Teoría asintótica

Gabriel V. Montes-Rojas

Convergencia de una secuencia

¿Cuán cerca es cerca?

Convergencia de una secuencia real: *La secuencia real* $\{a_N : N = 1, 2, ...\}$ *(ej.* $a_1, a_2, a_3 ...$*) tiene el número real a como su límite, o* **converge** *a a, si para cada número positivo* ϵ*, no importa cuan pequeño, hay un número entero positivo* N_ϵ *tal que para todo* $N > N_\epsilon$, $|a_N - a| < \epsilon$*. Escribimos* $a_N \to a$ *cuando* $N \to \infty$.

Estas definiciones se aplican a vectores y matrices elemento por elemento.

La secuencia real $\{a_N : N = 1, 2, ...\}$ *es* **acotada** *si y sólo si existe* $b < \infty$ *tal que* $|a_N| < b$ *para todo* $N = 1, 2, ...$*. De lo contrario decimos que es* **no acotada***.*

Convergencia en probabilidad (y con probabilidad 1)

Nuestro interés está en secuencias de variables aleatorias para las cuales necesitamos estudiar su comportamiento asintótico y en probabilidad.

Convergencia en probabilidad: *La secuencia* $\{x_N : N = 1, 2, ...\}$ *de variables aleatorias* **converge en probabilidad** *a la constante* a *si para todo* $\epsilon > 0$,

$$P[|x_N - a| > \epsilon] \to 0 \text{ cuando } N \to \infty.$$

En este caso a *se llama el límite en probabilidad (en inglés limit in probability o probability limit, plim) de la secuencia* $\{x_N\}$. *Escribimos*

$$\underset{n\to\infty}{plim}\, x_N = a \text{ o } plim\, x_N = a \text{ o } x_N \xrightarrow{p} a$$

Hay una versión fuerte de este tipo de convergencia, **convergencia con probabilidad 1** (en inglés **converge almost surely** (a.s.)), para lo cual requerimos que $P\left[\lim_{N\to\infty} x_N = a\right] = 1$. Convergencia con probabilidad 1 implica convergencia en probabilidad.

Convergencia en probabilidad (y con probabilidad 1)

Estas definiciones también se aplican elemento por elemento a vectores y matrices.

Si $\boldsymbol{x}_N \xrightarrow{p} \boldsymbol{a}$*, donde* $\boldsymbol{a}$ *es un vector* $K \times 1$*, esto es equivalente a usar la norma euclidiana* $\|\boldsymbol{x}_N - \boldsymbol{a}\| \xrightarrow{p} 0$*.*

Si $\boldsymbol{Z}_N \xrightarrow{p} \boldsymbol{B}$*, donde* $\boldsymbol{B}$ *es una matriz y ambos son de dimensión* $M \times K$*, esto es equivalente a usar la norma euclidiana* $\|\boldsymbol{Z}_N - \boldsymbol{B}\| \xrightarrow{p} 0$*, donde* $\|\boldsymbol{A}\| \equiv [tr(\boldsymbol{A}'\boldsymbol{A})]^{1/2}$ *y* $tr(\cdot)$ *es la traza de* $\cdot$*.*

Desigualdad de Chebyshev

Supongamos que X es una variable aleatoria y $E(X^2) < \infty$, entonces

$$P[|X - \mu_X| \geq r\sigma_X] = P[(X - \mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2}$$

(donde $\mu_X = E(X)$ y $\sigma_X^2 = Var(X)$)

La prueba de la desigualdad no es difícil. Supongamos que $g(X)$ es una función no negativa de la variable aleatoria X con dominio en los reales; entonces para cualquier k

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}$$

Prueba:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx =$$

$$\int_{x:g(x)\geq k} g(x)f_X(x)dx + \int_{x:g(x)\leq k} g(x)f_X(x)dx$$

$$\geq \int_{x:g(x)\geq k} g(x)f_X(x)dx \geq \int_{x:g(x)\geq k} kf_X(x)dx$$

$$= kP[g(X) \geq k]$$

Tomemos $g(X) = (x - \mu_X)^2$ y $k = r^2\sigma_X^2$, y la desigualdad de Chebyshev se prueba.

Desigualdad de Chebyshev

Nota:

$$P[|X-\mu_X| \leq r\sigma_X] \geq 1-\frac{1}{r^2}$$

tal que

$$P[\mu_X - r\sigma_x < X < \mu_X + r\sigma_x] \geq 1-\frac{1}{r^2},$$

o sea, la probabilidad de que X caiga entre una distancia $r\sigma_X$ de μ_X es mayor o igual que $1-\frac{1}{r^2}$. Para $r=2$ obtenemos:

$$P[\mu_X - 2\sigma_x < X < \mu_X + 2\sigma_x] \geq \frac{3}{4},$$

o, para toda variable aleatoria con varianza finita, al menos 3/4 de la masa de probabilidad cae entre dos errores estándar de la media.
La desigualdad de Chebyshev nos da un límite, el cual no depende de la distribución de X, para la probabilidad de eventos en términos de la media y la varianza.

Convergencia en probabilidad: Ejemplo

Definamos una variable aleatoria y_i para tirar una moneda que toma valores 0 si cae cara y 1 si cae ceca. Entonces después de N tiros la proporción de cecas será

$$x_N = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

y consideremos la secuencia $\{x_N\}$ de variables aleatorias.
Calculemos media y varianza,

$$E(x_N) = E\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) = \frac{1}{N}\sum_{i=1}^{n} E(y_i) = \frac{1}{2},$$

$$Var(x_N) = \frac{1}{N^2}\sum_{i=1}^{N} Var(y_i) = \frac{1}{4N}.$$

La última igualdad es informativa porque nos dice que $\lim_{N\to\infty} Var(x_N) = 0$ y que su límite no es aleatorio.
Ahora apliquemos Chebyshev usando $r = 2\sqrt{N}\epsilon$ para obtener que para cualquier $\epsilon > 0$

$$P[|x_N - \frac{1}{2}| > \epsilon) < \frac{1}{4N\epsilon^2}.$$

Tomemos el límite para $N \to \infty$. Así probamos formalmente que

$$\underset{N\to\infty}{plim}\; x_N = \frac{1}{2}.$$

Convergencia en probabilidad: Ejemplo

Definamos $z_i = y_i - \frac{1}{2}$ and $w_N = \frac{1}{N}\sum_{i=1}^{N} z_i$. Esto se llama una **secuencia centrada** porque $E(z_i) = 0$. Notemos que

$$E(w_N) = E\left(\frac{1}{N}\sum_{i=1}^{N} z_i\right) = \frac{1}{N}\sum_{i=1}^{N} E(z_i) = 0$$

$$Var(w_N) = \frac{1}{N^2}\sum_{i=1}^{N} Var(y_i) = \frac{1}{4N} = Var(x_N)$$

Aplicando Chebyshev usando $r = 2\sqrt{N}\epsilon$ obtenemos que para cualquier $\epsilon > 0$

$$Pr(|w_N - 0| > \epsilon) < \frac{1}{4N\epsilon^2}.$$

Tomemos el límite para $N \to \infty$. Así probamos formalmente que

$$\underset{N\to\infty}{plim}\; w_N = 0.$$

Relaciones de orden

O-grande *La secuencia* $\{a_N\}$ *es* $O(N^\lambda)$ *(al menos de orden* N^λ*) si* $N^{-\lambda}a_N$ *es acotada. Cuando* $\lambda = 0$, $\{a_N\}$ *es acotada y escribimos* $a_N = O(1)$.

o-chica *La secuencia* $\{a_N\}$ *es* $o(N^\lambda)$ *si* $N^{-\lambda}a_N \to 0$. *Cuando* $\lambda = 0$, *escribimos* $a_N = o(1)$.

Relaciones de orden estocásticas

O_p La secuencia de variables aleatorias $\{x_N\}$ es **acotada en probabilidad** *si y sólo si para cada $\epsilon > 0$, existe $b_\epsilon < \infty$ y un número entero N_ϵ tal que*

$$P[|x_N| \geq b_\epsilon] < \epsilon \text{ para todo } N \geq N_\epsilon.$$

Escribimos $x_N = O_p(1)$.

La secuencia de variables aleatorias $\{x_N\}$ es $O_p(a_N)$, donde $\{a_N\}$ es una secuencia positiva no aleatoria si $x_N/a_N = O_p(1)$. Usamos $x_N = O_p(a_N)$.

La secuencia de variables aleatorias $\{x_N\}$ es $O_p(N^\delta)$ si $N^{-\delta}x_N = O_p(1)$.

Si $x_n \xrightarrow{p} a$, entonces $x_N = O_p(1)$.

Relaciones de orden estocásticas

o_p

La secuencia de variables aleatorias $\{x_N\}$ *es* $o_p(1)$ *si* $x_n \xrightarrow{p} 0$.

La secuencia de variables aleatorias $\{x_N\}$ *es* $o_p(a_N)$, *donde* $\{a_N\}$ *es una secuencia positiva no aleatoria si* $x_N/a_N = o_p(1)$. *Usamos* $x_N = o_p(a_N)$

La secuencia de variables aleatorias $\{x_N\}$ *es* $o_p(N^{\delta})$ *si* $N^{-\delta}x_N = o_p(1)$.

Relaciones de orden estocásticas

Algunas propiedades

Si $w_N = o_p(1)$, $x_N = o_p(1)$, $y_N = O_p(1)$ y $z_N = O_p(1)$.

- $w_N + x_N = o_p(1) + o_p(1) = o_p(1)$.
- $y_N + z_N = O_p(1) + O_p(1) = O_p(1)$.
- $y_N + x_N = O_p(1) + o_p(1) = O_p(1)$.
- $y_N z_N = O_p(1) O_p(1) = O_p(1)$.
- $w_N x_N = o_p(1) o_p(1) = o_p(1)$.
- $y_N x_N = O_p(1) o_p(1) = o_p(1)$.
- Si $z_N = O_p(N^{\delta})$ entonces $z_N = o_p(N^{\delta+\epsilon}), \epsilon > 0$.

Relaciones de orden estocásticas: Ejemplo

Para el ejemplo de las monedas:

- $x_N = O_p(1)$ porque $plim\ x_N = \frac{1}{2}$.
- $w_N = o_p(1)$ porque $plim\ w_N = 0$.
- $Var(x_N) = Var(w_N) = O(N^{-1})$ porque $Var(x_N) = \frac{1}{4}N^{-1}$.

Teorema de Slutsky

Tomemos $\mathbf{g} : \mathbb{R}^K \to \mathbb{R}^J$ *como una función continua en el punto* $\mathbf{c} \in \mathbb{R}$. *Tomemos la secuencia* $\{\mathbf{x}_N : N = 1, 2, ...\}$ *de vectores aleatorios* $K \times 1$ *tal que* $\mathbf{x}_N \xrightarrow{p} \mathbf{c}$. *Entonces* $\mathbf{g}(\mathbf{x}_N) \xrightarrow{p} \mathbf{g}(\mathbf{c})$.

Notar que $E[\mathbf{g}(\mathbf{x}_N)]$ *puede no converger a* $E[\mathbf{g}(\mathbf{c})]$ *a menos que* $\mathbf{g}$ *sea lineal (la esperanza es un operador lineal). Esta es la principal ventaja de la teoría de muestras grandes sobre las de muestras finitas.*

Ley débil de los grandes números

Ley débil de los grandes números (Khinchin): *Tomemos* $\{\boldsymbol{w}_i : i = 1, 2...\}$ *una secuencia de vectores aleatorios* $G \times 1$ *que son* **independientes** *y* **distribuidos identicamente** *(i.i.d. independent and identically distributed en inglés) tal que* $E(|w_{ig}|) < \infty, g = 1, 2, ..., G$. *Entonces una secuencia satisface la* **ley débil de los grandes números** *si:*

$$N^{-1} \sum_{i=1}^{N} \boldsymbol{w}_i \xrightarrow{p} \boldsymbol{\mu}_w,$$

donde $\boldsymbol{\mu}_w \equiv E(\boldsymbol{w}_i)$.

Ley débil de los grandes números (Chebyshev): *Tomemos* $\{\boldsymbol{w}_i : i = 1, 2...\}$ *una secuencia de de vectores aleatorios* $G \times 1$ *que son* **independientes** *con media y varianza finita (no necesariamente la misma),* $E(w_{ig}) = \mu_{ig}, Var(w_{ig}) = v_{ig}, g = 1, 2, ..., G$ *tal que* $\sum_{i=1}^{N} v_i = o(N^2)$. *Entonces una secuencia satisface la* **ley débil de los grandes números** *si:*

$$N^{-1} \sum_{i=1}^{N} \boldsymbol{w}_i \xrightarrow{p} \boldsymbol{\mu}_w,$$

donde $\boldsymbol{\mu}_w \equiv lim\ N^{-1} \sum_{i=1}^{N} \boldsymbol{\mu}_i$.

Ley débil de los grandes números: Ejemplo

Consideremos una secuencia de variables aleatorias independientes $\{x_i\}$ tal que $x_i \sim Normal(\mu_i, \sigma_i^2)$, $i = 1, 2, ..., N$. Definamos la secuencia $\{\overline{x}_N\}$ como $\overline{x}_N = \frac{1}{n}\sum_{i=1}^N x_i$. Notar que $x_N \sim N(\overline{\mu}_N, N^{-1}\overline{\sigma}_N^2)$ donde $\overline{\mu}_N \equiv \frac{1}{N}\sum_{i=1}^N \mu_i$ y $\overline{\sigma}_N^2 \equiv \frac{1}{N}\sum_{i=1}^N \sigma_i^2$.

1. $\sigma_i^2 = \sigma^2$, $i = 1, 2, ..., N$, con la misma varianza. Entonces $\overline{\sigma}_N^2 = \sum_{i=1}^N \sigma^2 = N\sigma^2 = O(N) = o(N^2)$. Así $plim(\overline{x}_N - \overline{\mu}_N) = 0$.
2. $\sigma_i^2 = i$, $i = 1, 2, ..., N$. Entonces $\overline{\sigma}_N^2 = \sum_{i=1}^N i = N(N+1)/2 = O(N^2)$. En este caso no hay convergencia en probabilidad.
3. $\sigma_i^2 = N^{-1}$, $i = 1, 2, ..., N$. Entonces $\overline{\sigma}_N^2 = N^{-1} = o(N^2)$. Así $plim_{N\to\infty}(\overline{x}_N - \overline{\mu}_N) = 0$.
4. $\sigma_i^2 = N^{-i}$, $i = 1, 2, ..., N$. Entonces $\overline{\sigma}_N^2 = \sum_{i=1}^N \frac{1}{N^i} = \frac{1}{1-\frac{1}{N}} = \frac{N}{N-1} = O(1) = o(N^2)$. Así $plim_{N\to\infty}(\overline{x}_N - \overline{\mu}_N) = 0$.

Función de distribución

Definición: Función de distribución acumulada. La función de distribución acumulada (en inglés cumulative distribution function, *cdf*) de una variable aleatoria X, que se denota por $F_X(\cdot)$, es una función cuyo dominio son los reales y la imagen es el intervalo $[0,1]$ que satisface $F_X(x) = P[X \leq x] = P[\omega : X(\omega) \leq x]$ para cada número real x.

- $F_X(-\infty) = 0$ y $F_X(+\infty) = 1$
- $F_X(\cdot)$ es una función monótona no decreciente, $F_X(a) \leq F_X(b)$ si $a < b$
- $F_X(\cdot)$ es contínua de la derecha, $\lim_{0<h\to 0} F_X(x+h) = F_X(x)$

Variables aleatorias discretas

Definición: Variable aleatoria discreta. Una variable aleatoria es discreta si el dominio de X es contable $(\mathcal{X})$. Si la variable aleatoria es discreta entonces la función de distribución acumulada se define como discreta.

Definición: Función de probabilidad o densidad discreta X es una variable aleatoria discreta con valores $x_1, x_2, ...x_n, ...$ si la función $f_X(x) = P[X = x_j]$ si $x = x_j, j = 1, ...n, ...$ y 0 para otros valores de x define la función de densidad discreta de X.

Toda función $f(\cdot)$ con dominio en los reales e imagen $[0, \infty)$ es una función de probabilidad si:

- $f(x) \geq 0$ para todo $x \in \mathcal{X}$
- $\sum_{x \in \mathcal{X}} f(x) = 1$

Variables aleatorias contínuas

Definición: Variable aleatoria contínua. Una variable aleatoria es contínua si el dominio de X no es contable (ej. son los números reales). Si la variable aleatoria es contínua entonces la función de distribución acumulada es contínua.

Definición: Función de densidad contínua. Si X es una variable aleatoria contínua, existe una función $f_X(\cdot)$ tal que $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ para cada valor de x. $f_X(\cdot)$ es la función de densidad contínua.

Nota: f no es una función de probabilidad porque no se puede "medir" la probabilidad que $X = x$. Lo que se mide es la probabilidad de que X tenga valores en un intervalo (a, b) dada por $\int_a^b f(x)dx$.

Toda función $f(\cdot)$ con dominio en los reales e imagen $[0, \infty)$ es una función de densidad si:

- $f(x) \geq 0$ para todo x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Convergencia en distribución

La secuencia de variables aleatorias $\{x_N : N = 1, 2, \ldots\}$ **converge en distribución** *a la variable* x *si y sólo si*

$$F_N(\xi) \to F(\xi) \text{ as } N \to \infty \text{ para todo } \xi \in \mathbb{R},$$

donde F_N *es la función de distribución acumulada de* x_N *y* F *es la de* x*. Escribimos*

$$x_N \xrightarrow{d} x \text{ o } x_N \overset{a}{\sim} x.$$

La secuencia de vectores aleatorios $\{\boldsymbol{x}_N : N = 1, 2, \ldots\}$ *de dimensión* $K \times 1$ **converge en distribución** *a la variable aleatoria* $\boldsymbol{x}$ *si y sólo si para todo vector no aleatorio* $K \times 1$ $\mathbf{c}$ *tal que* $\mathbf{c}'\mathbf{c} = 1$, $\mathbf{c}'\boldsymbol{x}_N \xrightarrow{d} \mathbf{c}'\boldsymbol{x}_N$*. Escribimos*

$$\boldsymbol{x}_N \xrightarrow{d} \boldsymbol{x}.$$

Cuando $x \sim Normal(\mu, \sigma^2)$, x_N *es* **asintóticamente normal**.

Convergencia en distribución

Si $\mathbf{x}_N \xrightarrow{d} \mathbf{x}$ *entonces* $\mathbf{x}_N = O_p(1)$.

Si $N^{\delta}\mathbf{x}_N \xrightarrow{d} \mathbf{x}$, $\mathbf{x}$ *es un vector aleatorio y* $\mathbf{x}_N = O_p(N^{-\delta})$, *entonces* N^{δ} *es la* **tasa de convergencia** *de* x_N.

Teorema de mapeo contínuo (continuous mapping theorem): *La secuencia de vectores aleatorios* $K \times 1$ $\{\mathbf{x}_N\}$ *satisface* $\mathbf{x}_N \xrightarrow{d} \mathbf{x}$. *Si* $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^J$ *es una función contínua, entonces* $\mathbf{g}(\mathbf{x}_N) \xrightarrow{d} \mathbf{g}(\mathbf{x})$.

Lema de equivalencia asintótica: *Sean* $\{\mathbf{x}_N\}$ *y* $\{\mathbf{z}_N\}$ *dos secuencias de vectores aleatorios* $K \times 1$. *Si* $\mathbf{z}_N \xrightarrow{d} \mathbf{z}$ *y* $\mathbf{x}_N \xrightarrow{p} \mathbf{z}_N$, *entonces* $\mathbf{x}_N \xrightarrow{d} \mathbf{z}$.

Teorema central del límite

Teorema central del límite (Lindeberg-Lèvy): *Sea* $\{\boldsymbol{w}_i : i = 1, 2, ...\}$ *una secuencia de vectores aleatorios* $G \times 1$ *independientes y distribuidos idénticamente tal que* $E(w_{ig}) < \infty, g = 1, 2, ..., G,$ *y* $E(\boldsymbol{w}_i) = \mathbf{0}$*. Entonces* $\{\boldsymbol{w}_i : i = 1, 2, ...\}$ *satisface el teorema central del límite, esto es,*

$$N^{-1/2} \sum_{i=1}^{N} \boldsymbol{w}_i \xrightarrow{d} N(\mathbf{0}, \boldsymbol{B}),$$

donde $\boldsymbol{B} = Var(\boldsymbol{w}_i) = E(\boldsymbol{w}_i \boldsymbol{w}_i')$ *es necesariamente definida positiva.*

Otros teoremas no requieren la misma distribución e imponen restricciones en la varianza y otros momentos.

Consistencia

Consistencia débil: *Sea* $\{\widehat{\theta}_N : N = 1, 2, ...\}$ *una secuencia de estimadores de dimensión* $P \times 1$ *con* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, donde* N *indexa el tamaño de la muestra. Si*

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$$

para cualquier valor de $\boldsymbol{\theta}$*, entonces* $\hat{\boldsymbol{\theta}}$ *es un* **estimador consistente** *de* $\boldsymbol{\theta}$*.*

Como estamos considerando convergencia en probabilidad, y usamos la ley débil de los grandes números, entonces tenemos consistencia débil. Convergencia fuerte se refiere a convergencia con probabilidad 1.
Nota: en todas las aplicaciones empíricas no sabemos el valor de $\boldsymbol{\theta}$. Por ello tenemos que considerar todos los valores posibles de $\boldsymbol{\theta}$ conjuntamente.

Normalidad asintótica

Normalidad asintótica: *Sea* $\{\widehat{\boldsymbol{\theta}}_N : N = 1, 2, ...\}$ *una secuencia de estimadores* $P \times 1$ *con* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *donde* N *indexa el tamaño de la muestra. Si*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$$

donde $\boldsymbol{V}$ *es una matriz semidefinida positiva* $P \times P$ *para todo valor de* $\boldsymbol{\theta}$, *entonces* $\hat{\boldsymbol{\theta}}$ *es* $\sqrt{N}$-**distribuido asintóticamente normal** *y* $\boldsymbol{V}$ *es la* **varianza asintótica** *de* $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$, *que se denota por* $Avar\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{V}$ *y* $Avar(\hat{\boldsymbol{\theta}}) = \boldsymbol{V}/N$.

En la práctica, $\boldsymbol{V}$ *también tiene que ser estimada. Tomemos* $\hat{\boldsymbol{V}}_N$ *un estimador consistente de* $\boldsymbol{V}$ *tal que* $\widehat{Avar(\hat{\boldsymbol{\theta}}_N)} = \hat{\boldsymbol{V}}_N/N$.

Referencias

Estas notas están basadas en

- *Capítulo 2 Wooldridge.*
- *White, H. (1984,2001) Asymptotic Theory for Econometricians, Academic Press, London.*