Capítulo 1

Introducción

1.1. Introducción

Sea $\{X_n, n \geq 1\}$ una sucesión de v.a.i.i.d. con f.d. común F y definamos

$$M_n = \max_{i \leq n} X_i.$$

La distribución de esta variable aleatoria es

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = F^n(x).$$

Definimos

$$\alpha(F) = \inf\{x : F(x) > 0\} \geq \infty, \qquad \omega(F) = \sup\{x : F(x) < 1\} \leq \infty.$$

Entonces M_n es una sucesión creciente con límite $\omega(F)$ c.p.1: Si $x < \omega(F)$ entonces $F(x) < 1$ y en consecuencia

$$P(M_n \leq x) = F^n(x) \to 0.$$

Por lo tanto $M_n \to \omega(F)$ en probabilidad y como la sucesión es creciente, convergencia en probabilidad implica convergencia con probabilidad 1.

En principio, si conocemos F la distribución de M_n es conocida, pero las expresiones analíticas para $F^n(x)$ pueden ser complicadas. Usualmente F es desconocida, pero aún así quisiéramos tener alguna idea (al menos aproximada) de la distribución de M_n, es decir, buscamos una distribución límite que sirva de aproximación a F^n, así como la distribución normal sirve de aproximación a la distribución de una suma de variables independientes con gran generalidad.

Sin embargo, el resultado sobre la convergencia de M_n que acabamos de demostrar implica que no es posible obtener una distribución límite no-degenerada a menos que normalicemos M_n de manera adecuada. Algo similar ocurre en el caso del Teorema Central del Límite: Por la LFGN, el promedio $\overline{X}_n = (1/n)\sum_{i=1}^{n} X_i$ converge a la media de la población $\mu = \mathrm{E}(X_i)$, pero si hacemos una transformación lineal

$$\frac{\bar{X}_n - \mu_n}{\sigma_n},$$

donde $\mu_n = \mu$ y $\sigma_n = \sigma/\sqrt{n}$, entonces hay convergencia débil a una variable con distribución $N(0, 1)$.

Buscamos un teorema del tipo

$$P(\frac{M_n - b_n}{a_n} \leq x) = F^n(a_n x + b_n) \rightarrow G(x)$$

débilmente cuando $n \rightarrow \infty$ para una distribución límite no-degenerada y nos planteamos las siguientes preguntas

- ¿Cuáles son las distribuciones límite posibles?
- ¿Cuáles son las constantes a_n y b_n? ¿Son únicas?
- ¿Qué condiciones debe satisfacer F para que se cumpla un resultado así?
- Si hay varias G posibles, ¿Cómo sabemos, conociendo F, cuál de ellas es el límite? ¿Es único?

Responderemos todas estas preguntas en los capítulos sucesivos. El resto del presente capítulo lo dedicaremos a presentar algunos resultados previos.

1.2. Funciones de Distribución

1.2.1. Función de Distribución Empírica

Sea $X_1 \ldots, X_n$ una muestra aleatoria simple de una población con función de distribución (f.d.) F. Dada la muestra, definimos la función de distribución empírica (f.d.e.) $\widehat{F}_n$ por

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(X_i) = \frac{1}{n} \sharp\{i : X_i \leq x\}$$

$\widehat{F}_n$ es aleatoria y por el teorema de Glivenko-Cantelli sabemos que $\widehat{F}_n$ converge uniformemente a $F(x)$.

Teorema 1.1 (Glivenko-Cantelli) *Sea $X_1, \ldots, X_n$ una colección de v.a.i. con distribución común F y sea $\widehat{F}_n(x) = \widehat{F}_n(x, \omega)$ la función de distribución empírica correspondiente. Entonces*

$$\sup_x |\widehat{F}_n(x) - F(x)| \rightarrow 0$$

con probabilidad 1.

Para x fijo,

$$\widehat{F}_n(x, \omega) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_j(\omega) \leq x\}}$$

es una variable aleatoria y podemos calcular su valor esperado:

$$\begin{aligned} \mathrm{E}(\widehat{F}_n(x, \omega)) &= \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}(\mathbf{1}_{\{X_j(\omega) \leq x\}}) \\ &= \frac{1}{n} \sum_{j=1}^{n} P(X_j(\omega) \leq x) = F(x) \end{aligned}$$

de modo que, para cada x, $\widehat{F}_n(x)$ es un estimador insesgado de $F(x)$. Mas aún, por la Ley Fuerte de los Grandes Números, para cada x existe un conjunto nulo A_x tal que

$$\lim_{n\to\infty} \widehat{F}_n(x,\omega) = F(x)$$

siempre que $\omega \notin A_x$.

El teorema de Glivenko-Cantelli dice más. Dice que la convergencia vale para todo x siempre que ω esté fuera de un conjunto nulo (común) A y además que la convergencia es uniforme. La demostración no es complicada y puede hallarse en muchos libros de Probabilidades, por ejemplo, en el libro de Billingsley.

La función de supervivencia, a veces denotada por $\overline{F}$, es $\overline{F}(c) = 1 - F(c) = P(X > c)$.

1.2.2. La Función de cuantiles

La función de cuantiles (f.c.) Q es la inversa (generalizada) de F:

$$Q(p) = F^{\leftarrow}(p) = \inf\{s : F(s) \geq p\},$$

para cualquier $p \in (0,1)$.

Propiedades

1. Q es creciente (en sentido amplio) y continua por la izquierda.

   Supongamos $x < y$ ambos en $(0,1)$, entonces

   $$Q(y) = \inf\{s : F(s) \geq y\} \geq \inf\{s : F(s) \geq x\} = Q(x).$$

   Para ver que es continua por la izquierda supongamos que $x \in (0,1)$, $x_n \uparrow x$ pero $Q(x_n) \uparrow Q(x^-) < Q(x)$. Entonces existen $\delta > 0$ e y tales que para todo n,

   $$Q(x_n) < y < Q(x) - \delta$$

   La primera desigualdad y la definición de Q dicen que $F(y) \geq x_n$ para todo n, y haciendo $n \to \infty$ obtenemos $F(y) \geq x$, de donde, por la definición de Q, obtenemos $y \geq Q(x)$ pero $y < Q(x) - \delta$, lo cual es una contradicción.

2. $F(Q(y)) \geq y$.

   Comenzamos por ver que el conjunto $A(y) = \{s : F(s) \geq y\}$ es cerrado. Si $s_n \in A(y)$ y $s_n \downarrow s$ entonces $y \leq F(s_n) \downarrow F(s)$, de modo que $F(s) \geq y$ y $s \in A(y)$. Si $s_n \uparrow s$ y $s_n \in A(y)$ entonces $y \leq F(s_n) \uparrow F(s^-) \leq F(s)$ y $F(s) \geq y$, de modo que $s \in A(y)$ de nuevo y $A(y)$ es cerrado. Como $A(y)$ es cerrado, $\inf A(y) \in A(y)$, es decir, $Q(y) \in A(y)$, lo que implica que $F(Q(y)) \geq y$.

3. $Q(y) \leq t$ sii $y \leq F(t)$; $Q(y) > t$ sii $y > F(t)$.

   Esto es consecuencia de la definición de Q.

4. Sea $X \sim F$ y definamos $Y = aX + b$. La f.d. de Y es

   $$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$$

Una transformación de este tipo de un *cambio de ubicación y escala*; a es el parámetro de escala y b el de ubicación. La función de cuantiles de Y es

$$Q_Y(p) = aQ_X(p) + b$$

(Esto es fácil de ver si F_X es continua y estrictamente creciente, de modo que Q_X es la inversa de F_X). Por lo tanto, una transformación lineal de la variable X, produce una transformación lineal del mismo tipo en la función de cuantiles.

5. Sea $([0,1], \mathcal{B}, m)$ el espacio de Lebesgue y sea U la función identidad en $[0,1]$ (U tiene distribución uniforme). Si F es una f.d. y Q es la f.c. correspondiente, entonces $Q(U(\cdot))$ es una variable aleatoria en $[0,1]$ con f.d. F:
$$m(Q(U) \leq t) = m(U \leq F(t)) = F(t).$$

- Variante: Sea E una v.a. con distribución exponencial: $P(E > x) = e^{-x},\ x > 0$. Dada una f.d. F sea $R(x) = -\log(1 - F(x))$. Llamemos $R^{\leftarrow}$ a la inversa generalizada de R, entonces $R^{\leftarrow}(E)$ tiene f.d. F:
$$P(R^{\leftarrow}(E) > x) = P(E > R(x)) = \exp\{-R(x)\} = 1 - F(x).$$

La función de cuantiles empírica (f.c.e.) $\widehat{Q}_n$ es la inversa generalizada de $\widehat{F}_n$, es decir, es la función definida sobre $(0,1)$ con valores en $\mathbb{R}$ tal que, dado $p,\ 0 < p < 1$, $\widehat{Q}_n(p)$ es el menor valor a la izquierda del cual se encuentra un porcentaje $100p$ de los datos.

Si llamamos $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ a los estadísticos de orden de la muestra, entonces

$$\widehat{Q}_n(p) = X_{(i)} \quad \text{cuando } \frac{i-1}{n} < p \leq \frac{i}{n}.$$

1.3. Transformaciones

Con frecuencia es útil hacer transformaciones sobre los datos usando una función (diferenciable) g con la idea de que los datos transformados $Y_i = g(X_i)$, $1 \leq i \leq n$, satisfagan un modelo conocido. Típicamente, las funciones que se usan en la práctica son $ax+b,\ x^a,\ \exp(x)$ y $\log(x)$ para $a, b \in \mathbb{R}$. Según la monotonía de la transformación, ésta conserva los estadísticos de orden, los invierte o los permuta.

Bajo una transformación $Y_i = g(X_i)$ tenemos

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \quad \text{para } y \in g(\mathbb{R})$$

y

$$f_Y(y) = f_X(g^{-1}(y)) \Big| \frac{dg^{-1}(y)}{dy} \Big|$$

donde g^{-1} es la función inversa de g.

Veamos algunos ejemplos de interés para el curso.

1. Comenzamos con $X \sim \mathcal{E}xp(\alpha)$ y sea $g(x) = e^x$, $g^{-1}(y) = \log y$. Obtenemos la distribución de Pareto (estricta) con parámetro α:
$$1 - F_Y(y) = y^{-\alpha} \quad y \quad f_Y(y) = \alpha y^{-\alpha-1} \quad \text{para } y > 1.$$
α es el índice de Pareto.

2. Consideremos ahora la transformación $g(x) = x^{1/\tau}$ con $\tau > 0$. Si tenemos inicialmente una distribución exponencial $\mathcal{E}xp(\lambda)$, la transformación produce una distribución de Weibull,

$$1 - F(y) = \exp(-\lambda y^{\tau}).$$

3. Si partimos de la distribución $\mathcal{E}xp(1)$ podemos obtener las llamadas *distribuciones de valores extremos* G_γ, donde γ se conoce como el *índice de valor extremo*, usando la transformación $g(x) = (x^{-\gamma} - 1)/\gamma$, con $\gamma \in \mathbb{R}$.

$$G_\gamma(x) = \exp\big(-(1+\gamma x)^{-1/\gamma}\big) \quad \text{para } 1 + \gamma x > 0.$$

   El caso particular $\gamma = 0$ es la llamada distribución de Gumbel G_0 y se obtiene tomando el límite cuando $\gamma \to 0$ en la transformación anterior para obtener $g(x) = -\log x$ y

$$G_0(x) = \exp(-e^{-x})$$

4. Si añadimos una transformación lineal $g(x) = \sigma x + \mu$ a la transformación anterior obtenemos las llamadas *distribuciones de valores extremos generalizadas* o con parámetros de ubicación y escala,

$$F(x) = \exp\big(-[1+\gamma(x-\mu)/\sigma]^{-1/\gamma}\big)$$

   para $1+\gamma(x-\mu)/\sigma > 0$. Para el caso particular $\gamma = \sigma$ y $\mu = 0$ obtenemos la distribución de Fréchet,

$$F(x) = \exp\big(-x^{-1/\sigma}\big) \quad \text{para } x > 0.$$

5. Observamos finalmente que cualquier distribución positiva que tiene dominio infinito a la derecha, puede ser transformada a una distribución con extremo derecho finito arbitrario x_+ usando la transformación $g(x) = x_+ - 1/x$. Por ejemplo, usando esta transformación con la distribución estricta de Pareto $Pa(\alpha)$ obtenemos

$$1 - F(x) = (x_+ - x)^{\alpha} \qquad \text{para } x_+ - 1 < x < x_+.$$

   En particular, tomando $x_+ = 1$ y $\alpha = 1$ obtenemos la distribución uniforme en $(0, 1)$.

1.4. Convergencia de Funciones Monótonas

Para cualquier función H escribimos

$$\mathcal{C}(H) = \{x \in \mathbb{R} : H \text{ es finita y continua en } x\}$$

Una sucesión de funciones no-decrecientes $\{H_n, n \geq 0\}$ en $\mathbb{R}$ converge *débilmente* o *en ley* a H_0 cuando $n \to \infty$ si,

$$H_n(x) \to H_0(x)$$

para todo $x \in \mathcal{C}(H_0)$. Notación $H_n \to_w H_0$ o simplemente $H_n \to H_0$ si no hay lugar a confusión. Si las variables $\{X_n, n \geq 0\}$ tienen funciones de distribución $\{F_n, n \geq 0\}$ entonces $X_n \Rightarrow X_0$ quiere decir $F_n \to F_0$. En este caso también escribimos $\mathcal{L}(X_n) \to \mathcal{L}(X)$ donde $\mathcal{L}(X)$ denota la ley de X.

Proposición 1.1 *Si $\{H_n, n \geq 0\}$ son funciones no-decrecientes y $H_n \to_w H_0$, entonces $H_n^{\leftarrow} \to_w H_0^{\leftarrow}$.*

Demostración. Fijamos $t \in \mathcal{C}(H_0^{\leftarrow})$ y sea $\varepsilon > 0$. Como las discontinuidades de la función monótona H_0 son a los sumo numerables, existe $x \in (H_0^{\leftarrow}(t) - \varepsilon, H_0^{\leftarrow}(t))$ y $x \in \mathcal{C}(H_0)$. Como $x < H_0^{\leftarrow}(t)$, por la definición de $H_0^{\leftarrow}$ tenemos que $H_0(x) < t$. Como $x \in \mathcal{C}(H_0)$ implica que $H_n(x) \to H_0(x)$, tenemos para n grande que $H_n(x) < t$, y usando de nuevo la definición de la inversa generalizada obtenemos que $x \leq H_n^{\leftarrow}(t)$ para n grande. Por lo tanto

$$H_0^{\leftarrow}(t) - \varepsilon < x \leq H_n^{\leftarrow}(t)$$

para n grande, lo cual implica, como $\varepsilon > 0$ es arbitrario, que

$$H_0^{\leftarrow}(t) \leq \liminf_{n\to\infty} H_n^{\leftarrow}(t).$$

(para esta mitad no hemos usado que t es punto de continuidad de $H_0^{\leftarrow}$).

Para obtener la desigualdad contraria observamos que para cualquier $t' > t$ podemos hallar $y \in \mathcal{C}(H_0)$ con

$$H_0^{\leftarrow}(t') < y < H_0^{\leftarrow}(t') + \varepsilon. \tag{1.1}$$

La desigualdad izquierda en (1.1) y la definición de la inversa generalizada dan

$$t < t' \leq H_0(y).$$

Como $y \in \mathcal{C}(H_0)$ tenemos $H_n(y) \to H_0(y)$, y para n grande, $t \leq H_n(y)$ y en consecuencia $H_n^{\leftarrow}(t) \leq y$. Usando (1.1)

$$H_n^{\leftarrow}(t) \leq y < H_0^{\leftarrow}(t') + \varepsilon$$

para n grande y por lo tanto

$$\limsup_{n\to\infty} H_n^{\leftarrow}(t) \leq H_0^{\leftarrow}(t')$$

ya que ε es arbitrario. Hacemos $t' \to t$ y usamos la continuidad de $H_0^{\leftarrow}$ en t para obtener

$$\limsup_{n\to\infty} H_n^{\leftarrow}(t) \leq H_0^{\leftarrow}(t).$$

Esto completa la demostración. ■

La proposición anterior permite probar fácilmente la versión unidimensional del siguiente teorema debido a Skorohod

Teorema 1.2 (Skorohod) *Para $n \geq 0$ sea X_n una v. a. real sobre $(\Omega_n, \mathcal{B}_n, P_n)$ tal que $X_n \Rightarrow X_0$. Entonces existen variables aleatorias $\{X_n', n \geq 0\}$ definidas sobre el espacio de Lebesgue $([0,1], \mathcal{B}, m)$ tales que*

i) Para cada $n \geq 0$, $\mathcal{L}(X_n') = \mathcal{L}(X_n)$.

ii) $X_n' \to X_0'$ casi seguramente respecto a m.

Observación 1.1 Las distribuciones conjuntas de las nuevas variables (y en particular la ley de toda la sucesión) no necesariamente coinciden con las de las viejas. Sólo las distribuciones marginales.

Demostración. Sea U la identidad en $[0,1]$, supongamos que la f.d. de X_n es F_n, $Q_n = F_n^{\leftarrow}$ $n \geq 0$, y definamos

$$X_n' = Q_n(U),$$

entonces $\mathcal{L}(X_n') = \mathcal{L}(X_n)$ para $n \geq 0$.

Para ver (ii) observemos que $F_n \to F_0$ implica $Q_n \to Q_0$ y por lo tanto,

$$\begin{aligned} 1 &\geq m\{0 \leq u \leq 1 : X'_n(u) \to X'_0(u)\} \\ &= m\{u : Q_n(u) \to Q_0(u)\} \\ &\geq m\{u : u \in \mathcal{C}(Q_0)\} = 1 \end{aligned}$$

ya que las discontinuidades de Q_0 son, a lo sumo, numerables. ■

1.5. Teorema de Convergencia a Familias

Muchos resultados de convergencia de variables aleatorias son del siguiente tipo: Para una sucesión de v.a. $\xi_n, n \geq 1$ y constantes $a_n > 0$ y $b_n \in \mathbb{R}$, se demuestra que

$$\frac{\xi_n - b_n}{a_n} \Rightarrow Y,$$

donde Y es una v. a. no-degenerada. Usando esto tenemos

$$P(\frac{\xi_n - b_n}{a_n} \leq x) \approx P(Y \leq x) = G(x),$$

o poniendo $y = a_n x + b_n$,

$$P(\xi_n \leq y) \approx G(\frac{y - b_n}{a_n}).$$

Esto permite aproximar la distribución de ξ_n por una familia de distribuciones con parámetros de ubicación y escala.

La pregunta ahora es ¿Hasta qué punto son únicas estas constantes de normalización a_n y b_n? La respuesta la encontramos en el teorema de convergencia a familias de distribuciones: Las constantes están determinadas salvo por equivalencias asintóticas y la distribución límite está determinada salvo por parámetros de ubicación y escala.

Definición 1.1 Dos distribuciones F y G son *del mismo tipo* o *pertenecen a la misma familia* si para algunas constantes $a > 0, b \in \mathbb{R}$,

$$G(x) = F(ax + b), \qquad \text{para todo } x \in \mathbb{R}.$$

En términos de variables aleatorias, si $\mathcal{L}(X) = F$ y $\mathcal{L}(Y) = G$ entonces

$$\mathcal{L}(Y) = \mathcal{L}(\frac{X - b}{a})$$

Por ejemplo, podemos considerar la familia gaussiana. Si $X_{0,1}$ tiene distribución $N(0,1)$ y $X_{\mu,\sigma}$ tiene distribución $N(\mu, \sigma^2)$, entonces $\mathcal{L}(X_{\mu,\sigma}) = \mathcal{L}(\sigma X_{0,1} + \mu)$.

Teorema 1.3 (Convergencia a familias, Gnedenko & Khinchin) *Sean $G(x)$ y $H(x)$ dos funciones de distribución propias, ninguna de las cuales está concentrada en un punto. Supongamos que para $n \geq 0$, X_n son v.a. con funciones de distribución F_n, U y V son v.a. con f.d. G y H, respectivamente. Sean, además, constantes $a_n > 0$, $\alpha_n > 0$, $b_n \in \mathbb{R}, \beta_n \in \mathbb{R}$.*

a) Si

$$F_n(a_nx+b_n)\to G(x),\quad F_n(\alpha_nx+\beta_n)\to H(x) \tag{1.2}$$

o equivalentemente

$$\frac{X_n-b_n}{a_n}\Rightarrow U,\quad \frac{X_n-\beta_n}{\alpha_n}\Rightarrow V, \tag{1.3}$$

entonces existen constantes $A>0$ y $B\in\mathbb{R}$ tales que, cuando $n\to\infty$,

$$\frac{\alpha_n}{a_n}\to A>0,\quad \frac{\beta_n-b_n}{a_n}\to B, \tag{1.4}$$

y

$$H(x)=G(Ax+B),\quad \mathcal{L}(V)=\mathcal{L}(\frac{U-B}{A}). \tag{1.5}$$

b) Recíprocamente, si (1.4) vale, entonces cualquiera de las relaciones en (1.2) implica la otra y (1.5) vale.

Demostración

Veamos primero la demostración de (b). Supongamos que

$$G_n(x):=F_n(a_nx+b_n)\to G(x)$$

y

$$\frac{\alpha_n}{a_n}\to A>0,\quad \frac{\beta_n-b_n}{a_n}\to B,$$

entonces

$$F_n(\alpha_nx+\beta_n)=G_n(\frac{\alpha_n}{a_n}x+\frac{\beta_n-b_n}{a_n}).$$

Escogemos $x\in\mathcal{C}(G(A\cdot+B))$. Supongamos que $x>0$, un argumento similar sirve si $x\le 0$. Dado $\varepsilon>0$ para n grande tenemos

$$(A-\varepsilon)x+B-\varepsilon\le\frac{\alpha_n}{a_n}x+\frac{\beta_n-b_n}{a_n}\le(A+\varepsilon)x+B+\varepsilon$$

y entonces

$$\limsup_{n\to\infty}F_n(\alpha_nx+\beta_n)\le\limsup_{n\to\infty}G_n((A+\varepsilon)x+B+\varepsilon).$$

Por lo tanto, para cualquier $z\in\mathcal{C}(G)$ con $z>(A+\varepsilon)x+B+\varepsilon$ tenemos

$$\limsup_{n\to\infty}F_n(\alpha_nx+\beta_n)\le\limsup_{n\to\infty}G_n(z)=G(z).$$

En consecuencia,

$$\limsup_{n\to\infty}F_n(\alpha_nx+\beta_n)\le\inf\{G(z):z>(A+\varepsilon)x+B+\varepsilon,z\in\mathcal{C}(G)\}.$$

y como $\varepsilon>0$ es arbitrario y G es continua por la derecha

$$\limsup_{n\to\infty}F_n(\alpha_nx+\beta_n)\le\inf\{G(z):z>Ax+B\}=G(Ax+B)$$

De manera similar,

$$\begin{aligned}\liminf_{n\to\infty} F_n(\alpha_n x+\beta_n) &\geq \liminf_{n\to\infty} G_n((A-\varepsilon)x+B-\varepsilon)\\ &\geq \liminf_{n\to\infty} G_n(z) = G(z)\end{aligned}$$

para cualquier $z < (A-\varepsilon)x+B-\varepsilon$, $z \in \mathcal{C}(G)$. Como estas desigualdades valen para todo $\varepsilon > 0$,

$$\liminf_{n\to\infty} F_n(\alpha_n x+\beta_n) \geq \sup\{G(z) : z < Ax+B,\ z\in\mathcal{C}(G)\} = G(Ax+B)$$

porque $Ax+B \in \mathcal{C}(G)$.

Veamos ahora la demostración de la parte (a). Supongamos que

$$F_n(a_n x+b_n) \to G(x), \quad F_n(\alpha_n x+\beta_n) \to H(x).$$

Usando la proposición 1.1 tenemos

$$\begin{aligned}\frac{F_n^{\leftarrow}(y)-b_n}{a_n} &\to G^{\leftarrow}(y), \quad y\in\mathcal{C}(G^{\leftarrow}),\\ \frac{F_n^{\leftarrow}(y)-\beta_n}{\alpha_n} &\to H^{\leftarrow}(y), \quad y\in\mathcal{C}(H^{\leftarrow}).\end{aligned}$$

Como las f.d. $G(x)$ y $H(x)$ no están concentradas en un punto, podemos hallar $y_1 < y_2$ con $y_i \in \mathcal{C}(G^{\leftarrow})\cap\mathcal{C}(H^{\leftarrow})$ para $i=1,2$, tales que

$$-\infty < G^{\leftarrow}(y_1) < G^{\leftarrow}(y_2) < \infty, \qquad -\infty < H^{\leftarrow}(y_1) < H^{\leftarrow}(y_2) < \infty.$$

Por lo tanto, para $i=1,2$ tenemos

$$\frac{F_n^{\leftarrow}(y_i)-b_n}{a_n} \to G^{\leftarrow}(y_i), \quad \frac{F_n^{\leftarrow}(y_i)-\beta_n}{\alpha_n} \to H^{\leftarrow}(y_i). \tag{1.6}$$

En las expresiones anteriores restamos las ecuaciones con $i=1$ de las expresiones con $i=2$ para obtener

$$\frac{F_n^{\leftarrow}(y_2)-F_n^{\leftarrow}(y_1)}{a_n} \to G^{\leftarrow}(y_2)-G^{\leftarrow}(y_1), \tag{1.7}$$

$$\frac{F_n^{\leftarrow}(y_2)-F_n^{\leftarrow}(y_1)}{\alpha_n} \to H^{\leftarrow}(y_2)-H^{\leftarrow}(y_1). \tag{1.8}$$

Ahora dividimos (1.7) entre (1.8) y obtenemos

$$\frac{\alpha_n}{a_n} \to \frac{G^{\leftarrow}(y_2)-G^{\leftarrow}(y_1)}{H^{\leftarrow}(y_2)-H^{\leftarrow}(y_1)} := A > 0.$$

Además, de (1.6) vemos que

$$\begin{aligned}\frac{F_n^{\leftarrow}(y_1)-b_n}{a_n} &\to G^{\leftarrow}(y_1),\\ \frac{F_n^{\leftarrow}(y_1)-\beta_n}{a_n} = \frac{F_n^{\leftarrow}(y_1)-\beta_n}{\alpha_n}\frac{\alpha_n}{a_n} &\to H^{\leftarrow}(y_1)A,\end{aligned}$$

y restando obtenemos

$$\frac{\beta_n-b_n}{a_n} \to H^{\leftarrow}(y_1)A - G^{\leftarrow}(y_1) := B,$$

como queríamos ver, de modo que (1.4) vale. Por la parte (b) obtenemos (1.5).

■

Observación 1.2 Una consecuencia de la demostración es que, a partir de (1.6), (1.7) y (1.8), una posible selección de las constantes de normalización es

$$a_n = F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1), \qquad b_n = F_n^{\leftarrow}(y_1).$$

El siguiente ejemplo muestra la importancia de la hipótesis de que las distribuciones límite no están concentradas en un punto.

Ejemplo 1.1
Sea

$$G(x) = \begin{cases} 0, & \text{si } t < c, \\ 1, & \text{si } t \geq c. \end{cases}$$

Entonces,

$$G^{\leftarrow}(t) = \text{ínf}\{y : G(y) \geq t\} = \begin{cases} -\infty, & \text{si } t \leq 0 \\ c, & \text{si } 0 < t \leq 1, \\ \infty, & \text{si } t > 1. \end{cases}$$

El siguiente corolario es consecuencia inmediata del teorema de convergencia a familias.

Corolario 1.1 *Sea F_n una sucesión de f. d. y $a_n > 0$ y b_n sucesiones de constantes tales que*

$$F_n(a_n x + b_n) \to G(x) \tag{1.9}$$

en todo punto de continuidad de G, que es una f.d. propia y no está concentrada en un punto. Sean $c_n > 0$ y d_n sucesiones de constantes tales que

$$\frac{a_n}{c_n} \to 1, \qquad \frac{d_n - b_n}{a_n} \to 0.$$

Entonces (1.9) vale con c_n y d_n en lugar de a_n y b_n.

1.6. La Ecuación de Cauchy

Definición 1.2 La función f satisface la ecuación de Cauchy (o es aditiva) si

$$f(x+y) = f(x) + f(y), \quad x, y \in \mathbb{R}.$$

Una función $g : (0, \infty) \to (0, \infty)$ satisface la ecuación de Hamel (o es multiplicativa) si

$$g(\lambda\mu) = g(\lambda)g(\mu) \quad \lambda, \mu > 0.$$

Si $f(x) = \log g(e^x)$ entonces f es aditiva si y sólo si g es multiplicativa.

Teorema 1.4 *(Bingham et al., teoremas 1.1.8, 1.1.9) a) Si f es aditiva y medible entonces $f(x) = cx$ para algún c.*
b) Si g es multiplicativa y medible entonces $g(\lambda) = \lambda^c$ $(\lambda > 0)$ para algún real c.