

Convergencia de funciones evaluadas en estimadores

y su aplicación al estadístico de razón de verosimilitud

Contents

1	Introducción	1
2	Un teorema de convergencia para sucesiones $g_n(\hat{\theta}_n)$	2
2.1	Antecedentes: el Teorema de la aplicación continua	2
2.2	El resultado principal	2
3	Distribución asintótica de $-2 \ln \Lambda$: Teorema de Wilks	3
3.1	Marco y notación	3
3.2	Condiciones de regularidad	4
3.3	Enunciado	4
3.4	Demostración	4
3.5	Generalización	6

1 Introducción

Un problema recurrente en estadística asintótica consiste en determinar el comportamiento límite de una sucesión de la forma $g_n(\hat{\theta}_n)$, donde $\hat{\theta}_n$ es un estimador que converge en probabilidad al verdadero parámetro θ_0 , y g_n es una sucesión de funciones que aproxima cierta función límite g . El caso más elemental, en que $g_n = g$ es fija y continua, queda resuelto por el *Teorema de la aplicación continua* (Continuous Mapping Theorem, CMT). Sin embargo, cuando la función misma varía con n — como ocurre con la segunda derivada normalizada de la log-verosimilitud — se necesita un resultado adicional que autorice la sustitución del argumento fijo θ_0 por el argumento aleatorio $\hat{\theta}_n$.

El objetivo de este documento es doble. En la Sección 2 se enuncia y demuestra el teorema general de convergencia para sucesiones $g_n(\hat{\theta}_n)$. En la Sección 3 se aplica dicho resultado para derivar la distribución asintótica del estadístico $-2 \ln \Lambda$ en el caso más

simple: hipótesis simple $H_0: \theta = \theta_0$ frente a alternativa unidimensional $H_1: \theta \neq \theta_0$. Este resultado se conoce como *Teorema de Wilks* [5].

2 Un teorema de convergencia para sucesiones $g_n(\hat{\theta}_n)$

2.1 Antecedentes: el Teorema de la aplicación continua

Recordamos el resultado clásico del que partimos.

Teorema 2.1 (Teorema de la aplicación continua [4, Teor. 2.3]). *Sea $X_n \xrightarrow{p} X$ y sea $g: \mathbb{R}^k \rightarrow \mathbb{R}^m$ continua en x para casi todo x bajo la distribución de X . Entonces $g(X_n) \xrightarrow{p} g(X)$.*

El Teorema 2.1 cubre únicamente el caso de función fija g . Cuando la función también cambia con n , la conclusión puede fallar si g_n no aproxima g con suficiente uniformidad.

2.2 El resultado principal

Teorema 2.2 (Convergencia de $g_n(\hat{\theta}_n)$). *Sea $\Theta \subseteq \mathbb{R}$ un espacio paramétrico y $\theta_0 \in \Theta$. Supóngase que:*

$$(C1) \quad \hat{\theta}_n \xrightarrow{p} \theta_0.$$

(C2) *Existe un entorno abierto $U \ni \theta_0$ tal que*

$$\sup_{t \in U} |g_n(t) - g(t)| \xrightarrow{n \rightarrow \infty} 0.$$

(C3) *g es continua en θ_0 .*

Entonces $g_n(\hat{\theta}_n) \xrightarrow{p} g(\theta_0)$.

Observación 2.3. La condición (C3) suele ser consecuencia de (C2): si las g_n son equicontinuas en θ_0 y convergen puntualmente allí, su límite g hereda la continuidad. En la práctica, (C2) se establece mediante una *ley uniforme de los grandes números* (LUGN); véase van der Vaart [4, Cap. 5].

Proof. Descomponemos el error mediante la desigualdad triangular:

$$|g_n(\hat{\theta}_n) - g(\theta_0)| \leq \underbrace{|g_n(\hat{\theta}_n) - g(\hat{\theta}_n)|}_{(I)} + \underbrace{|g(\hat{\theta}_n) - g(\theta_0)|}_{(II)}.$$

Sea $\varepsilon > 0$ arbitrario.

Control del término (II). Por (C1), $\hat{\theta}_n \xrightarrow{p} \theta_0$, y por (C3), g es continua en θ_0 . El Teorema 2.1 garantiza que $g(\hat{\theta}_n) \xrightarrow{p} g(\theta_0)$, luego para n suficientemente grande

$$\Pr\left(|g(\hat{\theta}_n) - g(\theta_0)| > \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{2}.$$

Control del término (I). Por (C2), existe N_1 tal que para todo $n \geq N_1$, $\sup_{t \in U} |g_n(t) - g(t)| < \varepsilon/2$. Por (C1), existe N_2 tal que para todo $n \geq N_2$, $\Pr(\hat{\theta}_n \notin U) < \varepsilon/2$. En el evento $\{\hat{\theta}_n \in U\}$ se tiene

$$|g_n(\hat{\theta}_n) - g(\hat{\theta}_n)| \leq \sup_{t \in U} |g_n(t) - g(t)| < \frac{\varepsilon}{2},$$

de modo que, para $n \geq \max(N_1, N_2)$,

$$\Pr\left(|g_n(\hat{\theta}_n) - g(\hat{\theta}_n)| > \frac{\varepsilon}{2}\right) \leq \Pr(\hat{\theta}_n \notin U) < \frac{\varepsilon}{2}.$$

Conclusión. Por la subaditividad de la probabilidad,

$$\Pr\left(|g_n(\hat{\theta}_n) - g(\theta_0)| > \varepsilon\right) \leq \Pr\left(|g_n(\hat{\theta}_n) - g(\hat{\theta}_n)| > \frac{\varepsilon}{2}\right) + \Pr\left(|g(\hat{\theta}_n) - g(\theta_0)| > \frac{\varepsilon}{2}\right) < \varepsilon,$$

para n suficientemente grande. Como ε es arbitrario, $g_n(\hat{\theta}_n) \xrightarrow{p} g(\theta_0)$. \square

3 Distribución asintótica de $-2 \ln \Lambda$: Teorema de Wilks

3.1 Marco y notación

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con densidad $f(\cdot; \theta)$, donde $\theta \in \Theta \subseteq \mathbb{R}$. Definimos la log-verosimilitud

$$\ell(\theta) = \sum_{i=1}^n \ln f(X_i; \theta),$$

y el *estadístico de razón de verosimilitud* para el contraste $H_0: \theta = \theta_0$ frente a $H_1: \theta \neq \theta_0$ es

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta}_n)},$$

donde $\hat{\theta}_n$ denota el estimador de máxima verosimilitud (EMV) y $L(\theta) = e^{\ell(\theta)}$. El estadístico de la prueba es

$$W_n = -2 \ln \Lambda = 2[\ell(\hat{\theta}_n) - \ell(\theta_0)].$$

La información de Fisher se define como

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \ln f(X; \theta) \right)^2 \right] = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta) \right].$$

3.2 Condiciones de regularidad

Adoptamos las condiciones de regularidad estándar [3, Cap. 9]:

1. El soporte de $f(\cdot; \theta)$ no depende de θ .
2. $\ell(\theta)$ es tres veces diferenciable en un entorno de θ_0 .
3. $\mathcal{I}(\theta_0)$ existe, es finita y estrictamente positiva.
4. Las derivadas hasta tercer orden de $\ln f(X; \theta)$ son dominadas por una función integrable uniformemente en un entorno de θ_0 (condición de dominación de Cramér).

3.3 Enunciado

Teorema 3.1 (Wilks, 1938 [5]). *Bajo H_0 y las condiciones de regularidad anteriores,*

$$W_n = -2 \ln \Lambda \xrightarrow{d} \chi_1^2.$$

3.4 Demostración

La prueba se estructura en cuatro pasos.

Paso 1: Expansión de Taylor.

Expandimos $\ell(\theta_0)$ en torno al EMV $\hat{\theta}_n$:

$$\ell(\theta_0) = \ell(\hat{\theta}_n) + \ell'(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n) + \frac{1}{2} \ell''(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n)^2 + \frac{1}{6} \ell'''(\theta_n^*)(\theta_0 - \hat{\theta}_n)^3,$$

donde θ_n^* está entre θ_0 y $\hat{\theta}_n$. Como el EMV satisface la ecuación de puntuación $\ell'(\hat{\theta}_n) = 0$, obtenemos

$$W_n = 2[\ell(\hat{\theta}_n) - \ell(\theta_0)] = -\ell''(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 - \frac{1}{3} \ell'''(\theta_n^*)(\hat{\theta}_n - \theta_0)^3.$$

El tercer sumando es $o_p(1)$ bajo las condiciones de dominación [3, Lem. 9.14], de modo que

$$W_n = -\ell''(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 + o_p(1). \quad (1)$$

Paso 2: Factorización.

Multiplicamos y dividimos por $n\mathcal{I}(\theta_0)$ para aislar límites conocidos:

$$W_n = \underbrace{\frac{-\ell''(\hat{\theta}_n)}{n}}_{g_n(\hat{\theta}_n)} \times \underbrace{n(\hat{\theta}_n - \theta_0)^2}_{T_n} + o_p(1). \quad (2)$$

Paso 3: Convergencia en probabilidad de $g_n(\hat{\theta}_n)$.

Definimos $g_n(\theta) = -n^{-1}\ell''(\theta) = -n^{-1}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} \ln f(X_i; \theta)$.

Lema 3.2. $g_n(\hat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta_0)$.

Proof. Verificamos las tres condiciones del Teorema 2.2:

(C1) $\hat{\theta}_n \xrightarrow{p} \theta_0$: consistencia estándar del EMV bajo las condiciones de regularidad [4, Teor. 5.1].

(C2) Por la LUGN de Glivenko–Cantelli [4, Teor. 19.4], bajo la condición de dominación de Cramér,

$$\sup_{t \in U} \left| g_n(t) - \left(-\mathbb{E}_{\theta_0} \left[\frac{\partial^2}{\partial\theta^2} \ln f(X; t) \right] \right) \right| \xrightarrow{p} 0$$

para algún entorno U de θ_0 .

(C3) La aplicación $\theta \mapsto -\mathbb{E}_{\theta_0} \left[\frac{\partial^2}{\partial\theta^2} \ln f(X; \theta) \right]$ es continua en θ_0 por el teorema de convergencia dominada, y su valor en θ_0 es precisamente $\mathcal{I}(\theta_0)$ por la identidad de la información [3, Lem. 7.3].

El Teorema 2.2 concluye que $g_n(\hat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta_0)$. □

Este paso es el que hace *indispensable* el Teorema 2.2: no es suficiente aplicar la LUGN en el punto fijo θ_0 y argumentar por aproximación, ya que $\hat{\theta}_n$ es aleatorio y se requiere controlar g_n en un entorno de θ_0 de forma uniforme.

Paso 4: Distribución límite de T_n y conclusión.

Por el Teorema Central del Límite, la puntuación normalizada satisface

$$\frac{1}{\sqrt{n}} \ell'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_\theta \ln f(X_i; \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)).$$

La expansión de Taylor de la puntuación en torno a θ_0 , junto con la LUGN aplicada a ℓ'' , permite mostrar que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1}) \quad [4, \text{Teor. 5.39}],$$

de donde

$$T_n = n(\hat{\theta}_n - \theta_0)^2 \xrightarrow{d} \mathcal{I}(\theta_0)^{-1} \chi_1^2.$$

Combinando el Lema 3.2 con la convergencia de T_n mediante el *Teorema de Slutsky* [1, Teor. 2.8]:

$$W_n = g_n(\hat{\theta}_n) \cdot T_n + o_p(1) \xrightarrow{d} \mathcal{I}(\theta_0) \cdot \mathcal{I}(\theta_0)^{-1} \chi_1^2 = \chi_1^2,$$

lo que completa la demostración del Teorema 3.1. □

3.5 Generalización

El Teorema 3.1 se extiende al caso d -dimensional: si $\theta \in \mathbb{R}^d$ y se contrasta $H_0: \theta \in \Theta_0$ (un subespacio de dimensión $d-k$) frente a H_1 , entonces $-2 \ln \Lambda \xrightarrow{d} \chi_k^2$ bajo H_0 [3, Teor. 16.1]. La prueba sigue la misma estructura pero requiere álgebra matricial y la inversión de la información de Fisher multidimensional.

Conclusión

El Teorema 2.2 actúa como el eslabón técnico que hace posible el Teorema de Wilks. La dificultad esencial es que la segunda derivada de la log-verosimilitud debe evaluarse en el EMV, que es aleatorio; la convergencia uniforme de g_n cerca de θ_0 es lo que permite trasladar la convergencia puntual en θ_0 (que proviene de la ley de los grandes números) a una convergencia en el punto aleatorio $\hat{\theta}_n$ (que es lo que efectivamente aparece en el estadístico). Sin este argumento el paso queda como una brecha en la demostración.

References

- [1] Billingsley, P. (1999). *Convergence of Probability Measures*, 2.^a ed. Wiley, Nueva York.
- [2] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- [3] Lehmann, E. L. y Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3.^a ed. Springer, Nueva York.
- [4] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- [5] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.