INTRODUCCIÓN A LA CONVERGENCIA DE SUCESIONES DE VARIABLES ALEATORIAS

Julián de la Horra
Departamento de Matemáticas U.A.M.

1 Introducción

Se pueden utilizar diferentes conceptos de convergencia para las sucesiones de variables aleatorias: convergencia casi segura, en probabilidad, en media cuadrática o de orden p, en distribución. Cada uno de estos conceptos de convergencia pone el énfasis en un aspecto diferente, pero aquí no vamos a ver con detalle las definiciones de estos conceptos, ni las relaciones entre ellos. En este capítulo, nos vamos a limitar a introducir un par de ellos, para ver las consecuencias que podemos obtener en dos aspectos: aproximaciones útiles de la distribución de algunas variables aleatorias, y aplicaciones a la estimación de parámetros de interés en la Inferencia Estadística.

2 Algunos conceptos de convergencia de sucesiones de variables aleatorias

Definición.- Una sucesión de variables aleatorias $T_1, ..., T_n, ...$ *converge en probabilidad* a una variable aleatoria T cuando:

$$\text{Para todo } \varepsilon > 0 : \qquad \lim_n P(|T_n - T| \geq \varepsilon) = 0$$

En particular, una sucesión de variables aleatorias $T_1, ..., T_n, ...$ *converge en probabilidad* a una constante c cuando:

$$\text{Para todo } \varepsilon > 0 : \qquad \lim_n P(|T_n - c| \geq \varepsilon) = 0 \qquad \bullet$$

Definición.- Una sucesión de variables aleatorias $T_1, ..., T_n, ...$ *converge en distribución* a una variable aleatoria T cuando:

$$\lim_n F_{T_n}(x) = F_T(x)$$

en todos los puntos x en los que F_T es continua.

En particular, una sucesión de variables aleatorias $T_1, ..., T_n, ...$ *converge en distribución* a una constante c (es decir, a una distribución degenerada que concentra toda la probabilidad en c) cuando:

$$\begin{aligned} &\text{Para todo } x < c : \ \lim_n F_{T_n}(x) = 0 \\ &\text{Para todo } x > c : \ \lim_n F_{T_n}(x) = 1 \qquad \bullet \end{aligned}$$

La convergencia en probabilidad es más fuerte (siempre) que la convergencia en distribución, pero ambas son equivalentes cuando hablamos de convergencia a una constante. Probamos estos resultados a continuación:

Teorema.- Si $T_1, ..., T_n, ...$ converge en probabilidad a T, entonces $T_1, ..., T_n, ...$ converge en distribución a T.

Demostración.- Fijamos x, punto de continuidad de F_T, y fijamos $\varepsilon > 0$.
Por un lado:

$$\begin{aligned} F_{T_n}(x) &= P(T_n \leq x) \leq P(T \leq x + \varepsilon) + P(|T_n - T| \geq \varepsilon) \\ &= F_T(x + \varepsilon) + P(|T_n - T| \geq \varepsilon) \end{aligned}$$

Por otro lado:

$$\begin{aligned} F_T(x - \varepsilon) &= P(T \leq x - \varepsilon) \leq P(T_n \leq x) + P(|T_n - T| \geq \varepsilon) \\ &= F_{T_n}(x) + P(|T_n - T| \geq \varepsilon) \end{aligned}$$

Despejando $F_{T_n}(x)$ en la segunda desigualdad, y enlazando los resultados, tenemos:

$$F_T(x - \varepsilon) - P(|T_n - T| \geq \varepsilon) \leq F_{T_n}(x) \leq F_T(x + \varepsilon) + P(|T_n - T| \geq \varepsilon)$$

Tomando límites en n, y aplicando la convergencia en probabilidad:

$$F_T(x - \varepsilon) \leq \liminf_n F_{T_n}(x) \leq \limsup_n F_{T_n}(x) \leq F_T(x + \varepsilon)$$

Haciendo ahora que ε tienda a cero, y aplicando que x es punto de continuidad de $F_T(x)$, tenemos:

$$\lim_n F_{T_n}(x) = F_T(x) \qquad \bullet$$

Teorema.- Si $T_1, ..., T_n, ...$ converge en distribución a c, entonces $T_1, ..., T_n, ...$ converge en probabilidad a c.

Demostración.- Fijamos $\varepsilon > 0$, y tenemos:

$$\begin{aligned} \lim_n P(|T_n - c| \geq \varepsilon) &= \lim_n [P(T_n \leq c - \varepsilon) + P(T_n \geq c + \varepsilon)] \\ &= \lim_n [F_{T_n}(c - \varepsilon) + 1 - P(T_n < c + \varepsilon)] \\ &\leq \lim_n [F_{T_n}(c - \varepsilon) + 1 - P(T_n \leq c + \varepsilon/2)] \\ &= \lim_n [F_{T_n}(c - \varepsilon) + 1 - F_{T_n}(c + \varepsilon/2)] \\ &= 0 + 1 - 1 = 0 \qquad \bullet \end{aligned}$$

3 Aproximaciones útiles de la distribución de algunas variables aleatorias

En esta sección, vamos a dar (sin demostración) un resultado muy útil para obtener aproximaciones de distribuciones. La demostración de este resultado se obtiene a partir del concepto de función característica, y a partir de las relaciones que hay entre funciones de distribución y funciones características.

Teorema (Central limit theorem).- Sean $X_1, ..., X_n, ...$ variables aleatorias independientes e idénticamente distribuidas, con $E[X_i] = \mu$ y $V(X_i) = \sigma^2$. Entonces:

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \text{ converge en distribución a la } N(0;1) \qquad \bullet$$

También se puede expresar de la siguiente forma equivalente (dividiendo numerador y denominador por n):

$$\frac{\frac{\sum_{i=1}^{n} X_i}{n} - \mu}{\sigma/\sqrt{n}} \text{ converge en distribución a la } N(0;1)$$

Aplicación práctica:

Sabíamos de resultados anteriores que cuando consideramos variables aleatorias, $X_1, ..., X_n$, independientes e idénticamente distribuidas, con distribución $N(\mu, \sigma^2)$, tenemos:

$$\sum_{i=1}^{n} X_i \sim N(n\mu; n\sigma^2)$$

Lo que nos permite el resultado anterior es utilizar la siguiente aproximación:

Cuando consideramos un número n grande (digamos, $n \geq 30$) de variables aleatorias, $X_1, ..., X_n$, independientes e idénticamente distribuidas, con $E[X_i] = \mu$ y $V(X_i) = \sigma^2$:

$$\sum_{i=1}^{n} X_i \sim N(n\mu; n\sigma^2) \qquad \text{(aproximadamente)}$$

Es decir, eliminamos la necesidad de que las X_i se distribuyan según una Normal pero, a cambio, necesitamos que el número de variables aleatorias que sumamos sea grande.

Otra resultado muy interesante es el siguiente:

Corolario (Teorema de De Moivre).- Sea $T_1, ..., T_n, ...$ una sucesión de variables aleatorias, donde T_n tiene una distribución $Bin(n; p)$. Entonces:

$$\frac{T_n - np}{\sqrt{np(1-p)}} \text{ converge en distribución a la } N(0; 1)$$

Demostración.- Basta con definir una sucesión de variables aleatorias $X_1, ..., X_n, ...$ independientes y todas ellas con distribución Bernoulli (p), con lo que $E[X_i] = p$ y $V(X_i) = p(1-p)$. Ahora, cada T_n se puede considerar como $T_n = \sum_{i=1}^n X_i$, y no queda nada más que aplicar el teorema anterior. •

Aplicación práctica:
Cuando estamos trabajando con una variable aleatoria T_n con distribución $Bin(n; p)$, y n es grande (digamos, $n \geq 30$), podemos recurrir a las siguientes aproximaciones para la distribución de T_n:

(a) Cuando p es próximo a cero (digamos $p \leq 0,10$), podemos aproximar el modelo $Bin(n; p)$ mediante el modelo de Poisson($\lambda = np$), como ya se explicó al presentar el modelo de Poisson.

(b) Cuando p no es próximo ni a cero ni a uno (digamos $0,10 \leq p \leq 0,90$), podemos aproximar el modelo $Bin(n; p)$ mediante el modelo $N(\mu = np; \sigma^2 = np(1-p))$, aplicando el teorema de De Moivre.

(c) Cuando p es próximo a uno (digamos $p \geq 0,90$), cambiamos éxitos por fracasos, y estaremos en el caso (a).

4 Aplicaciones a la estimación de parámetros en la inferencia estadística paramétrica

En esta Sección, vamos a ver tres resultados sencillos que nos aseguran la convergencia en probabilidad bajo ciertas condiciones:

Teorema (Ley débil de los grandes números).- Sean $X_1, ..., X_n, ...$ variables aleatorias independientes e idénticamente distribuidas, con $E[X_i] = \mu$ y $V(X_i) = \sigma^2$. Entonces, la sucesión de variables aleatorias $T_n = \frac{\sum_{i=1}^n X_i}{n}$ converge en probabilidad a μ.

Demostración.- En primer lugar, tenemos:

$$E[T_n] = E\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^n X_i\right] = \frac{n\mu}{n} = \mu$$

$$V[T_n] = V\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n^2}V\left[\sum_{i=1}^n X_i\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

A continuación, fijamos $\varepsilon > 0$, y aplicamos la desigualdad de Chebychev:

$$\begin{aligned} P(|T_n - \mu| \geq \varepsilon) &= P((T_n - \mu)^2 \geq \varepsilon^2) \leq \frac{E[(T_n - \mu)^2]}{\varepsilon^2} \\ &= \frac{V(T_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \end{aligned}$$

Tomando límites:

$$\lim_n P(|T_n - \mu| \geq \varepsilon) \leq \lim_n \frac{\sigma^2}{n\varepsilon^2} = 0 \qquad \bullet$$

Teorema.- Sea $T_1, ..., T_n, ...$ una sucesión de variables aleatorias tales que:

$$\lim_n E[T_n] = \theta \qquad y \qquad \lim_n V(T_n) = 0$$

Entonces, T_n converge en probabilidad a θ.

Demostración.- Fijamos $\varepsilon > 0$, y aplicamos la desigualdad de Chebychev:

$$\begin{aligned} P(|T_n - \theta| \geq \varepsilon) &= P((T_n - \theta)^2 \geq \varepsilon^2) \leq \frac{E[(T_n - \theta)^2]}{\varepsilon^2} \\ &= \frac{E[(T_n - E[T_n] + E[T_n] - \theta)^2]}{\varepsilon^2} \\ &= \frac{E[(T_n - E[T_n])^2] + (E[T_n] - \theta)^2}{\varepsilon^2} \\ &= \frac{V(T_n) + (E[T_n] - \theta)^2}{\varepsilon^2} \end{aligned}$$

Tomando límites:

$$\lim_n P(|T_n - \theta| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \lim_n [V(T_n) + (E[T_n] - \theta)^2] = 0 \qquad \bullet$$

Teorema.- Sea $T_1, ..., T_n, ...$ una sucesión de variables aleatorias tales que:

$$\begin{array}{ll} \text{Para todo } x < \theta : & \lim_n F_{T_n}(x) = 0 \\ \text{Para todo } x > \theta : & \lim_n F_{T_n}(x) = 1 \end{array}$$

Entonces, T_n converge en probabilidad a θ.

Demostración.- Es consecuencia inmediata de la equivalencia entre las convergencias en distribución y en probabilidad a una constante. •

Estos tres teoremas tienen mucho interés por sí mismos pero, además, tienen mucho interés en la Inferencia Estadística Paramétrica que tiene el siguiente planteamiento:

Se desea estudiar una característica numérica en una población. Esta característica numérica se formaliza mediante una variable aleatoria X, con un modelo de probabilidad $P_\theta(x)$, donde el valor del parámetro θ es desconocido.

Para conseguir información sobre X en toda la población, se obtienen observaciones experimentales $X_1, ..., X_n, ...$ de dicha característica X, las cuales, formalmente, serán consideradas como variables aleatorias independientes y con la misma distribución que la variable aleatoria X. Abreviadamente, decimos que el vector aleatorio $(X_1, ..., X_n)$ constituye una muestra aleatoria de tamaño n de X. Uno de los objetivos primordiales es estimar (aproximar lo mejor posible) el valor desconocido del parámetro θ, mediante alguna función de la muestra, $T_n = T(X_1, ..., X_n)$, que suele recibir el nombre de estimador.

En estas condiciones, resulta interesante poder asegurar que una sucesión de estimadores T_n converge en probabilidad al parámetro θ que se desea estimar. Los tres teoremas anteriores proporcionan herramientas asequibles para estudiar si una sucesión de estimadores $T_1, ..., T_n, ...$ converge en probabilidad al parámetro θ, o utilizando la terminología estadística, para estudiar si un estimador T_n es *consistente* para estimar el parámetro θ.