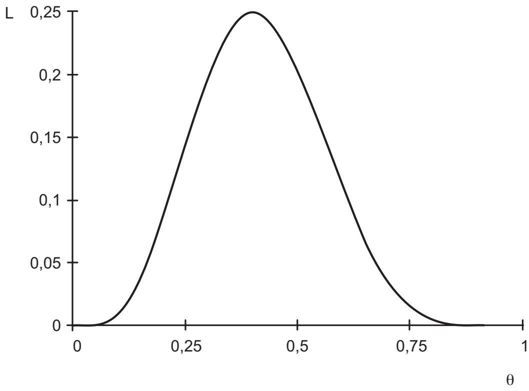

Figura 6.1.1: *Función de verosimilitud del modelo Binomial(10, θ) cuando se ha obtenido $s = 4$.*

> *Principio de verosimilitud*: Si para dos modelos, la combinación de los datos conducen a funciones de verosimilitud equivalentes, las inferencias sobre el parámetro desconocido deben ser idénticas.

Este principio impone que cualquier cosa que queramos decir sobre el valor desconocido de θ debe basarse sólo en $L(\cdot\,|s)$. Para muchos estadísticos esto constituye una restricción muy severa. Considere el siguiente ejemplo.

EJEMPLO 6.1.3 Suponga que lanza una moneda diversas veces de forma independiente hasta obtener cuatro caras y el número de cruces observado hasta que aparece la cuarta cara es $s = 6$. La distribución de s es por tanto una distribución Binomial Negativa(4, θ), y la función de verosimilitud para los datos obtenidos es

$$L(\theta|4) = \binom{9}{3}\theta^4(1-\theta)^6.$$

Observe que esta función de verosimilitud es un múltiplo positivo de (6.1.1).

Por tanto, el principio de verosimilitud afirma que para estos dos modelos, la combinación de los datos debe rendir las mismas inferencias sobre el valor desconocido de θ. En consecuencia, el principio de verosimilitud establece que podemos ignorar el hecho de que los datos se hayan obtenido de formas completamente diferentes. Sin embargo, si tenemos en cuenta algunas condiciones adicionales aparte de la función de verosimilitud, puede ocurrir que podamos derivar inferencias diferentes para las dos situaciones. En concreto, si se tiene en cuenta el método de muestreo, la verificación de un hipotético valor $\theta = \theta_0$ puede realizarse de formas diferentes. Muchos estadísticos opinan que esta información adicional debe utilizarse cuando se pretende derivar inferencias. □

Como ejemplo de una inferencia derivada de la función de verosimilitud, considere un conjunto de la forma

$$C(s) = \{\theta : L(\theta\,|\,s) \geq c\},$$

para algún valor $c \geq 0$. El conjunto $C(s)$ se denomina *región de verosimilitud*, y contiene todos aquellos valores de θ cuya verosimilitud es al menos c. Una región de verosimilitud, para algún valor de c, es un conjunto razonable en el que consideramos que se encuentra posiblemente el verdadero valor de θ. Si $\theta^* \notin C(s)$, entonces $L(\theta^*\,|\,s) < L(\theta\,|\,s)$ para cada $\theta \in C(s)$ y por tanto no resulta sostenible considerarlo, en base a los valores observados, como un valor de $C(s)$. Por lo tanto, puede considerarse que el tamaño de $C(s)$ constituye una medida de nuestra incertidumbre respecto al valor verdadero de θ.

Nos queda, sin embargo, el problema de escoger un valor adecuado de c y, como parece indicar el Ejemplo 6.1.1, la verosimilitud exclusivamente no parece ser una forma natural de realizar esta elección. En el Apartado 6.3.2 expondremos un método para seleccionar c basado en propiedades adicionales del modelo aparte de la función de verosimilitud.

En lo expuesto anteriormente en este apartado hemos supuesto que nuestros modelos estadísticos consisten en distribuciones discretas. La definición de verosimilitud resulta bastante natural, ya que $L(\theta\,|\,s)$ es simplemente la probabilidad de que ocurra s cuando θ es el verdadero valor. Sin embargo, claramente esta interpretación no es derivable directamente para un modelo continuo, puesto que en tal caso cada dato tiene una probabilidad nula de ocurrir. No obstante, considere que $f_{\theta_1}(s) > f_{\theta_2}(s)$ y que $s \in R^1$. En tal caso, suponiendo la continuidad de f_θ en s, tenemos

$$P_{\theta_1}(V) = \int_a^b f_{\theta_1}(x)\,dx > P_{\theta_2}(V) = \int_a^b f_{\theta_2}(x)\,dx$$

para cada intervalo $V = (a, b)$, lo suficientemente pequeño, que contenga a s. Esto significa que la probabilidad de que ocurra s cuando θ_1 es el verdadero valor es mayor que la probabilidad de que ocurra s cuando el verdadero valor es θ_2, por lo que el dato s respalda a θ_1 más que a θ_2. Una interpretación similar se aplica cuando $s \in R^n$ para $n > 1$ y V es una región que contiene a s.

Por lo tanto, para el caso continuo definiremos de nuevo la función de verosimilitud como $L(\theta\,|\,s) = f_\theta(s)$ e interpretaremos los valores verosimilitud de θ que se obtienen de ella exactamente igual que en el caso discreto.[1] De nuevo, también dos funciones de verosimilitud se considerarán equivalentes si una es un múltiplo positivo de la otra.

Vamos a considerar ahora un ejemplo muy importante.

EJEMPLO 6.1.4 *Posición de un modelo Normal* Suponga que $\{x_1, \ldots, x_n\}$ es una muestra observada independiente e idénticamente distribuida (i.i.d.) de una distribución $N(\theta, \sigma_0^2)$, donde $\theta \in \Omega = R^1$ es desconocida y $\sigma_0^2 > 0$ es conocida.

1. Tenga presente, sin embargo, que siempre que tengamos una situación en la que $f_{\theta_1}(s) = f_{\theta_2}(s)$, todavía puede ocurrir que $P_{\theta_1}(V) > P_{\theta_2}(V)$ para cualquier V que contenga a s y sea lo suficientemente pequeño. Esto implica que θ_1 estará más respaldado que θ_2 en lugar de que ambos estén respaldados por igual, como implicaría simplemente la verosimilitud. Este hecho no ocurre en los ejemplos que vamos a exponer y, por lo tanto, lo ignoraremos.