Probabilidad II
Tema 5: Convergencias estocásticas y leyes de los grandes números

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Estructura de este tema

- Definiciones: convergencias casi segura, en probabilidad, en media y en distribución.
- Relaciones entre los distintos tipos de convergencia.
- Leyes de los grandes números:
  - LDGN de Chebychev y Khintchine.
  - LFGN de Cantelli y de Kolmogorov.

Convergencias casi segura, en prob. y en media-r

Sean $X, X_1, X_2, \ldots$ variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$. Entonces:

Convergencia casi segura:

$$X_n \xrightarrow{\text{c.s.}} X \quad \text{si} \quad \mathrm{P}(X_n \to X) = 1.$$

Convergencia en probabilidad:

$$X_n \xrightarrow{\mathrm{P}} X \quad \text{si} \quad \forall \epsilon > 0,\ \mathrm{P}(|X_n - X| \geq \epsilon) \to 0.$$

Convergencia en media de orden r:

$$X_n \xrightarrow{\text{m-}r} X \quad \text{si} \quad \mathrm{E}|X_n - X|^r \to 0 \qquad (r > 0).$$

Convergencia en distribución

Sean $X, X_1, X_2, \ldots$ variables aleatorias *no necesariamente definidas sobre el mismo espacio de probabilidad* y sean F y F_n las funciones de distribución de X y X_n, respectivamente. Entonces:

$$X_n \xrightarrow{\mathrm{d}} X \quad \text{si} \quad \forall x \in \mathrm{Cont}(F),\ F_n(x) \to F(x),$$

donde $\mathrm{Cont}(F)$ es el conjunto de puntos de continuidad de F.

Pregunta: ¿Por qué no $\forall x \in \mathbb{R}$ en la definición anterior?

Observaciones

- La definición de convergencia c.s. tiene sentido.
- $X_n \to X$ si y solo si $X_n - X \to 0$ (para todas las convergencias).
- Cuando todas las variables involucradas son degeneradas, cada una de las cuatro convergencias coincide con la convergencia usual de sucesiones de números reales.
- Los límites son "únicos": si X e Y son dos variables límite, $\mathrm{P}(X = Y) = 1$ para las tres primeras convergencias, y $F_X = F_Y$ para la convergencia en distribución.

Esquema de relaciones entre las convergencias

$$
\begin{tikzcd}
 & X_n \xrightarrow{\text{c.s.}} X \arrow[dr, Rightarrow] & & \\
 & & X_n \xrightarrow{\text{P}} X \arrow[r, Rightarrow] & X_n \xrightarrow{\text{d}} X \\
X_n \xrightarrow{\text{m-}q} X \, (q > p) \arrow[r, Rightarrow] & X_n \xrightarrow{\text{m-}p} X \arrow[ur, Rightarrow] & &
\end{tikzcd}
$$

Nnguna de las implicaciones recíprocas es cierta en general.

Convergencia c.s. y convergencia en probabilidad

Lema: Son equivalentes:

(a) $X_n \xrightarrow{\text{c.s.}} X$.

(b) Para todo $\epsilon > 0$, $\mathrm{P}\left(\limsup\{|X_k - X| \geq \epsilon\}\right) = 0$.

(c) Para todo $\epsilon > 0$, $\lim_{n\to\infty} \mathrm{P}\big(\sup_{k\geq n} |X_k - X| \geq \epsilon\big) = 0$.

Corolario: Se tiene:

$$\sum_{n\geq 1} \mathrm{P}(|X_n - X| \geq \epsilon) < \infty, \quad \epsilon > 0 \quad \Rightarrow \quad X_n \xrightarrow{\text{c.s.}} X.$$

Cuando la serie anterior converge, se dice que hay convergencia completa de X_n a X. ¿Bajo qué condiciones es la convergencia completa equivalente a la convergencia c.s.?

Corolario: Si $X_n \xrightarrow{\text{c.s.}} X$, entonces $X_n \xrightarrow{\mathrm{P}} X$.
(El recíproco no es cierto en general.)

Convergencia c.s. y convergencia en probabilidad

Proposición
X_n converge en probabilidad a X si y solo si toda subsucesión de X_n tiene a su vez una subsucesión que converge a X c.s.

Corolario
Si $X_n \xrightarrow{\text{P}} X$ y $g : \mathbb{R} \to \mathbb{R}$ es continua, entonces $g(X_n) \xrightarrow{\text{P}} g(X)$.

(Se tiene en cuenta que si $X_n \xrightarrow{\text{c.s.}} X$ y $g : \mathbb{R} \to \mathbb{R}$ es continua, entonces $g(X_n) \xrightarrow{\text{c.s.}} g(X)$).

Convergencia en media

Proposición: Si $q > r > 0$ y $X_n \xrightarrow{\text{m-}q} X$, entonces $X_n \xrightarrow{\text{m-}r} X$. (Lyapunov)

Proposición: Si $X_n \xrightarrow{\text{m-}r} X$ para algún $r > 0$, entonces $X_n \xrightarrow{\text{P}} X$. (Chebychev)

Proposición: Si $X_n \xrightarrow{\text{m-}r} X$ para $r \geq 1$, entonces $\text{E}|X_n|^r \to \text{E}|X|^r$. (Minkowski)

Observaciones:

(1) La convergencia en media no implica la convergencia casi segura.

(2) Ni la convergencia en probabilidad ni la casi segura implican la convergencia en media.

Convergencia en probabilidad y en distribución

Proposición: Si $X_n \xrightarrow{\text{P}} X$, entonces $X_n \xrightarrow{\text{d}} X$.

Observación: En general, $X_n \xrightarrow{\text{d}} X \not\Rightarrow X_n \xrightarrow{\text{P}} X$. Sin embargo hay un caso particular importante en el que sí se da la implicación.

Proposición: Si $X_n \xrightarrow{\text{d}} a$ (constante), entonces $X_n \xrightarrow{\text{P}} a$.

Leyes fuertes y débiles de los grandes números

Las llamadas **leyes de los grandes números** son teoremas de convergencia que tienen la forma siguiente:

Sea (X_n) una sucesión de v.a. que cumplen las hipótesis $H_1, H_2, \ldots$. Entonces,

$$\frac{S_n - a_n}{b_n} \xrightarrow{?} 0.$$

- $S_n = \sum_{i=1}^{n} X_i$.
- $\xrightarrow{?}$ indica uno de los modos de convergencia anteriores.
- La ley se llama **fuerte** cuando la convergencia es casi segura, y **débil** cuando la convergencia es en probabilidad o en distribución. (Toda ley fuerte implica la correspondiente ley débil.)

Leyes fuertes y débiles de los grandes números

Sea (X_n) una sucesión de v.a. que cumplen las hipótesis $H_1, H_2, \ldots$. Entonces,

$$\frac{S_n - a_n}{b_n} \xrightarrow{?} 0.$$

Las hipótesis H_i, suelen ser de varios tipos:

(a) Sobre los momentos. Es decir si las variables son integrables o de cuadrado integrable, etc. Cuanto menor sea la exigencia de integrabilidad mayor generalidad tendrá el teorema (o la ley).

(b) Sobre la dependencia de las variables X_i. Es decir, si se trata de variables mutuamente independientes, o basta con que sean independientes dos a dos, o incorreladas, etc.

(c) Sobre la distribución de probabilidad de las X_i. Es decir si las variables tienen o no la misma distribución, etc.

(d) Sobre los requisitos que tienen que cumplir las sucesiones a_n y b_n.

LDGN de Chebychev

Proposición: Sean $X_1, X_2, \dots$ v.a. idénticamente distribuidas e independientes dos a dos, con varianza finita. Sea $S_n := X_1 + \cdots + X_n$. Entonces:

$$\bar{X}_n := \frac{S_n}{n} \xrightarrow{\mathrm{P}} \mu, \quad \text{donde} \quad \mu \in \mathbb{R}.$$

Observaciones:

(1) La misma demostración no vale para demostrar la convergencia casi segura.

(2) ¿Es posible eliminar la hipótesis de que la varianza sea finita?

(3) Generalización: sean $X_1, X_2, \dots$ v.a. tales que $\mathrm{Var}(X_n) \leq K$, para todo $n \geq 1$, y $\mathrm{Cov}(X_i, X_j) \leq 0$, si $i \neq j$. Entonces, si $a > 1/2$,

$$\frac{S_n - \mathrm{E}(S_n)}{n^a} \xrightarrow{\text{m-2}} 0.$$

Truncamiento

Se dice que dos sucesiones de v.a. X_n y X_n' son **equivalentes** si $\sum_{n=1}^{\infty} \mathrm{P}(X_n \neq X_n') < \infty$.

Dos sucesiones equivalentes tienen un comportamiento asintótico similar:

(1) $\sum_{n=1}^{\infty}(X_n - X_n')$ converge c.s.

(2) $\sum_{n=1}^{\infty} X_n$ converge c.s. si y solo si $\sum_{n=1}^{\infty} X_n'$ converge c.s.

(3) Si existe $a_n \to \infty$ y una v.a. X tal que $a_n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{\text{c.s.}} X$, entonces también $a_n^{-1} \sum_{i=1}^{n} X_i' \xrightarrow{\text{c.s.}} X$.

Técnica de truncamiento (aparece en muchas demostraciones):

$$X_n = X_n I_{\{|X_n| \leq c_n\}} + X_n I_{\{|X_n| > c_n\}} := X_n' + X_n''.$$

Si $\sum_{n=1}^{\infty} \mathrm{P}(|X_n| > c_n) < \infty$, entonces X_n y X_n' son equivalentes. Trabajar con X_n' es más fácil porque es acotada y tiene momentos de todos los órdenes.

LDGN de Khintchine

Proposición: Sean $X_1, X_2, \ldots$ v.a. idénticamente distribuidas e independientes dos a dos, con media finita μ. Sea $S_n := X_1 + \cdots + X_n$. Entonces:

$$\bar{X}_n := \frac{S_n}{n} \xrightarrow{\mathrm{P}} \mu.$$

Observaciones:

(1) Vamos a demostrar $\mathrm{E}|\bar{X}_n - \mu| \to 0$, que es un poco más fuerte.

(2) Basta demostrarlo para el caso $\mu = 0$.

(3) Se usa la técnica de truncamiento:

$$X_n = X_n I_{\{|X_n| \leq \sqrt{n}\}} + X_n I_{\{|X_n| > \sqrt{n}\}} := Y_n + Z_n.$$

LFGN de Cantelli

Proposición: Sean $X_1, X_2, \dots$ v.a. independientes, con $\mathrm{E}(X_i^4) < \infty$. Supongamos que $\mathrm{E}|X_n - \mathrm{E}(X_n)|^4 \leq C$ para todo $n \geq 1$. Sea $S_n := X_1 + \cdots + X_n$. Entonces:

$$\frac{S_n - \mathrm{E}(S_n)}{n} \xrightarrow{\text{c.s.}} 0.$$

Observaciones:

(1) Basta demostrarlo para el caso $\mathrm{E}(X_n) = 0$, $n \geq 1$.

(2) Por el primer lema de Borel-Cantelli, basta demostrar

$$\sum_{n=1}^{\infty} \mathrm{P}\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \infty, \quad \text{para todo } \epsilon > 0.$$

(3) Se acota cada término mediante un uso astuto de la desigualdad de Markov.

(4) Si $|X_n| \leq K$ para todo $n \geq 1$ se cumplen las hipótesis del teorema. En este caso se puede también usar la desigualdad de Hoeffding.

LFGN de Kolmogorov (caso i.i.d.)

Proposición: Sean $X_1, X_2, \ldots$ v.a. independientes e idénticamente distribuidas, con $\mathrm{E}|X_1| < \infty$. Sea $S_n := X_1 + \cdots + X_n$. Entonces:

$$\bar{X}_n := \frac{S_n}{n} \xrightarrow{\text{c.s.}} \mu, \quad \text{donde } \mu = \mathrm{E}(X_1).$$

Lema (desigualdad maximal): Sean $X_1, X_2, \ldots$ v.a. independientes e idénticamente distribuidas, con $\mathrm{E}(X_1) = 0$. Sea $S_n := X_1 + \cdots + X_n$. Entonces, para todo $\epsilon > 0$,

$$\mathrm{P}\left(\sup_{1 \leq n < \infty} \frac{S_n}{n} \geq \epsilon \right) \leq \frac{\mathrm{E}|X_1|}{\epsilon}.$$

Observación: Si $\bar{X}_n \xrightarrow{\text{c.s.}} c$ para algún $c \in \mathbb{R}$, entonces $\mu = \mathrm{E}(X_1)$ es finita y $c = \mu$.

LFGN de Kolmogorov (caso de varianza finita)

Proposición: Sean $X_1, X_2, \ldots$ v.a. independientes con varianza finita y sea $0 < b_n \uparrow \infty$. Entonces, si $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)/b_n^2 < \infty$,

$$\frac{S_n - \mathrm{E}(S_n)}{b_n} \xrightarrow{\text{c.s.}} 0.$$

La demostración está basada en el siguiente lema:

Lema (desigualdad maximal de Kolmogorov): Sean $X_1, X_2, \ldots$ v.a. independientes, con esperanza finita. Para todo $\epsilon > 0$,

$$\mathrm{P}\left(\max_{1 \leq j \leq n} |S_j - \mathrm{E}(S_j)| \geq \epsilon\right) \leq \frac{\mathrm{Var}(S_n)}{\epsilon^2}.$$