00365

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

POSGRADO EN CIENCIAS MATEMATICAS

FACULTAD DE CIENCIAS

PROPIEDADES GENERALES DE LOS PROCESOS EMPIRICOS

T E S I S

QUE PARA OBTENER EL GRADO ACADEMICO DE:

MAESTRO EN CIENCIAS

P R E S E N T A :

RAFAEL TORRES SIMON

DIRECTOR DE TESIS: DR. JOSE MARIA GONZALEZ BARRIOS MURGUIA

MEXICO, D. F. JUNIO DE 2004

Propiedades generales de los procesos empíricos

Rafael Torres Simón

Agradecimientos

De manera muy especial al Dr. José María, por haber aceptado ser mi asesor de tesis, por apoyarme de la mejor manera para la realización de la misma y por toda la atención que siempre me brindó. Muchas gracias.

A mis sinodales: Dra. Begoña Fernández, Dr. Alberto Contreras, Dr. Raúl Rueda, Dr. Mogens Bladt. Muchas gracias por su valioso tiempo y dedicación para el mejoramiento de este trabajo.

A todos mis amigos y todo el grupo que formamos en el IIMAS por compartir momentos gratos.

A mis padres por enseñarme a hacer
los sueños realidad.
A mis hermanos: Ignacio, Ciry, Martha,
Andrés, Lety. Los quiero mucho.
A NAZ con mucho cariño.

Índice general

Introducción

En 1933, Glivenko y Cantelli, publicaron un resultado fundamental sobre la convergencia de la función de distribución empírica.

Dadas $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F,

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i), \quad -\infty < t < \infty.$$

que asigna igual masa $\frac{1}{n}$ a cada X_i, se llama la función de distribución empírica. Aquí, 1_A denota la función indicadora del conjunto A.

El resultado publicado es sobre la convergencia uniforme de F_n a F, para el caso contínuo, Glivenko y después en general, Cantelli.

Si $X_1, \ldots, X_n$ son variables aleatorias independientes con función de distribución arbitraria F entonces, cuando $n \to \infty$,

$$D_n := \sup_{t \in \mathrm{R}} \mid F_n(t) - F(t) \mid \to 0 \quad \text{con probabilidad uno.}$$

Este resultado ha sido uno de los teoremas fundamentales de la estadística y parte central en el estudio de la inferencia estadística.

El estudio de los procesos empíricos cuya definicíon se dará mas adelante y de la función de distribución empírica, son uno de los temas con mayor seguimiento en su investigación. Las aplicaciones para las diferentes ramas de la estadística se han multiplicado, especialmente porque muchos procedimientos estadísticos, pueden ser vistos como funcionales del proceso empírico, y la conducta de tales procedimientos pueden ser deducidos del proceso empírico.

En este trabajo se dan algunas propiedades generales de los procesos empíricos. La mayoría de los resultados que se presentan aquí, se obtuvieron de [15].

Para entender este trabajo se necesitan conocimientos generales de Probabilidad, así como ideas básicas de Análisis e Inferencia Estadística. Como ayuda para el lector, se incluye un apéndice con los resultados principales que se usan.

Este trabajo, se organizó de la siguiente manera:

En el Capítulo 1 se da una introducción a lo que es la función de distribución empírica, sin hacer suposición alguna del carácter distribucional en los datos. También se dan algunos resultados principales de las transformaciones libres de distribución, que simplifican algunos resultados referentes a la distribución empírica, al caso uniforme.

En el Capítulo II, se dan algunas propiedades de la función de distribución empírica, donde los datos constituyen una muestra aleatoria con función de distribución F. En este capítulo se demuestra la convergencia uniforme de F_n a F y de algunos resultados de estadísticas relacionadas con D_n, llamada también estadística de Kolmogorov-Smirnov. De igual manera, se dan algunos resultados fundamentales que conllevan a la convergencia uniforme de la función característica empírica a la función característica usual.

En el Capítulo III, se da una introducción a la estimación de la función de densidad, propuesta por Rosenblatt [10]. Se dice que una sucesión W_n de pesos es consistente si siempre que $(X,Y),(X_1,Y_1),(X_2,Y_2),\ldots$ sean independientes e identicamente distribuidos (i.i.d.), donde Y toma valores reales y para $r \geq 1$, $\mathbb{E} \mid Y \mid^r < \infty$, entonces $\mathbb{E}_n(Y \mid X) \to \mathbb{E}(Y \mid X)$ en L_r ($Z_n \to Z$ en L_r, signifíca que $\mathbb{E} \mid Z_n - Z \mid^r \to 0$). Aquí, $X, X_1, X_2, \ldots$ es una sucesión de variables aleatorias i.i.d. con valores en $\mathbb{R}^d$, sobre un espacio de probabilidad Ω. Se probará la consistencia de $\{W_{in}\}$.

Capítulo 1

Definiciones básicas

1.1. Preliminares

Asumiremos que estamos en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, i.e., Ω un conjunto no vacío, $\mathcal{F}$ una σ-álgebra de subconjuntos de Ω y P una medida de probabilidad definida sobre $\mathcal{F}$. Para la notación de este capítulo y lo que sigue, ver [8].
La función $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ es una *variable aleatoria* (v.a.) si

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}).$$

$\mathcal{B}(\mathbb{R})$ es la σ-álgebra de Borel de $\mathbb{R}$.
En otras palabras, una variable aleatoria es una función medible de $(\Omega, \mathcal{F}, P)$ a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Un ejemplo de variable aleatoria muy útil, es la **función indicadora** de un conjunto $A \in \mathcal{F}$:

$$1_A(\omega) = \begin{cases} 1, & \text{si} \quad \omega \in A \\ 0, & \text{si} \quad \omega \notin A. \end{cases}$$

La variable aleatoria X definida sobre $(\Omega, \mathcal{F}, P)$, induce una medida P_X sobre $\mathcal{B}(\mathbb{R})$, definida por

$$P_X(E) = P\{X^{-1}(E)\} \quad (E \in \mathcal{B}(\mathbb{R})).$$

P_X es una medida de probabilidad sobre $\mathcal{B}(\mathbb{R})$ y es llamada la *distribución* o *ley* de X. Cuando se requiera, se denotará por $\mathcal{L}$ la distribución o ley de una variable aleatoria.
Para cualesquieras X y Y variables aleatorias, definidas sobre un espacio de

probabilidad común, la igualdad $X \stackrel{d}{=} Y$ o $\mathcal{L}(X) = \mathcal{L}(Y)$ significa que X y Y tienen la misma distribución.

1.1.1. Definición. *La función de distribución (f.d.) de la variable aleatoria X, es la función $F = F_X$ definida por*

$$F_X(x) = P_X(-\infty, x] = P(\omega \in \Omega : X(\omega) \leq x), \quad x \in \mathbb{R}.$$

En lo que sigue, escribiremos $\{X \leq x\}$ en lugar de $\{\omega \in \Omega : X(\omega) \leq x\}$, y cuando X tenga por f.d. F, se escribirá $X \sim F$.
Se dice que $f : \mathbb{R} \to \mathbb{R}$ es una función de densidad si $f(x) \geq 0$ para todo $x \in \mathbb{R}$, y

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

En este caso, la medida de probabilidad asociada a f, es la medida de probabilidad P_f sobre $\mathcal{B}(\mathbb{R})$, definida por

$$P_f(A) = \int_A f(x)dx \quad \text{para todo } A \in \mathcal{B}(\mathbb{R}).$$

1.1.2. Teorema. *La función de distribución de una variable aleatoria X es una función creciente, continua por la derecha y satisface*

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$

y

$$F(+\infty) = \lim_{x \to \infty} F(x) = 1.$$

Demostración: Sean $x \in \mathbb{R}$ y $h > 0$. Notemos que

$$\begin{aligned} F(x+h) - F(x) &= \mathbb{P}(X \leq x+h) - \mathbb{P}(X \leq x) \\ &= \mathbb{P}(-\infty < X \leq x \cup x < X \leq x+h) - \mathbb{P}(X \leq x) \\ &= \mathbb{P}(-\infty < X \leq x) + \mathbb{P}(x < X \leq x+h) - \mathbb{P}(X \leq x) \\ &= \mathbb{P}(x < X \leq x+h). \end{aligned}$$

Luego,

$$F(x+h) - F(x) = \mathbb{P}(x < X \leq x+h) \geq 0.$$

Por lo tanto, F es creciente.
Demostremos ahora que F es continua por la derecha, i.e..

$$\lim_{y \downarrow x} F(y) = F(x).$$

Puesto que F es creciente,

$$\lim_{y \downarrow x} F(y) \text{ existe y es igual a } F(x+).$$

Entonces debemos probar que $F(x+) = F(x)$.
Así, sea $\{a_n\} \downarrow 0$ y consideremos $\{a_n + x\}_{n \geq 1}$. Entonces $\{a_n + x\} \downarrow x$.
Definamos

$$A_n = \{X \leq x + a_n\} \quad \text{para todo } n = 1, 2, \ldots.$$

Entonces

$$A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$$

y

$$\bigcap_{n=1}^{\infty} A_n = \{X \leq x\}.$$

Así que,

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \mathbb{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

$$= \lim_{n\to\infty} \mathbb{P}(X \leq x + a_n) = \lim_{n\to\infty} F(x + a_n) = F(x+).$$

Por lo tanto,

$$F(x) = F(x+)$$

y

$$F \text{ es continua por la derecha.}$$

Finalmente, para todo $n \geq 1$,

$$F(n) - F(-n) = \mathbb{P}(-n < X \leq n).$$

Tomando límites en ambos lados cuando $n \to \infty$, se concluye

$$F(+\infty) - F(-\infty) = 1.$$

Puesto que $0 \leq F(x) \leq 1$ para todo $x \in \mathbb{R}$, se sigue que

$$F(+\infty) = 1 \text{ y } F(-\infty) = 0$$

□

1.1.3. Definición. *Para una función de distribución F, la función cuantil F^{-1}, se define como*

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1.$$

1.1.4. Lema. *Sea F función de distribución. Entonces $F^{-1}(t)$, $0 < t < 1$ es creciente, continua por la izquierda y satisface*

(i) $F^{-1}(F(x)) \leq x; \quad -\infty < x < \infty.$

(ii) $F(F^{-1}(t)) \geq t$ *si* $0 < t < 1.$

(iii) $F(x) \geq t$ *si y sólo si* $x \geq F^{-1}(t).$

Demostración: Mostremos que F^{-1} es creciente. Sea $0 < p < q < 1$, entonces

$$\{x \in \mathbb{R} : F(x) \geq q\} \subset \{x \in \mathbb{R} : F(x) \geq p\}.$$

Así que

$$\inf\{x \in \mathbb{R} : F(x) \geq p\} \leq \inf\{x \in \mathbb{R} : F(x) \geq q\}$$

y

$$F^{-1}(p) \leq F^{-1}(q).$$

Por lo tanto, F^{-1} es creciente.

(i) Sea $x \in \mathbb{R}$ tal que $0 < F(x) < 1$.

$$x \in \{y \in \mathbb{R} : F(y) \geq F(x)\}.$$

Entonces

$$\inf\{y \in \mathbb{R} : F(y) \geq F(x)\} \leq x.$$

Por lo tanto,

$$F^{-1}(F(x)) \leq x.$$

(ii) Sea $0 < t < 1$ y $C = \{x \in \mathbb{R} : F(x) \geq t\}$.
Observemos que si $y \in C$ entonces $F(y) \geq t$.
Sea

$$\alpha := \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

Entonces, si $h > 0$. $F(\alpha + h) \geq t$, y por ser F continua por la derecha,

$$F(\alpha) = \lim_{h \to 0} F(\alpha + h) \geq t.$$

Entonces $\alpha \in C$ y $F(F^{-1}(t)) \geq t$.

(iii) Supongamos $F(x) \geq t$. Entonces

$$x \in \{y \in \mathbb{R} : F(y) \geq t\} \text{ y } F^{-1}(t) = \inf\{y \in \mathbb{R} : F(y) \geq t\} \leq x.$$

Por lo tanto,

$$F(x) \geq t \text{ implica } F^{-1}(t) \leq x.$$

Por otro lado, si $x \geq F^{-1}(t)$ entonces, al ser F creciente y por la relación (ii),

$$F(x) \geq F(F^{-1}(t)) \geq t.$$

Por lo tanto,

$$x \geq F^{-1}(t) \text{ implica } F(x) \geq t.$$

□

1.2. La función de distribución empírica

Para un conjunto de datos $X_1, X_2, \ldots, X_n$ con escala nominal u ordinal, o de tipo univariados o multivariados y, $A \subset S$ un conjunto de interés asociado con los n datos, la cantidad

$$\mu_n(A) := \frac{1}{n} \sum_{i=1}^{n} 1_A(X_i) \tag{1.1}$$

se llama la **distribución empírica (medida)** de A. Notemos que $\mu_n(A)$ es la proporción de datos que caen en el conjunto A.

Cuando los datos tienen escala métrica, los intervalos extendidos $(-\infty, t]$ con $t \in \mathbb{R}$, son los conjuntos de interés. En este caso

$$F_n(t) \equiv \mu_n(-\infty, t] = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq t\}}, \tag{1.2}$$

se le llama la **función de distribución empírica** de $X_1, X_2, \ldots, X_n$.

Notemos que si $s \leq t$, $\{X_i \leq s\} \subset \{X_i \leq t\}$ para todo $i = 1, 2, \ldots, n$. Entonces $1_{\{X_i \leq s\}} \leq 1_{\{X_i \leq t\}}$ para todo $i = 1, 2, \ldots, n$. Por lo tanto

$$\frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq s\}} \leq \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq t\}}$$

y

$$F_n(s) \leq F_n(t) \quad \text{para } s \leq t,$$

i.e., F_n es creciente. Además, si $s \downarrow -\infty$ y $t \uparrow \infty$, obtenemos en el límite

$$F_n(-\infty) = 0 \ \text{ y } \ F_n(\infty) = 1.$$

La función indicadora de un conjunto $A \subset S$, está intimamente relacionada con la **medida de Dirac**. Para cada $x \in S$ y $A \subset S$.

$$\delta_x(A) = \begin{cases} 1, & \text{si} \quad x \in A \\ 0, & \text{si} \quad x \notin A. \end{cases}$$

En este caso,

$$\delta_x(A) = 1_A(x).$$

Notemos que en δ_x, x permanece fijo y A varía sobre la clase de conjuntos de S, mientras que para la función indicadora, el que permanece fijo es el conjunto A, variando x sobre S.

δ_x es una medida de probabilidad, en efecto:

Sea S un conjunto y $A \subset S$. Consideremos $\mathcal{A} = \{A, A^c, \emptyset, S\}$ la σ-álgebra generada por A. Entonces

(i)

$$\delta_x(\Omega) = \delta_x(S) = \begin{cases} 1, & \text{si} \quad x \in S \\ 0, & \text{si} \quad x \notin S. \end{cases}$$

Como $S \subset S$ y $x \in S$, entonces $\delta_x(S) = 1$.

(ii) Por definición de δ_x, $\delta_x(A) \geq 0$ para todo $A \in \mathcal{A}$.

(iii) Demostremos que

$$\delta_x(A \cup A^c) = \delta_x(A) + \delta_x(A^c).$$

Observemos que

$$\delta_x(A \cup A^c) = \delta_x(S) = 1.$$

Por otra parte, como $x \in S$, entonces $x \in A$ o $x \in A^c$.

Entonces, si $x \in A$, $\delta_x(A) = 1$ y $\delta_x(A^c) = 0$. Por lo tanto,

$$\delta_x(A) + \delta_x(A^c) = 1.$$

Si $x \in A^c$, $\delta_x(A^c) = 1$ y $\delta_x(A) = 0$. Por lo tanto,

$$\delta_x(A) + \delta_x(A^c) = 1.$$

Así que,

$$\delta_x(A \cup A^c) = \delta_x(A) + \delta_x(A^c)$$

y δ_x es una medida de probabilidad.

La expresión (1.1) ahora se puede escribir como

$$\mu_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) \tag{1.3}$$

o simplemente como

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Ahora introducimos lo que llamaremos **integrales empíricas**, que serán de gran utilidad en lo que sigue. Empezamos observando que

$$\delta_x(A) = \int 1_A d\delta_x = 1_A(x).$$

Entonces, si ponemos $\varphi = 1_A$, la segunda igualdad de la expresión anterior se puede escribir como

$$\int \varphi d\delta_x = \varphi(x).$$

Esta expresión se puede extender a otras $\varphi' s$, que no son necesariamente funciones indicadoras como veremos mas adelante. Puesto que μ_n es una combinación lineal de las masas puntuales δ_{X_i}, $1 \leq i \leq n$. obtenemos

$$\int \varphi d\mu_n = \int \varphi d(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}) = \frac{1}{n}\sum_{i=1}^{n} \int \varphi d\delta_{X_i} = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i).$$

Así, tenemos la expresión

$$\int \varphi d\mu_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i), \tag{1.4}$$

i.e., una integral empírica es justamente el promedio de la φ evaluada en los datos.

La medida producto $\mu_n \otimes \mu_n$ está unicamente determinada a través de conjuntos rectangulares $A_1 \times A_2$. Entonces la expresión (1.3) se escribe ahora

$$\begin{aligned}\mu_n \otimes \mu_n(A_1 \times A_2) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{X_i}(A_1)\delta_{X_j}(A_2)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}1_{\{X_i\in A_1, X_j\in A_2\}}.\end{aligned}$$

Así que para una función arbitraria de dos variables $\varphi = \varphi(x,y)$ la extensión de (1.4) se escribe

$$\int \varphi d(\mu_n \otimes \mu_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi(X_i, X_j).$$

Veamos un ejemplo.

1.2.1. Ejemplo. Si $X_1, X_2, \ldots, X_n$ son observaciones que toman valores reales, y si $\varphi(x,y) = \frac{1}{2}(x-y)^2$, entonces al evaluar $\int \varphi d(\mu_n \otimes \mu_n)$, obtenemos la varianza muestral

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Demostración:

$$\begin{aligned}\int \varphi d(\mu_n \otimes \mu_n) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}(x_i - x_j)^2\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}(x_i^2 - 2x_ix_j + x_j^2)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{n}{2}x_i^2 - x_i\sum_{j=1}^{n}x_j + \frac{1}{2}\sum_{j=1}^{n}x_j^2\right)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{n}{2}x_i^2 - nx_i\bar{x} + \frac{1}{2}\sum_{j=1}^{n}x_j^2\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}x_i^2 - x_i\bar{x} + \frac{1}{2n}\sum_{j=1}^{n}x_j^2\right)\\ &= \frac{1}{n}\left(\frac{1}{2}\sum_{i=1}^{n}x_i^2 - n\bar{x}^2 + \frac{1}{2}\sum_{j=1}^{n}x_j^2\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.
\end{aligned}
$$

Por lo tanto,

$$\int \varphi d(\mu_n \otimes \mu_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma_n^2$$

la varianza muestral. ■

1.3. Distribución empírica con datos reales

En esta sección estudiamos a las distribuciones empíricas, en el caso en que los datos toman valores reales. Así, sean $X_1, X_2, \ldots, X_n$ observaciones con valores reales. Ordenando estos datos en forma creciente obtenemos

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}.$$

La cantidad $X_{i:n}$ se llama la i-ésima **estadística de orden**. En lo que sigue supondremos que los datos no se repiten. Así que ahora se tiene

$$X_{1:n} < X_{2:n} < \ldots < X_{n:n}.$$

Si $X_j = X_{i:n}$, el índice i, determina la posición de X_j en la muestra. Este índice se llama el **rango** de X_j. Si denotamos por R_j al rango de X_j vemos que

$$X_j = X_{R_j:n}.$$

Notemos que el rango de X_j, se puede escribir en terminos de F_n como sigue:

$$
\begin{aligned}
F_n(X_j) &= \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq X_j\}} \\
&= \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq X_{R_j:n}\}} \\
&= \frac{1}{n}R_j.
\end{aligned}
$$

Así que se tiene la relación

$$R_j = nF_n(X_j).$$

De aquí también se deduce que al evaluar la i-ésima estadística de orden en F_n:

$$\frac{i}{n} = F_n(X_{i:n}).$$

Se vio en la sección anterior que F_n es creciente con

$$F_n(-\infty) = 0 \quad \text{y} \quad F_n(\infty) = 1.$$

Ahora mostremos que F_n es continua por la derecha, con la propiedad de que entre dos estadísticas de orden, ésta se mantiene constante.
Mostremos que

$$\lim_{x \downarrow t} F_n(x) = F_n(t).$$

Puesto que

$$\lim_{x \downarrow t} F_n(x) = \lim_{x \downarrow t} \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}},$$

(i) Si $t > X_{n:n}$,

$$\lim_{x \downarrow t} F_n(x) = F_n(t).$$

(ii) Si $t = X_{j:n}$,

$$\lim_{x \downarrow t} F_n(x) = F_n(t).$$

(iii) Si $X_{j:n} < t < X_{j+1:n}$,

$$\lim_{x \downarrow t} F_n(x) = F_n(t).$$

Además, si $X_{j:n} < t < s < X_{j+1:n}$,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq t\}} = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq s\}} = F_n(s).$$

Por lo tanto, F_n es continua por la derecha y se mantiene constante entre dos estadísticas de orden consecutivos. En otras palabras, la función de distribución empírica F_n, es una función de distribución.

Sean $y_1 < y_2 < \ldots < x$. Entonces $F_n(y_1) \leq F_n(y_2) \leq \ldots \leq F_n(x)$. Por lo tanto

$$\{F_n(y_k)\}_{k \geq 1} \quad \text{es una sucesión creciente y acotada,}$$

por eso

$$F_n(y_k) \;\text{ converge a }\; \lim_{k \to \infty} F_n(y_k).$$

A este límite se denotará por $F_n(x-)$, i.e.,

$$F_n(x-) = \lim_{y \uparrow x} F_n(y),$$

el límite por la izquierda de F_n en x.

Definamos la función

$$G_n(x) := F_n(x) - F_n(x-)$$

llamada el **incremento** de F_n en x. Observemos que

$$G_n : \mathbb{R} \longrightarrow \{0, \frac{1}{n}\}$$

definida por

$$G_n(x) = \begin{cases} \frac{1}{n}, & \text{si} \quad x = X_{i:n};\ 1 \leq i \leq n \\ 0, & \text{si} \quad x \neq X_{i:n}. \end{cases}$$

También definimos la **función cuantil empírica** de F_n como sigue:

1.3.1. Definición. *La función cuantil empírica F_n^{-1}, se define como*

$$F_n^{-1}(u) = \inf\{t : F_n(t) \geq u\}, \quad 0 < u \leq 1.$$

La restricción de u en $(0, 1]$, se debe a que si $u = 0$,

$$\begin{aligned} F_n^{-1}(0) &= \inf\{t : F_n(t) \geq 0\} \\ &= \inf(-\infty, \infty) \end{aligned}$$

el cual, no existe. Observemos que

$$F_n^{-1} : (0, 1] \longrightarrow \{X_{1:n}, X_{2:n}, \ldots, X_{n:n}\} \subset \mathbb{R}.$$

en efecto:

Si $\frac{i-1}{n} < u \leq \frac{i}{n}$ para todo $1 \leq i \leq n$, entonces

$$\begin{aligned} F_n^{-1}(u) &= \inf\{t : F_n(t) \geq u\} \\ &= \inf[X_{i:n}, \infty) = X_{i:n}. \end{aligned}$$

Así que F_n^{-1} es creciente y continua por la izquierda. Además, para todo $i = 1, 2 \ldots, n$, si

$$\frac{i-1}{n} < u < \frac{i}{n}, \ F_n(F_n^{-1}(u)) = F_n(X_{i:n}) = \frac{i}{n} > u.$$

y si

$$u = \frac{i}{n},\ F_n(F_n^{-1}(u)) = F_n(X_{i:n}) = \frac{i}{n} = u.$$

En otras palabras,

$$F_n(F_n^{-1}(u)) \geq u \quad \text{para } 0 < u \leq 1,$$

y la igualdad se da, sólo cuando $u = \frac{i}{n}$, $1 \leq i \leq n$.

Por otro lado, $F_n^{-1}(F_n(x))$ está definida unicamente para aquellas $x's$ con $x \geq X_{1:n}$, debido a que $0 \in DomF_n$, y $F_n(x) = 0$ si, y sólo si $x < X_{1:n}$.

Luego, para todo $i = 2, 3, \ldots, n$, si $X_{i-1:n} \leq x < X_{i:n}$,

$$\begin{aligned} F_n^{-1}(F_n(x)) &= F_n^{-1}\left(\frac{i-1}{n}\right) \\ &= \inf\left\{t \in \mathbb{R} : F_n(t) \geq \frac{i-1}{n}\right\} \\ &= \inf[X_{i-1:n}, \infty) = X_{i-1:n}. \end{aligned}$$

Así que para todo $i = 2, 3, \ldots, n$, si

$$X_{i-1:n} < x < X_{i:n},\ F_n^{-1}(F_n(x)) = F_n^{-1}\left(\frac{i-1}{n}\right) = X_{i-1:n} < x.$$

Si

$$x > X_{n:n},\ F_n^{-1}(F_n(x)) = F_n^{-1}(1) = X_{n:n} < x,$$

y para todo $i = 1, 2 \ldots, n$, si

$$x = X_{i:n},\ F_n^{-1}(F_n(x)) = F_n^{-1}\left(\frac{i}{n}\right) = X_{i:n} = x.$$

Por lo tanto se concluye que

$$F_n^{-1}(F_n(x)) \leq x \quad \text{para todo } x \geq X_{1:n},$$

y la igualdad se da, sólo cuando $x = X_{i:n}$.

Veamos algunos ejemplos de cómo las estadísticas de orden y rangos se pueden usar para construir expresiones estadísticas.

1.3.2. Ejemplo. Sean $X_1, X_2, \ldots, X_n$ observaciones con valores reales, y J una función de peso definida sobre el intervalo unitario. Entonces

$$\int xJ(F_n(x))dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i J\left(\frac{R_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) X_{i:n} \tag{1.5}$$

Demostración:

$$\int xJ(F_n(x))dF_n(x) = \int_{-\infty}^{\infty} xJ(F_n(x))dF_n(x)$$

$$\begin{aligned}
&= 0 + \int_{[X_{1:n},X_{2:n})} xJ(F_n(x))dF_n(x) + \ldots + \int_{[X_{n:n},\infty)} xJ(F_n(x))dF_n(x) \\
&= \int_{[X_{1:n},X_{2:n})} xJ\left(\frac{1}{n}\right)dF_n(x) + \ldots + \int_{[X_{n:n},\infty)} xJ(1)dF_n(x) \\
&= J\left(\frac{1}{n}\right)\int_{[X_{1:n},X_{2:n})} xdF_n(x) + \ldots + J(1)\int_{[X_{n:n},\infty)} xdF_n(x) \\
&= J\left(\frac{1}{n}\right)\left(\frac{X_{1:n}}{n}\right) + \ldots + J(1)\left(\frac{X_{n:n}}{n}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right)X_{i:n}.
\end{aligned}$$

Por lo tanto,

$$\int xJ(F_n(x))dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right)X_{i:n}.$$

El cero que aparece en el primer sumando de la segunda igualdad, se debe a que

$$\begin{aligned}
\int_{(-\infty,X_{1:n})} xJ(F_n(x))dF_n(x) &= \int_{(-\infty,X_{1:n})} xJ(0)dF_n(x) \\
&= J(0)\int_{(-\infty,X_{1:n})} xdF_n(x) \\
&= J(0)(0) = 0.
\end{aligned}$$

Falta comprobar que

$$\frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right)X_{i:n} = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{R_i}{n}\right)X_i.$$

En efecto:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{R_i}{n}\Big)X_i &= \frac{1}{n}\Big[J\Big(\frac{R_1}{n}\Big)X_1 + J\Big(\frac{R_2}{n}\Big)X_2 + \ldots + J\Big(\frac{R_n}{n}\Big)X_n\Big] \\
&= \frac{1}{n}\Big[J\Big(\frac{R_1}{n}\Big)X_{R_1:n} + J\Big(\frac{R_2}{n}\Big)X_{R_2:n} + \ldots + J\Big(\frac{R_n}{n}\Big)X_{R_n:n}\Big] \\
&= \frac{1}{n}\Big[J\Big(\frac{1}{n}\Big)X_{1:n} + J\Big(\frac{2}{n}\Big)X_{2:n} + \ldots + J(1)X_{n:n}\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{i}{n}\Big)X_{i:n}.
\end{aligned}$$

La segunda igualdad es porque $X_j = X_{R_j:n}$. En la tercera igualdad, simplemente se ordena la suma. Por lo tanto

$$\frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{i}{n}\Big)X_{i:n} = \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{R_i}{n}\Big)X_i$$

y

$$\int xJ(F_n(x))dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i J\Big(\frac{R_i}{n}\Big) = \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{i}{n}\Big)X_{i:n}.$$

■

Usando directamente este ejemplo vemos que si J es una función de peso definida sobre $(0,1)$,

$$\begin{aligned}
\int xJ\Big(\frac{n}{n+1}F_n(x)\Big)dF_n(x) &= \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{n}{n+1}\frac{i}{n}\Big)X_{i:n} \\
&= \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{i}{n+1}\Big)X_{i:n}.
\end{aligned}$$

Por lo tanto,

$$\int xJ\Big(\frac{n}{n+1}F_n(x)\Big)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{i}{n+1}\Big)X_{i:n}.$$

Observemos que si $J \equiv 1$, esta expresión es la media muestral.
Veamos otro ejemplo.

1.3.3. Ejemplo. Sean $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_m$. dos conjuntos de datos, con F_n y G_n sus f.d.'s empíricas respectivas. Si $H_{n,m}$ es la f.d. empírica de la muestra combinada de tamaño $n+m$, entonces para una función de peso J,

$$\int J\Big(\frac{n+m}{n+m+1}H_{n,m}(x)\Big)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} J\Big(\frac{R_i}{n+m+1}\Big),$$

donde ahora, R_i es el rango de X_i en la muestra combinada.

Demostración: Sea

$$Z_1 = X_1, Z_2 = X_2, \ldots, Z_n = X_n, Z_{n+1} = Y_1, Z_{n+2} = Y_2, \ldots, Z_{n+m} = Y_m$$

y

$$Z_{1:n+m} < Z_{2:n+m} < \ldots < Z_{n:n+m} < \ldots < Z_{n+m:n+m}$$

las estadísticas de orden. Entonces

$$H_{n,m}(z) = \frac{1}{n+m}\sum_{i=1}^{n+m} 1_{\{Z_i\le z\}} = \frac{1}{n+m}\sum_{i=1}^{n+m}\delta_{Z_i}(-\infty, z),$$

y

$$H_{n,m}(z) = \begin{cases} 0, & \text{si } \; z < Z_{1:n+m} \\ \frac{i}{n+m}, & \text{si } \; Z_{i:n+m} \le z < Z_{i+1:n+m} \quad \text{para todo } 1 \le i \le n+m-1 \\ 1, & \text{si } z \ge Z_{n+m:n+m}. \end{cases}$$

$$F_n(x) = \begin{cases} 0, & \text{si } \; x < X_{1:n} \\ \frac{i}{n}, & \text{si } \; X_{i:n} \le x < X_{i+1:n} \quad \text{para todo } 1 \le i \le n-1 \\ 1, & \text{si } \; x \ge X_{n:n}. \end{cases}$$

Luego,

$$\int J\Big(\frac{n+m}{n+m+1}H_{n,m}(x)\Big)dF_n(x) = \int_{-\infty}^{\infty} J\Big(\frac{n+m}{n+m+1}H_{n,m}(x)\Big)dF_n(x).$$

Separando esta integral en los intervalos

$$(-\infty, Z_{1:n+m}), [Z_{1:n+m}, Z_{2:n+m}), \ldots, [Z_{n+m:n+m}, \infty),$$

vemos que el primer sumando de esta integral es cero. En efecto:

$$\begin{aligned} \int_{(-\infty, Z_{1:n+m})} J\Big(\frac{n+m}{n+m+1}H_{n,m}(x)\Big)dF_n(x) &= \int_{(-\infty, Z_{1:n+m})} J(0)dF_n(x) \\ &= J(0)\int 1_{(-\infty, Z_{1:n+m})}(x)dF_n(x) \\ &= J(0)(0) = 0. \end{aligned}$$

La expresión de la integral es ahora

$$\int_{[Z_{1:n+m}, Z_{2:n+m})} J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x) + \dots$$
$$+ \int_{[Z_{n+m:n+m}, \infty)} J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x).$$

Por otro lado, observemos que para todo $1 \le j \le n+m$,

$$\int_{[Z_{j:n+m}, Z_{j+1:n+m})} J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x)$$

$$= J\left(\frac{n+m}{n+m+1}\frac{j}{n+m}\right) \int 1_{[Z_{j:n+m}, Z_{j+1:n+m})}(x) dF_n(x)$$
$$= J\left(\frac{j}{n+m+1}\right) \int 1_{[Z_{j:n+m}, Z_{j+1:n+m})}(x) dF_n(x),$$

donde $[Z_{j:n+m}, Z_{j+1:n+m}) = [Z_{n+m:n+m}, \infty)$, para $j = n+m$.
Luego, para todo $j = 1, 2, \dots, n+m$,

$$\begin{aligned} \int 1_{[Z_{j:n+m}, Z_{j+1:n+m})}(x) dF_n(x) &= \int \delta_x [Z_{j:n+m}, Z_{j+1:n+m}) dF_n(x) \\ &= \begin{cases} \frac{1}{n}, & \text{si} \quad Z_{j:n+m} = X_i \quad \text{p.a. } 1 \le i \le n \\ 0, & \text{si} \quad Z_{j:n+m} \ne X_i \quad \text{para todo } 1 \le i \le n. \end{cases} \end{aligned}$$

Por lo tanto, para todo $j = 1, 2, \dots, n+m$,

$$\int_{[Z_{j:n+m}, Z_{j+1:n+m})} J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x)$$

$$= \begin{cases} J\left(\frac{j}{n+m+1}\right)\frac{1}{n}, & \text{si} \quad Z_{j:n+m} = X_i \quad \text{p.a. } 1 \le i \le n \\ 0, & \text{si} \quad Z_{j:n+m} \ne X_i \quad \text{para todo } 1 \le i \le n. \end{cases}$$

Así que habrá n integrales diferentes de cero, ya que hay n datos $X_1, X_2, \dots, X_n$.
Pero

$$Z_i = X_i = Z_{R_i:n+m} \quad \text{para todo } i = 1, 2 \dots, n.$$

Entonces, las integrales donde no son cero, son donde se integran sobre

$$[Z_{j:n+m}, Z_{j+1:n+m}) = [Z_{R_i:n+m}, Z_{j+1:n+m}).$$

Entonces, para todo $j = 1, 2, \dots, n+m$, podemos escribir

$$\int_{[Z_{j:n+m}, Z_{j+1:n+m})} J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x)$$

$$= \begin{cases} J\left(\frac{R_i}{n+m+1}\right)\frac{1}{n}, & \text{si} \quad Z_{j:n+m} = Z_{R_i:n+m} \quad \text{p.a. } 1 \le i \le n \\ 0, & \text{si} \quad Z_{j:n+m} \ne Z_{R_i:n+m} \quad \text{para todo } 1 \le i \le n. \end{cases}$$

Finalmente, al sumar las n integrales diferentes de cero obtenemos

$$\int J\left(\frac{n+m}{n+m+1} H_{n,m}(x)\right) dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{R_i}{n+m+1}\right).$$

■

Hemos visto con algunos ejemplos que las integrales empíricas pueden ser de gran ayuda para expresar estadísticas de interés. La función φ de (1.4) puede ser independiente de los datos, como por ejemplo $\varphi(x) = x$, o puede depender directamente de los datos de una forma complicada, lo importante es que φ se puede extender a casos generales.

En muchas aplicaciones, se supone que los datos provienen de una misma función de distribución F. La desventaja es que F generalmente se desconoce y necesita ser estimada a través de los datos, única fuente de información.

Si se reemplaza F_n, por F en (1.5), se obtiene

$$\int xJ(F(x))dF(x).$$

Para $J \equiv 1$, la expresión anterior es ahora la esperanza de X:

$$\int xJ(F(x))dF(x) = \int xdF(x) = \mathbb{E}X \equiv \int X d\mathbb{P},$$

cuidando que la integral exista. Este sencillo ejemplo, muestra la posibilidad de tener dos o más opciones para representar cantidades de interés, una representación en términos de variables aleatorias y otra en términos de distribuciones. De esta manera se puede tener a un estimador empírico, para una representación en términos de variables aleatorias, eligiendo adecuadamente una función φ. Veamos algunos ejemplos.

1.3.4. Ejemplo. Si $\varphi(x) = x^k$, el k-ésimo momento de X es

$$\mathbb{E}X^k = \int x^k dF(x).$$

El estimador empírico es entonces

$$\int x^k dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} X_j^k.$$

Demostración:

$$\begin{aligned}
\int_{-\infty}^{\infty} x^k dF_n(x) &= \int_{(-\infty, X_{1:n})} x^k dF_n(x) + \int_{[X_{1:n}, X_{2:n})} x^k dF_n(x) + \dots \\
&+ \int_{[X_{n-1:n}, X_{n:n})} x^k dF_n(x) + \int_{[X_{n:n}, \infty)} x^k dF_n(x) \\
&= 0 + \frac{1}{n}X_{1:n}^k + \dots + \frac{1}{n}X_{n-1:n}^k + \frac{1}{n}X_{n:n}^k \\
&= \frac{1}{n}\sum_{i=1}^{n} X_{i:n}^k = \frac{1}{n}\sum_{j=1}^{n} X_j^k.
\end{aligned}$$

■

Otros ejemplos:

Si $\varphi_t(x) = 1_{\{x \leq t\}}$,

$$\int \varphi_t(x) dF(x) = F(t)$$

respectivamente,

$$\int \varphi_t(x) dF_n(x) = F_n(t).$$

Si $\varphi_t(x) = \exp[tx]$, la transormación de Laplace de X con respecto a F, es

$$t \to \int \exp[tx] dF(x) = \mathbb{E}e^{tX}.$$

El estimador empírico de la transformación de Laplace es

$$t \to \int \exp[tx] dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} \exp[tX_j].$$

Si $\varphi_t(x) = \exp[itx]$, la transformación de Fourier de X con respecto a F es

$$t \to \int \exp[itx]dF(x) = \mathbb{E}e^{itX}.$$

Su estimador empírico es

$$t \to \int \exp[itx]dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} \exp[itX_j].$$

Las transformaciones de Laplace y Fourier, determinan únicamente la distribución de una variable aleatoria X. Por lo tanto, sus versiones empíricas son candidatos para detectar el carácter distribucional de X.

1.4. Transformaciones libres de distribución

Generalmente, el propósito de una transformación libre de distribución, es transformar una variable (observación) X a otra que tenga una distribución conocida. En esta sección sólo consideraremos aquellas variales X que toman valores reales. La transformación más importante es la que transforma una variable con distribución uniforme.

1.4.1. Definición. ***La distribución uniforme*** *(sobre el intervalo [0,1]) está dada por su f.d.*

$$F_U(t) = \begin{cases} 0, & si \quad t < 0 \\ t, & si \quad 0 \leq t \leq 1 \\ 1, & si \quad 1 < t. \end{cases}$$

Por lo que una variable aleatoria de esta distribución solo toma valores en (0,1) con probabilidad uno.

Otra distribución importante, es la **distribución exponencial**.

1.4.2. Definición. *La función de distribuciíon exponencial con parámetro $\lambda > 0$ está dada por*

$$1 - F(t) = \begin{cases} 1, & \text{si} \quad t < 0 \\ exp(-\lambda t), & \text{si} \quad t \geq 0. \end{cases} \tag{1.6}$$

Escribiremos $X \sim Exp(\lambda)$ siempre que X tenga la distribución (1.6). Notemos que X es no negativa con probabilidad uno. Otra cantidad importante es la **función de riesgo acumulativa** de F, definida por

$$\Lambda_F(t) = \int_{-\infty}^{t} \frac{dF(x)}{1 - F(x-)} = \int_{-\infty}^{\infty} \frac{1_{\{x \leq t\}}}{1 - F(x-)} dF(x).$$

La función de riesgo acumulativa juega un papel importante en análisis de supervivencia y confiabilidad, cuando los datos son tiempos de vida (tiempos de fallo) de individuos o componentes físicos. Si F tiene densidad f, la función de riesgo acumulativa Λ_F es

$$\Lambda_F(t) = \int_{-\infty}^{t} \frac{f(x)dx}{1 - F(x)}.$$

La función

$$\lambda(x) = \frac{f(x)}{1 - F(x)},$$

es la **función de riesgo** de F.

La función de Hazard para la f.d. (1.6), es la función constante $\lambda(x) \equiv \lambda$, en efecto:

$$\lambda(x) = \frac{f(x)}{1 - F(x)}.$$

Por definición, $1 - F(x) = exp(-\lambda x)$. Entonces $F(x) = 1 - exp(-\lambda x)$.
Así

$$\frac{dF(x)}{dx} = f(x) = -(-\lambda)exp(-\lambda x) = \lambda exp(-\lambda x).$$

Por lo que

$$\lambda(x) = \frac{\lambda exp(-\lambda x)}{exp(-\lambda x)} = \lambda.$$

El siguiente lema revela una manera de como generar una variable X con f.d. F.

1.4.3. Lema. *Sea U una v.a. uniforme en el [0,1] y F una función de distribución arbitraria. Entonces*

$$X := F^{-1}(U) \sim F. \tag{1.7}$$

Demostración: Sea $t \in \mathbb{R}$. Aplicando el lema 1.1.4(iii), se tiene

$$\begin{aligned}\{X \leq t\} &= \{F^{-1}(U) \leq t\}\\ &= \{U \leq F(t)\}.\end{aligned}$$

Luego

$$P(X \leq t) = P(U \leq F(t)) = F(t).$$

Por lo tanto

$$X \sim F.$$

□

1.4.4. Ejemplo. Si F es la distribución exponencial con parámetro λ, entonces

$$X = -\frac{1}{\lambda}\ln(1-U) \sim Exp(\lambda).$$

Demostración: Por definición

$$1 - F(t) = exp(-\lambda t).$$

Entonces,

$$F(t) = 1 - exp(-\lambda t).$$

Sea

$$u = 1 - exp(-\lambda t).$$

Despejando t en esta igualdad, se tiene:

$$t = \frac{-\ln(1-u)}{\lambda} = F^{-1}(u).$$

Finalmente

$$X = F^{-1}(U) = -\frac{1}{\lambda}\ln(1-U) \sim Exp(\lambda) \tag{1.8}$$

■

1.4.5. Lema. *Supongamos que $X \sim F$ y F es continua. Entonces*

(i) $F(X) \sim F_U$.

(ii) $\Lambda_F(X) \sim Exp(1)$.

Demostración: *(i)* Por (1.7), podemos suponer que $X = F^{-1}(U)$. Entonces por la continuidad de F,

$$(F \circ F^{-1})(u) = F(F^{-1}(u)) = u \quad \text{para todo } 0 < u < 1.$$

Entonces

$$F(X) = U,$$

y

$$F(X) \sim F_U.$$

(ii) Como F es continua, $F(x-) := \lim_{y \uparrow x} F(y) = F(x)$.
Entonces

$$\Lambda_F(t) = \int_{\infty}^{t} \frac{dF(x)}{1 - F(x)} = -\ln[1 - F(t)].$$

Así, sea

$$g(x) = -\ln[1 - F(x)].$$

Entonces

$$\frac{dg(x)}{dx} = \frac{dF(x)}{1 - F(x)},$$

por lo que

$$\int_{-\infty}^{t} \frac{dg(x)}{dx} dx = g(t) = -\ln[1 - F(t)].$$

Luego, $X = F^{-1}(U)$ si y sólo si $F(X) = U$. Se sigue que

$$\Lambda_F(X) = -\ln[1 - F(X)] = -\ln[1 - U].$$

Así, en vista de (1.8)

$$\Lambda_F(X) \sim Exp(1).$$

□

La representación (1.7) es de suma importancia para derivar propiedades distribucionales. Para observaciones independientes X_i, $1 \leq i \leq n$, la distribución de una estadística $S = S(X_1, X_2, \ldots, X_n)$, está únicamente determinada por las distribuciones marginales de las $X_i^{'}s$. Por ejemplo, cuando $X_i \sim F$,

$$S \stackrel{d}{=} S(F^{-1}(U_1), \ldots, F^{-1}(U_n)),$$

para una muestra de observaciones independientes $U_1, \ldots, U_n$, con distribución F_U.

Capítulo 2

El Proceso Empírico Univariado

2.1. Hechos Básicos

En el capítulo anterior, estudiamos propiedades de la función de distribución empírica sin requerir de la suposición distribucional en las observaciones. Ahora será crucial conocer la estructura distribucional. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Recordemos que $X_1, X_2, \ldots, X_n$, constituyen una muestra aleatoria de tamaño n con función de distribución F. si $X_1, X_2, \ldots, X_n$, son variables aleatorias independientes e identicamente distribuidas (v.a.i.i.d.), con función de distribución común F, y con valores en algún espacio medible $(S, \mathcal{S})$. En este capítulo denotaremos por μ a la función de distribución de X_1 sobre $\mathcal{S}$, i.e.,

$$\mu(A) = \mathbb{P}(X_1 \in A) \quad \text{para todo } A \in \mathcal{S}.$$

Para todo $A \in \mathcal{S}$, $\mu_n(A)$ denotará a la medida empírica de A, i.e.,

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} 1_A(X_i).$$

El primer resultado en este capítulo, habla acerca de la distribución de $\mu_n(A)$ para observaciones i.i.d.

2.1.1. Lema. *Sean $X_1, \ldots, X_n$ observaciones i.i.d. con distribución μ. En-*

tonces para cualquier conjunto medible A,

$$\mathbb{P}(n\mu_n(A) = k) = \binom{n}{k} \mu^k(A)(1-\mu(A))^{n-k}, \quad k = 0, 1, \ldots, n.$$

Demostración: $n\mu_n(A)$ es la suma de n Bernoulli's independientes con parámetro $\mu(A)$. Por lo tanto, $n\mu_n(A)$ se distribuye Binomial con parámetros n y $\mu(A)$, i.e.,

$$\mathbb{P}(n\mu_n(A) = k) = \binom{n}{k} \mu^k(A)(1-\mu(A))^{n-k}.$$

□

Para intervalos extendidos $A = (-\infty, t]$, este lema toma una forma especial.

2.1.2. Lema. *Sean $X_1, \ldots, X_n$ i.i.d. con distribución F. Entonces*

(i) $\mathbb{P}(nF_n(t) = k) = \binom{n}{k} F^k(t)(1-F(t))^{n-k}, \quad k = 0, 1, \ldots, n.$

(ii) $\mathbb{E}F_n(t) = F(t)$.

(iii) $VarF_n(t) = \frac{1}{n}F(t)(1-F(t))$.

Demostración: (i). Es un caso particular del lema anterior.
(ii). De (i) se sigue que, $\mathbb{E}(nF_n(t)) = nF(t)$.

Por otra parte, $\mathbb{E}(nF_n(t)) = n\mathbb{E}(F_n(t))$. Entonces se concluye que

$$\mathbb{E}F_n(t) = F(t).$$

(iii). Por (i), $Var(nF_n(t)) = nF(t)(1-F(t))$.

Por otra parte, $Var(nF_n(t)) = n^2 VarF_n(t)$. De estas dos igualdades se sigue que

$$VarF_n(t) = \frac{1}{n}F(t)(1-F(t).$$

□

La afirmación (ii) establece que $F_n(t)$ es un **estimador insesgado** de $F(t)$.

Observemos que de (ii) y (iii), si $n \to \infty$,

$$F_n(t) \to F(t) \quad \text{en } L^2(\Omega, F, P),$$

en efecto

$$\begin{aligned}\mathbb{E}|F_n(t) - F(t)|^2 &= \mathbb{E}(F_n(t) - F(t))^2 \\ &= VarF_n(t) \\ &= \frac{1}{n}F(t)(1 - F(t)).\end{aligned}$$

Esta última igualdad converge a cero, cuando $n \to \infty$. Por lo tanto, $F_n(t)$ es un estimador que converge en L^2 a $F(t)$, si $n \to \infty$. Aquí,

$$L^2(\Omega, F, P) := \{\xi : \Omega \to \mathbb{R} \mid \xi \text{ es medible y} \int \mid \xi \mid^2 dP < \infty\},$$

con la **L^2-norma**

$$\parallel \xi \parallel_2 := \left[\int \xi^2 d\mathbb{P}\right]^{1/2}, \quad \xi \in L^2(\Omega, F, \mathbb{P}).$$

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias i.i.d. y $t \in \mathbb{R}$. Entonces

$$1_{\{X_1 \le t\}}, 1_{\{X_2 \le t\}}, \ldots, 1_{\{X_n \le t\}} \quad \text{son i.i.d..}$$

Además

$$nF_n(t) = \sum_{i=1}^{n} 1_{\{X_i \le t\}},$$

y

$$\mathbb{E}(1_{\{X_i \le t\}}) = \mathbb{P}(X_i \le t) = F(t) \quad \text{para todo } i = 1, 2, \ldots, n.$$

Entonces, aplicando la **ley fuerte de los grandes números** (LFGN), ver apéndice C.4, para cada $t \in \mathbb{R}$ fijo, se tiene:

$$\frac{nF_n(t)}{n} \xrightarrow{c.s.} F(t)$$

es decir,

$$F_n \xrightarrow{c.s.} F(t).$$

Así que

$$\lim_{n \to \infty} F_n(t) = F(t) \text{ con probabilidad uno.} \tag{2.1}$$

En otras palabras, $F_n(t)$ es un estimador **fuertemente consistente** de $F(t)$.

Finalmente, si $t \in \mathbb{R}$ y $S_n = nF_n(t)$ entonces, por el lema anterior,

$$\mathbb{E}S_n = nF(t) \quad \text{y} \quad VarS_n = nF(t)(1 - F(t)).$$

Entonces, la versión de De Moivre-Laplace del **Teorema del Límite Central** (TLC), ver apéndice C.4, se tiene que si $n \to \infty$,

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{VarS_n}} \longrightarrow \mathcal{N}(0,1).$$

En otras palabras,

$$\frac{nF_n(t) - nF(t)}{\sqrt{nF(t)(1 - F(t))}} \longrightarrow \mathcal{N}(0,1).$$

La expresión del lado izquierdo se puede simplificar a,

$$\frac{nF_n(t) - nF(t)}{\sqrt{nF(t)(1 - F(t))}} = \frac{n^{1/2}(F_n(t) - F(t))}{[F(t)(1 - F(t))]^{1/2}}.$$

Entonces, aplicando el Teorema de Slutsky, ver Serfling, R.(1980)

$$n^{1/2}(F_n(t) - F(t)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2), \tag{2.2}$$

donde $\mathcal{L}$ denota la **convergencia en ley (o en distribución)**, y $\mathcal{N}(0, \sigma^2)$ es la **distribución normal** con esperanza cero y varianza σ^2. En nuestro caso y de acuerdo al lema 2.1.2(iii),

$$\sigma^2 = \sigma_t^2 = F(t)(1 - F(t)).$$

2.1.3. Lema. *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con f.d. F, y sean $\varphi, \varphi_1, \varphi_2 : \mathbb{R} \to \mathbb{R}$ funciones arbitrarias (Borel medibles). Suponiendo que todas las integrales del lado derecho existen, se tiene:*

(i) $\mathbb{E}[\int \varphi dF_n] = \int \varphi dF$

(ii) $Var[\int \varphi dF_n] = \frac{1}{n}\left[\int \varphi^2 dF - (\int \varphi dF)^2\right]$

(iii) $Cov[\int \varphi_1 dF_n, \int \varphi_2 dF_n] = \frac{1}{n}[\int \varphi_1\varphi_2 dF - \int \varphi_1 dF \int \varphi_2 dF]$

(iv) $\lim_{n\to\infty} \int \varphi dF_n = \int \varphi dF$ *con probabilidad uno*

(v) $n^{1/2}[\int \varphi dF_n - \int \varphi dF] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, *con*

$$\sigma^2 = \int \varphi^2 dF - \left(\int \varphi dF\right)^2.$$

Demostración: (i)

$$\begin{aligned}\mathbb{E}\Big[\int \varphi dF_n\Big] &= \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i)\Big] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\varphi(X_i)\\ &= \mathbb{E}\varphi(X_1) = \int \varphi dF,\end{aligned}$$

La última igualdad se sigue de una simple transformación de integrales.

La afirmación (ii) es un caso especial de (iii), con $\varphi = \varphi_1 = \varphi_2$. Así, por (i), $Cov\Big(\int \varphi_1 dF_n, \int \varphi_2 dF_n\Big)$

$$\begin{aligned} &= \mathbb{E}\Big[\int \varphi_1 dF_n \int \varphi_2 dF_n\Big] - \int \varphi_1 dF \int \varphi_2 dF\\ &= \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\varphi_1(X_i)\frac{1}{n}\sum_{j=1}^{n}\varphi_2(X_j)\Big] - \int \varphi_1 dF \int \varphi_2 dF\\ &= \frac{1}{n^2}\mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi_1(X_i)\varphi_2(X_j)\Big] - \int \varphi_1 dF \int \varphi_2 dF\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(\varphi_1(X_i)\varphi_2(X_j)) - \int \varphi_1 dF \int \varphi_2 dF\\ &= \frac{1}{n}\Big[\int \varphi_1\varphi_2 dF - \int \varphi_1 dF \int \varphi_2 dF\Big].\end{aligned}$$

La última igualdad se debe a que $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(\varphi_1(X_i)\varphi_2(X_j))$

$$\begin{aligned} &= \sum_{i=1}^{n}\mathbb{E}(\varphi_1(X_i)\varphi_2(X_i)) + \sum\sum\nolimits_{i\neq j}\mathbb{E}(\varphi_1(X_i)\varphi_2(X_j))\\ &= \sum_{i=1}^{n}\mathbb{E}(\varphi_1(X_i)\varphi_2(X_i)) + \sum\sum\nolimits_{i\neq j}\mathbb{E}(\varphi_1(X_i))\mathbb{E}(\varphi_2(X_j))\\ &= n\int \varphi_1\varphi_2 dF + n(n-1)\int \varphi_1 dF \int \varphi_2 dF.\end{aligned}$$

Para la afirmaciones (iv) y (v), llamemos $S_n = \sum_{i=1}^{n}\varphi(X_i)$.
Por hipótesis $X_1, \ldots, X_n$ son i.i.d., entonces $\varphi(X_1), \ldots \varphi(X_n)$ son i.i.d.
Además, por (i)

$$\mu := \mathbb{E}(\varphi(X_1)) = \int \varphi dF < \infty \quad \text{y}.$$

aplicando LFGN, se tiene:

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu.$$

En otras palabras,

$$\lim_{n\to\infty} \int \varphi dF_n = \int \varphi dF.$$

Así se tiene (iv).

Por (ii)

$$\sigma^2 = Var\varphi(X_1) = \int \varphi^2 dF - \Big(\int \varphi dF \Big)^2 < \infty.$$

Aplicando TLC, se tiene:

$$n^{1/2}\Big[\frac{S_n}{n} - \mu\Big] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

en otras palabras,

$$n^{1/2}\Big[\int \varphi dF_n - \int \varphi dF \Big] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Así, hemos demostrado (v) □

Veamos algunas aplicaciones.

2.1.4. Ejemplo. Si $\varphi_1 = 1_{(-\infty,s]}$ y $\varphi_2 = 1_{(-\infty,t]}$, entonces

$$Cov[F_n(s), F_n(t)] = \frac{1}{n}[F(s \wedge t) - F(s)F(t)], \tag{2.3}$$

donde $s \wedge t := min(s,t)$.

Demostración:

$$\int \varphi_1 dF_n = \frac{1}{n}\sum_{i=1}^{n} \varphi_1(X_i) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq s\}} = F_n(s),$$

$$\int \varphi_2 dF_n = \frac{1}{n}\sum_{i=1}^{n} \varphi_2(X_i) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq t\}} = F_n(t).$$

Entonces

$$Cov\Big(\int \varphi_1 dF_n, \int \varphi_2 dF_n \Big) = Cov[F_n(s), F_n(t)].$$

También se tiene:

$$\varphi_1\varphi_2 = 1_{(-\infty,s]}1_{(-\infty,t]} = 1_{(-\infty,s\wedge t]}.$$

Así, por la parte (iii) lema 2.1.3,

$$\begin{aligned} Cov\Big(\int \varphi_1 dF_n, \int \varphi_2 dF_n\Big) &= \frac{1}{n}\Big[\int \varphi_1\varphi_2 dF - \int \varphi_1 dF \int \varphi_2 dF\Big] \\ &= \frac{1}{n}[F(s\wedge t) - F(s)F(t)]. \end{aligned}$$

■

2.1.5. Ejemplo. Si $\varphi_1 = 1_A$ y $\varphi_2 = 1_B$ con $A \cap B = \emptyset$, entonces,

$$Cov[\mu_n(A), \mu_n(B)] = -\frac{1}{n}\mu(A)\mu(B).$$

Demostración:

$$\int \varphi_1 dF_n = \frac{1}{n}\sum_{i=1}^{n}\varphi_1(X_i) = \frac{1}{n}\sum_{i=1}^{n}1_A(X_i) = \mu_n(A),$$

$$\int \varphi_2 dF_n = \frac{1}{n}\sum_{i=1}^{n}\varphi_2(X_i) = \frac{1}{n}\sum_{i=1}^{n}1_B(X_i) = \mu_n(B).$$

Entonces,

$$\begin{aligned} Cov[\mu_n(A), \mu_n(B)] &= Cov\Big(\int \varphi_1 dF_n, \int \varphi_2 dF_n\Big) \\ &= \frac{1}{n}\Big[\int \varphi_1\varphi_2 dF - \int \varphi_1 dF \int \varphi_2 dF\Big] \\ &= -\frac{1}{n}[\mu(A)\mu(B)]. \end{aligned}$$

La última igualdad se sigue del hecho de que,

$$\int \varphi_1\varphi_2 dF = 0 \quad \text{debido a que} \quad \varphi_1\varphi_2 = 1_A 1_B = 0,$$

y por definición,

$$\int \varphi_1 dF = \int 1_A dF = \mu(A) \quad \text{y} \quad \int \varphi_2 dF = \int 1_B dF = \mu(B).$$

■

Observemos que si $s = t$ en (2.3), se obtiene (iii) del lema 2.1.2. Mas generalmente, (2.3) revela la **estructura de covarianza** de F_n, cuando es considerado como un preceso estocástico. Ahora introducimos el proceso de estandarización

$$\alpha_n(t) := n^{1/2}[F_n(t) - F(t)], \quad t \in \mathbb{R}.$$

Este proceso es llamado **proceso empírico**. Notemos que $\alpha_n(t)$, es función de $F_n(t)$, por lo que es continua por la derecha y tiene límites por la izquierda. Además, puesto que F_n y F son funciones de distribución, se tiene

$$\lim_{t\downarrow-\infty} F_n(t) = 0 = \lim_{t\downarrow-\infty} F(t)$$

y

$$\lim_{t\uparrow\infty} F_n(t) = 1 = \lim_{t\uparrow\infty} F(t).$$

Entonces

$$\lim_{t\downarrow-\infty} \alpha_n(t) = 0 = \lim_{t\uparrow\infty} \alpha_n(t). \tag{2.4}$$

Así que podemos extender continuamente α_n a $\pm\infty$:

$$\alpha_n(-\infty) = 0 = \alpha_n(\infty). \tag{2.5}$$

Este proceso se caracteríza comunnmente como un **tipo de puente**, debido a la relación (2.5) y puesto que la gráfica de este proceso va oscilando en la recta real, con discontinuidades en las observaciones $X_i's$. El siguiente resultado junto con TLC, muestra que el proceso empírico $\alpha_n(t)$ converge debilmente a la distribución normal, con media cero y varianza $F(t)(1 - F(t))$.

2.1.6. Lema. *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con f.d. F. Entonces*

(i) $\mathbb{E}\alpha_n(t) = 0$ *para todo* $t \in \mathbb{R}$

(ii) $Cov[\alpha_n(s), \alpha_n(t)] = F(s)(1 - F(t)), \quad s \le t$

(iii) $\alpha_n(t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(t)(1 - F(t)))$, *cuando* $n \to \infty$.

Demostración: (i) Sea $t \in \mathbb{R}$. Entonces,

$$\begin{aligned}\mathbb{E}\alpha_n(t) &= \mathbb{E}[n^{1/2}(F_n(t) - F(t))] = n^{1/2}\mathbb{E}[F_n(t) - F(t)] \\ &= n^{1/2}[F(t) - F(t)] = 0.\end{aligned}$$

Por lo tanto, $\mathbb{E}\alpha_n(t) = 0$ para todo $t \in \mathbb{R}$.

(ii)

$$\begin{aligned}
Cov[\alpha_n(s), \alpha_n(t)] &= Cov[n^{1/2}(F_n(s) - F(s)), n^{1/2}(F_n(t) - F(t))] \\
&= \mathbb{E}[n^{1/2}(F_n(s) - F(s))n^{1/2}(F_n(t) - F(t))] \\
&= n\mathbb{E}[(F_n(s) - F(s))(F_n(t) - F(t))] \\
&= n\{\mathbb{E}[F_n(s)F_n(t) - F_n(s)F(t) - F(s)F_n(t) + F(s)F(t)]\} \\
&= n\{\mathbb{E}[F_n(s)F_n(t)] - F(s)F(t) - F(s)F(t) + F(s)F(t)\} \\
&= n\{\mathbb{E}[F_n(s)F_n(t)] - F(s)F(t)\} \\
&= n[\mathbb{E}(F_n(s)F_n(t)) - \mathbb{E}(F_n(s))\mathbb{E}(F_n(t))] \\
&= n[Cov(F_n(s), F_n(t))] \\
&= n\left[\frac{1}{n}(F(s \wedge t) - F(s)F(t))\right] \\
&= F(s) - F(s)F(t) \quad \text{ya que } F(s \wedge t) = F(s) \\
&= F(s)(1 - F(t)).
\end{aligned}$$

Por lo tanto $Cov[\alpha_n(s), \alpha_n(t)] = F(s)(1 - F(t))$, si $s \leq t$.

(iii)

$$\alpha_n(t) = n^{1/2}[F_n(t) - F(t)] = n^{1/2}\left[\int 1_{\{X \leq t\}} dF_n - \int 1_{\{X \leq t\}} dF\right]$$

Luego, aplicando (v), lema 2.1.3, con $\varphi(t) = 1_{\{X \leq x\}}$,

$$\alpha_n(t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

donde

$$\sigma^2 = \int 1_{\{X \leq t\}} dF - \left(\int 1_{\{X \leq t\}} dF\right)^2 = F(t) - F(t)^2 = F(t)(1 - F(t)).$$

Por lo tanto

$$\alpha_n(t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(t)(1 - F(t)).$$

□

Retomando a las transformaciones libres de distribución introducidas en la sección 1.4, se tiene que bajo la independencia de $X_1, \ldots, X_n$, podemos escribir (en distribución)

$$X_i \stackrel{d}{=} F^{-1}(U_i) \quad \text{para } i = 1, \ldots, n,$$

para una muestra independiente $U_1, \ldots, U_n$, provenientes de F_U. Luego, por el lema 1.1.4,

$$\{X_i \leq t\} = \{F^{-1}(U_i) \leq t\} = \{U_i \leq F(t)\}.$$

Entonces

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq t\}} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \leq F(t)\}}.$$

Concluímos que

$$F_n(t) = \bar{F}_n(F(t)), \tag{2.6}$$

donde

$$\bar{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \leq u\}}, \quad 0 \leq u \leq 1,$$

es la función de distribución empírica de la **muestra uniforme** $U_1, \ldots, U_n$. La representación (2.6), es de suma importancia para obtener algunas propiedades distribucionales relacionadas con F_n y α_n. Cuando F no es continua, crean problemas. Con la relación (2.6), quedan resueltos, debido a que F_U es continua y además con soporte compacto $[0, 1]$.

El siguiente lema, caracteríza una representación de la estadística de orden X, en términos de una estadística de orden uniforme.

2.1.7. Lema. *Para* $0 < u < 1$,

$$F_n^{-1}(u) = F^{-1}(\bar{F}_n^{-1}(u)). \tag{2.7}$$

En particular, para $u = i/n$, *se tiene*

$$X_{i:n} = F^{-1}(U_{i:n}), \quad 1 \leq i \leq n.$$

Demostración: Sea $0 < u < 1$. Sabemos que

$$F_n(F_n^{-1}(u)) \geq u \quad \text{para todo } 0 < u < 1.$$

En particular

$$\bar{F}_n(\bar{F}_n^{-1}(u)) \geq u \quad \text{para todo } 0 < u < 1.$$

Por otro lado, puesto que $\bar{F}_n$ es creciente,

$$F(t) \geq \bar{F}_n^{-1}(u) \text{ si y sólo si } \bar{F}_n(F(t)) \geq \bar{F}_n(\bar{F}_n^{-1}(u)) \geq u$$

Luego, por definición de función cuantil y de la relación (2.6),

$$\begin{aligned}(F^{-1} \circ \bar{F}_n^{-1})(u) &= F^{-1}(\bar{F}_n^{-1}(u)) \\ &= \inf\{t : F(t) \geq \bar{F}_n^{-1}(u)\} \\ &= \inf\{t : \bar{F}_n(F(t)) \geq u\} \\ &= \inf\{t : (\bar{F}_n \circ F)(t) \geq u\} \\ &= \inf\{t : F_n(t) \geq u\} \\ &= F_n^{-1}(u).\end{aligned}$$

Por lo tanto,

$$(F^{-1} \circ \bar{F}_n^{-1})(u) = F_n^{-1}(u).$$

La otra parte se sigue, puesto que

$$F_n^{-1}(i/n) = X_{i:n} \quad \text{y} \quad \bar{F}_n^{-1}(i/n) = U_{i:n}.$$

□

Como se vio en el capítulo 1, una medida para describir la posición de una observación dentro de la muestra, se denominó el rango de la observación. Sea

$$R = (R_1, \cdots, R_n)$$

el vector de rangos de la muestra $X_1, \ldots, X_n$. El siguiente resultado muestra que R es libre de distribución, cuando F, la función de distribución de la muestra es continua.

2.1.8. Lema. *Sea R el vector de rangos para una sucesión de observaciones i.i.d. con f.d. continua F. Entonces la distribución de R, es libre de distribución, i.e., no depende de F.*

Demostración: Puesto que F es continua,

$$(F \circ F^{-1})(u) = u \quad \text{para todo } 0 < u < 1.$$

Sea $X_i = F^{-1}(U_i)$ para todo $1 \leq i \leq n$. Entonces para todo $1 \leq i \leq n$, por (2.6) y el hecho de que $X_i \stackrel{d}{=} F^{-1}(U_i)$, se tiene

$$\begin{aligned}R_i &= nF_n(X_i) \\ &= n\bar{F}_n(F(X_i)) \\ &= n\bar{F}_n(F(F^{-1}(U_i))) \\ &= n\bar{F}_n(U_i).\end{aligned}$$

Luego,

$$R = (R_1, \ldots, R_n) = (n\bar{F}_n(U_1), \ldots, n\bar{F}_n(U_n)) \quad \text{no depende de } F.$$

□

Este lema, no especifíca la distribución del vector de rangos. Sólo garantiza que es suficiente considerar observaciones que estén distribuidas uniformemente.

2.1.9. Lema. *Sean $R = (R_1, \ldots, R_n)$ y $r = (r_1, \ldots, r_n)$. Para observaciones i.i.d. con f.d. continua F, el vector de rangos R. está distribuida uniformemente sobre el conjunto de permutaciones $1, \ldots, n$. Es decir,*

$$\mathbb{P}(R = r) = \frac{1}{n!}.$$

Demostración: Es claro que no puede haber repeticiones en los datos, debido a la continuidad de F. Se sigue entonces que cada rango está bien definido. El vector R toma valores en el conjunto de todas las permutaciones $\{1, \ldots, n\}$. Así, al evento $\{R = r\}$ le corresponde una única ordenación de las $U_i's$. Luego, por la independencia de las $U_i's$, cada una de las $n!$ posibles ordenaciones, tiene probabilidad

$$\begin{aligned}
\mathbb{P}(R = r) &= \int \cdots \int_{0<u_1<\ldots<u_n<1} 1 du_1 \ldots du_n \\
&= \int_0^1 \cdots \int_0^{u_4} \int_0^{u_3} \int_0^{u_2} 1 du_1 du_2 du_3 \ldots du_n \\
&= \int_0^1 \cdots \int_0^{u_4} \int_0^{u_3} u_2 du_2 du_3 \ldots du_n \\
&= \int_0^1 \cdots \int_0^{u_4} \frac{u_3^2}{2} du_3 \ldots du_n \\
&= \int_0^1 \frac{u_n^{n-1}}{(n-1)!} du_n \\
&= \frac{1}{n!}
\end{aligned}$$

Por lo tanto,

$$\mathbb{P}(R = r) = \frac{1}{n!}.$$

□

El presente lema tiene una aplicación a la construcción de intervalos de predicción, para posibles valores de X de una muestra $X_1, \ldots, X_n$. Para ello, ordenando $X_1, \ldots, X_n$, se obtienen las estadísticas de orden

$$-\infty < X_{1:n} < X_{2:n} < \ldots < X_{n:n} < \infty.$$

Estos n puntos dividen a la recta real en $n+1$ intervalos aleatorios $I_1 = (-\infty, X_{1:n}], I_2 = (X_{1:n}, X_{2:n}], \cdots, I_{n+1} = (X_{n:n}, \infty)$. Sea X_{n+1} otra observación X proveniente de F siendo independiente de las $X_i^{'}s$ previas. Entonces el evento $\{X_{n+1} \in I_i\}$ es equivalente al evento de que el rango de X_{n+1} de la muestra $X_1, \ldots, X_n, X_{n+1}$ sea igual a i. Entonces se tiene

$$\mathbb{P}(X_{n+1} \in I_i) = \frac{1}{n+1},$$

en efecto

$$\mathbb{P}(X_{n+1} \in I_i) = \mathbb{P}(R_{n+1} = i).$$

Por otro lado,

$$\begin{aligned} 1 &= \mathbb{P}\left(X_{n+1} \in \bigcup_{i=1}^{n+1} I_i\right) = \sum_{i=1}^{n+1} \mathbb{P}(X_{n+1} \in I_i) \\ &= \sum_{i=1}^{n+1} \mathbb{P}(R_{n+1} = i). \end{aligned}$$

Luego, si $R = (R_1, \ldots, R_{n+1})$ y $r = (r_1, \ldots, r_{n+1})$, por el lema anterior,

$$\mathbb{P}(R = r) = \frac{1}{(n+1)!}.$$

Ahora, como R_{n+1} puede tomar valores $1, 2, \ldots, n+1$, cada uno es igualmente probable, i.e.,

$$\mathbb{P}(R_{n+1} = i) = \mathbb{P}(R_{n+1} = j).$$

Por lo tanto,

$$\mathbb{P}(X_{n+1} \in I_i) = \frac{1}{n+1} \quad \text{para cada } 1 \leq i \leq n+1.$$

El próximo resultado se debe a Glivenko y Cantelli, el cual constituye una extensión de (2.1), la convergencia uniforme de F_n a F. Este resultado ha sido una parte central en la inferencia estadística moderna, resultado fundamental sobre la convergencia en funciones de distribución empíricas.

2.1.10. Teorema. ***(Glivenko-Cantelli).*** *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con f.d. arbitraria F. Entonces, cuando $n \to \infty$,*

$$D_n := \sup_{t \in \mathbb{R}} | F_n(t) - F(t) | \to 0 \quad \textit{con probabilidad uno.}$$

Demostración: Puesto que $F_n(t) = \bar{F}_n(F(t))$, se tiene

$$D_n = \sup_{t \in \mathbb{R}} | \bar{F}_n(F(t)) - F(t) | . \tag{2.8}$$

Sea $u = F(t)$. Entonces $0 \leq u \leq 1$ y,

$$D_n \leq \sup_{0 \leq u \leq 1} | \bar{F}_n(u) - u | \equiv \bar{D}_n. \tag{2.9}$$

Luego, es suficiente acotar $\bar{D}_n$. Así, sea $\varepsilon > 0$ y $0 = u_0 < u_1 < u_2 < \ldots < u_k = 1$ una partición de $[0,1]$ tal que,

$$\max_{0 \leq i \leq k} (u_{i+1} - u_i) \leq \varepsilon.$$

Afirmación 1.

$$\bar{F}_n(u) - u \leq \bar{F}_n(u_{i+1}) - u_i \leq | \bar{F}_n(u_{i+1}) - u_{i+1} | + \varepsilon.$$

En fecto,
sea u tal que $u_i \leq u$ y $u \leq u_{i+1}$. Entonces

$$-u \leq -u_i \quad \text{y} \quad \bar{F}_n(u) \leq \bar{F}_n(u_{i+1}).$$

Sumando las desigualdades, obtenemos

$$\bar{F}_n(u) - u \leq \bar{F}_n(u_{i+1}) - u_i \qquad \text{(I)}.$$

Por otro lado,

$$\max_{0 \leq i \leq k-1} (u_{i+1} - u_i) \leq \varepsilon.$$

Entonces, para todo $0 \leq i \leq k-1$,

$$u_{i+1} - u_i \leq \varepsilon,$$

es decir, $-u_i - \varepsilon \leq -u_{i+1}$. De aquí se sigue que

$$\bar{F}_n(u_{i+1}) - u_i - \varepsilon \leq \bar{F}_n(u_{i+1}) - u_{i+1} \leq | \bar{F}_n(u_{i+1}) - u_{i+1} | .$$

Por lo que

$$\bar{F}_n(u_{i+1}) - u_i \leq \mid \bar{F}_n(u_{i+1}) - u_{i+1} \mid + \varepsilon \qquad \text{(II)}.$$

Por lo tanto, de (I) y (II),

$$\bar{F}_n(u) - u \leq \bar{F}_n(u_{i+1}) - u_i \leq \mid \bar{F}_n(u_{i+1}) - u_{i+1} \mid + \varepsilon.$$

Afirmación 2.

$$-\varepsilon - \mid \bar{F}_n(u_i) - u_i \mid \leq \bar{F}_n(u) - u.$$

Como

$$u_{i+1} - u_i \leq \varepsilon \quad \text{y} \quad u \leq u_{i+1} \quad \text{para todo } 0 \leq i \leq k-1,$$

se sigue que

$$u - \varepsilon \leq u_i \quad \text{para todo } 0 \leq i \leq k-1.$$

Por otro lado, como $u_i \leq u$,

$$\bar{F}_n(u_i) \leq \bar{F}_n(u) \ \text{ si, y sólo si } \ -\bar{F}_n(u) \leq -\bar{F}_n(u_i).$$

Entonces

$$u - \varepsilon - \bar{F}_n(u) \leq u_i - \bar{F}_n(u_i) \leq \mid u_i - \bar{F}_n(u_i) \mid = \mid \bar{F}_n(u_i) - u_i \mid.$$

De aquí se sigue

$$-\varepsilon - \mid \bar{F}_n(u_i) - u_i \mid \leq \bar{F}_n(u) - u.$$

Combinando las dos afirmaciones se tiene,

$$-\varepsilon - \mid \bar{F}_n(u_i) - u_i \mid \leq \bar{F}_n(u) - u \leq \mid \bar{F}_n(u_{i+1}) - u_{i+1} \mid + \varepsilon.$$

Entonces,

$$\begin{aligned} \mid \bar{F}_n(u) - u \mid &\leq \text{máx}\{\mid \bar{F}_n(u_{i+1}) - u_{i+1} \mid + \varepsilon, \mid \bar{F}_n(u_i) - u_i \mid + \varepsilon\} \\ &\leq \max_{0 \leq i \leq k} \mid \bar{F}_n(u_i) - u_i \mid + \varepsilon. \end{aligned}$$

Por lo tanto,

$$\mid \bar{F}_n(u) - u \mid \leq \max_{0 \leq i \leq k} \mid \bar{F}_n(u_i) - u_i \mid + \varepsilon.$$

Luego, sea

$$\mid \bar{F}_n(u_{i_0}) - u_{i_0} \mid + \varepsilon = \max_{0 \leq i \leq k} \mid \bar{F}_n(u_i) - u_i \mid + \varepsilon \quad i_0 \in \{0, \ldots, k\}.$$

Entonces

$$| \bar{F}_n(u) - u | \leq | \bar{F}_n(u_{i_0}) - u_{i_0} | + \varepsilon \quad \text{donde } 0 \leq u_{i_0} \leq 1.$$

Si $u_{i_0} = 0$, $| \bar{F}_n(u_{i_0}) - u_{i_0} | \leq \frac{1}{n}$ porque $\bar{F}_n(0) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{U_i \leq 0\}} \leq \frac{1}{n}$. Así,

$$| \bar{F}_n(u) - u | \leq \frac{1}{n} + \varepsilon.$$

Entonces

$$\bar{D}_n = \sup_{0 \leq u \leq 1} | \bar{F}_n(u) - u | \leq \frac{1}{n} + \varepsilon.$$

Por lo que

$$\lim_{n \to \infty} D_n \leq \lim_{n \to \infty} \bar{D}_n \leq \lim_{n \to \infty} \frac{1}{n} + \varepsilon = \varepsilon.$$

Entonces

$$\lim_{n \to \infty} D_n \leq \varepsilon \quad \text{para todo } \varepsilon > 0.$$

Por lo tanto

$$\sup_{t \in \mathbb{R}} | F_n(t) - F(t) | \longrightarrow 0 \quad \text{cuando } n \to \infty.$$

Si $u_{i_0} \neq 0$, sea $u_{i_0} = F(t_0)$, donde $t_0 = \inf\{t : F(t) \geq u_{i_0}\}$.
Entonces

$$\begin{aligned} | \bar{F}_n(u_{i_0}) - u_{i_0} | &= | \bar{F}_n(F(t_0)) - F(t_0) | \\ &= | F_n(t_0) - F(t_0) | . \end{aligned}$$

Así que

$$| \bar{F}_n(u) - u | \leq | F_n(t_0) - F(t_0) | + \varepsilon.$$

Entonces

$$D_n \leq \bar{D}_n = \sup_{0 \leq u \leq 1} | \bar{F}_n(u) - u | \leq | F_n(t_0) - F(t_0) | + \varepsilon.$$

Luego, puesto que $| F_n(t_0) - F(t_0) | \longrightarrow 0$ cuando $n \to \infty$.

$$\lim_{n \to \infty} D_n \leq \lim_{n \to \infty} \bar{D}_n \leq \varepsilon \quad \text{para todo } \varepsilon > 0.$$

Por lo tanto

$$\sup_{t \in \mathbb{R}} | F_n(t) - F(t) | \longrightarrow 0 \quad \text{cuando } n \to \infty.$$

□

La variable D_n, proporciona la **distancia** o **discrepancia** entre F_n y F. D_n se conoce como la **distancia K-S**, debido a que en una prueba de hipótesis, cuando F se reemplaza por una distribución conocida, la prueba resultante se llama la prueba Kolmogorov-Smirnov, ver [4].

Cuando F es continua, se tiene la igualdad en (2.9), esto es.

$$D_n = \bar{D}_n.$$

Se sigue entonces que

$$D_n \text{ es libre de distribución cuando } F \text{ es continua.} \tag{2.10}$$

El próximo lema, presenta un algoritmo del cálculo de D_n en un número finito de pasos.

2.1.11. Lema. *Para una F continua,*

$$D_n = D_n^* := \max_{1\leq i\leq n}\left\{ \max\left[F(X_{i:n}) - \frac{i-1}{n}, \frac{i}{n} - F(X_{i:n})\right]\right\}.$$

Demostración: Sea $X_{i:n} \leq t < X_{i+1:n}$, $1 \leq i < n$. Entonces

$$F_n(t) - F(t) \leq F_n(X_{i:n}) - F(X_{i:n}) = \frac{i}{n} - F(X_{i:n}) \leq D_n^*,$$

en efecto

$$F_n(X_{i:n}) = F_n(t)$$

y,

$$F(X_{i:n}) \leq F(t) \text{ si, y sólo si } -F(t) \leq -F(X_{i:n}).$$

Restando, obtenemos

$$\begin{aligned} F_n(t) - F(t) &\leq F_n(X_{i:n}) - F(X_{i:n}) \\ &= \frac{i}{n} - F(X_{i:n}) \leq D_n^*. \end{aligned}$$

Por lo tanto,

$$F_n(t) - F(t) \leq F_n(X_{i:n}) - F(X_{i:n}) = \frac{i}{n} - F(X_{i:n}) \leq D_n^*.$$

También se tiene

$$F_n(t) - F(t) \geq F_n(X_{i:n}) - F(X_{i+1:n}) \geq -D_n^*.$$

en efecto

$$F(t) \leq F(X_{i+1:n}) \text{ si, y sólo si } -F(t) \geq -F(X_{i+1:n}).$$

Entonces

$$\begin{aligned} F_n(t) - F(t) &\geq F_n(X_{i:n}) - F(X_{i+1:n}) \\ &= \frac{i}{n} - F(X_{i+1:n}) \geq -D_n^*. \end{aligned}$$

Por lo tanto,

$$F_n(t) - F(t) \geq F_n(X_{i:n}) - F(X_{i+1:n}) \geq -D_n^*.$$

Ahora, sea $t < X_{1:n}$, entonces

$$-D_n^* \leq -F(X_{1:n}) \leq F_n(t) - F(t) \leq 0 \leq D_n^*,$$

en efecto, para $i = 1$,

$$D_n^* \geq \text{máx}[F(X_{1:n}), \frac{1}{n} - F(X_{1:n})] \geq F(X_{1:n}).$$

Entonces

$$-D_n^* \leq -F(X_{1:n}).$$

Luego, como $t < X_{1:n}$, $F_n(t) = 0 \leq F_n(X_{1:n})$ y puesto que F es creciente $F(t) \leq F(X_{1:n})$ si, y sólo si $-F(X_{1:n}) \leq -F(t)$. Finalmente, combinando las desigualdades se tiene

$$-D_n^* \leq -F(X_{1:n}) \leq F_n(t) - F(t) \leq 0 \leq D_n^*.$$

Mientras que para $t > X_{n:n}$,

$$-D_n^* \leq 0 \leq F_n(t) - F(t) \leq 1 - F(X_{n:n}) \leq D_n^*,$$

en efecto, para $i = 1$,

$$D_n^* \geq \text{máx}[F(X_{1:n}), \frac{1}{n} - F(X_{1:n})] \geq F(X_{1:n}) \geq 0,$$

entonces

$$-D_n^* \leq 0.$$

Por otro lado, como $t > X_{n:n}$,

$$F_n(X_{n:n}) = 1 = F_n(t) \text{ y } F(X_{n:n}) \leq F(t) \text{ si, y sólo si } -F(t) \leq -F(X_{n:n}).$$

Sumando $F_n(t)$ a esto último se tiene

$$F_n(t) - F(t) \leq 1 - F(X_{n:n}).$$

También se tiene $0 \leq F_n(t) - F(t)$ pues $F_n(t) = 1$. Luego, para $i = n$,

$$\frac{i}{n} - F(X_{i:n}) = 1 - F(X_{n:n}) \leq D_n^*.$$

Finalmente, combinando las desigualdades obtenemos

$$-D_n^* \leq 0 \leq F_n(t) - F(t) \leq 1 - F(X_{n:n}) \leq D_n^*.$$

Resumiendo, obtenemos

$$D_n \leq D_n^*.$$

Para la otra desigualdad; para todo $1 \leq i \leq n$,

$$\begin{aligned} \frac{i}{n} - F(X_{i:n}) = F_n(X_{i:n}) - F(X_{i:n}) &\leq \mid F_n(X_{i:n}) - F(X_{i:n}) \mid \\ &\leq \sup_{t \in \mathbb{R}} \mid F_n(t) - F(t) \mid = D_n. \end{aligned}$$

Por lo tanto,

$$\frac{i}{n} - F(X_{i:n}) \leq D_n \quad \text{para todo } 1 \leq i \leq n.$$

luego, como F es continua y F_n tiene límites por la izquierda,

$$\begin{aligned} F(X_{i:n}) - \frac{i-1}{n} = \lim_{t \uparrow X_{i:n}} [F(t) - F_n(t)] &\leq \lim_{t \uparrow X_{i:n}} \mid F_n(t) - F(t) \mid \\ &= \mid F_n(X_{i:n}-) - F(X_{i:n}-) \mid \\ &\leq \sup_{t \in \mathbb{R}} \mid F_n(t) - F(t) \mid = D_n. \end{aligned}$$

Por lo tanto,

$$F(X_{i:n}) - \frac{i-1}{n} \leq D_n \quad \text{para todo } 1 \leq i \leq n.$$

Luego

$$D_n^* \leq D_n.$$

Por lo tanto,

$$D_n = D_n^*.$$

□

Se sigue ahora del lema 2.1.11 que

$$D_n = \text{máx}(D_n^+, D_n^-),$$

donde

$$D_n^+ = \sup_{t\in\mathbb{R}}[F_n(t) - F(t)] = \max_{1\le i\le n}\left[\frac{i}{n} - F(X_{i:n})\right]$$

y

$$D_n^- = \sup_{t\in\mathbb{R}}[F(t) - F_n(t)] = \max_{1\le i\le n}\left[F(X_{i:n}) - \frac{i-1}{n}\right],$$

en efecto,

$$D_n^* = \text{máx}\left\{\max_{1\le i\le n}\left[F(X_{i:n}) - \frac{i-1}{n}\right], \max_{1\le i\le n}\left[\frac{i}{n} - F(X_{i:n})\right]\right\}.$$

Entonces, si

$$D_n^* = \max_{1\le i\le n}\left[F(X_{i:n}) - \frac{i-1}{n}\right].$$

$$D_n^* = \sup_{t\in\mathbb{R}}[F(t) - F_n(t)] := D_n^-.$$

debido a que

$$\max_{1\le i\le n}\left[F(X_{i:n}) - \frac{i-1}{n}\right] = \max_{1\le i\le n}[F(X_{i:n}-) - F_n(X_{i:n}-)] = D_n.$$

Si

$$D_n^* = \max_{1\le i\le n}\left[\frac{i}{n} - F(X_{i:n})\right],$$

$$D_n^* = \sup_{t\in\mathbb{R}}[F_n(t) - F(t)] := D_n^+.$$

debido a que

$$\max_{1\le i\le n}\left[\frac{i}{n} - F(X_{i:n})\right] = \max_{1\le i\le n}[F_n(X_{i:n}) - F(X_{i:n})] = D_n.$$

Por lo tanto,

$$D_n = \text{máx}(D_n^+, D_n^-).$$

D_n^+ y D_n^- denotan las desviaciones entre F_n y F.

Observemos que si $X_i = F^{-1}(U_i)$ para todo $1 \leq i \leq n$, donde $U_1, \ldots, U_n$ es una muestra de variables aleatorias independientes uniformes $(0,1)$, implica

$$F(X_i) = U_i \quad \text{para todo } 1 \leq i \leq n.$$

Entonces

$$D_n^+ = \max_{1 \leq i \leq n} \left[\frac{i}{n} - U_{i:n} \right]$$

y

$$D_n^- = \max_{1 \leq i \leq n} \left[U_{i:n} - \frac{i-1}{n} \right].$$

Ambos no dependen de F. Así, D_n^+ y D_n^- son libres de distribución, mientras F sea continua. Además, ambos tienen la misma distribución. Para ver esto, como antes, pongamos $U_i = F(X_i)$ para todo $1 \leq i \leq n$, y sea $U_i^* = 1 - U_i$.
Veamos que U_i^* son i.i.d. con f.d. F_U.
Como las $U_i's$ son independientes para todo $1 \leq i \leq n$, entonces $U_i^* = 1 - U_i$ son independientes para todo $1 \leq i \leq n$. Luego,

$$\begin{aligned} \mathbb{P}(U_i^* \leq u^*) &= \mathbb{P}(1 - U_i \leq u^*) \\ &= \mathbb{P}(U_i \geq 1 - u^*) \\ &= 1 - \mathbb{P}(U_i < 1 - u^*) \\ &= 1 - (1 - u^*) \\ &= u^*. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(U_i^* \leq u^*) = u^* \quad \text{para todo } 1 \leq i \leq n.$$

Así,

$$U_i^* \sim F_U \quad \text{para todo } 1 \leq i \leq n.$$

Luego,

$$D_n^{+*} = D_n^-,$$

en efecto

$$\begin{aligned} D_n^{+*} &= \max_{1 \leq i \leq n} \left[\frac{i}{n} - U_{i:n}^* \right] \\ &= \max_{1 \leq i \leq n} \left[\frac{i}{n} - (1 - U_{n-i+1:n}) \right] \\ &= \max_{1 \leq i \leq n} \left[U_{n-i+1:n} - \frac{n-i}{n} \right]. \end{aligned}$$

Sea $j = n - i + 1$. Entonces,

$$D_n^{+*} = \max_{1\leq j\leq n}\left[U_{j:n} - \frac{j-1}{n}\right] = D_n^{-}.$$

Por lo tanto,

$$D_n^{+*} = \max_{1\leq i\leq n}\left[\frac{i}{n} - U_{i:n}^{*}\right] = D_n^{-}.$$

Finalmente, $U_{i:n}^{*}$ está en función de $1 - U_{n-i+1:n}$ y

$$D_n^{+} = D_n^{-} \quad \text{en distribución.}$$

Los siguientes resultados son dedicados a la consistencia de los cuantiles empíricos. Estos se definen con F_n^{-1} y F^{-1}.

El siguiente lema, muestra que la distancia K-S entre $\bar{F}_n^{-1}$ y la función identidad, es igual a la distancia K-S entre $\bar{F}_n$ y la función identidad.

2.1.12. Lema.

$$\bar{D}_n := \sup_{0\leq u\leq 1} \mid \bar{F}_n(u) - u \mid = \sup_{0<u<1} \mid \bar{F}_n^{-1}(u) - u \mid =: \bar{D}_n^{**}.$$

Demostración: Si F es continua,

$$D_n = \sup_{t\in\mathbb{R}} \mid F_n(t) - F(t) \mid = \sup_{0\leq u\leq 1} \mid \bar{F}_n(u) - u \mid = \bar{D}_n.$$

Luego, por el lema 2.1.11,

$$\bar{D}_n = \max_{1\leq i\leq n}\left\{ \text{máx}[F(X_{i:n}) - \frac{i-1}{n}, \frac{i}{n} - F(X_{i:n})]\right\}.$$

Si $X_i = F^{-1}(U_i)$ entonces $F(X_i) = U_i$ para todo $1 \leq i \leq n$. Así que

$$\bar{D}_n = \max_{1\leq i\leq n}\left\{ \text{máx}[U_{i:n} - \frac{i-1}{n}, \frac{i}{n} - U_{i:n}]\right\}.$$

Ahora, puesto que

$$U_{i:n} - \frac{i-1}{n} = \lim_{u\downarrow\frac{i-1}{n}} [\bar{F}_n^{-1}(u) - u]$$

y

$$\frac{i}{n} - U_{i:n} = \frac{i}{n} - \bar{F}_n^{-1}\left(\frac{i}{n}\right),$$

entonces,

$$\frac{i}{n} - \bar{F}_n^{-1}\left(\frac{i}{n}\right) \leq \mid \bar{F}_n^{-1}\left(\frac{i}{n}\right) - \frac{i}{n} \mid \leq \sup_{0<u<1} \mid \bar{F}_n^{-1}(u) - u \mid = \bar{D}_n^{**}$$

y

$$\lim_{u \downarrow \frac{i-1}{n}} [\bar{F}_n^{-1}(u) - u] \leq \lim_{u \downarrow \frac{i-1}{n}} \mid \bar{F}_n^{-1}(u) - u \mid \leq \sup_{0<u<1} \left|\bar{F}_n^{-1}(u) - u\right| = \bar{D}_n^{**}.$$

Por lo tanto,

$$\bar{D}_n \leq \bar{D}_n^{**}.$$

Para la otra desigualdad, si

$$\frac{i-1}{n} < u \leq \frac{i}{n} \quad \text{entonces} \quad \bar{F}_n^{-1}(u) = U_{i:n}.$$

Entonces,

$$-\bar{D}_n \leq U_{i:n} - \frac{i}{n} \leq \bar{F}_n^{-1}(u) - u \leq U_{i:n} - \frac{i-1}{n} \leq \bar{D}_n.$$

En efecto, por definición de $\bar{D}_n$,

$$-\bar{D}_n \leq U_{i:n} - \frac{i}{n}.$$

Luego, $u \leq \frac{i}{n}$ implica $-\frac{i}{n} \leq -u$. Entonces,

$$U_{i:n} - \frac{i}{n} \leq \bar{F}_n^{-1}(u) - u.$$

Por otra parte,

$$\frac{i-1}{n} < u \quad \text{implica} \quad -u < -\frac{i-1}{n}.$$

Entonces,

$$\bar{F}_n^{-1}(u) - u \leq U_{i:n} - \frac{i-1}{n}.$$

Juntando las desigualdades obtenidas, se tiene:

$$-\bar{D}_n \leq U_{i:n} - \frac{i}{n} \leq \bar{F}_n^{-1}(u) - u \leq U_{i:n} - \frac{i-1}{n} \leq \bar{D}_n.$$

Entonces,

$$\mid \bar{F}_n^{-1}(u) - u \mid \leq \bar{D}_n.$$

Así que

$$\sup_{0<u<1} \mid \bar{F}_n^{-1}(u) - u \mid \leq \bar{D}_n.$$

Por lo tanto,

$$\bar{D}_n^{**} \leq \bar{D}_n$$

y

$$\bar{D}_n = \bar{D}_n^{**}.$$

□

Una aplicación del Teorema Glivenko-Cantelli para una muestra uniforme, junto con el teorema 2.1.12, es el siguiente:

$$\sup_{0<u<1} \mid \bar{F}_n^{-1}(u) - u \mid = \bar{D}_n^{**} \longrightarrow 0 \quad \text{con probabilidad uno,}$$

en efecto

$$\sup_{0<u<1} \mid \bar{F}_n^{-1}(u) - u \mid = \sup_{0\leq u\leq 1} \mid \bar{F}_n(u) - u \mid = \sup_{t\in\mathbb{R}} \mid F_n(t) - F(t) \mid .$$

El siguiente lema, proporciona la consistencia puntual entre F_n^{-1} y F^{-1}, cuando F^{-1} es continua.

2.1.13. Lema. *Sea F^{-1} continua en $0 < u < 1$. Entonces*

$$\lim_{n\to\infty} F_n^{-1}(u) = F^{-1}(u) \quad \textit{con probabilidad uno.}$$

Demostración: Sea $0 < u < 1$. Aplicando el lema 2.1.7. se tiene

$$F_n^{-1}(u) = F^{-1}(\bar{F}_n^{-1}(u)).$$

Luego,

$$\begin{aligned}
\lim_{n\to\infty} F_n^{-1}(u) &= \lim_{n\to\infty} F^{-1}(\bar{F}_n^{-1}(u)) \\
&= F^{-1}(\lim_{n\to\infty} \bar{F}_n^{-1}(u)) \\
&= F^{-1}(u).
\end{aligned}$$

Por lo tanto,

$$\lim_{n\to\infty} F_n^{-1}(u) = F^{-1}(u) \quad \text{con probabilidad uno.}$$

□

Luego,

$$F^{-1} \text{ es continua en } u, \text{ si y sólo si } F(t) > u, \text{ para cada } t > F^{-1}(u).$$

en efecto:

Supongamos que F^{-1} es continua en u, y sea $t > F^{-1}(u)$. Entonces, por el lema 1.1.4(ii),

$$F(t) \geq F(F^{-1}(u)) \geq u.$$

Así,

$$F(t) = u \quad \text{o} \quad F(t) > u.$$

Si $F(t) = u$,

$$u = F(t) \geq F(F^{-1}(u)) \geq u.$$

Entonces

$$F(F^{-1}(u)) = u.$$

Por otra parte, como F es creciente, existe $s > t$ tal que $F(s) > F(t) = u$. Luego,

$$F^{-1}(u + \varepsilon) = \inf\{x : F(x) \geq u + \varepsilon\} \geq t > F^{-1}(u), \quad \text{para todo } \varepsilon > 0.$$

Así que

$$F^{-1}(u + \varepsilon) - F^{-1}(u) > 0, \quad \text{para todo } \varepsilon > 0.$$

Entonces, F^{-1} no es continua en u, y esto contradice la hipótesis. Por lo tanto,

$$F(t) > u.$$

La proposición anterior dice que F no puede ser constante en un intervalo no-degenerado(intervalo con un sólo punto) $[F^{-1}(u), a]$. Dicho en otras palabras, F^{-1} no puede ser discontinua en todo un intervalo. De aquí que F^{-1} tiene a lo mas, un número numerable de discontinuidades. Si F^{-1} fuera continua en algún subintervalo compacto $[u_0, u_1]$ de $(0, 1)$, entonces, una modificación de la prueba del teorema 2.1.10, dá la convergencia uniforme sobre $[u_0, u_1]$:

$$\sup_{u_0 \leq u \leq u_1} \mid F_n^{-1}(u) - F^{-1}(u) \mid \longrightarrow 0 \quad \text{con probabilidad uno.}$$

El argumento de la compacidad, es crucial para obtener la convergencia uniforme. Por ejemplo, si F tiene soporte no acotado, entonces $X_{1:n} \leq F_n^{-1}(u) \leq X_{n:n}$ y F^{-1} no está acotada. Por eso,

$$\sup_{0<u<1} \mid F_n^{-1}(u) - F^{-1}(u) \mid = \infty.$$

Un ejemplo es la función de riesgo acumulativa. En el lema 1.1.4 de la sección 1.4, vimos que la función de Hazard acumulativa cuando F es continua tiene la expresión:

$$\Lambda_F(t) = -\ln[1 - F(t)].$$

entonces,

$$\Lambda_F(t) \uparrow \infty \quad \text{cuando } t \to \infty.$$

Luego, su estimador empírico

$$\Lambda_n(t) = \int_{-\infty}^{t} \frac{dF_n x}{1 - F_n(x-)},$$

es una función acotada, por eso

$$\sup_t \mid \Lambda_n(t) - \Lambda_F(t) \mid = \infty.$$

Otro ejemplo cuya función de distribución tiene soporte no acotado es la distribución normal.

2.2. Distribuciones finito-dimensionales

En la sección anterior vimos que para variables aleatorias i.i.d. con distribución μ,

$$\mathbb{P}(n\mu_n(A) = k) = \binom{n}{k} \mu^k(A)(1 - \mu(A))^{n-k}, \quad k = 0, 1, \ldots, n,$$

i.e., $n\mu_n(A) \sim Bin(n, \mu(A))$, una **distribución Binomial** con parámetros n y $p = \mu(A)$.

La colección $\{A, A^c\}$ forma una partición del espacio muestral S, donde A^c denota el complemento de A. Luego, debido a que el número de observaciones n es fijo, dado el evento $\{n\mu_n(A) = k\}$, se tiene automaticamente $\{n\mu_n(A^c) = n - k\}$. En otras palabras,

$$\{n\mu_n(A) = k\} = \{n\mu_n(A) = k,\ n\mu_n(A^c) = n - k\}.$$

Generalizamos esta idea a más de dos conjuntos. Sea $\{A_1, A_2, \ldots, A_m\}$ una partición de conjuntos medibles del espacio muestral, i.e.,

$$A_i \cap A_j = \emptyset \ \text{ para } i \neq j, \ \ i,j = 1, \ldots, m \text{ y } \bigcup_{i=1}^{m} A_i = S.$$

El siguiente lema dice que la distribución conjunta de $n\mu_n(A_i)$, para todo $i = 1, \ldots, m$ sigue una **distribución multinomial**, como era de esperarse.

2.2.1. Lema. *Para observaciones i.i.d. con distribución μ,*

$$\mathbb{P}(n\mu_n(A_i) = k_i;\ 1 \leq i \leq m) = \binom{n}{k_1, \ldots, k_m} \prod_{i=1}^{m} [\mu(A_i)]^{k_i},$$

para $k_i = 1, \ldots, n$ tal que $\sum_{i=1}^{m} k_i = n$.

Demostración:

$$n\mu_n(A_j) = \sum_{i=1}^{n} 1_{A_j}(X_i) = k_j \quad \text{para todo } j = 1, 2, \ldots, m.$$

Entonces,

$$\begin{aligned} \sum_{j=1}^{m} k_j &= \sum_{j=1}^{m} \Big(\sum_{i=1}^{n} 1_{A_j}(X_i) \Big) \\ &= \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} 1_{A_j}(X_i) \Big) \\ &= \sum_{i=1}^{n} 1_S(X_i) = n. \end{aligned}$$

En la tercera igualdad, se debe a que $\sum_{j=1}^{m} 1_{A_j}(X_i) = 1_S(X_i)$, donde $S = \bigcup_{j=1}^{m} A_j$. Por lo tanto,

$$\sum_{i=1}^{m} k_i = n.$$

Luego,

$$\mathbb{P}(n\mu_n(A_1) = k_1, \ldots, n\mu_n(A_m) = k_m) = \binom{n}{k_1, \ldots, k_m} \prod_{i=1}^{m} [\mu(A_i)]^{k_i},$$

donde

$$\begin{aligned}\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\ldots-k_{m-1}}{k_m} &= \frac{n!}{k_1!k_2!\ldots k_m!}\\ &= \binom{n}{k_1,\ldots,k_m}\end{aligned}$$

□

También se tiene un resultado con conjuntos crecientes en vez de ajenos.

2.2.2. Lema. *Sean $C_1 \subset C_2 \subset \ldots \subset C_m$ y enteros $0 \leq k_1 \leq k_2 \leq \ldots \leq k_m \leq n$. Entonces*

$$\mathbb{P}(n\mu_n(C_i) = k_i;\ 1 \leq i \leq m) = \binom{n}{n_1, n_2, \ldots, n_{m+1}} \prod_{i=1}^{m+1} [\mu(B_i)]^{n_i},$$

donde $n_1 = k_1, n_i = k_i - k_{i-1}$ para $2 \leq i \leq m$ y $n_{m+1} = n - k_m$. $B_1 = C_1, B_i = C_i \setminus C_{i-1}$ para $2 \leq i \leq m$ y $B_{m+1} = S \setminus C_m$.

Demostración: Es claro que

$$\{B_1, B_2, \ldots, B_m, B_{m+1}\} \text{ es una partición de } S.$$

$$\sum_{i=1}^{m+1} n_i = n,$$

en efecto:

$$\begin{aligned}\sum_{i=1}^{m+1} n_i &= k_1 + \sum_{i=2}^{m} n_i + n_{m+1}\\ &= k_1 + \sum_{i=2}^{m}(k_i - k_{i-1}) + n - k_m\\ &= k_1 + (k_m - k_1) + n - k_m = n.\end{aligned}$$

Aplicando el lema 2.2.1, se tiene

$$\mathbb{P}(n\mu_n(B_i) = n_i;\ 1 \leq i \leq m+1) = \binom{n}{n_1, n_2, \ldots, n_{m+1}} \prod_{i=1}^{m+1} [\mu(B_i)]^{n_i},$$

Luego, el evento

$\{n\mu_n(B_i) = n_i;\ 1 \leq i \leq m+1\}$ es igual al evento $\{n\mu_n(C_i) = k_i;\ 1 \leq i \leq m\}$.

y esto completa la prueba. □

El mismo lema puede reescribirse mediante condicionales como sigue:

2.2.3. Lema. *Con las mismas hipótesis del lema 2.2.2,*

$$\begin{aligned}
&\mathbb{P}(n\mu_n(C_m) = k_m \mid n\mu_n(C_i) = k_i;\ 1 \le i \le m-1) \\
&= \mathbb{P}(n\mu_n(C_m) = k_m \mid n\mu_n(C_{m-1}) = k_{m-1}) \\
&= \binom{n-k_{m-1}}{k_m - k_{m-1}} \left[\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{k_m - k_{m-1}} \left[1 - \frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{n-k_m}.
\end{aligned}$$

Demostración: $\mathbb{P}(n\mu_n(C_m) = k_m \mid n\mu_n(C_i) = k_i;\ 1 \le i \le m-1)$

$$= \frac{\mathbb{P}(n\mu_n(C_i) = k_i,\ 1 \le i \le m)}{\mathbb{P}(n\mu_n(C_i) = k_i,\ 1 \le i \le m-1)} = *$$

Para el numerador se aplica el lema 2.2.2. Para el denominador también se aplica el mismo lema, donde $n_1 = k_1, n_i = k_i \setminus k_{i-1}$ para $2 \le i \le m-1$ y $n_m' = n - k_{m-1}$. $B_1 = C_1$, $B_i = C_i \setminus C_{i-1}$ para $2 \le i \le m-1$ y $B_m' = S \setminus C_{m-1}$. Entonces

$$\begin{aligned}
* &= \frac{\binom{n}{n_1, n_2 \ldots, n_{m+1}} \prod_{i=1}^{m+1} [\mu(B_i)]^{n_i}}{\binom{n}{n_1, n_2 \ldots, n_m'} \prod_{i=1}^{m-1} [\mu(B_i)]^{n_i} [\mu(B_m')]^{n_m'}} \\
&= \frac{\frac{n!}{n_1! n_2! \ldots n_m! n_{m+1}!}}{\frac{n!}{n_1! n_2! \ldots n_{m-1}! n_m'!}} \frac{[\mu(B_m)]^{n_m} [\mu(B_{m+1})]^{n_{m+1}}}{[\mu(B_m')] n_m'} \\
&= \frac{n_m'!}{n_m! n_{m+1}!} \frac{[\mu(C_m \setminus C_{m-1})]^{k_m - k_{m-1}} [\mu(S \setminus C_m)]^{n-k_m}}{[\mu(S \setminus C_{m-1})]^{n-k_{m-1}}}.
\end{aligned}$$

Observemos que

$$\frac{n_m'!}{n_m! n_{m+1}!} = \frac{(n-k_{m-1})!}{(k_m - k_{m-1})!(n-k_m)!} = \binom{n-k_{m-1}}{k_m - k_{m-1}}$$

y

$$\frac{[\mu(C_m \setminus C_{m-1})]^{k_m - k_{m-1}} [\mu(S \setminus C_m)]^{n-k_m}}{[\mu(S \setminus C_{m-1})]^{n-k_{m-1}}}$$

$$
\begin{aligned}
&= \frac{[\mu(C_m \setminus C_{m-1})]^{k_m-k_{m-1}}[1-\mu(C_{m-1})-\mu(C_m \setminus C_{m-1})]^{n-k_m}}{[1-\mu(C_{m-1})]^{n-k_{m-1}}} \\
&= \left[\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{k_m-k_{m-1}}\left[1-\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{n-k_m}.
\end{aligned}
$$

Por lo tanto,

$$
\begin{aligned}
&\mathbb{P}(n\mu_n(C_m)=k_m \mid n\mu_n(C_i)=k_i;\ 1\leq i\leq m-1) \\
&= \binom{n-k_{m-1}}{k_m-k_{m-1}}\left[\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{k_m-k_{m-1}}\left[1-\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}\right]^{n-k_m}.
\end{aligned}
$$

Por otra parte:

$$
\begin{aligned}
&\mathbb{P}(n\mu_n(C_m)=k_m \mid n\mu_n(C_{m-1})=k_{m-1}) \\
&= \frac{\mathbb{P}(n\mu_n(C_{m-1})=k_{m-1}, n\mu_n(C_m)=k_m)}{\mathbb{P}(n\mu_n(C_{m-1})=k_{m-1})} \\
&= \frac{\binom{n}{n_{m-1},n_m,n_{m+1}}[\mu(B_{m-1})]^{n_{m-1}}[\mu(B_m)]^{n_m}[\mu(B_{m+1})]^{n_{m+1}}}{\binom{n}{n_{m-1},n_m'}[\mu(B_{m-1})]^{n_{m-1}}[\mu(B_m')]^{n_m'}} \\
&= \frac{\frac{n!}{n_{m-1}!n_m!n_{m+1}!}}{\frac{n!}{n_{m-1}!n_m'!}}\frac{[\mu(B_m)]^{n_m}[\mu(B_{m+1})]^{n_{m+1}}}{[\mu(B_m')]^{n_m'}} \\
&= \frac{n_m'!}{n_m!n_{m+1}!}\frac{[\mu(C_m \setminus C_{m-1})]^{k_m-k_{m-1}}[\mu(S\setminus C_m)]^{n-k_m}}{[\mu(S\setminus C_{m-1})]^{n-k_{m-1}}} \\
&= \mathbb{P}(n\mu_n(C_m)=k_m \mid n\mu_n(C_i)=k_i;\ 1\leq i\leq m-1).
\end{aligned}
$$

□

El último lema establece que $n\mu_n$ tiene la **propiedad de Markov** cuando se evalua sobre conjuntos crecientes. las **probabilidades de transición** estan dadas a través de la distribución binomial con parámetros $n-k_{m-1}$ y $\mu(\cdot \mid C_{m-1}^c)$, i.e., dado $n\mu_n(C_{m-1})=k_{m-1}$, $n\mu_n(C_m)$ tiene la misma distribución que $k_{m-1}+(n-k_{m-1})\mu_{n-k_{m-1}}(C_m \setminus C_{m-1})$,
en efecto: Como $C_{m-1}\subset C_m$,

$$
\begin{aligned}
n\mu_n(C_m) &= n\mu_n(C_{m-1}\cup(C_m \setminus C_{m-1})) \\
&= n\mu_n(C_{m-1})+n\mu_n(C_m \setminus C_{m-1}) \\
&= k_{m-1}+(n-k_{m-1})\mu_{n-k_{m-1}}(C_m \setminus C_{m-1}).
\end{aligned}
$$

La última igualdad se debe a que estamos calculando la medida sobre el nuevo espacio muestral $S \setminus C_{m-1}$, dado que $n\mu_n(C_{m-1}) = k_{m-1}$.

En vista de la propiedad de Markov, obtenemos para la esperanza condicional:

$$\begin{aligned}\mathbb{E}[n\mu_n(C_m) \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})] &= \mathbb{E}[n\mu_n(C_m) \mid \mu_n(C_{m-1})] \\ &= n\mu_n(C_{m-1}) + [n - n\mu_n(C_{m-1})]\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}.\end{aligned}$$

Así, se tiene la siguiente igualdad:

$$\frac{\mathbb{E}[1-\mu_n(C_m) \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})]}{1-\mu(C_m)} = \frac{1-\mu_n(C_{m-1})}{1-\mu(C_{m-1})}. \tag{2.11}$$

En efecto:

$$\frac{\mathbb{E}[1-\mu_n(C_m) \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})]}{1-\mu(C_m)}$$

$$\begin{aligned}
&= \frac{1-\mathbb{E}[\mu_n(C_m) \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})]}{1-\mu(C_m)} \\
&= \frac{1-[\mu_n(C_{m-1}) + (1-\mu_n(C_{m-1}))\frac{\mu(C_m \setminus C_{m-1})}{1-\mu(C_{m-1})}]}{1-\mu(C_m)} \\
&= \frac{1-[\mu_n(C_{m-1}) + (1-\mu_n(C_{m-1}))\frac{\mu(C_m)-\mu(C_{m-1})}{1-\mu(C_{m-1})}]}{1-\mu(C_m)} \\
&= \frac{1-\mu_n(C_{m-1}) + \frac{(1-\mu_n(C_{m-1}))[1-\mu(C_m)-(1-\mu(C_{m-1}))]}{1-\mu(C_{m-1})}}{1-\mu(C_m)} \\
&= \frac{1-\mu_n(C_{m-1})}{1-\mu(C_m)} + \frac{\frac{(1-\mu_n(C_{m-1}))(1-\mu(C_m))-(1-\mu_n(C_{m-1}))(1-\mu(C_{m-1}))}{1-\mu(C_{m-1})}}{1-\mu(C_m)} \\
&= \frac{1-\mu_n(C_{m-1})}{1-\mu(C_m)} + \frac{1-\mu_n(C_{m-1})}{1-\mu(C_{m-1})} - \frac{1-\mu_n(C_{m-1})}{1-\mu(C_m)} \\
&= \frac{1-\mu_n(C_{m-1})}{1-\mu(C_{m-1})}
\end{aligned}$$

Por lo tanto,

$$\frac{\mathbb{E}[1-\mu_n(C_m) \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})]}{1-\mu(C_m)} = \frac{1-\mu_n(C_{m-1})}{1-\mu(C_{m-1})}.$$

Luego, si definimos

$$H_{n,m} = \frac{1-\mu_n(C_m)}{1-\mu(C_m)},$$

se tiene que

$$\mathbb{E}[H_{n,m} \mid \mu_n(C_1), \ldots, \mu_n(C_{m-1})] = H_{n,m-1},$$

es decir, $\{H_{n,m}\}$ es una **martingala** con respecto a la filtración usual.

Cuando los conjuntos son intervalos extendidos de la forma $(-\infty, t]$, F_n resulta ser un proceso de Markov.

2.2.4. Teorema. *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con f.d. F. Entonces F_n es un proceso de Markov, tal que para $t_1 < t_2 < \ldots < t_m$ y $0 \le k_1 \le k_2 \le \ldots \le k_m \le n$:*

$$\mathbb{P}(nF_n(t_m) = k_m \mid nF_n(t_{m-1}) = k_{m-1})$$

$$= \binom{n-k_{m-1}}{k_m - k_{m-1}} \left[\frac{F(t_m)-F(t_{m-1})}{1-F(t_{m-1})}\right]^{k_m-k_{m-1}} \left[1-\frac{F(t_m)-F(t_{m-1})}{1-F(t_{m-1})}\right]^{n-k_m}.$$

Demostración: Sean $C_1 = (-\infty, t_1], C_2 = (-\infty, t_2], \ldots, C_m = (-\infty, t_m]$. Entonces

$$C_1 \subset C_2 \subset \ldots \subset C_m.$$

Luego, definamos $n_1 = k_1, n_i = k_i - k_{i-1}$ para $2 \le i \le m$ y $n_{m+1} = n - k_m$. $B_1 = C_1, B_i = C_i \setminus C_{i-1}$ para $2 \le i \le m$ y $B_{m+1} = \mathbb{R} \setminus (-\infty, t_m]$. Entonces

$$nF_n(t_m) = n\mu_n(-\infty, t_m] = n\mu_n(C_m)$$

y

$$nF_n(t_i) = n\mu_n(-\infty, t_i] = n\mu_n(C_i); \;\; 1 \le i \le m-1.$$

Así,

$$\mathbb{P}(nF_n(t_m) = k_m \mid nF_n(t_i) = k_i; \; 1 \le i \le m-1)$$

$$\begin{aligned}
&= \mathbb{P}(n\mu_n(C_m) = k_m \mid n\mu_n(C_i) = k_i; \; 1 \le i \le m-1) \\
&= \mathbb{P}(nF_n(t_m) = k_m \mid nF_n(t_{m-1}) = k_{m-1}) \\
&= \binom{n-k_{m-1}}{k_m - k_{m-1}} \left[\frac{F(t_m)-F(t_{m-1})}{1-F(t_{m-1})}\right]^{k_m-k_{m-1}} \left[1-\frac{F(t_m)-F(t_{m-1})}{1-F(t_{m-1})}\right]^{n-k_m}.
\end{aligned}$$

Las dos ultimas igualdades se deben al lema 2.2.3 □

Si definimos
$$M_n^\circ(t_i) = \frac{1 - F_n(t_i)}{1 - F(t_i)}, \quad i = 1, 2, \ldots, m$$
entonces, M_n° es una **martingala** con respecto a la **filtración natural**
$$\mathcal{F}_j = \sigma(F_n(t_i) : 1 \leq i \leq j).$$
Aquí, $\sigma(\ldots)$ denota el $\sigma -$ álgebra generado por las variables dentro del paréntesis. Notemos que
$$\mathbb{E}M_n^\circ(t_i) = 1 \quad \text{para cada } t_i,$$
y para $s \leq t$,
$$Cov(M_n^\circ(s), M_n^\circ(t)) = \frac{F(s)}{n(1 - F(s))}.$$
En efecto:
$$\begin{aligned}
Cov(M_n^\circ(s), M_n^\circ(t)) &= \mathbb{E}(M_n^\circ(s)M_n^\circ(t)) - \mathbb{E}(M_n^\circ(s))\mathbb{E}(M_n^\circ(t)) \\
&= \mathbb{E}\left[\frac{(1 - F_n(s))}{1 - F(s)}\frac{(1 - F_n(t))}{1 - F(t)}\right] - \mathbb{E}\left(\frac{1 - F_n(s)}{1 - F(s)}\right)\mathbb{E}\left(\frac{1 - F_n(t)}{1 - F(t)}\right) \\
&= \frac{\mathbb{E}(1 - F_n(t) - F_n(s) + F_n(s)F_n(t)) - (1 - \mathbb{E}(F_n(s)))(1 - \mathbb{E}(F_n(t)))}{(1 - F(s))(1 - F(t))} \\
&= \frac{\mathbb{E}(F_n(s)F_n(t)) - \mathbb{E}(F_n(s))\mathbb{E}(F_n(t))}{(1 - F(s))(1 - F(t))} \\
&= \frac{Cov(F_n(s), F_n(t))}{(1 - F(s))(1 - F(t))} \\
&= \frac{\frac{1}{n}(F(s \wedge t) - F(s)F(t))}{(1 - F(s))(1 - F(t))} \\
&= \frac{\frac{1}{n}(F(s) - F(s)F(t))}{(1 - F(s))(1 - F(t))} \\
&= \frac{\frac{1}{n}[F(s)(1 - F(t))]}{(1 - F(s))(1 - F(t))} \\
&= \frac{F(s)}{n(1 - F(s))}.
\end{aligned}$$
Observemos que la varianza de esta martingala crece a infinito, conforme $F(s)$ tiende a uno.

El siguiente teorema es una versión reversa del teorema 2.2.4.

2.2.5. Teorema. *Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con f.d. F. Entonces F_n es un proceso de Markov reverso, tal que para $t_1 < t_2 < \ldots < t_m$ y $0 \leq k_1 \leq k_2 \leq \ldots \leq k_m \leq n$:*

$$\mathbb{P}(nF_n(t_1) = k_1 \mid nF_n(t_2) = k_2, \ldots, nF_n(t_m) = k_m)$$

$$\begin{aligned} &= \mathbb{P}(nF_n(t_1) = k_1 \mid nF_n(t_2) = k_2) \\ &= \begin{pmatrix} k_2 \\ k_1 \end{pmatrix} \left[\frac{F(t_1)}{F(t_2)}\right]^{k_1} \left[1 - \frac{F(t_1)}{F(t_2)}\right]^{k_2 - k_1}. \end{aligned}$$

Demostración: Como en la demostración del teorema 2.2.4, sean $C_1 = (-\infty, t_1], C_2 = (-\infty, t_2], \ldots, C_m = (-\infty, t_m]$. Entonces,

$$C_1 \subset C_2 \subset \ldots \subset C_m.$$

Luego, definamos $n_1 = k_1, n_i = k_i \setminus k_{i-1}$ para $2 \leq i \leq m$ y $n_{m+1} = n - k_m$. $B_1 = C_1, B_i = C_i \setminus C_{i-1}$ para $2 \leq i \leq m$ y $B_{m+1} = \mathbb{R} \setminus C_m$. Entonces,

$$nF_n(t_m) = n\mu_n(-\infty, t_m] = n\mu_n(C_m)$$

y

$$nF_n(t_i) = n\mu_n(-\infty, t_i] = n\mu_n(C_i); \quad 1 \leq i \leq m-1.$$

$\mu(B_1) = \mu(C_1) = F(t_1)$
$\mu(B_i) = \mu(C_i \setminus C_{i-1}) = \mu(t_i - t_{i-1}) = F(t_i) - F(t_{i-1})$
$\mu(B_{m+1}) = \mu(\mathbb{R} \setminus C_m) = F(\infty - t_m) = 1 - F(t_m)$. Luego.

$$\mathbb{P}(nF_n(t_1) = k_1 \mid nF_n(t_2) = k_2, \ldots, nF_n(t_m) = k_m)$$

$$\begin{aligned} &= \frac{\mathbb{P}(nF_n(t_1) = k_1, nF_n(t_2) = k_2, \ldots, nF_n(t_m) = k_m)}{\mathbb{P}(nF_n(t_2) = k_2, \ldots, nF_n(t_m) = k_m)} \\ &= \frac{\begin{pmatrix} n \\ n_1, n_2, \ldots, n_{m+1} \end{pmatrix} \prod_{i=1}^{m+1} [\mu(B_i)]^{n_i}}{\begin{pmatrix} n \\ k_2, n_3, \ldots, n_{m+1} \end{pmatrix} [\mu(C_2)]^{k_2} \prod_{i=3}^{m+1} [\mu(B_i)]^{n_i}} \\ &= \frac{\frac{n!}{n_1! n_2! \ldots n_{m+1}!} [\mu(B_1)]^{n_1} [\mu(B_2)]^{n_2}}{\frac{n!}{k_2! n_3! \ldots n_{m+1}!} [\mu(C_2)]^{k_2}} \\ &= \frac{k_2!}{n_1! n_2!} \frac{[F(t_1)]^{k_1} [F(t_2) - F(t_1)]^{k_2 - k_1}}{[F(t_2)]^{k_2}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{k_2!}{k_1!(k_2-k_1)!}\frac{[F(t_1)]^{k_1}[F(t_2)-F(t_1)]^{k_2-k_1}}{[F(t_2)]^{k_2}}\frac{[F(t_2)]^{k_1}}{[F(t_2)]^{k_1}} \\
&= \binom{k_2}{k_1}\left[\frac{F(t_1)}{F(t_2)}\right]^{k_1}\left[1-\frac{F(t_1)}{F(t_2)}\right]^{k_2-k_1}.
\end{aligned}
$$

□

Hemos enunciado propiedades acerca de F_n para el caso discreto. Para el caso continuo, F_n se considera como un proceso estocástico sobre la recta real total, i.e., para todo $t \in \mathbb{R}$. Para ello, necesitamos considerar $\sigma-$álgebras como

$$\mathcal{F}_t := \sigma(F_n(s) : s \leq t)$$

dicho de otra manera,

$$\mathcal{F}_t = \sigma(1_{\{X_i \leq s\}} : s \leq t, i = 1, \ldots, n).$$

Entonces, un generador para la primera $\mathcal{F}_t$, está dada por la familia de eventos

$$\{nF_n(t_1) = k_1, \ldots, nF_n(t_{m-1}) = k_{m-1}\}.$$

donde $t_1 \leq \ldots \leq t_{m-1} \leq t$ y $m \geq 1$ son arbitrarios. Debido a que la familia es cerrada bajo uniones e intersecciones, ésta determina la distribución de F_n. Conjuntamente con los teoremas 2.2.4 y 2.2.5, el carácter distribucional de F_n, se puede caracterizar a tráves de una f.d. empírica uniforme, como sigue:

(i) F_n es un proceso de Markov, tal que $\mathcal{L}(nF_n(t) : t_0 \leq t \mid nF_n(t_0) = k)$

$$= \mathcal{L}\left(k + (n-k)\bar{F}_{n-k}\left[\frac{F(t)-F(t_0)}{1-F(t_0)}\right] : t_0 \leq t\right).$$

(ii) F_n es un proceso de Markov reverso, tal que

$$\mathcal{L}(nF_n(t) : t \leq t_0 \mid nF_n(t_0) = k) = \mathcal{L}\left(k\bar{F}_k\left[\frac{F(t)}{F(t_0)}\right] : t \leq t_0\right).$$

Aquí $\mathcal{L}$ denota ley o distribución, y $\bar{F}_k$ y $\bar{F}_{n-k}$ denotan las f.d.'s empíricas de una muestra uniforme.

(iii) El proceso $t \to \frac{1-F_n(t)}{1-F(t)}$, es una martingala con respecto a la filtración natural $\mathcal{F}_t = \sigma(F_n(s) : s \leq t)$.

(iv) El proceso $t \to \frac{F_n(t)}{F(t)}$, es una martingala reversa con respecto a la filtración natural $\mathcal{F}_t^0 := \sigma(F_n(s) : t \leq s)$.

La última proposición de esta sección, habla acerca de la distribución de α_n en un número finito de puntos. Aquí $\mathcal{N}(0, \Sigma)$ denota a la distribución normal con esperanza $0 \in \mathbb{R}^m$ y matriz de varianzas y covarianzas Σ de $m \times m$. El simbolo T es la **transpuesta**.

2.2.6. Lema. *Recordemos que*

$$\alpha_n(t) = n^{1/2}[F_n(t) - F(t)], \quad t \in \mathbb{R},$$

es el proceso empírico. Entonces, para cada $t_1 \leq t_2 \leq \ldots \leq t_m$,

$$[\alpha_n(t_1), \ldots, \alpha_n(t_m)]^T \xrightarrow{\mathcal{L}} \mathcal{N}_m(0, \Sigma) \text{ cuando } n \to \infty$$

con

$$\Sigma = (\sigma_{ij})_{1 \leq i,j \leq m}$$

y

$$\sigma_{ij} = F(t_i \wedge t_i) - F(t_i)F(t_j). \tag{2.12}$$

Este lema, es una consecuencia inmediata de TLC multivariado.

2.3. La descomposición Doob-Meyer

En la sección anterior se vio que el proceso

$$t \longrightarrow \frac{1 - F_n(t)}{1 - F(t)}$$

es una martingala con media uno, con respecto a la filtración natural $\mathcal{F}_t = \sigma(F_n(s) : s \leq t)$. Un ejemplo de martingala relacionada con ésta, es la siguiente:

2.3.1. Ejemplo. Si a la expresión anterior se le sustrae uno y se multiplica todo por $-n^{1/2}$, se tiene una martingala con media cero, varianza $\frac{F(t)}{1-F(t)}$, para la misma filtración.

Demostración: Sea

$$\beta_n(t) = -n^{1/2}\left[\frac{1 - F_n(t)}{1 - F(t)} - 1\right].$$

Entonces,

$$\beta_n(t) = n^{1/2}\frac{[F_n(t) - F(t)]}{1 - F(t)} = \frac{\alpha_n(t)}{1 - F(t)}.$$

Luego

$$\begin{aligned}
\mathbb{E}[\beta_n(t) \mid \mathcal{F}_{t-1}] &= \mathbb{E}\left[n^{1/2}\frac{[F_n(t) - F(t)]}{1 - F(t)} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[- n^{1/2}\left[\frac{1 - F_n(t)}{1 - F(t)} - \left(\frac{1 - F(t)}{1 - F(t)}\right)\right] \mid \mathcal{F}_{t-1}\right] \\
&= -n^{1/2}\left[\frac{1 - F_n(t-1)}{1 - F(t-1)} - \left(\frac{1 - F(t)}{1 - F(t)}\right)\right] \\
&= -n^{1/2}\left[\frac{1 - F_n(t-1)}{1 - F(t-1)} - \left(\frac{1 - F(t-1)}{1 - F(t-1)}\right)\right] \\
&= \frac{\alpha_n(t-1)}{1 - F(t-1)} = \beta_n(t-1).
\end{aligned}$$

Por lo tanto,

$$\mathbb{E}[\beta_n(t) \mid \mathcal{F}_{t-1}] = \beta_n(t-1)$$

y β_n es una martingala.
Además,

$$\mathbb{E}[\beta_n(t)] = \mathbb{E}\left[\frac{\alpha_n(t)}{1 - F(t)}\right] = \frac{1}{1 - F(t)}\mathbb{E}[\alpha_n(t)] = 0.$$

Por último,

$$\begin{aligned}
Var(\beta_n(t)) = Var\left(\frac{\alpha_n(t)}{1 - F(t)}\right) &= \frac{1}{(1 - F(t))^2}Var(\alpha_n(t)) \\
&= \frac{1}{(1 - F(t))^2}F(t)(1 - F(t)) \\
&= \frac{F(t)}{1 - F(t)}.
\end{aligned}$$

■

Observemos que la varianza de β_n tiende a infinito cuando $F(t) \to 1$. En otras palabras, el proceso β_n, dá un mal ajuste para F_n y F.

Una martingala que evita la presencia del factor $(1-F)^{-1}$. resuelve el problema anterior. Esta martingala, juega un papel importante en la **descomposición aditiva** de F_n, Esta martingala es parte de la **descomposición Doob-Meyer** de F_n.

Primero se discutirá la descomposición en el caso de un proceso de tiempo discreto.

2.3.2. Lema. *Sean S_k, $k = 0, 1, \ldots$ una sucesión de variables aleatorias integrables, adaptadas a las filtraciones $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots$. i.e., S_k es $\mathcal{F}_k-$ medible. Entonces, dado un número real x, existe una única descomposición*

$$S_k = M_k + h_k,$$

donde $(M_k, \mathcal{F}_k)_{k=0,1,\ldots}$ es una martingala:

$$\mathbb{E}[M_k \mid \mathcal{F}_{k-1}] = M_{k-1} \quad \text{para } k = 1, 2, \ldots$$

*y $(h_k, \mathcal{F}_k)$ es **predecible**, i.e.,*

$$h_k \text{ es } \mathcal{F}_{k-1} - \text{medible para } k = 1, 2, \ldots$$

con condición inicial

$$h_0 = x.$$

Demostración: Definamos

$$\begin{aligned} h_k &= \begin{cases} x, & \text{para } k = 0 \\ h_{k-1} - S_{k-1} + \mathbb{E}(S_k \mid \mathcal{F}_{k-1}), & \text{para } k \geq 1. \end{cases} \\ M_k &= \begin{cases} S_0 - x, & \text{para } k = 0 \\ M_{k-1} + S_k - \mathbb{E}(S_k \mid \mathcal{F}_{k-1}), & \text{para } k \geq 1. \end{cases} \end{aligned}$$

Así, veamos que $(M_k, \mathcal{F}_k)_{k=0,1,\ldots}$ es una martingala.

(i) $\mathbb{E} \mid M_k \mid < \infty$ para $k \geq 0$:

Por inducción sobre k,

$$\mathbb{E} \mid M_0 \mid = \mathbb{E} \mid S_0 - x \mid \leq \mathbb{E} \mid S_0 \mid + \mid x \mid < \infty.$$

Para $k = 1$,

$$\begin{aligned}
\mathbb{E} \mid M_1 \mid &= \mathbb{E} \mid M_0 + S_1 - \mathbb{E}(S_1 \mid \mathcal{F}_0) \mid \\
&\leq \mathbb{E} \mid M_0 \mid + \mathbb{E} \mid S_1 \mid + \mathbb{E} \mid \mathbb{E}(S_1 \mid \mathcal{F}_0) \mid \\
&\leq \mathbb{E} \mid M_0 \mid + \mathbb{E} \mid S_1 \mid + \mathbb{E}(\mathbb{E}(\mid S_1 \mid \mid \mathcal{F}_0)) \\
&= \mathbb{E} \mid M_0 \mid + \mathbb{E} \mid S_1 \mid + \mathbb{E} \mid S_1 \mid \\
&= \mathbb{E} \mid M_0 \mid + 2\mathbb{E} \mid S_1 \mid < \infty.
\end{aligned}$$

Supongamos que $\mathbb{E} \mid M_k \mid < \infty$. Entonces,

$$\begin{aligned}
\mathbb{E} \mid M_{k+1} \mid &= \mathbb{E}|M_k + S_{k+1} - \mathbb{E}(S_{k+1} \mid \mathcal{F}_k)| \\
&\leq \mathbb{E} \mid M_k \mid + \mathbb{E} \mid S_{k+1} \mid + \mathbb{E}|\mathbb{E}(S_{k+1} \mid \mathcal{F}_k)| \\
&\leq \mathbb{E} \mid M_k \mid + \mathbb{E} \mid S_{k+1} \mid + \mathbb{E}(\mathbb{E}(\mid S_{k+1} \mid \mid \mathcal{F}_k)) \\
&= \mathbb{E} \mid M_k \mid + \mathbb{E} \mid S_{k+1} \mid + \mathbb{E} \mid S_{k+1} \mid \\
&= \mathbb{E} \mid M_k \mid + 2\mathbb{E} \mid S_{k+1} \mid < \infty.
\end{aligned}$$

Por lo tanto,

$$\mathbb{E} \mid M_{k+1} \mid < \infty.$$

y $\mathbb{E} \mid M_k \mid < \infty$ para todo $k \geq 0$.

(ii) M_k es $\mathcal{F}_k$ – medible para todo $k \geq 0$:
Por inducción sobre k,

$$M_0 = S_0 - x \text{ es } \mathcal{F}_0 - \text{medible.}$$

$$M_1 = M_0 + S_1 - \mathbb{E}(S_1 \mid \mathcal{F}_0) \text{ es } \mathcal{F}_1 - \text{medible.}$$

puesto que los tres sumandos son $\mathcal{F}_1$ – medibles.

Supongamos que M_k es $\mathcal{F}_k$ – medible. Entonces,

$$M_{k+1} = M_k + S_{k+1} - \mathbb{E}(S_{k+1} \mid \mathcal{F}_k) \text{ es } \mathcal{F}_{k+1} - \text{medible,}$$

ya que por H.I., si M_k es $\mathcal{F}_k$ – medible entonces es $\mathcal{F}_{k+1}$ – medible. Además, por hipótesis, S_{k+1} es $\mathcal{F}_{k+1}$ – medible, y como $\mathbb{E}(S_{k+1} \mid \mathcal{F}_k)$ es $\mathcal{F}_k$ – medible, $\mathbb{E}(S_{k+1} \mid \mathcal{F}_k)$ es $\mathcal{F}_{k+1}$ – medible.

Por lo tanto,

$$M_{k+1} \text{ es } \mathcal{F}_{k+1} - \text{medible}$$

y

$$M_k \text{ es } \mathcal{F}_k - \text{medible para todo } k = 0, 1, 2, \ldots$$

(iii)

$$\begin{aligned}
\mathbb{E}[M_1 \mid \mathcal{F}_0] &= \mathbb{E}[M_0 + S_1 - \mathbb{E}(S_1 \mid \mathcal{F}_0) \mid \mathcal{F}_0] \\
&= \mathbb{E}[M_0 \mid \mathcal{F}_0] + \mathbb{E}[S_1 \mid \mathcal{F}_0] - \mathbb{E}[S_1 \mid \mathcal{F}_0] \\
&= \mathbb{E}[M_0 \mid \mathcal{F}_0] = M_0.
\end{aligned}$$

Y para $k \geq 1$,

$$\begin{aligned}
\mathbb{E}[M_k \mid \mathcal{F}_{k-1}] &= \mathbb{E}[M_{k-1} + S_k - \mathbb{E}(S_k \mid \mathcal{F}_{k-1}) \mid \mathcal{F}_{k-1}] \\
&= \mathbb{E}[M_{k-1} \mid \mathcal{F}_{k-1}] + \mathbb{E}[S_k \mid \mathcal{F}_{k-1}] - \mathbb{E}[S_k \mid \mathcal{F}_{k-1}] \\
&= \mathbb{E}[M_{k-1} \mid \mathcal{F}_{k-1}] = M_{k-1}.
\end{aligned}$$

Por lo tanto,

$$\mathbb{E}[M_k \mid \mathcal{F}_{k-1}] = M_{k-1} \quad \text{para } k = 1, 2, \ldots$$

y

$$(M_k, \mathcal{F}_k)_k \quad \text{es una martingala.}$$

(iii) Mostremos que $(h_k, \mathcal{F}_k)$ es predecible.

$$h_0 = x \text{ es } \mathcal{F}_0 - \text{medible.}$$

Para $k = 1$,

$$h_1 = h_0 + \mathbb{E}[S_1 \mid \mathcal{F}_0) \text{ es } \mathcal{F}_0 - \text{medible,}$$

puesto que cada sumando es $\mathcal{F}_0$ – medible.

Ahora supongamos que h_k es $\mathcal{F}_{k-1}$ – medible. Entonces

$$h_{k+1} = h_k - S_k + \mathbb{E}(S_{k+1} \mid \mathcal{F}_k) \text{ es } \mathcal{F}_k - \text{medible,}$$

puesto que por hipótesis, si h_k es $\mathcal{F}_{k-1}$ – medible, h_k es $\mathcal{F}_k$ – medible.
Por lo tanto,

$$h_k \text{ es } \mathcal{F}_{k-1} - \text{medible} \quad \text{para todo } k \geq 1.$$

Observemos que para $k \geq 1$, $M_k + h_k = M_{k-1} + h_{k-1} + S_k - S_{k-1}$. También se tiene

$$M_k + h_k = S_k,$$

en efecto,

$$M_0 + h_0 = S_0 - x + x = S_0.$$

Para $k = 1$,

$$M_1 + h_1 = M_0 + h_0 + S_1 - S_0 = S_0 + S_1 - S_0 = S_1.$$

Supongamos que $M_k + h_k = S_k$. Entonces,

$$M_{k+1} + h_{k+1} = M_k + h_k + S_{k+1} - S_k = S_k + S_{k+1} - S_k = S_{k+1}.$$

Por lo tanto,

$$M_k + h_k = S_k \quad \text{para } k = 0, 1, 2 \ldots$$

Por último, veamos que esta descomposición es única.

Supongamos que $S_k = M_k^{'} + h_k^{'}$ es otra descomposición, con $(M_k^{'}, \mathcal{F}_k)_{k=0,1,\ldots}$ martingala, $(h_k^{'}, \mathcal{F}_k)$ predecible, con condición inicial $h_0^{'} = x$. Entonces,

$$M_k + h_k = M_k^{'} + h_k^{'}.$$

Así que

$$h_k - h_k^{'} = M_k^{'} - M_k \quad \text{y } \mathbb{E}[h_k - h_k^{'} \mid \mathcal{F}_{k-1}] = \mathbb{E}[M_k^{'} - M_k \mid \mathcal{F}_{k-1}]$$

Así,

$$h_k - h_k^{'} = M_{k-1}^{'} - M_{k-1} = h_{k-1} - h_{k-1}^{'}.$$

Luego se tiene

$$h_k - h_k^{'} = 0 \quad \text{para todo } k \geq 1,$$

en efecto:
Para $k = 1$, $h_1 - h_1^{'} = h_0 - h_0^{'} = x - x = 0$.

Supongamos que $h_k - h_k^{'} = 0$. Entonces,

$$h_{k+1} - h_{k+1}^{'} = h_k - h_k^{'} = 0.$$

Por lo tanto,

$$h_k - h_k^{'} = 0 \quad \text{para todo } k \geq 1.$$

Así

$$M_{k-1}^{'} - M_{k-1} = 0 \quad \text{para todo } k \geq 1.$$

Luego,

$$h_k = h_k^{'} \quad \text{y} \quad M_k = M_k^{'}$$

Por lo tanto,

$$S_k = M_k + h_k, \text{ y esta descomposición es única}$$

□

Si $(S_k, \mathcal{F}_k)_k$ es una submartingala, i.e., si

$$\mathbb{E}(S_k \mid \mathcal{F}_{k-1}) \geq S_{k-1},$$

entonces

$$h_1 = h_0 - S_0 + \mathbb{E}(S_1 \mid \mathcal{F}_0) \geq h_0 - S_0 + S_0 = h_0$$

Por lo tanto,

$$h_1 \geq h_0.$$

Para $k \geq 2$,

$$h_k = h_{k-1} - S_{k-1} + \mathbb{E}(S_k \mid \mathcal{F}_{k-1}) \geq h_{k-1} - S_{k-1} + S_{k-1} = h_{k-1}.$$

Por lo tanto,

$$h_k \geq h_{k-1} \quad \text{para todo } k \geq 2.$$

Así que, $h_0 \leq h_1 \leq \ldots \leq h_k \leq \ldots$, es un **proceso no decreciente**. El proceso $(h_k)_k$ se le llama compensador, debido a que compensa cualquier desviación de un proceso con tendencia libre. Las submartingalas tienen tendencia positiva y las martingalas tienen tendencia libre. La fuente de aleatoriedad (o de ruido) en S_k se debe a M_k, debido a que h_k es predecible.

Ahora calculamos la descomposición Doob-Meyer de $nF_n(t)$. F_n es no decreciente, por lo que el compensador será no decreciente. Entonces, es suficiente considerar una muestra de tamaño $n = 1$. Escribimos $X = X_1$. Sean $t_1 < \ldots < t_m$. La variable S_k del lema anterior es ahora:

$$S_k = 1_{\{X \leq t_k\}}, \quad k = 1, \ldots, m$$

con filtración

$$\mathcal{F}_k = \sigma(1_{\{X \leq t_j\}} : j \leq k).$$

Así que,

$$\mathbb{E}(S_k \mid \mathcal{F}_{k-1}) = \mathbb{E}(1_{\{X \leq t_k\}} \mid 1_{\{X \leq t_j\}} : j \leq k-1)$$

$$= \mathbb{E}(1_{\{X \leq t_k\}} \mid 1_{\{X \leq t_{k-1}\}}) = 1_{\{X \leq t_{k-1}\}} + 1_{\{X > t_{k-1}\}} \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}.$$

La segunda igualdad se debe al teorema 2.2.4 con $n = 1$. La comprobación de la última igualdad es como sigue:

Si $1_{\{X \leq t_{k-1}\}} = 0$ entonces $\{X > t_{k-1}\}$. Así que,

$$\begin{aligned}
\mathbb{E}(1_{\{X\leq t_k\}} \mid 1_{\{X\leq t_{k-1}\}}) &= \mathbb{E}(1_{\{X\leq t_k\}} \mid X > t_{k-1}) \\
&= \mathbb{P}(X \leq t_k \mid X > t_{k-1}) \\
&= \frac{\mathbb{P}(X \leq t_k, X > t_{k-1})}{\mathbb{P}(X > t_{k-1})} \\
&= \frac{\mathbb{P}(t_{k-1} < X \leq t_k)}{1 - \mathbb{P}(X \leq t_{k-1})} \\
&= \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}.
\end{aligned}$$

y si, $1_{\{X \leq t_{k-1}\}} = 1$ entonces $\{X \leq t_{k-1}\}$. Así que,

$$\begin{aligned}
\mathbb{E}(1_{\{X\leq t_k\}} \mid 1_{\{X\leq t_{k-1}\}}) &= \mathbb{E}(1_{\{X\leq t_k\}} \mid X \leq t_{k-1}) \\
&= \mathbb{P}(X \leq t_k \mid X \leq t_{k-1}) \\
&= \frac{\mathbb{P}(X \leq t_k, X \leq t_{k-1})}{\mathbb{P}(X \leq t_{k-1})} \\
&= \frac{\mathbb{P}(X \leq t_{k-1})}{\mathbb{P}(X \leq t_{k-1})} = 1
\end{aligned}$$

Ahora, sea $S_0 = 0$ y $x = 0$, el cuál corresponde a $t_0 = -\infty$ y $\mathcal{F}_0 = \{\emptyset, \Omega\}$. De la prueba del lema 2.3.2, obtenemos

$$\begin{aligned}
M_k &= M_{k-1} + S_k - \mathbb{E}(S_k \mid \mathcal{F}_{k-1}) \\
&= M_{k-1} + 1_{\{X\leq t_k\}} - 1_{\{X\leq t_{k-1}\}} - 1_{\{X>t_{k-1}\}}\frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})} \\
&= M_{k-1} + 1_{\{t_{k-1}<X\leq t_k\}} - 1_{\{X>t_{k-1}\}}\frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}.
\end{aligned}$$

Y por inducción,

$$M_k = 1_{\{X\leq t_k\}} - \sum_{j=1}^{k} 1_{\{X>t_{j-1}\}}\frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}.$$

en efecto,
Para $k = 1$,

$$M_1 = M_0 + 1_{\{t_0<X\leq t_1\}} - 1_{\{X>t_0\}}\frac{F(t_1) - F(t_0)}{1 - F(t_0)}$$

$$\begin{aligned}
&= 0 + 1_{\{X \le t_1\}} - 1_{\{X > t_0\}} \frac{F(t_1) - F(t_0)}{1 - F(t_0)} \\
&= 1_{\{X \le t_1\}} - \sum_{j=1}^{1} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}
\end{aligned}$$

Por lo tanto,

$$M_1 = 1_{\{X \le t_1\}} - \sum_{j=1}^{1} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}.$$

Supongamos que

$$M_k = 1_{\{X \le t_k\}} - \sum_{j=1}^{k} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}.$$

Entonces,

$$\begin{aligned}
M_{k+1} &= M_k + \underbrace{1_{\{X \le t_{k+1}\}} - 1_{\{X \le t_k\}} - 1_{\{X > t_k\}} \frac{F(t_{k+1}) - F(t_k)}{1 - F(t_k)}}_{*} \\
&= 1_{\{X \le t_k\}} - \sum_{j=1}^{k} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})} + * \\
&= 1_{\{X \le t_{k+1}\}} - \sum_{j=1}^{k+1} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}.
\end{aligned}$$

Por lo tanto,

$$M_k = 1_{\{X \le t_k\}} - \sum_{j=1}^{k} 1_{\{X > t_{j-1}\}} \frac{F(t_j) - F(t_{j-1})}{1 - F(t_{j-1})}.$$

Esta ecuación nos da una idea de como escribir la descomposición Doob-Meyer, al proceso de tiempo continuo

$$t \to 1_{\{X \le t\}}.$$

Si fijamos $t = t_k$ e incrementamos el número de puntos menores que t, de tal manera que sobre la escala de F, la norma de la partición tienda a cero, obtenemos:

$$\begin{aligned}
M_t &= 1_{\{X \le t\}} - \int_{-\infty}^{t} \frac{1_{\{x \le X\}}}{1 - F(x-)} dF(x) \\
&= 1_{\{X \le t\}} - \Lambda_F(t \wedge X).
\end{aligned}$$

La segunda igualdad se debe a que

$$\begin{aligned}\int_{-\infty}^{t} \frac{1_{\{x\le X\}}}{1-F(x-)}dF(x) &= \int_{-\infty}^{\infty} \frac{1_{\{x\le t\}}1_{\{x\le X\}}}{1-F(x-)}dF(x)\\ &= \int_{-\infty}^{\infty} \frac{1_{\{x\le t\wedge X\}}}{1-F(x-)}dF(x)\\ &= \int_{-\infty}^{t\wedge X} \frac{dF(x)}{1-F(x-)} = \Lambda_F(t\wedge X).\end{aligned}$$

Ahora formulamos la descomposición Doob-Meyer de F_n en el tiempo continuo, para una muestra arbitraria.

2.3.3. Teorema. *Denotando*

$$\mathcal{F}_{n,t} = \sigma(1_{\{X_i\le s\}} : s \le t, i = 1, \ldots, n).$$

Entonces, la martingala de innovación nF_n es igual a

$$M_n(t) = \sum_{i=1}^{n} 1_{\{X_i\le t\}} - \sum_{i=1}^{n} \int_{-\infty}^{t} \frac{1_{\{x\le X_i\}}}{1-F(x-)}dF(x).$$

Para F_n, la parte martingala es

$$M_n^1(t) = F_n(t) - \int_{-\infty}^{t} \frac{1-F_n(x-)}{1-F(x-)}dF(x)$$

Demostración: Damos una prueba formal para $M_n(.)$. Definamos

$$\mathcal{F}_{n,x}^{-} = \sigma(1_{\{X_i\le s\}} : s < x,\ i = 1, \ldots, n).$$

Cada indicador

$$1_{\{x\le X_i\}} = 1 - 1_{\{X_i<x\}},$$

es medible con respecto a $\mathcal{F}_{n,x}^{-}$, debido a que

$$1_{\{X_i<x\}} = \lim_{s\uparrow x} 1_{\{X_i\le s\}}.$$

Ahora, debido a la independencia, es suficiente mostrar que M_1 es una martingala.

Ya hemos probado que para $s \le t$,

$$\mathbb{E}[1_{\{X\le t\}} \mid \mathcal{F}_{1,s}] = 1_{\{X\le s\}} + 1_{\{X>s\}}\frac{F(t)-F(s)}{1-F(s)}.$$

Por otro lado, por un argumento del tipo Fubini,

$$\mathbb{E}\left[\int_{-\infty}^{t} \frac{1_{\{x\leq X\}}}{1-F(x-)}dF(x) \mid \mathcal{F}_{1,s}\right]$$

$$\begin{aligned}
&= \int_{-\infty}^{t} \frac{\mathbb{E}[1_{\{x\leq X\}} \mid \mathcal{F}_{1,s}]}{1-F(x-)}dF(x) \\
&= \int_{(-\infty,s]} \frac{1_{\{x\leq X\}}}{1-F(x-)}dF(x) + \int_{(s,t]} \frac{\mathbb{E}[1_{\{x\leq X\}} \mid \mathcal{F}_{1,s}]}{1-F(x-)}dF(x).
\end{aligned}$$

El primer sumando de la última igualdad es porque

$$\mathbb{E}[1_{\{x\leq X\}} \mid \mathcal{F}_{1,s}] = 1_{\{x\leq X\}}.$$

Para $s < x \leq t$, se tiene

$$\mathbb{E}[1_{\{x\leq X\}} \mid \mathcal{F}_{1,s}] = \mathbb{E}[1_{\{x\leq X\}} \mid 1_{\{X\leq s\}}] = 1_{\{X>s\}}\frac{1-F(x-)}{1-F(s)},$$

en efecto:
Si $1_{\{X\leq s\}} = 1$, entonces $\{X \leq s\}$. Así,

$$\begin{aligned}
\mathbb{E}[1_{\{x\leq X\}} \mid 1_{\{X\leq s\}}] &= \mathbb{E}[1_{\{x\leq X\}} \mid X \leq s] \\
&= \mathbb{P}[x \leq X \mid X \leq s] \\
&= \frac{\mathbb{P}[x \leq X, X \leq s]}{\mathbb{P}[X \leq s]} \\
&= 0.
\end{aligned}$$

Por otra parte, si $1_{\{X\leq s\}} = 0$, entonces $\{X > s\}$. Así,

$$\begin{aligned}
\mathbb{E}[1_{\{x\leq X\}} \mid 1_{\{X\leq s\}}] &= \mathbb{E}[1_{\{x\leq X\}} \mid X > s] \\
&= \mathbb{P}[x \leq X \mid X > s] \\
&= \frac{\mathbb{P}[x \leq X, X > s]}{\mathbb{P}[X > s]} \\
&= \frac{1-\mathbb{P}[X < x]}{1-\mathbb{P}[X \leq s]}.
\end{aligned}$$

Por lo tanto,

$$\mathbb{E}[1_{\{x\leq X\}} \mid \mathcal{F}_{1,s}] = \mathbb{E}[1_{\{x\leq X\}} \mid 1_{\{X\leq s\}}] = 1_{\{X>s\}}\frac{1-F(x-)}{1-F(s)}.$$

Así se tiene que

$$\mathbb{E}[M_1(t) \mid \mathcal{F}_{1,s}] = 1_{\{X\leq s\}} + 1_{\{X>s\}}\frac{F(t)-F(s)}{1-F(s)}$$

$$-\int_{(-\infty,s]}\frac{1_{\{x\leq X\}}}{1-F(x-)}dF(x) - \int_{(s,t]}\frac{1_{\{X>s\}}}{1-F(s)}dF(x) = M_1(s).$$

□

Ahora, analizaremos la estructura de covarianza de $M_1(\cdot)$. Sea $s \leq t$, entonces

$$Cov(M_1(s), M_1(t)) = \mathbb{E}[M_1(s), M_1(t)] = \mathbb{E}M_1^2(s) := \gamma(s).$$

Luego, por el teorema de Fubini, con $x \vee y := \max(x, y)$. obtenemos

$$\begin{aligned}
\gamma(s) &= F(s) + \int_{-\infty}^{s}\int_{-\infty}^{s}\frac{1-F(x\vee y-)}{(1-F(x-))(1-F(y-))}dF(x)dF(y) \\
&- 2\int_{-\infty}^{s}\frac{F(s)-F(x-)}{1-F(x-)}dF(x) \\
&= F(s) - \int_{-\infty}^{s}\frac{F\{x\}}{1-F(x-)}dF(x) \\
&= F(s) \quad \text{cuando } F \text{ es continua.}
\end{aligned}$$

Para F arbitraria, se tiene $\gamma \leq F$. Por lo tanto, γ está acotada, cosa que no sucedio con la martingala M_n^0, cuya varianza fue no acotada.

Observemos también que si F es continua, se puede expresar M_n^1, como

$$\begin{aligned}
M_n^1(t) &= \bar{F}_n(F(t)) - \int_{-\infty}^{t}\frac{1-F_n(x)}{1-F(x)}dF(x) \\
&= \bar{F}_n(F(t)) - \int_{0}^{F(t)}\frac{1-\bar{F}_n(u)}{1-u}du \\
&\equiv \bar{M}_n^1(F(t)).
\end{aligned}$$

Esta representación tiene una analogía como sucedio con F_n, i.e., M_n^1 definida en toda la recta real, se expresa en terminos de $\bar{M}_n^1$ definida sobre el intervalo $[0,1]$. Entonces, una versión de la estadística **K-S**, es como sigue:

$$D_n^1 \equiv \sup_{t\in\mathbb{R}}\left|M_n^1(t)\right| = \sup_{0\leq v\leq 1}\left|\bar{M}_n^1(v)\right|.$$

el cual, es libre de distribución.

También se tiene el siguiente teorema, referente a esta martingala, el cual, tiende a cero uniformemente en $0 \leq v \leq 1$.

2.3.4. Teorema. *Cuando $n \to \infty$,*

$$\sup_{0 \leq v \leq 1} \left|\bar{M}_n^1(v)\right| \to 0, \quad \textit{con probabilidad uno.}$$

2.4. Representaciones Poisson

En esta sección discutiremos la relación entre la distribución empírica y otras cantidades importantes del **Proceso Poisson**.

2.4.1. Definición. *Un proceso puntual $N = N(t)_{t \geq 0}$ se llama Proceso Poisson (estándar) con parámetro $\lambda > 0$ ($Poisson(\lambda)$) si, y sólo si*

(i) $N(0) = 0$

(ii) *N tiene incrementos independientes, i.e., para $0 < t_0 < t_1 < \ldots < t_m < \infty$ las variables aleatorias $N(t_1) - N(t_0), \ldots, N(t_m) - N(t_{m-1})$ son independientes.*

(iii) *Cada incremento $N(t) - N(s)$ se distribuye Poisson con parámetro $\lambda(t-s)$:*

$$\mathbb{P}(N(t) - N(s) = k) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^k}{k!}, \quad k = 0, 1, \ldots$$

Un proceso Poisson se puede generar de la siguiente manera: sean $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes con distribución $Exp(\lambda)$. Si estamos interesados en la ocurrencia de eventos seguros, como la llegada de clientes a una estación de servicios, las $X_i's$ se pueden interpretar como los tiempos transcurridos del $(i-1)$-ésimo al i-ésimo evento. La suma

$$S_n = \sum_{i=1}^{n} X_i$$

es el tiempo de la n-ésima llegada o generalmente cuando la n-ésima llegada ocurre. El proceso

$$N(t) = sup\{n \geq 1 : S_n \leq t\}, \quad t \geq 0$$

es el número de eventos en el intervalo $[0,t]$, con $N(t) = 0$ si $\{n \geq 1 : S_n \leq t\} = \emptyset$. Este proceso, es un Proceso Poisson con parámetro $\lambda > 0$.
Sabemos que los incremetos de la función de distribución empírica son de tipo multinomial y también que los incrementos son generalmente correlacionados y por consiguiente dependientes. Los siguientes dos resultados muestran que existe una relación íntima entre ambos procesos.

2.4.2. Lema. *Sea N un Proceso Poisson con parámetro $\lambda > 0$. Entonces, condicionado a que $N(1) = n$, la distribución de $N(t)$ sobre $0 \leq t \leq 1$, es la misma que $n\bar{F}_n(t)$, $0 \leq t \leq 1$, la función de distribución uniforme renormalizada.*

Demostración: Mostremos que ambos procesos tienen la misma distribución finita dimensional. Para ello, fijemos $t_0 = 0 < t_1 < \ldots < t_m \leq 1 = t_{m+1}$ y enteros $k_0 = 0 \leq k_1 \leq \ldots \leq k_m \leq n = k_{m+1}$. Entonces obtenemos

$$\mathbb{P}(N(t_i) = k_i;\ 1 \leq i \leq m \mid N(1) = n)$$

$$\begin{aligned}
&= \frac{\mathbb{P}(N(t_i) - N(t_{i-1}) = k_i - k_{i-1};\ 1 \leq i \leq m+1)}{\mathbb{P}(N(1) = n)} \\
&= \frac{\prod_{i=1}^{m+1} e^{-\lambda(t_i - t_{i-1})} \frac{[\lambda(t_i - t_{i-1})]^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}}{e^{-\lambda}\frac{\lambda^n}{n!}} \\
&= \frac{n!}{e^{-\lambda}\lambda^n} \prod_{i=1}^{m+1} \frac{e^{-\lambda(t_i - t_{i-1})}\lambda^{k_i - k_{i-1}}(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} \\
&= \frac{n! e^{-\lambda}\lambda^n}{e^{-\lambda}\lambda^n} \prod_{i=1}^{m+1} \frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} \\
&= n! \prod_{i=1}^{m+1} (t_i - t_{i-1})^{k_i - k_{i-1}} / (k_i - k_{i-1})! \\
&= \mathbb{P}(n\bar{F}_n(t_i) = k_i;\ 1 \leq i \leq m).
\end{aligned}$$

La primera igualdad se debe a que los eventos $\{N(t_i) = k_i;\ 1 \leq i \leq m+1\}$ y $\{N(t_i) - N(t_{i-1}) = k_i - k_{i-1};\ 1 \leq i \leq m+1\}$ son iguales.
La última igualdad es como sigue:

$$n\bar{F}_n(t_i) = n\mu_n(0, t_i) := n\mu_n(C_i) \text{ para todo } 1 \leq i \leq m.$$

Entonces

$$C_1 \subset C_2 \subset \ldots \subset C_m.$$

Definamos $B_1 = C_1$, $B_i = C_i \backslash C_{i-1}$ para todo $2 \leq i \leq m$ y $B_{m+1} = S \backslash C_m$. $n_1 = k_1$, $n_i = k_i - k_{i-1}$ para todo $2 \leq i \leq m$ y $n_{m+1} = n - k_m$.
Aplicando ahora el lema 2.2.2,

$$\begin{aligned}
\mathbb{P}(n\bar{F}_n(t_i) = k_i;\ 1 \leq i \leq m) &= \binom{n}{n_1, n_2, \ldots, n_{m+1}} \prod_{i=1}^{m+1} (\bar{F}(t_i) - \bar{F}(t_{i-1}))^{n_i} \\
&= \binom{n}{n_1, n_2, \ldots, n_{m+1}} \prod_{i=1}^{m+1} (t_i - t_{i-1})^{n_i} \\
&= n! \prod_{i=1}^{m+1} (t_i - t_{i-1})^{k_i - k_{i-1}} / (k_i - k_{i-1})!.
\end{aligned}$$

□

Este lema muestra que $n\bar{F}_n$, puede ser vista como un proceso Poisson, dado que han ocurrido n eventos al tiempo $t = 1$. El siguiente lema, revela por otro lado que, un Proceso Poisson, se puede obtener a partir de una función de distribución empírica, simplemente aleatorizando el tamaño de muestra.

2.4.3. Lema. *Sea N una variable aleatoria Poisson con parámetro $\lambda > 0$. Sean $U_1, U_2, \ldots$ variables aleatorias independientes con distribución F_U, siendo independientes de N. Entonces $\{N\bar{F}_N(t) : 0 \leq t \leq 1\}$ es un Proceso Poisson(λ) sobre el intervalo unitario.*

Demostración: Puesto que las distribuciones de dimensión finita determinan la distribución de un proceso, nuevamente fijamos $t_0 = 0 < t_1 < \ldots < t_m < 1 = t_{m+1}$ y enteros no negativos $k_1, \ldots, k_{m+1}$. entonces

$$\mathbb{P}(N[\bar{F}_N(t_i) - \bar{F}_N(t_{i-1})] = k_i;\ 1 \leq i \leq m+1)$$

$$= \sum_{r=0}^{\infty} \mathbb{P}(r[\bar{F}_r(t_i) - \bar{F}_r(t_{i-1})] = k_i;\ 1 \leq i \leq m+1) e^{-\lambda} \frac{\lambda^r}{r!}.$$

En efecto:
Por la fórmula de la probabilidad total,

$$\mathbb{P}(N[\bar{F}_N(t_i) - \bar{F}_N(t_{i-1})] = k_i;\ 1 \leq i \leq m+1)$$

$$= \sum_{r=0}^{\infty} \mathbb{P}(N[\bar{F}_N(t_i) - \bar{F}_N(t_{i-1})] = k_i;\ 1 \leq i \leq m+1 \mid N = r) \mathbb{P}(N = r)$$

$$
\begin{aligned}
&= \sum_{r=0}^{\infty} \mathbb{P}(r[\bar{F}_r(t_i) - \bar{F}_r(t_{i-1})] = k_i;\ 1 \leq i \leq m+1) \frac{e^{-\lambda}\lambda^r}{r!} \\
&= \mathbb{P}(r_0[\bar{F}_{r_0}(t_i) - \bar{F}_{r_0}(t_{i-1})] = k_i;\ 1 \leq i \leq m+1) \frac{e^{-\lambda}\lambda^{r_0}}{r_0!} \\
&= \frac{r_0!}{k_1!k_2!\ldots k_m!} \prod_{i=1}^{m+1} (t_i - t_{i-1})^{k_i} \frac{e^{-\lambda}\lambda^{r_0}}{r_0!} \\
&= \prod_{i=1}^{m+1} \frac{e^{-\lambda(t_i - t_{i-1})}[\lambda(t_i - t_{i-1})]^{k_i}}{k_i!},
\end{aligned}
$$

donde $r_0 = k_1 + k_2 + \ldots + k_m$.

Esto prueba la independencia en los incrementos, y al mismo tiempo especifíca las distribuciones marginales univariados. Esto completa la prueba. □

2.5. Estadísticas de orden

En esta sección daremos algunas propiedades básicas de las estadísticas de orden, con especial énfasis en la distribución uniforme y exponencial.

2.5.1. Lema. *Sean $X_1, X_2, \ldots, X_n$ una muestra de observaciones independientes con distribución F. Entonces:*

$$G_i(x) := \mathbb{P}(X_{i:n} \leq x) = \sum_{k=i}^{n} \binom{n}{k} F^k(x)(1 - F(x))^{n-k};\ 1 \leq i \leq n.$$

Demostración:

$$
\begin{aligned}
\mathbb{P}(X_{i:n} \leq x) &= \mathbb{P}(F_n(X_{i:n}) \leq F_n(x)) \\
&= \mathbb{P}(\frac{i}{n} \leq F_n(x)) \\
&= \mathbb{P}(nF_n(x) \geq i) \\
&= \sum_{k=i}^{n} \binom{n}{k} F^k(x)(1 - F(x))^{n-k};\ 1 \leq i \leq n.
\end{aligned}
$$

□

Ahora determinaremos la distribución conjunta de las estadísticas de orden.

Por construcción $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. La distribución conjunta tiene soporte K, del conjunto de vectores $x = (x_1, x_2, \ldots, x_n)$ de dimensión n, que satisfacen $x_1 \leq x_2 \leq \ldots \leq x_n$. El conjunto de rectangulos $\prod_{i=1}^{n}(a_i, b_i]$, con $a_1 \leq b_1 \leq \ldots \leq a_n \leq b_n$, determinan unicamente una distribución sobre K. Debido a que las $X_i's$ son i.i.d. se tiene

$$
\begin{aligned}
\mathbb{P}[a_i < X_{i:n} \leq b_i;\ 1 \leq i \leq n] &= \mathbb{P}[\bigcup_{\sigma \in \Sigma}(a_i < X_{\sigma(i)} \leq b_i;\ 1 \leq i \leq n)] \\
&= \sum_{\sigma \in \Sigma} \mathbb{P}(a_i < X_{\sigma(i)} \leq b_i;\ 1 \leq i \leq n) \\
&= n! \prod_{i=1}^{n} \mathbb{P}(a_i < X_i \leq b_i) \\
&= n! \prod_{i=1}^{n} [F(b_i) - F(a_i)].
\end{aligned}
$$

esto es

$$
\mathbb{P}[a_i < X_{i:n} \leq b_i;\ 1 \leq i \leq n] = n! \prod_{i=1}^{n} [F(b_i) - F(a_i)].
$$

Cuando F admite una densidad de Lebesgue f, las $X_{i:n}'s$ también poseen una densidad de Lebesgue, con soporte K.

2.5.2. Lema. *Sean $X_1, X_2, \ldots, X_n$ independientes con densidad f. Entonces $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ tiene la densidad de Lebesgue*

$$
g_n(x_1, \ldots, x_n) = \begin{cases} n! \prod_{i=1}^{n} f(x_i), & \textit{para} \quad x_1 \leq x_2 \leq \ldots \leq x_n \\ 0, & \textit{e.o.c.} \end{cases}
$$

Demostración:

$$
\begin{aligned}
\mathbb{P}(X_{i:n} \leq x_i;\ 1 \leq i \leq n) &= n! \mathbb{P}(X_i \leq x_i;\ 1 \leq i \leq n) \\
&= n! \prod_{i=1}^{n} F(x_i) \\
&= G(x_1, x_2, \ldots, x_n)
\end{aligned}
$$

Entonces,

$$
\begin{aligned}
\frac{\partial^n G}{\partial x_1 \partial x_2 \ldots \partial x_n} &= n! \prod_{i=1}^{n} f(x_i) \\
&= g_n(x_1, x_2, \ldots, x_n) \ \textit{si} \ x_1 \leq x_2 \leq \ldots \leq x_n.
\end{aligned}
$$

□

Como casos especiales, damos los siguiente ejemplos:
$F = F_U$:

$$g_n(x_1, \ldots, x_n) = \begin{cases} n!, & \text{si} \quad 0 \leq x_1 \leq \ldots \leq x_n \leq 1 \\ 0, & \text{e.o.c.} \end{cases}$$

$F = Exp(1)$:

$$g_n(x_1, \ldots, x_n) = \begin{cases} n!exp(-x_1 - \ldots - x_n), & \text{si} \quad 0 \leq x_1 \leq \ldots \leq x_n < \infty \\ 0, & \text{e.o.c.} \end{cases}$$

Para conocer mas del carácter distribucional de las estadísticas de orden, recordemos algunos resultados de probabilidad:

Sea (V, W) un vector aleatorio $d_1 + d_2$-dimensional con densidad de Lebesgue g. Sea $\bar{w} \in \mathbb{R}^{d_2}$ tal que $\int g(\bar{v}, \bar{w}) d\bar{v} > 0$. Entonces, la distribución condicional de V dado $W = \bar{w}$ es:

$$\mathbb{P}(V \in A \mid W = \bar{w}) = \frac{\int_A g(\bar{v}, \bar{w}) d\bar{v}}{\int g(\bar{v}, \bar{w}) d\bar{v}}.$$

Sean $W = (U_{1:n}, \ldots, U_{k:n})$ y $V = (U_{k+1:n}, \ldots, U_{n:n})$ para $1 \leq k \leq n$. Entonces, para $0 < y_1 < \ldots < y_k < a_{k+1} < b_{k+1} < \ldots < a_n < b_n < 1$:

$$\mathbb{P}(a_i < U_{i:n} < b_i;\ k+1 \leq i \leq n \mid U_{1:n} = y_1, \ldots, U_{k:n} = y_k)$$

$$\begin{aligned} &= \frac{\int_A g(\bar{v}, \bar{w}) d\bar{v}}{\int g(\bar{v}, \bar{w}) d\bar{v}}, \quad \text{donde} \quad A = \prod_{i=k+1}^{n} (a_i, b_i) \\ &= \frac{n!\lambda\{(x_{k+1}, \ldots, x_n) :\ a_i < x_i < b_i;\ k+1 \leq i \leq n\}}{n!\lambda\{(x_{k+1}, \ldots, x_n) :\ y_k < x_{k+1} < \ldots < x_n < 1\}}. \end{aligned}$$

Desarrollando numerador y denominador de esto último se tiene:

$$\lambda\{(x_{k+1}, \ldots, x_n) :\ a_i < x_i < b_i;\ k+1 \leq i \leq n\}$$

$$\begin{aligned} &= \int_{a_n}^{b_n} \cdots \int_{a_{k+2}}^{b_{k+2}} \int_{a_{k+1}}^{b_{k+1}} dx_{k+1} dx_{k+2} \ldots dx_n \\ &= \prod_{i=k+1}^{n} (b_i - a_i). \end{aligned}$$

$$\begin{aligned}
&\lambda\{(x_{k+1},\ldots,x_n) : y_k < x_{k+1} < \ldots < x_n < 1\} \\
&= \int_{y_k}^{1}\cdots\int_{y_k}^{x_{k+3}}\int_{y_k}^{x_{k+2}} dx_{k+1}dx_{k+2}\ldots dx_n \\
&= \int_{y_k}^{1}\cdots\int_{y_k}^{x_{k+3}} (x_{k+2}-y_k)dx_{k+2}\ldots dx_n \\
&= \int_{y_k}^{1}\cdots\int_{y_k}^{x_{k+4}} \frac{(x_{k+2}-y_k)^2}{2}\bigg|_{y_k}^{x_{k+3}} dx_{k+2}\ldots dx_n \\
&= \int_{y_k}^{1}\cdots\int_{y_k}^{x_{k+4}} \frac{(x_{k+3}-y_k)^2}{2} dx_{k+2}\ldots dx_n \\
&= \ldots \\
&= \frac{(1-y_k)^{n-k}}{(n-k)!}
\end{aligned}$$

Así,

$$\begin{aligned}
&\mathbb{P}(a_i < U_{i:n} < b_i;\ k+1 \leq i \leq n \mid U_{1:n} = y_1,\ldots,U_{k:n} = y_k) \\
&\qquad = \frac{(n-k)!\prod_{i=k+1}^{n}(b_i - a_i)}{(1-y_k)^{n-k}}.
\end{aligned}$$

Observemos que en esta expresión no depende de $y_1,\ldots,y_{k-1}$, se sigue entonces que $U_{1:n},\ldots,U_{n:n}$ es una sucesión de Markov para cada $n \geq 1$ fija. Las probabilidades de transición se dan en el siguiente:

2.5.3. Lema. *Para cada $n \geq 1$, $U_{1:n},\ldots,U_{n:n}$ es una sucesión de Markov, tal que si $0 \leq y_1 \leq \ldots \leq y_k \leq 1$:*

$$\begin{aligned}
&\mathcal{L}\Big(\frac{U_{i:n}-y_k}{1-y_k},\ k+1 \leq i \leq n \mid U_{1:n} = y_1,\ldots,U_{k:n} = y_k\Big) \\
&= \mathcal{L}(U_{1:n-k},\ldots,U_{n-k:n-k}).
\end{aligned}$$

Demostración:

$$y_k < U_{k+1:n} < U_{k+2:n} < \ldots < U_{n:n} < 1$$

Entonces,

$$0 < U_{k+1:n} - y_k < U_{k+2:n} - y_k < \ldots < U_{n:n} - y_k < 1 - y_k$$

y

$$0 < \frac{U_{k+1:n}-y_k}{1-y_k} < \frac{U_{k+2:n}-y_k}{1-y_k} < \ldots < \frac{U_{n:n}-y_k}{1-y_k} < 1,$$

es otra muestra uniforme de estadísticas de orden, de tamaño $n-k$. Por último, definimos

$$U_{1:n-k} = \frac{U_{k+1:n} - y_k}{1-y_k},\ U_{2:n-k} = \frac{U_{k+2:n} - y_k}{1-y_k}, \ldots, U_{n-k:n-k} = \frac{U_{n:n} - y_k}{1-y_k}.$$

□

El siguiente lema presenta una propiedad de Markov reversa para las $U'_{i:n}s$.

2.5.4. Lema. *Para cada* $n \geq 1$ *y* $0 < y_{k+1} \leq \ldots \leq y_n < 1$:

$$\mathcal{L}(U_{i:n},\ 1 \leq i \leq k \mid U_{k+1:n} = y_{k+1}, \ldots, U_{n:n} = y_n) = \mathcal{L}(y_{k+1}U_{i:k},\ 1 \leq i \leq k).$$

Demostración:

$$0 < U_{1:n} < U_{n:2} < \ldots < U_{k:n} < y_{k+1}.$$

Entonces,

$$0 < \frac{U_{1:n}}{y_{k+1}} < \frac{U_{2:n}}{y_{k+1}} < \ldots < \frac{U_{k:n}}{y_{k+1}} < 1,$$

es una muestra uniforme de estadísticas de orden de tamaño k. Así, definiendo

$$U_{1:k} = \frac{U_{1:n}}{y_{k+1}},\ U_{2:k} = \frac{U_{2:n}}{y_{k+1}}, \ldots, U_{k:k} = \frac{U_{k:n}}{y_{k+1}},$$

se tiene,

$$U_{1:n} = y_{k+1}U_{1:k},\ U_{2:n} = y_{k+1}U_{2:k}, \ldots, U_{k:n} = y_{k+1}U_{k:k}.$$

□

2.5.5. Lema. *Para* $1 < r < n$ *y* $0 < x < 1$, *dado* $U_{r:n} = x$, *los vectores aleatorios* $(U_{1:n}, \ldots, U_{r-1:n})$ *y* $(U_{r+1:n}, \ldots, U_{n:n})$ *son independientes y distribuidos como* $(xU_{1:r-1}, \ldots, xU_{r-1:r-1})$ *y* $(x + (1-x)U_{1:n-r}, \ldots, x + (1-x)U_{n-r:n-r})$ *respectivamente.*

Demostración: P.D. que son independientes. Sean

$$0 \leq a_1 \leq b_1 \leq \ldots \leq a_{r-1} \leq b_{r-1} \leq a_{r+1} \leq b_{r+1} \leq \ldots \leq a_n \leq b_n \leq 1.$$

$$\mathbb{P}(a_i < U_{i:n} \leq b_i;\ 1 \leq i \leq n,\ i \neq r \mid U_{r:n} = x)$$

$$= \frac{n!\lambda\{(x_1, \ldots, x_{r-1}, x_{r+1}, \ldots, x_n):\ a_i < x_i \leq b_i;\ 1 \leq i \leq n,\ i \neq r\}}{n!\lambda\{(x_1, \ldots, x_{r-1}, x_{r+1}, \ldots, x_n): 0 < x_1 < \ldots < x_{r-1} < x < x_{r+1} < \ldots < x_n < 1\}}$$

Luego,

$$\begin{aligned}
&\lambda\{(x_1,\ldots,x_{r-1},x_{r+1},\ldots,x_n):\ a_i<x_i\le b_i;\ 1\le i\le n,\ i\ne r\}\\
&= \int_{a_1}^{b_1}\cdots\int_{a_{r-1}}^{b_{r-1}}\int_{a_{r+1}}^{b_{r+1}}\cdots\int_{a_n}^{b_n} dx_n\ldots dx_{r+1}dx_{r-1}\ldots dx_1\\
&= \prod_{i\ne r}(b_i-a_i).
\end{aligned}$$

y

$$\lambda\{(x_1,\ldots,x_{r-1},x_{r+1},\ldots,x_n):0<x_1<\ldots<x_{r-1}<x<x_{r+1}<\ldots<x_n<1\}$$

$$\begin{aligned}
&= \int_0^x\cdots\int_0^{x_2}\int_x^1\cdots\int_{x_{n-1}}^1 dx_n\ldots dx_{r+1}dx_1\ldots dx_{r-1}\\
&= \int_0^x\cdots\int_0^{x_2}\int_x^1\cdots\int_{x_{n-2}}^1 (1-x_{n-1})dx_{n-1}\ldots dx_{r+1}dx_1\ldots dx_{r-1}\\
&= \int_0^x\cdots\int_0^{x_2}\int_x^1\cdots\int_{x_{n-3}}^1 \left.\frac{-(1-x_{n-1})^2}{2}\right|_{x_{n-2}}^1 dx_{n-2}\ldots dx_{r+1}dx_1\ldots dx_{r-1}\\
&= \int_0^x\cdots\int_0^{x_2}\int_x^1\cdots\int_{x_{n-3}}^1 \frac{(1-x_{n-2})^2}{2}dx_{n-2}\ldots dx_{r+1}dx_1\ldots dx_{r-1}\\
&= \ldots\\
&= \int_0^x\cdots\int_0^{x_2}\frac{(1-x)^{n-r}}{(n-r)!}dx_1\ldots dx_{r-1}\\
&= \frac{(1-x)^{n-r}}{(n-r)!}\int_0^x\cdots\int_0^{x_3} x_2dx_2\ldots dx_{r-1}\\
&= \frac{(1-x)^{n-r}}{(n-r)!}\int_0^x\cdots\int_0^{x_4} \left.\frac{x_2^2}{2}\right|_0^{x_3} dx_3\ldots dx_{r-1}\\
&= \frac{(1-x)^{n-r}}{(n-r)!}\int_0^x\cdots\int_0^{x_4} \frac{x_3^2}{2}dx_3\ldots dx_{r-1}\\
&= \ldots\\
&= \frac{(1-x)^{n-r}}{(n-r)!}\frac{x^{r-1}}{(r-1)!}.
\end{aligned}$$

Así que,

$$\mathbb{P}(a_i<U_{i:n}\le b_i;\ 1\le i\le n,\ i\ne r\mid U_{r:n}=x)$$

$$
\begin{aligned}
&= \frac{n!\prod_{i\neq r}(b_i - a_i)}{n!x^{r-1}(1-x)^{n-r}/(r-1)!(n-r)!} \\
&= \frac{(r-1)!(n-r)!\prod_{i\neq r}(b_i - a_i)}{x^{r-1}(1-x)^{n-r}} \\
&= (r-1)!\prod_{i=1}^{r-1}\left[\frac{b_i - a_i}{x}\right] \times (n-r)!\prod_{i=r+1}^{n}\left[\frac{b_i - a_i}{1-x}\right].
\end{aligned}
$$

De aquí se sigue que los vectores son independientes. Ahora,

$$0 < U_{1:n} < \ldots < U_{r-1:n} < x.$$

Entonces,

$$0 < \frac{U_{1:n}}{x} < \ldots < \frac{U_{r-1:n}}{x} < 1.$$

Definamos

$$\frac{U_{1:n}}{x} = U_{1:r-1}, \ldots, \frac{U_{r-1:n}}{x} = U_{r-1:r-1}.$$

Se sigue entonces,

$$U_{1:n} = xU_{1:r-1}, \ldots, U_{r-1:n} = xU_{r-1:r-1}.$$

Por otra parte:

$$x < U_{r+1:n} < \ldots < U_{n:n} < 1.$$

Entonces,

$$0 < U_{r+1:n} - x < \ldots < U_{n:n} - x < 1 - x$$

y

$$0 < \frac{U_{r+1:n} - x}{1-x} < \ldots < \frac{U_{n:n} - x}{1-x} < 1.$$

Definamos

$$\frac{U_{r+1:n} - x}{1-x} = U_{1:n-r}, \ldots, \frac{U_{n:n} - x}{1-x} = U_{n-r:n-r}.$$

Se sigue entonces,

$$U_{r+1:n} = x + (1-x)U_{1:n-r}, \ldots, U_{n:n} = x + (1-x)U_{n-r:n-r}.$$

Por lo tanto, los vectores $(U_{1:n}, \ldots, U_{r-1:n})$ y $(U_{r+1:n}, \ldots, U_{n:n})$ se distribuyen como $(xU_{1:r-1}, \ldots, xU_{r-1:r-1})$ y $(x + (1-x)U_{1:n-r}, \ldots, x + (1-x)U_{n-r:n-r})$ respectivamente. □

Cuando F es derivable, $f = F'$. Entonces G_i también tiene una densidad de Lebesgue. Primero veamos que

$$G_n(x) = \mathbb{P}(U_{n:n} \leq x) = \binom{n}{n} x^n(1-x)^{n-n} = x^n, \quad \text{si } 0 \leq x \leq 1.$$

Entonces,

$$g_n(x) = nx^{n-1}, \quad \text{si} \quad 0 \leq x \leq 1.$$

2.5.6. Lema. *Para $1 \leq i \leq n$,*

$$g_i(x) = \begin{cases} x^{i-1}(1-x)^{n-i}/B(i, n-i+1), & si \quad 0 \leq x \leq 1 \\ 0, & e.o.c. \end{cases}$$

donde

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt, \quad a, b > 0,$$

*denota a la **función Beta.***

Recordemos que

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

Γ es la **función Gama** que satisface

$$\Gamma(a) = (a-1)!, \quad a \in \mathbb{N}.$$

Demostración: Del lema 2.5.1,

$$G_i(x) = \sum_{k=i}^{n} \binom{n}{k} x^k(1-x)^{n-k}.$$

Entonces,

$$\begin{aligned} g_i(x) &= \sum_{k=i}^{n} \binom{n}{k} [kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}] \\ &= \sum_{k=i}^{n} \frac{n!}{k!(n-k)!}[kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}] \\ &= \sum_{k=i}^{n} \left[\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k} - \frac{n!}{k!(n-k-1)!}x^k(1-x)^{n-k-1}\right] \\ &= \frac{n!}{(i-1)!(n-i)!}x^{i-1}(1-x)^{n-i} \end{aligned}$$

La última igualdad, se sigue del hecho de que se tiene una suma telescopica. Finalmente,

$$\begin{aligned}\frac{n!}{(i-1)!(n-i)!} &= \frac{1}{(i-1)!(n-i)!/n!} = \frac{1}{\Gamma(i)\Gamma(n-i+1)/\Gamma(n+1)!} \\ &= \frac{1}{B(i,n-i+1)}\end{aligned}$$

□

Para una función de distribución general F, con densidad f, debido a que

$$\mathbb{P}(X_{i:n} \leq x) = \mathbb{P}(U_{i:n} \leq F(x)),$$

$X_{i:n}$ tiene densidad

$$f_i(x) = f(x)F^{i-1}(x)(1-F(x))^{n-i}/B(i,n-i+1).$$

Ahora, calculamos los momentos de $U_{i:n}$.

2.5.7. Corolario. *Para $j \geq 1$ y $1 \leq i \leq n$,*

$$\mathbb{E}U_{i:n}^j = \frac{B(j+i,n-i+1)}{B(i,n-i+1)}.$$

Demostración:

$$\begin{aligned}\mathbb{E}U_{i:n}^j &= \int_0^1 x^j x^{i-1}(1-x)^{n-i}/B(i,n-i+1)dx \\ &= \frac{1}{B(i,n-i+1)}\int_0^1 x^{j+i-1}(1-x)^{n-i}dx \\ &= \frac{1}{B(i,n-i+1)}\int_0^1 x^{j+i-1}(1-x)^{(n-i+1)-1}dx \\ &= \frac{B(j+i,n-i+1)}{B(i,n-i+1)}.\end{aligned}$$

□

Como caso especial se tiene que:

$$\begin{aligned}\mathbb{E}U_{i:n} &= \frac{B(1+i,n-i+1)}{B(i,n-i+1)} = \frac{\Gamma(i+1)\Gamma(n-i+1)/\Gamma(n+2)}{\Gamma(i)\Gamma(n-i+1)/\Gamma(n+1)} \\ &= \frac{\Gamma(n+1)\Gamma(i+1)}{\Gamma(n+2)\Gamma(i)} = \frac{n!i!}{(n+1)!(i-1)!} = \frac{i}{n+1} \equiv \pi_{i,n}.\end{aligned}$$

y,

$$\begin{aligned}\mathbb{E}U_{i:n}^2 &= \frac{B(i+2,n-i+1)}{B(i,n-i+1)} = \frac{\Gamma(n+1)\Gamma(i+2)}{\Gamma(n+3)\Gamma(i)} \\ &= \frac{i(i+1)}{(n+1)(n+2)} = \frac{i+1}{n+2}\pi_{i,n}.\end{aligned}$$

De aquí se sigue que:

$$\begin{aligned}VarU_{i:n} &= \mathbb{E}U_{i:n}^2 - \mathbb{E}^2U_{i;n} \\ &= \frac{i+1}{n+2}\pi_{i,n} - \pi_{i:n}^2 \\ &= \pi_{i,n}\left(\frac{i+1}{2} - \pi_{i:n}\right).\end{aligned}$$

Para obtener las esperanzas condicionales de las estadísticas de orden uniformes, observemos primero que del corolario anterior,

$$\mathbb{E}U_{1:n} = \frac{1}{n+1},$$

y para $1 \leq i \leq n$, por la propiedad de Markov,

$$\mathbb{E}\left[U_{i:n} \mid U_{1:n}, \ldots, U_{i-1:n}\right] = \mathbb{E}\left[U_{i:n} \mid U_{i-1:n}\right].$$

Luego, del lema 2.5.3,

$$\mathcal{L}\left(\frac{U_{j:n} - y_{i-1}}{1 - y_{i-1}};\ i \leq j \leq n \mid U_{1:n} = y_1, \ldots, U_{i-1:n} = y_{i-1}\right)$$
$$= \mathcal{L}\left(U_{1:n-i+1}, \ldots, U_{n-i+1:n-i+1}\right).$$

Entonces,

$$\mathbb{E}\left[\frac{U_{i:n} - y_{i-1}}{1 - y_{i-1}} \mid U_{1:n} = y_1, \ldots, U_{i-1:n} = y_{i-1}\right]$$

$$\begin{aligned} &= \mathbb{E}\left[\frac{U_{i:n} - y_{i-1}}{1 - y_{i-1}} \mid U_{i-1:n} = y_{i-1}\right] \\ &= \mathbb{E}[U_{1:n-i+1}].\end{aligned}$$

Por otro lado,

$$\mathbb{E}\left[\frac{U_{i:n} - y_{i-1}}{1 - y_{i-1}} \mid U_{i-1:n} = y_{i-1}\right] = \frac{\mathbb{E}[U_{i:n} \mid U_{i-1:n} = y_{i-1}] - y_{i-1}}{1 - y_{i-1}}.$$

Así que,

$$\frac{\mathbb{E}[U_{i:n} \mid U_{i-1:n} = y_{i-1}] - y_{i-1}}{1 - y_{i-1}} = \mathbb{E}[U_{1:n-i+1}].$$

Entonces,

$$\mathbb{E}[U_{i:n} \mid U_{i-1:n} = y_{i-1}] = (1 - y_{i-1})\mathbb{E}[U_{1:n-i+1}] + y_{i-1}.$$

Por lo tanto,

$$\begin{aligned}
\mathbb{E}\left[U_{i:n} \mid U_{i-1:n}\right] &= (1 - U_{i-1:n})\mathbb{E}[U_{1:n-i+1}] + U_{i-1:n} \\
&= U_{i-1:n} + \frac{1 - U_{i-1:n}}{n-i+2} \\
&= \frac{(n-i+2)U_{i-1:n} + 1 - U_{i-1:n}}{n-i+2} \\
&= \frac{(n-i+1)U_{i-1:n} + 1}{n-i+2} \\
&= \frac{n-i+1}{n-i+2}U_{i-1:n} + \frac{1}{n-i+2}.
\end{aligned}$$

Como una consecuencia, tenemos el siguiente:

2.5.8. Lema. *La sucesión*

$$\frac{U_{i:n} - \frac{i}{n+1}}{n-i+1}, \quad 1 \leq i \leq n,$$

es una martingala con media cero, con respecto a la filtración natural

$$\mathcal{F}_i = \sigma(U_{j:n} : 1 \leq j \leq i), \quad 1 \leq i \leq n.$$

Demostración: Es claro que esta sucesión tiene media cero, debido a que $\mathbb{E}[U_{i:n}] = \frac{i}{n+1}$. Además,

$$\begin{aligned}
\mathbb{E}\left[\frac{U_{i:n} - \frac{i}{n+1}}{n-i+1} \mid \mathcal{F}_{i-1}\right] &= \frac{1}{n-i+1}\left(\mathbb{E}[U_{i:n} \mid \mathcal{F}_{i-1}] - \frac{i}{n+1}\right) \\
&= \frac{1}{n-i+1}\left(\mathbb{E}[U_{i:n} \mid U_{i-1:n}] - \frac{i}{n+1}\right) \\
&= \frac{1}{n-i+1}\left(\frac{n-i+1}{n-i+2}U_{i-1:n} + \frac{1}{n-i+2} - \frac{i}{n+1}\right) \\
&= \frac{U_{i-1:n}}{n-i+2} + \frac{1}{n-i+1}\left(\frac{1}{n-i+2} - \frac{i}{n+1}\right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{U_{i-1:n}}{n-i+2} + \frac{1}{n-i+1}\left(\frac{n+1-ni+i^2-2i}{(n-i+2)(n+1)}\right) \\
&= \frac{U_{i-1:n}}{n-i+2} + \frac{1}{n-i+1}\left(\frac{-(i-1)(n-i+1)}{(n-i+2)(n+1)}\right) \\
&= \frac{U_{i-1:n}}{n-i+2} - \frac{i-1}{(n-i+2)(n+1)} \\
&= \frac{U_{i-1:n} - \frac{i-1}{n+1}}{n-i+2}.
\end{aligned}$$

Por lo tanto, la sucesión

$$\frac{U_{i:n} - \frac{i}{n+1}}{n-i+1}, \quad 1 \leq i \leq n,$$

es una martingala. □

Para una martingala reversa, la tiene el siguiente:

2.5.9. Lema. *La sucesión*

$$(n+1)U_{i:n}/i, \quad 1 \leq i \leq n,$$

es una martingala reversa con media uno, con respecto a la filtración natural

$$\mathcal{F}_i = \sigma(U_{j:n};\ i \leq j \leq n).$$

Demostración: Por la propiedad de Markov reversa,

$$\mathbb{E}[U_{i:n} \mid U_{i+1:n}, \ldots, U_{n:n}] = \mathbb{E}[U_{i:n} \mid U_{i+1:n}].$$

Por otra parte, del lema 2.5.4,

$$\mathcal{L}(U_{j:n};\ 1 \leq j \leq i \mid U_{i+1:n} = y_{i+1}, \ldots, U_{n:n} = y_n) = \mathcal{L}(y_{i+1}U_{j:i};\ 1 \leq j \leq i).$$

Entonces,

$$\begin{aligned}
\mathbb{E}[U_{i:n} \mid U_{i+1:n} = y_{i+1}, \ldots, U_{n:n} = y_n] &= \mathbb{E}[U_{i:n} \mid U_{i+1:n} = y_{i+1}] \\
&= \mathbb{E}[y_{i+1}U_{i:i}] \\
&= y_{i+1}\mathbb{E}[U_{i:i}] \\
&= y_{i+1}\frac{i}{i+1}.
\end{aligned}$$

Por lo tanto,

$$\mathbb{E}[U_{i:n} \mid U_{i+1:n} = y_{i+1}] = y_{i+1}\frac{i}{i+1}.$$

Entonces,

$$\mathbb{E}[U_{i:n} \mid U_{i+1:n}] = U_{i+1:n}\frac{i}{i+1}.$$

Luego,

$$\begin{aligned}
\mathbb{E}\left[(n+1)U_{i:n}/i \mid \mathcal{F}_{i+1}\right] &= \frac{n+1}{i}\mathbb{E}[U_{i:n} \mid \mathcal{F}_{i+1}] \\
&= \frac{n+1}{i}\mathbb{E}[U_{i:n} \mid U_{i+1:n}] \\
&= \frac{n+1}{i}U_{i+1:n}\frac{i}{i+1} \\
&= (n+1)U_{i+1:n}/(i+1).
\end{aligned}$$

Por lo tanto, la sucesión

$$(n+1)U_{i:n}/i, \quad 1 \le i \le n,$$

es una martingala reversa. □

Ahora discutiremos el comportamiento distribucional de las estadísticas de orden $0 \le E_{1:n} \le \ldots \le E_{n:n}$, donde $E_1, E_2, \ldots, E_n$ son variables aleatorias independientes, con distribución $Exp(1)$.

Recordemos que el vector $(E_{1:n}, \ldots, E_{n:n})$, tiene densidad

$$g_n(x_1, \ldots, x_n) = \begin{cases} n!exp\left\{-\sum_{i=1}^n x_i\right\}, & \text{si } \ 0 \le x_1 < \ldots < x_n \\ 0, & \text{e.o.c.} \end{cases}$$

Definamos $E_{0:n} = 0$ y

$$Y_{in} \equiv Y_i = (n-i+1)(E_{i:n} - E_{i-1:n}).$$

Entonces,

$$E_{i:n} - E_{i-1:n} = \frac{Y_i}{n-i+1}.$$

Por lo tanto,

$$E_{r:n} = \sum_{i=1}^{r}(E_{i:n} - E_{i-1:n}) = \sum_{i=1}^{r}\frac{Y_i}{n-i+1}.$$

Lo interesante de esta representación, es que las $Y_i's$ tienen una estructura distribucional simple.

2.5.10. Lema. *Las variables aleatorias $Y_1, \ldots, Y_n$ son i.i.d. $Exp(1)$.*

Demostración:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} E_{1:n} \\ \vdots \\ E_{n:n} \end{pmatrix}$$

donde A es una matriz de $n \times n$, definida como

$$A = \begin{pmatrix} n & 0 & \ldots & \ldots & & 0 \\ -(n-1) & n-1 & \ldots & & & 0 \\ 0 & -(n-2) & n-2 & & & \\ \vdots & & & \ddots & & \\ & & & & 2 & 0 \\ 0 & \ldots & \ldots & \ldots & -1 & 1 \end{pmatrix}.$$

Como $detA = n! \neq 0$, entonces existe A^{-1}, y

$$\begin{pmatrix} E_{1:n} \\ \vdots \\ E_{n:n} \end{pmatrix} = A^{-1} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Luego,

$$J = detA^{-1} = \frac{1}{detA} = \frac{1}{n!}.$$

Entonces,

$$\begin{aligned} \varphi_n(y_1, \ldots, y_n) &= |J|\, g_n(\varphi_1(y_1, \ldots, y_n), \ldots, \varphi_n(y_1, \ldots, y_n)) \\ &= \frac{1}{n!} g_n\left(\frac{y_1}{n}, \frac{y_1}{n} + \frac{y_2}{n-1}, \ldots, \frac{y_1}{n} + \frac{y_2}{n-1} + \ldots + \frac{y_n}{1}\right) \\ &= \frac{1}{n!} n! \exp\left[-\sum_{i=1}^{n} y_i\right] \\ &= exp\left[-\sum_{i=1}^{n} y_i\right] \quad \text{si } y_1, y_2, \ldots, y_n \geq 0. \end{aligned}$$

Por lo tanto,

$$Y_1, \ldots, Y_n \text{ son v.a.i.i.d. } Exp(1).$$

□

Ahora, si representamos a las $E_i's$, en terminos de las variables aleatorias uniformes $U_1, \ldots, U_n$:

$$E_i = -\ln(1 - U_i).$$

Puesto que

$$1 - U_{n:n} < \ldots < 1 - U_{1:n},$$

$$-\ln(1 - U_{1:n}) < \ldots < -\ln(1 - U_{n:n}).$$

Entonces,

$$\begin{aligned} -\ln(1 - U_{r:n}) &= \sum_{i=1}^{r} \frac{Y_i}{n-i+1} \\ &= \sum_{i=1}^{r-1} \frac{Y_i}{n-i+1} + \frac{Y_r}{n-r+1} \\ &= -\ln(1 - U_{r-1:n}) + \frac{Y_r}{n-r+1}. \end{aligned}$$

Así, se tiene que

$$-Y_r = \ln \left[\frac{1 - U_{r-1:n}}{1 - U_{r:n}} \right]^{n-r+1}.$$

Luego,

$$Q_r \equiv exp(-Y_r) = \left[\frac{1 - U_{r-1:n}}{1 - U_{r:n}} \right]^{n-r+1}.$$

Junto con el lema 2.5.10, podemos derivar el siguiente:

2.5.11. Corolario. *Las variables aleatorias $Q_1, \ldots, Q_n$ son i.i.d. F_U.*

Demostración: para todo $1 \leq i \leq n$, Y_i tiene distribución $Exp(1)$. Entonces,

$$\begin{aligned} \mathbb{P}(Q_i \leq q_i) &= \mathbb{P}(\exp(-Y_i) \leq q_i) \\ &= \mathbb{P}(-Y_i \leq \ln q_i) \\ &= \mathbb{P}(Y_i \geq -\ln q_i) \\ &= 1 - \mathbb{P}(Y_i \leq \ln q_i) \\ &= 1 - (1 - \exp(\ln q_i)) = q_i. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(Q_i \leq q_i) = q_i \quad \text{para todo } 1 \leq i \leq n.$$

Además, las $Q_i's$ son independientes, puesto que $Q_i = \varphi(Y_i)$ para todo $1 \leq i \leq n$. Por lo tanto,

$$Q_1, \ldots, Q_n \text{ son i.i.d. } F_U.$$

□

Existe otra variante interesante de este corolario, al caso general de una función de distribución continua F, que habla de la función de riesgo acumulativa Λ_F.

2.5.12. Corolario. *Sean $X_1, \ldots, X_n$ v.a.i.i.d. con función de distribución F continua. Entonces, las variables aleatorias*

$$Z_i := (n - i + 1)\left[\Lambda_F(X_{i:n}) - \Lambda_F(X_{i-1:n})\right],$$

son i.i.d. Exp(1), para todo $1 \leq i \leq n$.

Demostración: Definamos

$$\Lambda_F(X_{0:n}) = 0.$$

Como F es continua, entonces

$$\Lambda_F(t) = \int_{-\infty}^{t} \frac{dF(x)}{1 - F(x)} = -\ln(1 - F(t)).$$

Luego,

$$\Lambda_F(X) = -\ln(1 - F(X)) = -\ln(1 - U).$$

Así,

$$\Lambda_F(X_{i:n}) = -\ln(1 - U_{i:n}) \quad \text{por ser } \Lambda_F \text{ creciente.}$$

Finalmente, aplicando el lema 2.5.10,

$$Z_i := (n - i + 1)\left[\Lambda_F(X_{i:n}) - \Lambda_F(X_{i-1:n})\right], \quad 1 \leq i \leq n \text{ son i.i.d. Exp(1).}$$

□

Este corolario es una extensión del lema 2.5.10, al caso general de una función de distribución F continua. Cuando las $X_i's$ son $Exp(1)$, entonces $\Lambda_F(t) = t$, por lo que las $Z_i's$ son iguales a las $Y_i's$ de antes.
Este corolario, revela que Λ_F, puede sevir como una herramienta para transformar la dependencia de las estadísticas de orden, a una muestra de variables aleatorias independientes con una distribución específica.

2.5.13. Lema. *El vector $(U_{1:n}, \ldots, U_{n:n})$ de estadísticas de orden uniformes, tiene la misma distribución que*

$$\left(\frac{Z_1}{Z_{n+1}}, \ldots, \frac{Z_n}{Z_{n+1}}\right),$$

donde $Z_i = E_1 + \ldots + E_i$, es la i-ésima suma parcial de una sucesión $E_1, \ldots, E_{n+1}$ de v.a.i.i.d. Exp(1).

Demostración: La densidad conjunta de $E_1, \ldots, E_{n+1}$ es:

$$g_{n+1}(x_1, \ldots, x_{n+1}) = \begin{cases} \prod_{i=1}^{n+1} \exp(-x_i), & \text{si} \quad x_i \geq 0 \quad \text{para todo } 1 \leq i \leq n \\ 0, & \text{e.o.c.} \end{cases}$$

Sean

$$z_1 = x_1, z_2 = x_1 + x_2, \ldots, z_{n+1} = x_1 + x_2 + \ldots + x_{n+1}.$$

Entonces,

$$x_1 = z_1, x_2 = z_2 - z_1, \ldots, x_{n+1} = z_{n+1} - z_n,$$

y

$$J = \begin{vmatrix} 1 & 0 & & \ldots & 0 \\ -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 \end{vmatrix} = 1.$$

Entonces, la densidad conjunta de $Z_1, \ldots, Z_{n+1}$, es:

$$\begin{aligned} \varphi_{n+1}(z_1, \ldots, z_{n+1}) &= |J|\, g_{n+1}(z_1, z_2 - z_1, \ldots, z_{n+1} - z_n) \\ &= \exp\left(-\sum_{i=1}^{n+1}(z_i - z_{i-1})\right) \\ &= \exp(-z_{n+1}) = exp\left(-\sum_{i=1}^{n+1} x_i\right). \end{aligned}$$

Luego, sean

$$0 \leq a_1 \leq b_1 \leq \ldots \leq a_{n+1} \leq b_{n+1}.$$

Entonces,

$$\mathbb{P}(a_i < Z_i \leq b_i;\ 1 \leq i \leq n+1)$$

$$\begin{aligned} &= \int_{a_1}^{b_1} \int_{a_2 - x_1}^{b_2 - x_1} \cdots \int_{a_{n+1} - x_1 - \ldots - x_n}^{b_{n+1} - x_1 - \ldots - x_n} \exp\left(-\sum_{i=1}^{n+1} x_i\right) \boldsymbol{dx}_{n+1} dx_n \ldots dx_1 \\ &= \int_{a_1}^{b_1} \int_{a_2 - x_1}^{b_2 - x_1} \cdots \int_{a_{n+1}}^{b_{n+1}} \exp(-z_{n+1}) dz_{n+1} dx_n \ldots \boldsymbol{dx}_1 \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_{n+1}}^{b_{n+1}} \exp(-z_{n+1}) dz_{n+1} dz_n \ldots dz_1. \end{aligned}$$

Esto muestra que $(Z_1, \ldots, Z_{n+1})$ tiene la densidad de Lebesgue:

$$(z_1, \ldots, z_{n+1}) \to \begin{cases} \exp(-z_{n+1}), & \text{si } \ 0 \leq z_1 < z_2 < \ldots < z_{n+1} \\ 0, & \text{e.o.c.} \end{cases}$$

Finalmente,

$$\mathbb{P}\left(a_i < \frac{Z_i}{Z_{n+1}} \leq b_i;\ 1 \leq i \leq n\right)$$

$$\begin{aligned} &= \int_{a_1 z_{n+1}}^{b_1 z_{n+1}} \cdots \int_{a_n z_{n+1}}^{b_n z_{n+1}} \int_0^{\infty} \exp(-z_{n+1}) dz_{n+1} \ldots dz_1 \\ &= \prod_{i=1}^{n}(b_i - a_i) \int_0^{\infty} x^n \exp(-x) dx \\ &= n! \prod_{i=1}^{n}(b_i - a_i). \end{aligned}$$

□

Los resultados presentados en esta sección, son herramientas fundamentales para probar algunos resultados acerca de las funciones de distribución empíricas.

2.6. Cotas de probabilidad de frontera

En esta sección, se enuncian algunos resultados que caracterizan la desviación entre F_n y F, para una n fija.

El primer resultado, dá la distribución exacta de **desviación relativa** máxima, entre F_n y F:

$$R_n = \sup_{0<F(t)} \frac{F_n(t)}{F(t)}.$$

2.6.1. Teorema. *Para una F continua y $0 < \varepsilon \leq 1$,*

$$G_n(\varepsilon) \equiv \mathbb{P}(R_n < \frac{1}{\varepsilon}) = 1 - \varepsilon.$$

Demostración: Observemos que G_n es libre de distribución, en efecto:

$$F_n(t) = \bar{F}_n(F(t)) \ \text{ para todo } t \in \mathbb{R}.$$

Sea

$$u = F(t).$$

Entonces,

$$0 < u \leq 1, \text{ y por la continuidad de } F.$$

$$R_n = \sup_{0<u\leq 1} \bar{F}_n(u)/u \equiv \bar{R}_n.$$

Luego, por inducción sobre n:

Para $n = 1$,

$$\frac{\bar{F}_1(u)}{u} = \frac{1_{\{U_1 \leq u\}}}{u}.$$

Por lo que,

$$\sup_{0<u\leq 1} \frac{\bar{F}_1(u)}{u} = \sup_{0<u\leq 1} \frac{1_{\{U_1 \leq u\}}}{u} = \frac{1}{U_1}.$$

Entonces,

$$\begin{aligned} \mathbb{P}\left(R_1 < \frac{1}{\varepsilon}\right) &= \mathbb{P}\left(\frac{1}{U_1} < \frac{1}{\varepsilon}\right) = \mathbb{P}(U_1 > \varepsilon) \\ &= 1 - \mathbb{P}(U_1 \leq \varepsilon) = 1 - \varepsilon. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}\left(R_1 < \frac{1}{\varepsilon}\right) = 1 - \varepsilon.$$

Supongamos que

$$\mathbb{P}\left(R_{n-1} < \frac{1}{\varepsilon}\right) = 1 - \varepsilon.$$

Observemos que

$$\bar{R}_n = \max_{1\leq k\leq n} k/nU_{k:n},$$

en efecto:

Si $u \in [U_{k:n}, U_{k+1:n})$, entonces

$$\sup_{u\in[U_{k:n},U_{k+1:n})} \frac{\bar{F}_n(u)}{u} = \frac{\bar{F}_n(U_{k:n})}{U_{k:n}} = \frac{k}{nU_{k:n}} \quad \text{para todo } 1 \leq k \leq n.$$

Por lo tanto,

$$\bar{R}_n = \max_{1\leq k\leq n} k/nU_{k:n}.$$

Luego,

$$\begin{aligned}
G_n(\varepsilon) &= \mathbb{P}\left(\max_{1\le k\le n} k/nU_{k:n} < \frac{1}{\varepsilon}\right)\\
&= \mathbb{P}\left(\frac{1}{\max_{1\le k\le n} k/nU_{k:n}} > \varepsilon\right)\\
&= \mathbb{P}(\min_{1\le k\le n} nU_{k:n}/k > \varepsilon).
\end{aligned}$$

El lema 2.5.4, establece que

$$\mathcal{L}(U_{i:n},\ 1\le i\le k \mid U_{k+1:n} = y_{k+1},\dots,U_{n:n}=y_n) = \mathcal{L}(y_{k+1}U_{i:k},\ 1\le i\le k).$$

Entonces se sigue que,

$$\mathcal{L}(U_{k:n},\ 1\le k\le n-1 \mid U_{n:n}=y) = \mathcal{L}(yU_{k:n-1},\ 1\le k\le n-1).$$

También recordemos que $U_{n:n}$ tiene la densidad de Lebesgue ny^{n-1}, si $0 < y < 1$. Entonces se tiene que:

$$\begin{aligned}
G_n(\varepsilon) &= \int_\varepsilon^1 \mathbb{P}(\min_{1\le k\le n-1} nU_{k:n}/k > \varepsilon \mid U_{n:n}=y)dG_n(y)\\
&= \int_\varepsilon^1 \mathbb{P}(\min_{1\le k\le n-1} nyU_{k:n-1}/k > \varepsilon)dG_n(y)\\
&= \int_\varepsilon^1 \mathbb{P}\left(\min_{1\le k\le n-1} (n-1)U_{k:n-1}/k > \frac{(n-1)\varepsilon}{ny}\right)dG_n(y)\\
&= \int_\varepsilon^1 \left[1-\frac{(n-1)\varepsilon}{ny}\right] ny^{n-1}dy\\
&= \int_\varepsilon^1 [ny^{n-1}-(n-1)\varepsilon y^{n-2}]dy\\
&= y^n - \varepsilon y^{n-1}\Big|_\varepsilon^1 = 1-\varepsilon.
\end{aligned}$$

□

En la sección 2.3 se vio que el proceso $t \to F_n(t)/F(t)$ es una martingala reversa con media uno. Además, por la LFGN,

$$\frac{F_n(t)}{F(t)} \to 1 \quad \text{con probabilidad uno.}$$

Sin embargo, si ponemos

$$\frac{1}{\varepsilon} = 1+\delta \quad \text{con } \delta \ge 0,$$

se sigue del teorema 2.6.1,

$$\mathbb{P}\left(\sup_{0<F(t)} \frac{F_n(t)}{F(t)} < 1+\delta\right) = \mathbb{P}\left(\sup_{0<F(t)} \left[\frac{F_n(t)}{F(t)} - 1\right] < \delta\right) = 1 - \frac{1}{1+\delta} = \frac{\delta}{1+\delta}.$$

De aquí se sigue que F_n/F no converge uniformemente a 1 en t. Por otro lado, para formular un resultado positivo al teorema 2.6.1, se analizarán algunas cotas relacionadas con R_n.

Sean $0 \leq s \leq n$ y $a, b > 0$ tales que $a + sb < 1$. Definamos

$$P_{ns}(a,b) = \mathbb{P}(U_{k:n} \leq a + (k-1)b;\ 1 \leq k \leq s \text{ y } U_{s+1:n} > a + sb),$$

con $U_{n+1:n} \equiv 1$.

2.6.2. Lema. ***(Dempster).***

$$P_{ns}(a,b) = \binom{n}{s} a(a+sb)^{s-1}(1-a-sb)^{n-s}.$$

Demostración: por inducción$/n$.

Para $n = 1$; $0 \leq s \leq 1$.

Si $s = 0$,

$$\begin{aligned} P_{10}(a,b) &= \mathbb{P}(U_{1:1} > a) = 1 - \mathbb{P}(U_{1:1} \leq a) \\ &= 1 - a. \end{aligned}$$

Por otra parte,

$$\binom{1}{0} a(a+0b)^{0-1}(1-a-0b)^{1-0} = 1 - a.$$

Por lo tanto,

$$P_{10}(a,b) = \binom{1}{0} a(a+0b)^{0-1}(1-a-0b)^{1-0}.$$

Si $s = 1$,

$$\begin{aligned} P_{11}(a,b) &= \mathbb{P}(U_{1:1} \leq a \text{ y } U_{2:1} = 1 > a + sb) \\ &= \mathbb{P}(U_{1:1} \leq a) = a. \end{aligned}$$

Por otra parte,
$$\binom{1}{1} a(1+b)^0(1-a-b)^0 = a.$$
Por lo tanto,
$$P_{11}(a,b) = \binom{1}{1} a(1+b)^0(1-a-b)^0.$$
La igualdad también se cumple para una n en general y $s=0$. En efecto:
$$\begin{aligned} P_{no}(a,b) &= \mathbb{P}(U_{1:n} > a) = \mathbb{P}(U_1 > a, \ldots, U_n > a) \\ &= \prod_{i=1}^{n} \mathbb{P}(U_1 > a) = \prod_{i=1}^{n}(1 - \mathbb{P}(U_1 \leq a)) \\ &= \prod_{i=1}^{n}(1-a) = (1-a)^n. \end{aligned}$$
Por otro lado,
$$\binom{n}{0} a(a+0b)^{0-1}(1-a-0b)^{n-0} = (1-a)^n.$$
Por lo tanto,
$$P_{no}(a,b) = \binom{n}{0} a(a+0b)^{0-1}(1-a-0b)^{n-0}.$$
Luego, sean $n \geq 2$ y $s \geq 1$. Entonces, $P_{ns}(a,b)$
$$\begin{aligned} &= \mathbb{P}(U_{k:n} \leq a+(k-1)b;\ 1 \leq k \leq s \ \text{ y } \ U_{s+1:n} > a+sb) \\ &= \int_0^a \mathbb{P}(U_{k:n} \leq a+(k-1)b\text{: } 2 \leq k \leq s \text{ y } U_{s+1:n} > a+sb \mid U_{1:n} = y)dG_1(y). \end{aligned}$$
Ahora observemos que si hacemos $k=1$ en el lema 2.5.3, se tiene:
$$\mathcal{L}\left(\frac{U_{i:n} - y_1}{1-y_1};\ 2 \leq i \leq n \mid U_{1:n} = y_1\right) = \mathcal{L}(U_{1:n-1}, \ldots, U_{n-1:n-1}).$$
Así que
$$\mathbb{P}(U_{k:n} \leq a+(k-1)b;\ 2 \leq k \leq s \text{ y } U_{s+1:n} > a+sb \mid U_{1:n} = y)$$
$$\begin{aligned} &= \mathbb{P}\left(\frac{U_{k:n}-y}{1-y} \leq \frac{a+(k-1)b-y}{1-y};\ 2 \leq k \leq s \text{ y } \frac{U_{s+1:n}-y}{1-y} > \frac{a+sb-y}{1-y} \mid U_{1:n} = y\right) \\ &= \mathbb{P}\left(U_{k-1:n-1} \leq \frac{a+(k-1)b-y}{1-y};\ 2 \leq k \leq s \text{ y } U_{s:n-1} > \frac{a+sb-y}{1-y}\right). \end{aligned}$$

Esta última expresión la podemos reescribir como

$$\mathbb{P}\left(U_{i:n-1} \leq \frac{a+b-y}{1-y} + \frac{(i-1)b}{1-y};\ 1 \leq i \leq s-1 \text{ y } U_{s:n-1} > \frac{a+b-y}{1-y} + \frac{(s-1)b}{1-y}\right).$$

Por lo tanto,

$$P_{ns}(a,b) = \int_0^a P_{n-1,s-1}\left(\frac{a+b-y}{1-y}, \frac{b}{1-y}\right) n(1-y)^{n-1} dy.$$

El siguiente paso es aplicar la hipótesis de inducción y resolver la integral para llegar al resultado. □

2.6.3. Corolario. ***(Rényi).*** *Sea* $0 < \varepsilon < 1$. *Entonces*

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq k\leq n} nU_{k:n}/k > \varepsilon\right) &= \sum_{s=0}^{n-1}\binom{n}{s}\left(\frac{\varepsilon}{n}\right)^s (1+s)^{s-1}\left[1 - \frac{(1+s)\varepsilon}{n}\right]^{n-s} \\ &\to \sum_{s=0}^{\infty} \varepsilon^s e^{-(s+1)\varepsilon}(s+1)^{s-1}/s! \ \text{ cuando } n \to \infty.\end{aligned}$$

Demostración: Observemos que

$$\begin{aligned}P_{ns}\left(\frac{\varepsilon}{n}, \frac{\varepsilon}{n}\right) &= \mathbb{P}\left(U_{k:n} \leq \frac{\varepsilon}{n} + (k-1)\frac{\varepsilon}{n};\ 1 \leq k \leq s \text{ y } U_{s+1:n} > \frac{\varepsilon}{n} + s\frac{\varepsilon}{n}\right) \\ &= \mathbb{P}(nU_{k:n} \leq \varepsilon + (k-1)\varepsilon;\ 1 \leq k \leq s \text{ y } nU_{s+1:n} > \varepsilon + s\varepsilon) \\ &= \mathbb{P}\left(nU_{k:n}/k \leq \varepsilon;\ 1 \leq k \leq s \text{ y } nU_{s+1:n}/(s+1) > \varepsilon\right).\end{aligned}$$

Luego, $\mathbb{P}\left(\max_{1\leq k\leq n} nU_{k:n}/k > \varepsilon\right)$

$$\begin{aligned}&= \sum_{s=0}^{n-1} P_{ns}\left(\frac{\varepsilon}{n}, \frac{\varepsilon}{n}\right) \\ &= \sum_{s=0}^{n-1}\binom{n}{s}\left(\frac{\varepsilon}{n}\right)\left(\frac{\varepsilon}{n} + s\frac{\varepsilon}{n}\right)^{s-1}\left(1 - \frac{\varepsilon}{n} - s\frac{\varepsilon}{n}\right)^{n-s} \\ &= \sum_{s=0}^{n-1}\binom{n}{s}\left(\frac{\varepsilon}{n}\right)^s (1+s)^{s-1}\left[1 - \frac{(s+1)\varepsilon}{n}\right]^{n-s} \\ &= \sum_{s=0}^{n-1}\frac{n!}{s!(n-s)!}\frac{\varepsilon^s}{n^s}(1+s)^{s-1}\left[1 - \frac{(s+1)\varepsilon}{n}\right]^n \frac{1}{\left[1 - \frac{(s+1)\varepsilon}{n}\right]^s} \\ &= \sum_{s=0}^{n-1}\frac{\varepsilon^s}{s!}(1+s)^{s-1}\left[1 - \frac{(s+1)\varepsilon}{n}\right]^n \frac{n!}{(n-s)!(n-(s+1)\varepsilon)^s}.\end{aligned}$$

Puesto que

$$\begin{aligned}\frac{n!}{(n-s)!(n-(s+1)\varepsilon)^s} &= \frac{n(n-1)\dots(n-s+1)}{(n-(s+1)\varepsilon)^s}\\ &= \frac{n}{(n-(s+1)\varepsilon)}\frac{n-1}{(n-(s+1)\varepsilon)}\cdots\frac{(n-s+1)}{(n+(s+1)\varepsilon)},\end{aligned}$$

$$\lim_{n\to\infty}\frac{n!}{(n-s)!(n-(s+1)\varepsilon)^s}=1.$$

También

$$\lim_{n\to\infty}\left[1-\frac{(s+1)\varepsilon}{n}\right]^n=e^{-(s+1)\varepsilon}.$$

Por lo tanto,

$$\mathbb{P}\left(\max_{1\le k\le n} nU_{k:n}/k>\varepsilon\right)\to\sum_{s=0}^{\infty}\varepsilon^s e^{-(s+1)\varepsilon}(s+1)^{s-1}/s! \text{ cuando } n\to\infty.$$

□

Ahora derivamos una fórmula para

$$\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k>\varepsilon\right).$$

Primero veamos que

$$\inf_{t>U_{1:n}}\bar{F}_n(t)/t=\left[\max_{1\le k\le n-1} nU_{k+1:n}/k\right]^{-1}:$$

Sea $t\in(U_{1:n},U_{2:n})$. Entonces,

$$\inf_{t\in(U_{1:n},U_{n:2})}\bar{F}_n(t)/t=\frac{\bar{F}_n(U_{2:n}-)}{U_{2:n}}=\frac{1/n}{U_{2:n}}=1/nU_{2:n}.$$

Si $t\in[U_{k:n},U_{k+1:n})$ para todo $2\le k\le n-1$,

$$\inf_{t\in[U_{k:n},U_{k+1:n})}\bar{F}_n(t)/t=\frac{\bar{F}_n(U_{k+1:n}-)}{U_{k+1:n}}=\frac{k/n}{U_{k+1:n}}=k/nU_{k+1:n}.$$

Por lo tanto,

$$\inf_{t>U_{1:n}}\bar{F}_n(t)/t=\min_{1\le k\le n-1}k/nU_{k+1:n}=\left[\max_{1\le k\le n-1}nU_{k+1:n}/k\right]^{-1}.$$

Luego,

$$\begin{aligned}\mathbb{P}\left(\inf_{t>U_{1:n}} \bar{F}_n(t)/t < \frac{1}{\varepsilon}\right) &= \mathbb{P}\left(\frac{1}{\max_{1\le k\le n-1} nU_{k+1:n}/k} < \frac{1}{\varepsilon}\right)\\ &= \mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right).\end{aligned}$$

Ahora, condicionando sobre $U_{1:n}$, obtenemos para $0 < \varepsilon < n$:

$$\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right) = \int_0^1 \mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y)dG_1(y).$$

Esta integral la podemos dividir en dos integrales:

$$\int_0^{\varepsilon/n} \mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y)dG_1(y)$$

y

$$\int_{\varepsilon/n}^1 \mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y)dG_1(y).$$

Luego, observemos que $\mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y) = 1$:

Para $k = 1$, $nU_{2:n}/1 > \varepsilon$ implica $U_{2:n} > \varepsilon/n$. Así, $U_{2:n} \ge U_{1:n} > \varepsilon/n$. Esto prueba que en efecto, lo anterior es cierto. Entonces, el segundo integrando se reduce a

$$\int_{\varepsilon/n}^1 dG_1(y) = \mathbb{P}(U_{1:n} > \varepsilon/n).$$

Por lo tanto,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right) &= \int_0^{\varepsilon/n} \mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y)dG_1(y)\\ &+ \mathbb{P}(U_{1:n} > \varepsilon/n).\end{aligned}$$

Ahora, veamos que

$$\int_0^{\varepsilon/n} \mathbb{P}(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon \mid U_{1:n} = y)dG_1(y)$$

$$= \int_0^{\varepsilon/n} n(1-y)^{n-1} \sum_{s=0}^{r} P_{n-1,s}\left[\frac{\frac{\varepsilon}{n} - y}{1-y}, \frac{\varepsilon}{n(1-y)}\right] dy \quad \text{donde } r = \left[\frac{n}{\varepsilon}\right].$$

Primero observemos que

$$P_{n-1,s}\left[\frac{\frac{\varepsilon}{n}-y}{1-y},\frac{\varepsilon}{n(1-y)}\right]$$

$$\begin{aligned}
&= \mathbb{P}\left(U_{k:n-1}\leq\frac{\frac{\varepsilon}{n}-y}{1-y}+\frac{(k-1)\varepsilon}{n(1-y)};\ 1\leq k\leq s \text{ y } U_{s+1:n-1}>\frac{\frac{\varepsilon}{n}-y}{1-y}+\frac{s\varepsilon}{n(1-y)}\right)\\
&= \mathbb{P}\left(U_{k:n-1}\leq\frac{\varepsilon-ny}{n(1-y)}+\frac{(k-1)\varepsilon}{n(1-y)};\ 1\leq k\leq s \text{ y } U_{s+1:n-1}>\frac{\varepsilon-ny}{n(1-y)}+\frac{s\varepsilon}{n(1-y)}\right)\\
&= \mathbb{P}\left(U_{k:n-1}\leq\frac{k\varepsilon-ny}{n(1-y)};\ 1\leq k\leq s \text{ y } U_{s+1:n-1}>\frac{(s+1)\varepsilon-ny}{n(1-y)}\right).
\end{aligned}$$

Por otra parte,

$$\mathbb{P}(\max_{1\leq k\leq n-1} nU_{k+1:n}/k>\varepsilon\mid U_{1:n}=y)$$

$$\begin{aligned}
&= \mathbb{P}\left(\max_{1\leq k\leq n-1} U_{k+1:n}>\frac{k\varepsilon}{n}\mid U_{1:n}=y\right)\\
&= \mathbb{P}\left(\max_{1\leq k\leq n-1} \frac{U_{k+1:n}-y}{1-y}>\frac{\frac{k\varepsilon}{n}-y}{1-y}\mid U_{1:n}=y\right)\\
&= \mathbb{P}\left(\max_{1\leq k\leq n-1} \frac{U_{k+1:n}-y}{1-y}>\frac{k\varepsilon-ny}{n(1-y)}\mid U_{1:n}=y\right)\\
&= \mathbb{P}\left(\max_{1\leq k\leq n-1} U_{k:n-1}>\frac{k\varepsilon-ny}{n(1-y)}\right)\\
&= \sum_{s=0}^{r} P_{n-1,s}\left[\frac{\frac{\varepsilon}{n}-y}{1-y},\frac{\varepsilon}{n(1-y)}\right].
\end{aligned}$$

La cuarta igualdad, se debe al lema 2.5.3, haciendo k=1. es decir:

$$\mathcal{L}\left(\frac{U_{k+1:n}-y}{1-y};\ 1\leq k\leq n-1\mid U_{1:n}=y\right)=\mathcal{L}(U_{1:n-1},\ldots,U_{n-1:n-1}).$$

El índice s, debe satisfacer: $0\leq s\leq n$ y $a,b>0$, tales que $a+sb<1$. En nuestro caso:

$$a=\frac{\frac{\varepsilon}{n}-y}{1-y},\ b=\frac{\varepsilon}{n(1-y)}.$$

Entonces,

$$\frac{\frac{\varepsilon}{n}-y}{1-y}+s\frac{\varepsilon}{n(1-y)}<1.$$

Resolviendo esta inecuación, se tiene que

$$s < \frac{n}{\varepsilon} - 1.$$

Por eso, $r = \left[\frac{n}{\varepsilon}\right]$. En resúmen, se tiene

$$\begin{aligned}\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right) &= \int_0^{\varepsilon/n} n(1-y)^{n-1}\sum_{s=0}^{r} P_{n-1,s}\left[\frac{\frac{\varepsilon}{n}-y}{1-y}, \frac{\varepsilon}{n(1-y)}\right] dy \\ &+ \mathbb{P}\left(U_{1:n} > \frac{\varepsilon}{n}\right).\end{aligned}$$

Aplicando la fórmula de Dempster, junto con esta igualdad, se tiene el siguiente:

2.6.4. Corolario. ***(Chang).*** *Sea $0 < \varepsilon < n$. Entonces,*

$$\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right) = \left(1-\frac{\varepsilon}{n}\right)^n + \sum_{i=1}^{r+1}\binom{n}{i}\left(\frac{\varepsilon}{n}\right)^i\left(1-\frac{i\varepsilon}{n}\right)^{n-i}(i-1)^{i-1}.$$

Observemos que

$$\sum_{i=1}^{r+1}\binom{n}{i}\left(\frac{\varepsilon}{n}\right)^i\left(1-\frac{i\varepsilon}{n}\right)^{n-i}(i-1)^{i-1}$$

$$\begin{aligned}&= \sum_{k=1}^{r+1}\frac{n!}{k!(n-k)!}\frac{\varepsilon^k}{n^k}\left(1-\frac{k\varepsilon}{n}\right)^{n-k}(k-1)^{k-1} \\ &= \sum_{k=1}^{r+1}\frac{\varepsilon^k(k-1)^{k-1}}{k!}\left(1-\frac{k\varepsilon}{n}\right)^n\frac{n!}{(n-k)!n^k}\left(\frac{n-k\varepsilon}{n}\right)^{-k} \\ &= \sum_{k=1}^{r+1}\frac{\varepsilon^k(k-1)^{k-1}}{k!}\left(1-\frac{k\varepsilon}{n}\right)^n\frac{n(n-1)\dots(n-k+1)}{(n-k\varepsilon)^k}.\end{aligned}$$

Haciendo tender $n\to\infty$, esta última suma converge a

$$\sum_{k=1}^{\infty}\frac{\varepsilon^k(k-1)^{k-1}}{k!}e^{-k\varepsilon} = \sum_{k=1}^{\infty}\frac{(\varepsilon e^{-\varepsilon})^k(k-1)^{k-1}}{k!}.$$

Por lo tanto, haciendo tender $n\to\infty$ para todo $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{1\le k\le n-1} nU_{k+1:n}/k > \varepsilon\right) \to e^{-\varepsilon} + \sum_{k=1}^{\infty}\frac{(\varepsilon e^{-\varepsilon})^k(k-1)^{k-1}}{k!}.$$

El siguiente resultado, dá la distribución exacta de

$$D_n^+ = \sup_{t\in\mathbb{R}}[F_n(t) - F(t)] = \sup_{0\le u\le 1}[\bar{F}_n(u) - u],$$

cuando F es continua, Birnbaum-Tingey [2].

2.6.5. Teorema. ***(Birnbaum-Tingey).*** *Sea* $0 \le x \le 1$ *y* $n \ge 1$. *Entonces,*

$$\mathbb{P}(D_n^+ \le x) = 1 - x \sum_{j=0}^{[n(1-x)]} \binom{n}{j} \left(1 - x - \frac{j}{n}\right)^{n-j} \left(x + \frac{j}{n}\right)^{j-1}.$$

Demostración: Primero observemos que $D_n^+ \le x$ si, y sólo si

$$\max_{1\le i\le n} \{\frac{i}{n} - F(X_{i:n})\} = \max_{1\le i\le n} \{\frac{i}{n} - U_{i:n}\} \le x,$$

si, y sólo si

$$\frac{i}{n} - U_{i:n} \le x$$

o equivalentemente,

$$\frac{i}{n} - x \le U_{i:n} \quad \text{para todo } i = n, n-1, \dots, K,$$

donde K es tal que

$$\frac{K}{n} - x \ge 0 > \frac{K-1}{n} - x.$$

En vista del lema 2.5.2, $\mathbb{P}(D_n^+ \le x)$ es igual a la integral

$$J(x, n, K) = n! \int_{1-x}^{1} \int_{1-x-\frac{1}{n}}^{y-n} \cdots \int_{1-x-\frac{n-K}{n}}^{y_{K+1}} \int_0^{y_K} \cdots \int_0^{y_2} dy_1 \dots dy_n,$$

en efecto:

Para $i = 1, 2, \dots, K-1$, las $y's$ son libres. Para $i = K$;

$$1 - x - \frac{n-K}{n} = \frac{K}{n} - x \le U_{K:n} \le U_{K+1:n}. \text{ entonces,}$$

$$1 - x - \frac{n-K}{n} \le y_K \le y_{K+1}.$$

Para $i = K + 1$;

$$1 - x - \frac{n-(K+1)}{n} = \frac{K+1}{n} - x \le U_{K+1:n} \le U_{K+2:n}. \text{ Entonces,}$$

$$1 - x - \frac{n-(K+1)}{n} \leq y_{K+1} \leq y_{K+2}.$$

Sucesivamente, para $i = n - 1$;

$$1 - x - \frac{1}{n} = \frac{n-1}{n} - x \leq U_{n-1:n} \leq U_{n:n}, \text{ entonces } 1 - x - \frac{1}{n} \leq y_{n-1} \leq y_n.$$

Para $i = n$;

$$1 - x = \frac{n}{n} - x \leq U_{n:n} \leq 1, \text{ entonces } 1 - x \leq y_n \leq 1.$$

Por otro lado,

$$\int_0^{y_K} \cdots \int_0^{y_2} dy_1 \ldots dy_{K-1} = \frac{y_K^{K-1}}{(K-1)!}.$$

Insertando esta expresión en $J(x, n, K)$ obtenemos: $J(x, n, K)$

$$\begin{aligned}
&= n! \int_{1-x}^{1} \cdots \int_{1-x-\frac{n-(K+1)}{n}}^{y_{K+2}} \int_{1-x-\frac{n-K}{n}}^{y_{K+1}} \frac{y_K^{K-1}}{(K-1)!} dy_K dy_{K+1} \ldots dy_n \\
&= n! \int_{1-x}^{1} \cdots \int_{1-x-\frac{n-(K+1)}{n}}^{y_{K+2}} \left. \frac{y_K^K}{K!} \right|_{1-x-\frac{n-K}{n}}^{y_{K+1}} dy_{K+1} \ldots dy_n \\
&= n! \int_{1-x}^{1} \cdots \int_{1-x-\frac{n-(K+1)}{n}}^{y_{K+2}} \left(\frac{y_{K+1}^K}{K!} - \frac{\left(1 - x - \frac{n-K}{n}\right)^K}{K!} \right) dy_{K+1} \ldots dy_n.
\end{aligned}$$

Esta última integral es igual a:

$$J(x, n, K+1) - \frac{\left(1 - x - \frac{n-K}{n}\right)^K}{K!} \times n! \int_{1-x}^{1} \cdots \int_{1-x-\frac{n-(K+1)}{n}}^{y_{K+2}} dy_{K+1} \ldots dy_n$$

$$\equiv J(x, n, K+1) - \frac{\left(1 - x - \frac{n-K}{n}\right)^K}{K!} I(x, n, K+1).$$

Por lo tanto,

$$J(x, n, K) = J(x, n, K+1) - \frac{\left(1 - x - \frac{n-K}{n}\right)^K}{K!} I(x, n, K+1).$$

Entonces,

$$J(x, n, K) - J(x, n, K+1) = -\frac{\left(1 - x - \frac{n-K}{n}\right)^K}{K!} I(x, n, K+1).$$

Sumando ambos lados, se tiene una suma telescopica del lado izquierdo:

$$\sum_{i=K}^{n-1}(J(x,n,i)-J(x,n,i+1)) = -\sum_{i=K}^{n-1}\frac{\left(1-x-\frac{n-i}{n}\right)^i}{i!}I(x,n,i+1).$$

Entonces,

$$J(x,n,K)-J(x,n,n) = -\sum_{i=K}^{n-1}\frac{\left(1-x-\frac{n-i}{n}\right)^i}{i!}I(x,n,i+1).$$

Por lo tanto,

$$J(x,n,K) = J(x,n,n) - \sum_{i=K}^{n-1}\frac{\left(1-x-\frac{n-i}{n}\right)^i}{i!}I(x,n,i+1).$$

Notemos que

$$\begin{aligned}
J(x,n,n) &= n!\int_{1-x}^{1}\int_0^{y_n}\dots\int_0^{y_2} dy_1\dots dy_{n-1}dy_n \\
&= n!\int_{1-x}^{1}\frac{y_n^{n-1}}{(n-1)!}dy_n \\
&= n!\left.\frac{y_n^n}{n!}\right|_{1-x}^{1} = 1-(1-x)^n.
\end{aligned}$$

Mientras que para las $I's$, obtenemos por inducción sobre i:

$$I(x,n,i+1) = n!\frac{x\left(x+\frac{n-i}{n}\right)^{n-i-1}}{(n-i)!}, \quad i=K,\dots,n-i.$$

Luego,

$$J(x,n,K) = 1-(1-x)^n - \sum_{i=K}^{n-1}\frac{\left(1-x-\frac{n-i}{n}\right)^i}{i!}n!\frac{x\left(x+\frac{n-i}{n}\right)^{n-i-1}}{(n-i)!}.$$

Sea $j=n-i$. Entonces.

$$\begin{aligned}
J(x,n,K) &= 1-(1-x)^n - \sum_{j=1}^{n-K}\binom{n}{j}\left(1-x-\frac{j}{n}\right)^{n-j}x\left(x+\frac{j}{n}\right)^{j-1} \\
&= 1-x\sum_{j=0}^{n-K}\binom{n}{j}\left(1-x-\frac{j}{n}\right)^{n-j}\left(x+\frac{j}{n}\right)^{j-1} \\
&= 1-x\sum_{j=0}^{[n(1-x)]}\binom{n}{j}\left(1-x-\frac{j}{n}\right)^{n-j}\left(x+\frac{j}{n}\right)^{j-1}.
\end{aligned}$$

La última igualdad se debe a que

$$n - K = [n(1-x)],$$

en efecto: como $\frac{K}{n} - x \geq 0 > \frac{K-1}{n} - x$, multiplicando por n y sumando nx en la doble desigualdad, se tiene $K \geq nx > K-1$. Luego, multiplicando por $(-)$ y sumando n, se obtiene $n - K \leq n(1-x) < n - (K-1)$. De aquí se sigue que $n - K = [n(1-x)]$. □

La distribución de $D_n = \max(D_n^+, D_n^-)$ es mucho mas complicado. Debido a que $D_n^+ = D_n^-$ en distribución (cuando F es continua), tenemos al menos una desigualdad, como sigue:

$$\begin{aligned}
\mathbb{P}(D_n > x) &= \mathbb{P}(\text{máx}\{D_n^+, D_n^-\} > x) \\
&\leq \mathbb{P}(D_n^+ > x) + \mathbb{P}(D_n^- > x) \\
&= \mathbb{P}(D_n^+ > x) + \mathbb{P}(D_n^+ > x) \\
&= 2\mathbb{P}(D_n^+ > x).
\end{aligned}$$

2.7. La cota D-K-W

En esta sección se analiza una cota muy especial de la desviación entre F_n y F. Esta cota fue originalmente propuesta por Dvoretzky-Kiefer-Wolfowitz(1956), ver [5]. Esta cota tiene algunas consecuencias interesantes.

2.7.1. Teorema. *Para todo $x \geq 0$ y $n \geq 1$, existe $c > 0$ tal que*

$$\mathbb{P}(n^{1/2} D_n^+ > x) \leq c \exp(-2x^2). \tag{2.13}$$

Demostración: En vista de (2.9), es suficiente estudiar el caso en que F sea continua. Este teorema se deduce del teorema 2.6.5, como sigue:

Si $x \geq \sqrt{n}$,

$$\mathbb{P}(n^{1/2} D_n^+ > x) \leq \mathbb{P}(n^{1/2} D_n^+ > \sqrt{n}) = \mathbb{P}(D_n^+ > 1) = 0.$$

Por lo que basta considerar $0 \leq x \leq \sqrt{n}$. Entonces,

$$\mathbb{P}(n^{1/2} D_n^+ > x) = 1 - \mathbb{P}(n^{1/2} D_n^+ \leq x) = 1 - \mathbb{P}\left(D_n^+ \leq \frac{x}{\sqrt{n}}\right)$$

$$
\begin{aligned}
&= \frac{x}{\sqrt{n}} \sum_{j=0}^{[n(1-\frac{x}{\sqrt{n}})]} \binom{n}{j} \left(1 - \frac{x}{\sqrt{n}} - \frac{j}{n}\right)^{n-j} \left(\frac{x}{\sqrt{n}} + \frac{j}{n}\right)^{j-1} \\
&= \left(1 - \frac{x}{\sqrt{n}}\right)^n + \frac{x}{\sqrt{n}} \sum_{j=0}^{[n(1-\frac{x}{\sqrt{n}})]} \binom{n}{j} \left(1 - \frac{x}{\sqrt{n}} - \frac{j}{n}\right)^{n-j} \left(\frac{x}{\sqrt{n}} + \frac{j}{n}\right)^{j-1} \\
&- \left(1 - \frac{x}{\sqrt{n}}\right)^n \\
&= \left(1 - \frac{x}{\sqrt{n}}\right)^n + \frac{x}{\sqrt{n}} \sum_{j=1}^{[n(1-\frac{x}{\sqrt{n}})]} \binom{n}{j} \left(1 - \frac{x}{\sqrt{n}} - \frac{j}{n}\right)^{n-j} \left(\frac{x}{\sqrt{n}} + \frac{j}{n}\right)^{j-1}.
\end{aligned}
$$

Sea $i = n - j$, entonces $j = n - i$. Lo anterior se escribe ahora

$$
\left(1 - \frac{x}{\sqrt{n}}\right)^n + \frac{x}{\sqrt{n}} \sum_{i=n-1}^{n-[n(1-\frac{x}{\sqrt{n}})]} \binom{n}{i} \left(1 - \frac{x}{\sqrt{n}} - \frac{n-i}{n}\right)^i \left(\frac{x}{\sqrt{n}} + \frac{n-i}{n}\right)^{n-i-1}.
$$

Observemos que

$$
\begin{aligned}
&\left(1 - \frac{x}{\sqrt{n}} - \frac{n-i}{n}\right)^i \left(\frac{x}{\sqrt{n}} + \frac{n-i}{n}\right)^{n-i-1} \\
&= \left(\frac{n - \sqrt{n}x - n + i}{n}\right)^i \left(\frac{\sqrt{n}x + n - i}{n}\right)^{n-i-1} \\
&= \frac{1}{n^{n-1}} (i - \sqrt{n}x)^i (\sqrt{n}x + n - i)^{n-i-1}.
\end{aligned}
$$

Por otro lado, observemos que

$$
\left[n(1 - \frac{x}{\sqrt{n}})\right] \leq n(1 - \frac{x}{\sqrt{n}}).
$$

Entonces,

$$
n - \left[n(1 - \frac{x}{\sqrt{n}})\right] \geq n - n(1 - \frac{x}{\sqrt{n}}) = \sqrt{n}x.
$$

Por lo tanto,

$$
n - \left[n(1 - \frac{x}{\sqrt{n}})\right] = [\sqrt{n}x] + 1.
$$

La expresión de la última igualdad se escribe ahora

$$\left(1-\frac{x}{\sqrt{n}}\right)^n+\sqrt{n}x\sum_{i=n-1}^{[n\sqrt{x}]+1}\binom{n}{i}(i-\sqrt{n}x)^i(\sqrt{n}x+n-i)^{n-i-1}n^{-n}.$$

Por lo tanto,

$$\mathbb{P}(n^{1/2}D_n^+>x)=(1-x/\sqrt{n})^n+x\sqrt{n}\sum_{j=[x\sqrt{n}]+1}^{n-1}Q_n(j,x) \qquad (2.14)$$

donde

$$Q_n(j,x)=\binom{n}{j}(j-x\sqrt{n})^j(n-j+x\sqrt{n})^{n-j-1}n^{-n}.$$

Definamos

$$f(x)=(1-x/\sqrt{n})^n\exp(2x^2).$$

Es claro que esta función alcanza su máximo en $x=0$. Además, $f(0)=1$, por lo tanto,

$$(1-x/\sqrt{n})^n\exp(2x^2)\leq 1.$$

Se sigue entonces que

$$(1-x/\sqrt{n})^n\leq\exp(-2x^2).$$

Ahora, sea $x\sqrt{n}<j<n$. Entonces,

$$\ln Q_n(j,x)=\ln\binom{n}{j}+j\ln(j-x\sqrt{n})+(n-j-1)\ln(n-j+x\sqrt{n})-n\ln n.$$

$$\begin{aligned}
\frac{d}{dx}\ln Q_n(j,x) &= \frac{-j\sqrt{n}}{j-x\sqrt{n}}+\frac{(n-j-1)\sqrt{n}}{n-j+x\sqrt{n}}\\
&= \frac{-j\sqrt{n}(n-j+x\sqrt{n})+(n-j+1)\sqrt{n}(j-x\sqrt{n})}{(j-x\sqrt{n})(n-j+x\sqrt{n})}\\
&= \frac{-n^2x-\sqrt{n}j+xn}{(j-x\sqrt{n})(n-j+x\sqrt{n})}\\
&= \frac{-n^2x}{(j-x\sqrt{n})(n-j+x\sqrt{n})}-\frac{\sqrt{n}}{n-j+x\sqrt{n}}\\
&< \frac{-n^2x}{(j-x\sqrt{n})(n-j+x\sqrt{n})}\\
&= \frac{-4x}{1-\frac{4}{n^2}(\frac{n}{2}-j+x\sqrt{n})^2}.
\end{aligned}$$

Ahora, como $x\sqrt{n} < j$ entonces $x\sqrt{n} - j < 0$. Sumando $\frac{n}{2}$ y elevando a la cuarta ambos lados de la desigualdad se tiene:

$$\left(\frac{n}{2} - j + x\sqrt{n}\right)^4 < \frac{n^4}{16}$$

o equivalentemente

$$0 < \frac{16}{n^4}\left(\frac{n}{2} - j + x\sqrt{n}\right)^4 < 1.$$

Así que

$$\frac{1}{1 - \frac{16}{n^4}\left(\frac{n}{2} - j + x\sqrt{n}\right)^4} > 1.$$

El denominador es una diferencia de cuadrados, por eso

$$\frac{1}{\left[1 - \frac{4}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2\right]\left[1 + \frac{4}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2\right]} > 1.$$

Multiplicando por $-4x$, $x > 0$, y por uno de los factores del denominador, se tiene:

$$\begin{aligned}\frac{-4x}{1 - \frac{4}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2} &< -4x\left[1 + \frac{4}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2\right]\\ &= -4x - \frac{16x}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2.\end{aligned}$$

Por lo tanto, si $x\sqrt{n} < j < n$,

$$\frac{d}{dx}\ln Q_n(j,x) < -4x - \frac{16x}{n^2}\left(\frac{n}{2} - j + x\sqrt{n}\right)^2, \quad x > 0.$$

Este último término tiene primitiva

$$h(x) = -2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}.$$

Por lo tanto,

$$\frac{d}{dx}\ln Q_n(j,x) < \frac{d}{dx}\left[-2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}\right].$$

Luego, integrando de 0 a x se tiene:

$$\ln Q_n(j,x) - \ln Q_n(j,0) \leq -2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}.$$

Así que

$$Q_n(j,x) \le Q_n(j,0)\exp\left[-2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}\right] \quad (2.15)$$

Ahora, integrando de 1 a x, i.e., $x \ge 1$,

$$\ln Q_n(j,x) - \ln Q_n(j,1) \le -2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n} + c_0,$$

donde

$$c_0 = 2 + \frac{8}{n^2}\left(\frac{n}{2} - j + \frac{2\sqrt{n}}{3}\right)^2 + \frac{4}{9n}.$$

Por lo tanto,

$$Q_n(j,x) \le c_1 Q_n(j,1)\exp\left[-2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}\right] \quad (2.16)$$

Para acotar $Q_n(j,0)$ y $Q_n(j,1)$, primero consideramos el caso $\left|j - \frac{n}{2}\right| \le \frac{n}{4}$. Recordando la fórmula de Stirling

$$k! \sim \left(\frac{k}{e}\right)^k \sqrt{2\pi k},$$

se tiene

$$\begin{aligned} Q_n(j,0) &= \frac{n!}{(n-j)!j!} j^j (n-j)^{n-j-1} n^{-n} \\ &\sim \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{n-j}{e}\right)^{n-j} \sqrt{2\pi(n-j)} \left(\frac{j}{e}\right)^j \sqrt{2\pi j}} j^j (n-j)^{n-j-1} n^{-n} \\ &= \frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-j)}\sqrt{2\pi j}} \frac{1}{(n-j)} = \frac{\sqrt{n}}{(n-j)^{3/2}\sqrt{2\pi}\sqrt{j}}. \end{aligned}$$

Como $\left|j - \frac{n}{2}\right| \le \frac{n}{4}$, se sigue que $\frac{n^{1/2}}{2} \le j^{1/2} \le \frac{(3n)^{1/2}}{2}$, y que $\frac{3n}{4} \ge n - j \ge \frac{n}{4}$. Entonces

$$\frac{\sqrt{n}}{(n-j)^{3/2}\sqrt{2\pi}\sqrt{j}} \le \frac{\sqrt{n}}{\left(\frac{n}{4}\right)^{3/2}\sqrt{2\pi}\frac{n^{1/2}}{2}} = \frac{2(4)^{3/2}}{\sqrt{2\pi}} n^{-3/2}.$$

Así, obtenemos

$$Q_n(j,0) \le c_2 n^{-3/2}, \quad \text{donde } c_2 = \frac{2(4)^{3/2}}{\sqrt{2\pi}}.$$

Ahora, denotemos con Σ' la suma sobre aquellas $j's$ en (2.14), que satisface $\left|j - \frac{n}{2}\right| \leq 4$, y con Σ'' el resto. Entonces, de (2.15) obtenemos

$$\begin{aligned}
\Sigma' Q_n(j,x) &\leq \Sigma' Q_n(j,0)\exp\left[-2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^2}{9n}\right] \\
&\leq c_2 n^{-3/2}\exp(-2x^2)\Sigma'\exp\left[-8x^2\left(\frac{1}{2} - \frac{j}{n} - \frac{2x}{3\sqrt{n}}\right)^2 - \frac{4x^2}{9n}\right] \\
&\leq c_2 n^{-3/2}\exp(-2x^2)\Sigma'\exp\left[-8x^2\left(\frac{1}{2} - \frac{j}{n} - \frac{2x}{3\sqrt{n}}\right)^2\right] \\
&\leq 2c_2 n^{-3/2}\exp(-2x^2)\sum_{j=0}^{\infty}\exp(-8x^2j^2/n^2) \\
&\leq 2c_2 n^{-1/2}\exp(-2x^2)\left[\frac{1}{n} + \int_0^{\infty}\exp(-8x^2t^2)dt\right] = *.
\end{aligned}$$

La última desigualdad, se debe a lo siguiente:

$$\frac{1}{n}\sum_{j=0}^{\infty}\exp(-8x^2j^2/n^2) \geq \int_0^{\infty}\exp(-8x^2t^2)dt$$

y

$$\frac{1}{n}\sum_{j=1}^{\infty}\exp(-8x^2j^2/n^2) \leq \int_0^{\infty}\exp(-8x^2t^2)dt.$$

Por otra parte,

$$\sum_{j=0}^{\infty}\exp(-8x^2j^2/n^2) = 1 + \sum_{j=1}^{\infty}\exp(-8x^2j^2/n^2).$$

Así,

$$\begin{aligned}
\frac{1}{n}\sum_{j=0}^{\infty}\exp(-8x^2j^2/n^2) &= \frac{1}{n} + \frac{1}{n}\sum_{j=1}^{\infty}\exp(-8x^2j^2/n^2) \\
&\leq \frac{1}{n} + \int_0^{\infty}\exp(-8x^2t^2)dt.
\end{aligned}$$

Ahora, observemos que

$$\int_0^{\infty}\exp(-8x^2t^2)dt = \frac{\sqrt{2\pi}}{4x}\int_0^{\infty}\frac{1}{\sqrt{2\pi}/4x}\exp\left(-\frac{t^2}{2/16x^2}\right)dt = \frac{\sqrt{2\pi}}{8x}.$$

Entonces,

$$\begin{aligned} * &= 2c_2 n^{-1/2}\exp(-2x^2)\left(\frac{1}{n}+\frac{\sqrt{2\pi}}{8x}\right) \\ &= 2c_2 n^{-3/2}\exp(-2x^2) + c_2 n^{-1/2}x^{-1}\frac{\sqrt{2\pi}}{4}\exp(-2x^2) \\ &\leq c_3 x^{-1} n^{-1/2}\exp(-2x^2). \end{aligned}$$

La última desigualdad es porque $x \leq \sqrt{n}$. Se sigue entonces que $nx^{-1} \geq \sqrt{n} \geq 1$. Por lo tanto,

$$\begin{aligned} 2c_2 n^{-3/2}\exp(-2x^2) &\leq 2c_2 n^{-3/2}\exp(-2x^2)nx^{-1} \\ &= 2c_2 n^{-1/2}x^{-1}\exp(-2x^2). \end{aligned}$$

Concluímos que

$$\Sigma' Q_n(j,x) \leq c_3 x^{-1} n^{-1/2}\exp(-2x^2),$$

o equivalentemente

$$\Sigma' x n^{1/2} Q_n(j,x) \leq c_3 \exp(-2x^2).$$

Ahora analizamos la segunda suma. Supongamos s.p.g. que $x > 1$. Si $2x\sqrt{n}/3 \leq n/8$ entonces el segundo término en el exponente de (2.16) no excede $-x^2/8$, en efecto:
Como estamos fuera de $|j - \frac{n}{2}| \leq \frac{n}{4}$ entonces $j < \frac{n}{4}$ o $\frac{3n}{4} < j$. Entonces,

$$\frac{n}{2} - j + \frac{2x\sqrt{n}}{3} \leq \frac{n}{2} - j + \frac{n}{8} = \frac{5n}{8} - j < \frac{5n}{8} - \frac{3n}{4} = -\frac{n}{8}.$$

Elevando al cuadrado y multiplicando por $\frac{-8x^2}{n^2}$, se tiene lo deseado.

Ahora, para $2x\sqrt{n}/3 > n/8$ o lo que es lo mismo $x > 3\sqrt{n}/16$, el último término es menor o igual que $-(4/9)(3/16)^2x^2$. En efecto:

$$\frac{4x^4}{9n} = \frac{4x^2x^2}{9n} > \frac{4}{9n}\left(\frac{3}{16}\right)^2 nx^2.$$

Así que

$$\left[-2x^2 - \frac{8x^2}{n^2}\left(\frac{n}{2} - j + \frac{2x\sqrt{n}}{3}\right)^2 - \frac{4x^4}{9n}\right] \leq -2x^2 - c_4 x^2.$$

Luego,

$$\begin{aligned} Q_n(j,x) &\leq c_1 Q_n(j,1)\exp(-2x^2 - c_4 x^2) \\ &= c_1 Q_n(j,1)\exp(-2x^2)\exp(-c_4 x^2) \\ &\leq c_5 x^{-1} Q_n(j,1)\exp(-2x^2). \end{aligned}$$

Por lo tanto,

$$Q_n(j,x) \leq c_5 x^{-1} Q_n(j,1)\exp(-2x^2).$$

Ahora, puesto que $\sqrt{n}\Sigma'' Q_n(j,1) \leq 1$, se sigue que

$$x\sqrt{n}\Sigma'' Q_n(j,x) \leq c_5 \exp(-2x^2)\sqrt{n}\Sigma'' Q_n(j,1) \leq c_5 \exp(-2x^2).$$

Por lo tanto,

$$x\sqrt{n}\Sigma' Q_n(j,x) \leq c_3 \exp(-2x^2)$$

y

$$x\sqrt{n}\Sigma'' Q_n(j,x) \leq c_5 \exp(-2x^2).$$

Finalmente

$$x\sqrt{n} \sum_{j=[x\sqrt{n}]+1}^{n-1} Q_n(j,x) \leq c_6 \exp(-2x^2).$$

Además,

$$(1 - x/\sqrt{n})^n \leq \exp(-2x^2).$$

Por lo tanto,

$$(1 - x/\sqrt{n})^n + x\sqrt{n} \sum_{j=[x\sqrt{n}]+1}^{n-1} Q_n(j,x) \leq c\exp(-2x^2)$$

y

$$\mathbb{P}(n^{1/2}D_n^+ > x) \leq c\exp(-2x^2).$$

□

Veamos que consecuencias tiene esta cota. Dada $\varepsilon > 0$,

$$\mathbb{P}(D_n > \varepsilon) \leq 2\mathbb{P}(D_n^+ > \varepsilon) = 2\mathbb{P}(n^{1/2}D_n^+ > n^{1/2}\varepsilon) \leq c\exp(-2n\varepsilon^2).$$

Luego.

$$\sum_{n=0}^{\infty} \exp(-2n\varepsilon^2)$$

es una serie geométrica con $x = \exp(-2\varepsilon^2) < 1$, el cuál converge a $\frac{1}{1-x}$. Así, aplicando el lema de Borel Cantelli, se tiene:

$$\mathbb{P}(\limsup_{n\to\infty} D_n > \varepsilon) = 0 \quad \text{para toda } \varepsilon > 0$$

lo cual implica,

$$D_n \to 0 \quad \text{con probabilidad uno.}$$

es decir, el Teorema de Glivenko-Cantelli. Con una modificación a este argumento, se obtiene el siguiente:

2.7.2. Corolario. *Con probabilidad uno,*

$$\limsup_{n\to\infty} \left[\frac{n}{\ln n}\right]^{1/2} D_n < \infty.$$

Demostración: Sea $x = (c \ln n)^{1/2}$ con $c > 1/2$. Entonces.

$$\begin{aligned} \mathbb{P}(D_n > x) &\le 2\mathbb{P}(D_n^+ < x) \le 2c_0 \exp(-2x^2) \\ &= 2c_0 \exp(-2c \ln n). \end{aligned}$$

Luego,

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left(D_n > (c\ln n)^{1/2}\right) &\le 2\sum_{n=1}^{\infty} c_0 \exp(-2c\ln n) \\ &= 2c_0 \sum_{n=1}^{\infty} \exp\left[\ln(n^{-2c})\right] \\ &= 2c_0 \sum_{n=1}^{\infty} n^{-2c} < \infty \quad \text{si } c > 1/2. \end{aligned}$$

Entonces,

$$\mathbb{P}(\limsup_{n\to\infty} D_n > (c\ln n)^{1/2}) = 0.$$

Se sigue que

$$\limsup_{n\to\infty} \left[\frac{n}{\ln n}\right]^{1/2} D_n < \infty.$$

□

2.8. Cotas Binomiales

Como antes, sean $X_1, \ldots, X_n$ v.a.i.i.d. con función de distribución F. En la sección 2.1 se probó que $nF_n(t) \sim Bin(n, F(t))$. Denotemos

$$\eta := nF_n(t).$$

Así que $\eta \sim Bin(n,p)$, con $p = F(t)$. Si $0 < p < 1$, se sigue del TLC:

$$\mathbb{P}\left(n^{1/2}\frac{F_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} \geq x\right) \to \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-u^2/2}du = \mathbb{P}(\xi \geq x), \quad (2.17)$$

donde $\xi \sim \mathcal{N}(0,1)$.

2.8.1. Lema. ***(Cociente de Mill).*** *Para toda $x > 0$,*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}}\exp(-x^2/2) \leq \mathbb{P}(\xi \geq x) \leq \frac{1}{x}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2). \quad (2.18)$$

Demostración: Sea $u = x + z$, se sigue que $du = dz$ y $u^2 = x^2 + 2xz + z^2$. Además, si $x \leq u < \infty$ entonces $0 \leq z < \infty$. Así,

$$\begin{aligned}
\mathbb{P}(\xi \geq x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-u^2/2)du \\
&= \frac{1}{\sqrt{2\pi}}\int_0^\infty \exp((-x^2 - 2xz - z^2)/2)dz \\
&= \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\int_0^\infty \exp(-xz)\exp(-z^2/2)dz \\
&\leq \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\int_0^\infty \exp(-xz)dz \\
&= \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\left[-\frac{1}{x}\exp(-xz)\right]_0^\infty = \frac{1}{x}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2).
\end{aligned}$$

Por lo tanto,

$$\mathbb{P}(\xi \geq x) \leq \frac{1}{x}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2).$$

Por otra parte, integrando y evaluando se tiene

$$(x^{-1} - x^{-3})\exp(-x^2/2) = \int_x^\infty (1 - 3y^{-4})\exp(-y^2/2)dy.$$

Luego, debido a que $1 - 3y^{-4} \leq 1$,

$$\int_x^\infty (1 - 3y^{-4}) \exp(-y^2/2) dy \leq \int_x^\infty \exp(-y^2/2) dy.$$

Entonces,

$$(x^{-1} - x^{-3}) \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \leq \mathbb{P}(\xi \geq x).$$

Juntando las dos desigualdades se obtiene lo deseado. □

Vamos a probar que se cumple una cota similar para una n finita, de una Binomial estandarizada.

Recordemos que $nF_n(t) \sim Bin(n, F(t))$. Pongamos

$$X = \frac{nF_n(t)}{\sqrt{nF(t)(1 - F(t))}}.$$

Observemos que

$$\mu := \mathbb{E}(X) = \frac{nF(t)}{\sqrt{nF(t)(1 - F(t))}} \text{ y } Var(X) = 1.$$

Aplicando la desigualdad de Chevichev, ver apéndice B.2, obtenemos:

$$\mathbb{P}(\mid X - \mu \mid \geq x) = \mathbb{P}\left(n^{1/2} \frac{F_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \geq x\right) \leq x^{-2},$$

lo cual es una peor cota de lo que se podía esperar de (2.17) y (2.18). Sin embargo, podemos obtener una mejor cota si incorporamos la **función generadora** de momentos de una variable aleatoria Binomial.

2.8.2. Lema. *Sea* $\eta \sim Bin(n, p)$, $0 < p < 1$. *Entonces, la función generadora de momentos de* η *es*

$$M(z) = \mathbb{E}[\exp(z\eta)] = [(1 - p) + p \exp z]^n.$$

Demostración: Puesto que η es igual en distribución a la suma de n v.a.i.'s Bernoullis con parámetro p, cuya función generadora de momentos es $1 - p + p \exp z$, la función generadora de momentos de η es:

$$M(z) = \mathbb{E}[\exp(z\eta)] = \prod_{i=1}^{n} [(1 - p) + p \exp z] = [(1 - p) + p \exp z]^n.$$

□

Ahora daremos una cota para $\mathbb{P}(\eta - np \geq \varepsilon)$ con $\varepsilon > 0$. Primero observemos que

$$\begin{aligned}
\mathbb{P}(\eta - np \geq \varepsilon) &= \mathbb{P}[z(\eta - np) \geq z\varepsilon] = \mathbb{P}(z\eta \geq z\varepsilon + znp) \\
&= \mathbb{P}[z\eta \geq z(\varepsilon + np)] = \mathbb{P}[\exp(z\eta) \geq \exp[z(\varepsilon + np)]] \\
&\leq \frac{1}{\exp[z(\varepsilon + np)]}\mathbb{E}[\exp(z\eta)] = M(z)\exp[-z(\varepsilon + np)].
\end{aligned}$$

La desigualdad anterior se sigue debido a la desigualdad de Markov, ver apéndice B.2. Por lo tanto,

$$\mathbb{P}(\eta - np \geq \varepsilon) \leq M(z)\exp[-z(np + \varepsilon)].$$

Así, obtenemos la siguiente cota:

2.8.3. Lema. *Para cada* $\varepsilon > 0$,

$$\mathbb{P}(\eta - np \geq \varepsilon) \leq \inf_{z \geq 0} M(z)\exp[-z(np + \varepsilon)] \equiv \rho.$$

Demostración: Se sigue del hecho de que

$$\mathbb{P}(\eta - np \geq \varepsilon) \leq M(z)\exp[-z(np + \varepsilon)].$$

□

Para determinar ρ, definamos f como sigue:

$$\exp[-f(u)] = \inf_{z \geq 0} M(z)\exp(-zu).$$

Entonces,

$$\exp[-f(np + \varepsilon)] = \inf_{z \geq 0} M(z)\exp[-z(np + \varepsilon)],$$

y

$$\rho = \exp[-f(np + \varepsilon)].$$

La función f se le conoce como la **función de Chernoff** de M. Es claro que f es no negativa y no decreciente. Así, enunciamos el siguiente lema.

2.8.4. Lema.

$$f(u) \equiv f(u, n, p) = \begin{cases} 0, & si \quad u \leq 0 \\ u \ln \frac{u}{np} + (n - u)\ln \frac{n-u}{n(1-p)}, & si \quad 0 < u < n \\ -n \ln p, & si \quad u = n \\ +\infty, & si \quad n < u. \end{cases}$$

Demostración: Sea $u \leq 0$. Entonces,

$$\inf_{z \geq 0}[(1-p) + p \exp z]^n \exp(-zu) = 1.$$

Se sigue que

$$\exp[-f(u)] = 1,$$

y por tanto,

$$f(u) = 0.$$

Sea $0 < u < n$ y $g(z) := [(1-p) + p\exp z]^n \exp(-zu)$. Entonces

$$\begin{aligned}
\frac{dg(z)}{dz} &= n\left[(1-p) + pe^z\right]^{n-1} pe^z e^{-zu} - \left[(1-p) + pe^z\right]^n e^{-zu} u \\
&= \left[(1-p) + pe^z\right]^{n-1} e^{-zu} \left(npe^z - \left[(1-p) + pe^z\right] u\right).
\end{aligned}$$

Igualando a cero esta derivada, tenemos

$$\left[(1-p) + pe^z\right]^{n-1} e^{-zu} = 0 \quad \text{o} \quad npe^z - \left[(1-p) + pe^z\right] u = 0.$$

Es claro que

$$\left[(1-p) + pe^z\right]^{n-1} e^{-zu} \neq 0,$$

y

$$npe^z - \left[(1-p) + pe^z\right] u = 0.$$

Despejando a z en esta igualdad se obtiene

$$z = \ln \frac{(1-p)u}{(n-u)p}.$$

Este valor está bien definido debido a la hipótesis $0 < u < n$. Entonces, sustituyendo z en f, se obtiene

$$\exp[-f(u)] = \left[(1-p) + p\frac{(1-p)u}{(n-u)p}\right]^n \exp\left[-u \ln \frac{(1-p)u}{(n-u)p}\right].$$

Despejando $f(u)$ en esta igualdad, se obtiene

$$\begin{aligned}
f(u) &= u \ln \frac{(1-p)u}{(n-u)p} - n \ln \frac{(1-p)n}{n-u} \\
&= u \ln \frac{u}{np} + (n-u) \ln \frac{n-u}{n(1-p)}.
\end{aligned}$$

Por lo tanto,

$$f(u) = u \ln \frac{u}{np} + (n-u) \ln \frac{n-u}{n(1-p)}, \quad \text{si } 0 < u < n.$$

Si $u = n$,

$$f(u) = u \ln \frac{1}{p} = -u \ln p = -n \ln p.$$

Por último, si $u > n$ entonces $u = n + \varepsilon$ con $\varepsilon > 0$. Entonces,

$$e^{-uz} = e^{-(n+\varepsilon)z},$$

y

$$\begin{aligned} g(z) &= [(1-p) + pe^z]^n \, e^{-zu} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^z)^k \, (1-p)^{n-k} e^{-zu} \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{z(k-(n+\varepsilon))}. \end{aligned}$$

Esta última suma converge a cero cuando $z \to \infty$, ya que $k - (n+\varepsilon) < 0$. Por lo tanto,

$$\inf_{z \geq 0} g(z) = 0,$$

y

$$f(u) = +\infty.$$

□

Antes de analizar más a fondo la función de Chernoff, notemos que

$$\mathbb{P}(\eta - np \leq -\varepsilon) = \mathbb{P}(\tilde{\eta} \geq \varepsilon + n(1-p)), \quad \text{donde } \tilde{\eta} \sim Bin(n, 1-p),$$

en efecto,

$$\begin{aligned} \mathbb{P}(\eta - np \leq -\varepsilon) &= \mathbb{P}(-\eta \geq \varepsilon - np) = \mathbb{P}(n - \eta \geq n + \varepsilon - np) \\ &= \mathbb{P}(n - \eta \geq \varepsilon + n(1-p)). \end{aligned}$$

Sea $\tilde{\eta} = n - \eta$. Entonces,

$$\begin{aligned} \mathbb{P}(\tilde{\eta} = x) &= \mathbb{P}(n - \eta = x) = \mathbb{P}(\eta = n - x) \\ &= \binom{n}{n-x} p^{n-x} (1-p)^x = \binom{n}{x} (1-p)^x p^{n-x}. \end{aligned}$$

Por lo tanto,

$$\tilde{\eta} \sim Bin(n, 1-p).$$

Esto, junto con el lema 2.8.3, se sigue que

$$\begin{aligned}\mathbb{P}(\eta - np \leq -\varepsilon) &= \mathbb{P}(\tilde{\eta} \geq \varepsilon + n(1-p)) = \mathbb{P}(\tilde{\eta} - n(1-p) \geq \varepsilon) \\ &\leq \rho = \exp[-f(n(1-p)+\varepsilon, n, 1-p)].\end{aligned}$$

Retomando nuevamente a la función de Chernoff, sea $u = np + \varepsilon$, sobre $0 < u < n$, i.e., $\varepsilon < n(1-p)$. Entonces,

$$\begin{aligned}f(n,u,p) &= (np+\varepsilon)\ln\frac{np+\varepsilon}{np} + (n-np-\varepsilon)\ln\frac{n-np-\varepsilon}{n(1-p)} \\ &= np\left[1+\frac{\varepsilon}{np}\right]\ln\left[1+\frac{\varepsilon}{np}\right] \\ &+ n(1-p)\left[1-\frac{\varepsilon}{n(1-p)}\right]\ln\left[1-\frac{\varepsilon}{n(1-p)}\right].\end{aligned}$$

Luego, expandiendo $f(x) := \ln(1+x)$, con $|x| < 1$ alrededor de $x = 0$, se tiene

$$f^k(x) = \frac{(-1)^{k-1}(k-1)!}{(1+x)^k} \text{ y } f^k(0) = (-1)^{k-1}(k-1)! \text{ para todo } k = 1, 2, \ldots$$

Usando la fórmula de Taylor, podemos escribir $\ln(1+x)$ como

$$\begin{aligned}\ln(1+x) &= \sum_{k=0}^{\infty}\frac{(-1)^{k-1}(k-1)!x^k}{k!} \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots\end{aligned}$$

Multiplicando esta suma por $1+x$, obtenemos

$$(1+x)\ln(1+x) = x + \frac{x^2}{1\cdot 2} - \frac{x^3}{2\cdot 3} + \frac{x^4}{3\cdot 4} - \ldots$$

Observemos que

$$f(-x) = \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \ldots$$

Multiplicando por $1-x$, obtenemos

$$(1-x)\ln(1-x) = -x + \frac{x^2}{1\cdot 2} + \frac{x^3}{2\cdot 3} + \frac{x^4}{3\cdot 4} + \ldots$$

Entonces, con $x_1 = \varepsilon/np$ y $x_2 = \varepsilon/n(1-p)$ obtenemos, $f(n,u,p)$

$$
\begin{aligned}
&= np\left[\frac{\varepsilon}{np} + \frac{\varepsilon^2}{2n^2p^2}\right] + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] \\
&+ n(1-p)\left[-\frac{\varepsilon}{n(1-p)} + \frac{\varepsilon^2}{2n^2(1-p)^2}\right] + n(1-p)\left[\frac{x_2^3}{2\cdot 3} + \frac{x_2^4}{3\cdot 4} + \dots\right]
\end{aligned}
$$

La expresión del lado derecho se puede simplificar a

$$
\frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] + \left[\frac{x_2^3}{2\cdot 3} + \frac{x_2^4}{3\cdot 4} + \dots\right]
$$

Por lo tanto,

$$
f(u,n,p) = \frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] + \left[\frac{x_2^3}{2\cdot 3} + \frac{x_2^4}{3\cdot 4} + \dots\right]
$$

Cuando $0 < x_1, x_2 < 1$, se sigue que

$$
\begin{aligned}
f(n,u,p) &\geq \frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] \\
&\geq \frac{\varepsilon^2}{2np(1-p)} + \frac{np}{1-p}\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] \\
&= \frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right),
\end{aligned}
$$

donde ψ está definida para $|x| < 1$, como

$$
\psi(x) = 1 - \frac{x}{3} + \frac{x^2}{6} - \dots + \frac{(-1)^k 2x^k}{(k+1)(k+2)} + \dots
$$

en efecto, $\frac{np}{1-p}[\dots]$

$$
\begin{aligned}
&= \frac{np}{1-p}\left[-\frac{\varepsilon^3}{2\cdot 3(np)^3} + \frac{\varepsilon^4}{3\cdot 4(np)^4} - \dots\right] \\
&= \frac{1}{1-p}\left[-\frac{\varepsilon^3}{2\cdot 3(np)^2} + \frac{\varepsilon^4}{3\cdot 4(np)^3} - \dots + \frac{(-1)^k\varepsilon^{(k+2)}}{(k+1)(k+2)(np)^{k+1}} + \dots\right]
\end{aligned}
$$

Así,

$$
\frac{\varepsilon^2}{2np(1-p)} + \frac{np}{1-p}[\dots]
$$

$$
\begin{aligned}
&= \frac{\varepsilon^2}{2np(1-p)}\left[1 - \frac{\varepsilon}{3(np)} + \frac{\varepsilon^2}{6(np)^2} - \ldots + \frac{(-1)^k 2\varepsilon^k}{(k+1)(k+2)(np)^k} + \ldots\right] \\
&= \frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right), \quad \text{con } x = \frac{\varepsilon}{np}.
\end{aligned}
$$

Notemos que

$$\psi(x) = 2h(1+x)/x^2 \tag{2.19}$$

con

$$h(x) = x(\ln x - 1) + 1, \tag{2.20}$$

en efecto,

$$
\begin{aligned}
h(1+x) &= (1+x)[\ln(1+x) - 1] + 1 = (1+x)\ln(1+x) - x \\
&= x + \frac{x^2}{1\cdot 2} - \frac{x^3}{2\cdot 3} + \frac{x^4}{3\cdot 4} - \ldots - x \\
&= \frac{x^2}{1\cdot 2} - \frac{x^3}{2\cdot 3} + \frac{x^4}{3\cdot 4} - \ldots
\end{aligned}
$$

Por lo tanto,

$$2h(1+x)/x^2 = 1 - \frac{x}{3} + \frac{x^2}{6} - \ldots$$

y

$$\psi(x) = 2h(1+x)/x^2, \quad \text{con } h(x) = x(\ln x - 1) + 1.$$

En vista de (2.19) y (2.20), ψ está definida para todo x positivo. Definamos

$$g_1(\varepsilon) = f(np+\varepsilon, n, p) \text{ y } g_2(\varepsilon) = \varepsilon^2\psi\left(\frac{\varepsilon}{np}\right)/2np(1-p).$$

Es claro que

$$g_1(0) = 0 = g_2(0).$$

Veamos también que

$$\partial g_1/\partial\varepsilon \geq \partial g_2/\partial\varepsilon.$$

$$\partial g_1/\partial\varepsilon = \ln\left(1 + \frac{\varepsilon}{np}\right) - \ln\left(1 - \frac{\varepsilon}{n(1-p)}\right)$$

y,

$$\partial g_2/\partial\varepsilon = \frac{\ln\left(1+\frac{\varepsilon}{np}\right)}{1-p}.$$

Luego,

$$\partial g_1/\partial\varepsilon \geq \partial g_2/\partial\varepsilon \text{ si, y sólo si } (1-p)\ln\left[1 - \frac{\varepsilon}{n(1-p)}\right] + p\ln\left[1 + \frac{\varepsilon}{np}\right] \leq 0.$$

Luego, esto último es cierto ya que si definimos

$$f(x) = \ln(1+x),$$

entonces, f es concava, y para todo $-1 < x < y$, $f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y)$. Entonces, si

$$x := -\frac{\varepsilon}{n(1-p)} \text{ y } y := \frac{\varepsilon}{np}, \text{ con } \alpha = 1-p \text{ y } 1-\alpha = p,$$

entonces

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= f\left(-\frac{\varepsilon}{n} + \frac{\varepsilon}{n}\right) = f(0) = \ln(1) = 0 \\ &\geq (1-p)\ln\left(1 - \frac{\varepsilon}{n(1-p)}\right) + p\ln\left(1 + \frac{\varepsilon}{np}\right). \end{aligned}$$

Por lo tanto,

$$(1-p)\ln\left(1 - \frac{\varepsilon}{n(1-p)}\right) + p\ln\left(1 + \frac{\varepsilon}{np}\right) \leq 0.$$

Se sigue entonces que

$$\partial g_1/\partial\varepsilon \geq \partial g_2/\partial\varepsilon.$$

Finalmente, integrando esto último obtenemos

$$g_1(\varepsilon) \geq g_2(\varepsilon).$$

Enunciamos el siguiente lema.

2.8.5. Lema. *Sea $\eta \sim Bin(n,p)$. Entonces para todo $\varepsilon > 0$,*

(i)

$$\begin{aligned} \mathbb{P}(\eta - np \geq \varepsilon) &\leq \exp\left[-\frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right)\right] \\ &= \exp\left[-\frac{np}{1-p}h\left(1 + \frac{\varepsilon}{n(1-p)}\right)\right]. \end{aligned}$$

(ii)

$$\mathbb{P}(\eta - np \leq -\varepsilon) \leq \exp\left[-\frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{n(1-p)}\right)\right].$$

Demostración: (i) Ya se demostró que

$$f(np+\varepsilon) \geq \frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right).$$

Entonces,

$$-f(np+\varepsilon) \leq \frac{-\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right).$$

Luego, aplicando el lema 2.8.3, se tiene

$$\begin{aligned}
\mathbb{P}(\eta - np \geq \varepsilon) &\leq \exp[-f(np+\varepsilon)] \\
&\leq \exp\left[-\frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right)\right] \\
&= \exp\left[\frac{-np}{1-p}h\left(1+\frac{\varepsilon}{n(1-p)}\right)\right].
\end{aligned}$$

(ii) Aplicando directamente el inciso anterior se tiene

$$\begin{aligned}
\mathbb{P}(\eta - np \leq -\varepsilon) &= \mathbb{P}(\tilde{\eta} \geq \varepsilon + n(1-p)) \quad \text{donde } \tilde{\eta} \sim Bin(n, 1-p) \\
&= \mathbb{P}(\tilde{\eta} - n(1-p)) \geq \varepsilon) \\
&\leq \exp\left[-\frac{\varepsilon^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{n(1-p)}\right)\right].
\end{aligned}$$

□

Se pueden obtener mejores cotas para $f(u,n,p)$. Por ejemplo, si $1-p \leq p$, i.e., $p \geq 1/2$, se obtiene para $u = np+\varepsilon$,

$$f(u,n,p) \geq \varepsilon^2/2np(1-p) \tag{2.21}$$

en efecto,

$$f(u,n,p) = \frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \ldots\right] + n(1-p)\left[\frac{x_2^3}{2\cdot 3} + \frac{x_2^4}{3\cdot 4} + \ldots\right].$$

Si $p = \frac{1}{2}$ entonces $1-p = p$ y $x_1 = x_2$. Luego,

$$np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \ldots\right] + n(1-p)\left[\frac{x_2^3}{2\cdot 3} + \frac{x_2^4}{3\cdot 4} + \ldots\right] \geq 0,$$

puesto que se cancelan todos los terminos con exponente impar. De forma análoga se prueba si $p > \frac{1}{2}$.

para una $0 < p < 1$ arbitraria, se tiene

$$f(u,n,p) \geq \frac{\varepsilon^2}{2np(1-p)}\left[1 - \frac{\varepsilon(1-p)}{3np}\right] \tag{2.22}$$

en efecto, $f(u,n,p)$

$$\begin{aligned}
&\geq \frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{x_1^3}{2\cdot 3} + \frac{x_1^4}{3\cdot 4} - \dots\right] \\
&= \frac{\varepsilon^2}{2np(1-p)} + np\left[-\frac{1}{6}\left(\frac{\varepsilon}{np}\right)^3 + \frac{1}{3\cdot 4}\left(\frac{\varepsilon}{np}\right)^4 - \dots\right] \\
&= \frac{\varepsilon^2}{2np(1-p)} - \frac{np}{6}\left(\frac{\varepsilon}{np}\right)^3 + np\left[\frac{1}{3\cdot 4}\left(\frac{\varepsilon}{np}\right)^4 - \frac{1}{4\cdot 5}\left(\frac{\varepsilon}{np}\right)^5 + \dots\right]
\end{aligned}$$

Luego, el último sumando de la última igualdad es positivo. Por lo tanto,

$$f(u,n,p) \geq \frac{\varepsilon^2}{2np(1-p)} - \frac{np}{6}\left(\frac{\varepsilon}{np}\right)^3 = \frac{\varepsilon^2}{2np(1-p)}\left[1 - \frac{\varepsilon(1-p)}{3np}\right].$$

Para una $0 < \delta < 1$, si

$$\frac{\varepsilon(1-p)}{3np} \leq \delta,$$

entonces

$$f(u,n,p) \geq \frac{\varepsilon^2}{2np(1-p)}\left[1 - \frac{\varepsilon(1-p)}{3np}\right] \geq \frac{\varepsilon^2(1-\delta)}{2np(1-p)}.$$

Por lo tanto,

$$f(u,n,p) \geq \frac{\varepsilon^2(1-\delta)}{2np(1-p)} \tag{2.23}$$

Finalmente, para la estandarización de $F_n(t)$, se obtiene para $x \geq 0$ y $0 < \delta < 1$,

$$\mathbb{P}\left(\frac{n^{1/2}[F_n(t) - F(t)]}{\sqrt{F(t)(1-F(t))}} \geq x\right) \leq \begin{cases} \exp[-x^2/2], & \text{si} \quad p = F(t) \geq 1/2 \\ \exp[-x^2(1-\delta)/2]. & \text{si} \quad 0 < p < 1, \end{cases}$$

Siempre y cuando $x \leq \frac{3\delta\sqrt{np}}{(1-p)^{3/2}}$, en efecto,

$$nF_n(t) \sim Bin(n,p), \quad \text{donde } p = F(t).$$

Luego, si $p \geq 1/2$ aplicamos el lema 2.8.3, usando (2.21) se tiene

$$\mathbb{P}(nF_n(t) - nF(t) \geq \varepsilon) \leq \exp\left[-\frac{\varepsilon^2}{2np(1-p)}\right].$$

Por otro lado,

$$\begin{aligned}
\mathbb{P}(nF_n(t) - nF(t) \geq \varepsilon) &= \mathbb{P}\left(\frac{nF_n(t) - nF(t)}{\sqrt{nF(t)(1-F(t))}} \geq \frac{\varepsilon}{\sqrt{nF(t)(1-F(t))}}\right) \\
&= \mathbb{P}\left(n^{1/2}\frac{F_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} \geq \frac{\varepsilon}{\sqrt{np(1-p)}}\right).
\end{aligned}$$

Por lo tanto, si $x = \frac{\varepsilon}{\sqrt{np(1-p)}}$, se tiene

$$\mathbb{P}\left(n^{1/2}\frac{F_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} \geq x\right) \leq \exp[-x^2/2], \quad \text{si } p = F(t) \geq 1/2.$$

Analogamente, para $p = F(t)$ en general, aplicamos lema 2.8.3, usando (2.22), para obtener

$$\mathbb{P}\left(n^{1/2}\frac{F_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} \geq x\right) \leq \exp[-x^2(1-\delta)/2].$$

cuidando que $\frac{\varepsilon(1-p)}{3np} \leq \delta$, i.e., $x \leq \frac{3\delta\sqrt{np}}{(1-p)^{3/2}}$.

2.9. Oscilación de los procesos empíricos

Consideremos el proceso empírico uniforme

$$\bar{\alpha}_n(t) = n^{1/2}[\bar{F}_n(t) - t], \quad 0 \leq t \leq 1.$$

sobre la muestra uniforme $U_1, \ldots, U_n$. Las discontinuidades de $\bar{\alpha}_n$, se dan en las $U_i's$. Así que el número de saltos en la gráfica de $\bar{\alpha}_n$, se incrementan conforme la muestra crece. Por otro lado, $n^{-1/2}$ converge a cero cuando n crece. De esta manera, surge un poco de complicación en el comportamiento de $\bar{\alpha}_n$. Una mejor medida para este caso es

$$\bar{\omega}_n(a) = \sup_{|t-s|\leq a} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)|.$$

llamada el **módulo de oscilación**. $\bar{w}_n$ mide las **oscilaciones locales** de $\bar{\alpha}_n$, por lo que a es pequeño, con $0 < a < 1$. Para tener un análisis completo de $\bar{w}_n$, primero analizaremos el caso en que $s = 0$ y $\sup_{0 \le t \le a} \bar{\alpha}_n(t)$.
Sea $0 = t_0 < t_1 < \ldots < t_m = a$ una partición de $[0, a]$, obtendremos una cota para $\bar{\alpha}_n$ sobre esta partición, posteriormente haciendo que la norma de la partición tienda a cero, obtendremos una cota de $\bar{\alpha}_n$ sobre $0 \le t \le a$.

Sea $x > 0$. Definamos

$$A_i := \{\bar{\alpha}_n(t_i) > x\}, \quad \text{para todo } i = 0, 1, \ldots, m.$$

Observemos que

$$\mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) = \mathbb{P}(\sup_{0 \le i \le m} \bar{\alpha}_n(t_i) > x).$$

en efecto,

$$\omega \in \bigcup_{i=0}^{m} A_i \quad \text{si, y sólo si existe } i \in \{0, 1, \ldots, m\} \text{ tal que } \omega \in A_i,$$

si, y sólo si existe $i \in \{0, 1, \ldots, m\}$ tal que $\bar{\alpha}_n(t_i)(\omega) > x$. Esto último es equivalente a decir que $\max_{0 \le i \le m} \bar{\alpha}_n(t_i)(\omega) > x$. Luego, $\max_{0 \le i \le m} \bar{\alpha}_n(t_i) = \sup_{0 \le i \le m} \bar{\alpha}_n(t_i)$. Por lo tanto,

$$\bigcup_{i=0}^{m} A_i = \sup_{0 \le i \le m} \bar{\alpha}_n(t_i)$$

y,

$$\mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) = \mathbb{P}(\sup_{0 \le i \le m} \bar{\alpha}_n(t_i) > x).$$

También definamos el evento

$$A_m^* = \{\bar{\alpha}_n(t_m) > x^*\} \quad \text{con } x^* \le x.$$

Se sigue entonces que $A_m \subset A_m^*$. Luego,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) &= \mathbb{P}\left(\bigcup_{i=0}^{m} A_i \cap (A_m^* \cup A_m^{*c})\right) \\ &= \mathbb{P}\left(\bigcup_{i=0}^{m} (A_i \cap A_m^*) \cup (A_i \cap A_m^{*c})\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left(\bigcup_{i=0}^{m}(A_i \cap A_m^*) \cup \bigcup_{i=0}^{m}(A_i \cap A_m^{*c})\right) \\
&= \mathbb{P}\left(\bigcup_{i=0}^{m}(A_i \cap A_m^*)\right) + \mathbb{P}\left(\bigcup_{i=0}^{m}(A_i \cap A_m^{*c})\right) = *
\end{aligned}
$$

Por otro lado,

$$\bigcup_{i=0}^{m} A_i \cap A_m^* \subset A_m^*,$$

y

$$
\begin{aligned}
\bigcup_{i=0}^{m} A_i \cap A_m^{*c} &= A_m^{*c} \cap (A_0 \cup A_1 \cup \ldots \cup A_m) \\
&= A_m^{*c} \cap (A_0 \cup (A_1 \cap A_0^c) \cup \ldots \cup (A_m \cap (A_0^c \cap A_1^c \cap \ldots \cap A_{m-1}^c))) \\
&= (A_m^{*c} \cap A_0) \cup \ldots \cup (A_m^{*c} \cap A_m \cap A_0^c \cap A_1^c \cap \ldots \cap A_{m-1}^c) \\
&= (A_m^{*c} \cap A_0) \cup \ldots \cup (A_m^{*c} \cap A_{m-1} \cap A_0^c \cap A_1^c \cap \ldots \cap A_{m-2}^c) \\
&= \bigcup_{i=0}^{m-1} A_m^{*c} \cap A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c.
\end{aligned}
$$

Entonces,

$$* \leq \mathbb{P}(A_m^*) + \sum_{i=0}^{m-1} \mathbb{P}(A_m^{*c} \cap A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c).$$

Por lo tanto,

$$\mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) \leq \mathbb{P}(A_m^*) + \sum_{i=0}^{m-1} \mathbb{P}(A_m^{*c} \cap A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c).$$

Observemos que

$$
\begin{aligned}
A_i \cap A_0^c \cap A_1^c \cap \ldots \cap A_{i-1}^c &= A_i \cap A_1^c \cap \ldots \cap A_{i-1}^c \\
&= \{\bar{\alpha}_n(t_i) > x\} \cap \{\bar{\alpha}_n(t_1) \leq x\} \cap \ldots \cap \{\bar{\alpha}(t_{i-1}) \leq x\}.
\end{aligned}
$$

Entonces, podemos descomponer cada conjunto $A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c$ en conjuntos finitos, de tal manera que las variables discretas $\bar{\alpha}_n(t_j)$, $j \leq i$, tomen valores específicos $\bar{\alpha}_n(t_j) = x_j$. Entonces,

$$
\begin{aligned}
\mathbb{P}(A_m^{*c}, \bar{\alpha}_n(t_j) &= x_j;\ 0 \leq j \leq i) \\
&= \mathbb{P}(A_m^{*c} \mid \bar{\alpha}_n(t_j) = x_j;\ 0 \leq j \leq i)\mathbb{P}(\bar{\alpha}_n(t_j) = x_j;\ 0 \leq j \leq i) \\
&= \mathbb{P}(A_m^{*c} \mid \bar{\alpha}_n(t_i) = x_i)\mathbb{P}(\bar{\alpha}_n(t_j) = x_j;\ 0 \leq j \leq i).
\end{aligned}
$$

La última igualdad se debe a la propiedad de Markov de $\bar{\alpha}_n$. Ahora supongamos que x^* ha sido escogido de tal manera que las probabilidades condicionales son menores o iguales a una constante c, $c < 1$. Entonces

$$\mathbb{P}(A_m^{*c}, \bar{\alpha}_n(t_j) = x_j;\ 0 \leq j \leq i) \leq c\mathbb{P}(\bar{\alpha}_n(t_j) = x_j;\ 0 \leq j \leq i).$$

Así,

$$\begin{aligned} \mathbb{P}(A_m^{*c} \cap A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c) &\leq c \sum_{x_0, \ldots, x_i} \mathbb{P}(\bar{\alpha}_n(t_j) = x_j) \\ &= c\mathbb{P}(A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c). \end{aligned}$$

Se sigue entonces que

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) &\leq \mathbb{P}(A_m^*) + c\sum_{i=0}^{m-1} \mathbb{P}(A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c) \\ &\leq \mathbb{P}(A_m^*) + c\sum_{i=0}^{m} \mathbb{P}(A_i \cap A_0^c \cap \ldots \cap A_{i-1}^c) \\ &= \mathbb{P}(A_m^*) + c\mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right). \end{aligned}$$

Factorizando y despejando se tiene

$$\mathbb{P}\left(\bigcup_{i=0}^{m} A_i\right) \leq \frac{1}{1-c}\mathbb{P}(A_m^*).$$

Dicho en otras palabras,

$$\mathbb{P}\left(\sup_{0 \leq i \leq m} \bar{\alpha}_n(t_i) > x\right) \leq \frac{1}{1-c}\mathbb{P}(\bar{\alpha}_n(a) > x^*).$$

Observemos que el lado derecho de esta desigualdad no depende de la partición, así que si hacemos tender tender la norma de la partición a cero, esta desigualdad se siguirá cumpliendo. Ahora, puesto que $\bar{\alpha}_n$ es continua por la derecha y tiene límites por la izquierda,

$$\mathbb{P}\left(\sup_{0 \leq t \leq a} \bar{\alpha}_n(t) > x\right) \leq \frac{1}{1-c}\mathbb{P}(\bar{\alpha}_n(a) > x^*).$$

Queda por investigar las condiciones sobre $x^* \leq x$ que garantizen

$$\mathbb{P}(A_m^{*c} \mid \bar{\alpha}_n(t_i) = x_i) \leq c.$$

Para cada $x_i > x$,

$$\begin{aligned}\mathbb{P}(A_m^{*c} \mid \bar{\alpha}_n(t_i) = x_i) &= \mathbb{P}(\bar{\alpha}_n(t_m) \leq x^* \mid \bar{\alpha}_n(t_i) = x_i) \\ &= \mathbb{P}(\sqrt{n}(\bar{F}_n(t_m) - t_m) \leq x^* \mid \sqrt{n}(\bar{F}_n(t_i) - t_i) = x_i) \\ &= \mathbb{P}(n\bar{F}_n(t_m) \leq \sqrt{n}x^* + nt_m \mid n\bar{F}_n(t_i) = \sqrt{n}x_i + nt_i) \\ &= \mathbb{P}(n\bar{F}_n(t_m) \leq \sqrt{n}x^* + nt_m \mid n\bar{F}_n(t_i) = N) = *\end{aligned}$$

donde $N = \sqrt{n}x_i + nt_i$.

Por otro lado, recordemos que F_n es un proceso de Markov tal que

$$\begin{aligned}\mathcal{L}(nF_n(t) : t_0 &\leq t \mid nF_n(t_0) = k) \\ &= \mathcal{L}\left(k + (n-k)\bar{F}_{n-k}\left[\frac{F(t) - F(t_0)}{1 - F(t_0)}\right] : t_0 \leq t\right).\end{aligned}$$

Aplicando este resultado para $\bar{F}_n$ con $t = t_m, t_0 = t_i$ y $N = k$, se tiene

$$\begin{aligned}\mathcal{L}(n\bar{F}_n(t_m) : t_i &\leq t_m \mid n\bar{F}_n(t_i) = N) \\ &= \mathcal{L}\left(N + (n-N)\bar{F}_{n-N}\left[\frac{t_m - t_i}{1 - t_i}\right] : t_i \leq t_m\right).\end{aligned}$$

Entonces,

$$\begin{aligned}* &= \mathbb{P}\left(N + (n-N)\bar{F}_{n-N}\left[\frac{t_m - t_i}{1 - t_i}\right] \leq \sqrt{n}x^* + nt_m\right) \\ &= \mathbb{P}\left((n-N)\bar{F}_{n-N}\left[\frac{t_m - t_i}{1 - t_i}\right] \leq \sqrt{n}x^* + nt_m - N\right).\end{aligned}$$

Por lo tanto,

$$\begin{aligned}\mathbb{P}(A_m^{*c} \mid \bar{\alpha}_n(t_i) &= x_i) \\ &= \mathbb{P}\left((n-N)\bar{F}_{n-N}\left[\frac{t_m - t_i}{1 - t_i}\right] \leq \sqrt{n}x^* + nt_m - N\right).\end{aligned}$$

donde $N = \sqrt{n}x_i + nt_i$. Pero

$$\sqrt{n}x^* + nt_m - N = \sqrt{n}\left[x^* - x_i\frac{1 - t_m}{1 - t_i}\right] + (n - N)\frac{t_m - t_i}{1 - t_i}.$$

Entonces, la probabilidad del lado derecho se escribe ahora

$$\mathbb{P}\left\{(n-N)\bar{F}_{n-N}\left(\frac{t_m - t_i}{1 - t_i}\right) - (n-N)\frac{t_m - t_i}{1 - t_i} \leq \sqrt{n}\left[x^* - x_i\frac{1 - t_m}{1 - t_i}\right]\right\}.$$

Ahora, puesto que $x_i > x$, se sigue que

$$\begin{aligned}\sqrt{n}\left[x^* - x_i\frac{1-t_m}{1-t_i}\right] &\leq \sqrt{n}\left[x^* - x\frac{1-t_m}{1-t_i}\right] \\ &\leq \sqrt{n}[x^* - x(1-a)].\end{aligned}$$

La última desigualdad se debe a que $\frac{1-a}{1-t_i} > 1-a$. Escribiendo $x^* = x(1-\delta)$ con $a < \delta/2$, se sigue que

$$\begin{aligned}\sqrt{n}[x^* - x(1-a)] &= \sqrt{n}[x(1-\delta) - x(1-a)] \\ &= \sqrt{n}[x(a-\delta)] \\ &\leq \sqrt{n}x(-\delta/2) = -\sqrt{n}x\delta/2.\end{aligned}$$

Bajo estas condiciones se tiene ahora

$$\mathbb{P}(\bar{\alpha}_n(t_m) \leq x^* \mid \bar{\alpha}_n(t_i) = x_i)$$

$$\begin{aligned}&\leq \mathbb{P}\left((n-N)\bar{F}_{n-N}\left(\frac{t_m - t_i}{1-t_i}\right) - (n-N)\frac{t_m-t_i}{1-t_i} \leq -\sqrt{n}x\delta/2\right) \\ &\leq \frac{1}{nx^2\delta^2/4}(n-N)\frac{t_m-t_i}{1-t_i}\left[1 - \frac{t_m-t_i}{1-t_i}\right] \\ &\leq \frac{1}{nx^2\delta^2/4}(n-N)\frac{t_m-t_i}{1-t_i}.\end{aligned}$$

La segunda desigualdad, se debe a la desigualdad de Chevichev, ver apéndice B.2.1. Ahora, debido a que $t_m < 1$, $t_m - 1 \leq 0$. Entonces se tiene $nt_i(t_m - 1) - N(t_m - t_i) \leq 0$. Desarrollando y agrupando terminos nos lleva a la desigualdad

$$\frac{(n-N)(t_m - t_i)}{n(1-t_i)} \leq t_m = a.$$

De aquí se sigue que

$$\frac{4(n-N)(t_m-t_i)}{nx^2\delta^2(1-t_i)} \leq \frac{4a}{x^2\delta^2}.$$

Por último, si suponemos $8a \leq x^2\delta^2$, entonces $\frac{4a}{x^2\delta^2} \leq \frac{1}{2}$. Se tiene entonces que

$$\mathbb{P}(\bar{\alpha}_n(t_m) \leq x^* \mid \bar{\alpha}_n(t_i) = x_i) \leq \frac{1}{2}.$$

2.9.1. Lema. *Supongamos que se cumplen la siguientes condiciones*

(i) $a < \delta/2$

(ii) $8a \le x^2\delta^2$.

Entonces,

$$\mathbb{P}\left(\sup_{0\le t\le a} \bar{\alpha}_n(t) > x\right) \le 2\mathbb{P}(\bar{\alpha}_n(a) > x(1-\delta)).$$

Demostración: Ya demostramos que

$$\mathbb{P}\left(\sup_{0\le t\le a} \bar{\alpha}_n(t) > x\right) \le \frac{1}{1-c}\mathbb{P}(\bar{\alpha}_n(a) > x^*).$$

Entonces, haciendo $c = 1/2$ y $x^* = x(1-\delta)$ se obtiene lo deseado. □

Por supuesto que también se cumple la otra cota:

$$\mathbb{P}\left(\sup_{0\le t\le a} -\bar{\alpha}_n(t) > x\right) \le 2\mathbb{P}(\bar{\alpha}_n(a) < x(1-\delta)).$$

Así, con las condiciones (i)-(ii) del lema 2.9.1, se tiene

$$\mathbb{P}\left(\sup_{0\le t\le a} |\bar{\alpha}_n(t)| > x\right) \le 2\mathbb{P}(|\bar{\alpha}_n(a)| > x(1-\delta)). \tag{2.24}$$

Ahora daremos una cota para el módulo de oscilación $\bar{\omega}_n$. Para ello, necesitaremos el siguiente lema

2.9.2. Lema. *Si η se distribuye Binomial con párametros $0 < p < 1$ y $n \in \mathbb{N}$, entonces para cada $z > 0$,*

$$\mathbb{P}(|\eta - np| \ge z) \le 2\exp -np[(1+z/np)\ln(1+z/np) - z/np]. \tag{2.25}$$

Demostración: $\mathbb{P}(|\eta - np| \ge z)$

$$\begin{aligned}
&= \mathbb{P}(\eta - np \ge z \text{ o } \eta - np \le -z) \\
&= \mathbb{P}(\eta - np \ge z) + \mathbb{P}(\eta - np \le -z) \\
&\le \exp\left[-\frac{z^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right)\right] + \exp\left[-\frac{z^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{n(1-p)}\right)\right].
\end{aligned}$$

La última desigualdad, se debe al lema 2.8.5. Ahora observemos que

$$\begin{aligned}
-\frac{z^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{np}\right) &= -\frac{z^2}{2np(1-p)}2h\left(1+\frac{z}{np}\right)/\left(\frac{z}{np}\right)^2 \\
&= -\frac{np}{1-p}h\left(1+\frac{z}{np}\right) \\
&= -\frac{np}{1-p}[(1+z/np)(\ln(1+z/np)-1)+1] \\
&= -\frac{np}{1-p}[(1+z/np)\ln(1+z/np)-z/np] \\
&\leq -np[(1+z/np)\ln(1+z/np)-z/np].
\end{aligned}$$

Analogamente,

$$-\frac{z^2}{2np(1-p)}\psi\left(\frac{\varepsilon}{n(1-p)}\right) \leq -np[(1+z/np)\ln(1+z/np)-z/np].$$

Aplicando exponencial las dos desigualdades obtenidas, se obtiene el resultado. □

Si definimos $y = z/np$, obtenemos para el exponente en (2.25) la siguiente expresión

$$-np[(1+y)\ln(1+y)-y] := H(y).$$

Observemos que

$$H(y) = -np\int_0^y \ln(1+x)dx.$$

El lema anterior tiene ahora la expresión

$$\begin{aligned}
\mathbb{P}(|\eta-np| \geq z) &\leq 2\exp\left[-np\int_0^y \ln(1+x)dx\right] \\
&= 2\exp\left[-np\int_0^y x\frac{\ln(1+x)}{x}dx\right].
\end{aligned}$$

Puesto que $\frac{\ln(1+x)}{x} \to 1$ cuando $x \downarrow 0$, si $0<\delta<1$. $0<1-\delta<1$. Entonces existe $x_\delta > 0$ tal que si $0 < x \leq x_\delta$, $1-\delta \leq \frac{\ln(1+x)}{x}$ i.e., $(1-\delta)x \leq \ln(1+x)$.

Luego, si $0 < y \leq x_\delta$,

$$\int_0^y (1-\delta)xdx \leq \int_0^y \ln(1+x)dx$$

o equivalentemente

$$\frac{(1-\delta)z^2}{2(np)^2} = \frac{(1-\delta)y^2}{2} \leq \int_0^y \ln(1+x)dx.$$

Multiplicando por a esto último por $-np$, se obtiene

$$-\frac{z^2(1-\delta)}{2np} \geq -np \int_0^y \ln(1+x)dx.$$

Así, se tiene

$$\mathbb{P}(|\eta - np| \geq z) \leq 2\exp[-(1-\delta)z^2/2np] \quad \text{si } y \leq x_\delta \ (z \leq npx_\delta) \tag{2.26}$$

2.9.3. Lema. *Sean* $0 < a,\ \delta < 1$ *y* $s > 0$ *tales que*

(i) $a < \delta/2$

(ii) $8 \leq [s\delta/(1+\delta)]^2$

(iii) $s \leq \delta x_\delta \sqrt{na}/4$ *p.a.* x_δ *que depende solo de* δ.

Entonces,

$$\mathbb{P}(\bar{\omega}_n(a) > s\sqrt{a}) \leq C_\delta a^{-1} \exp[-s^2(1-\delta)^5/2].$$

donde $C_\delta = 64\delta^{-2}$.

Demostración: Sea $x = s\sqrt{a}$ y R el entero más chico tal que $1/\sqrt{R} \leq \delta\sqrt{a}/2$. Entonces

$$\begin{aligned}\bar{\omega}_n(a) &\leq \max_{0\leq i\leq R-1} \sup_{0\leq t\leq a} \left| \bar{\alpha}_n\left(\frac{i}{R}+t\right) - \bar{\alpha}_n\left(\frac{i}{R}\right) \right| \\ &+ 2 \max_{0\leq i\leq R-1} \sup_{0\leq \tau\leq 1/R} \left| \bar{\alpha}_n\left(\frac{i}{R}+\tau\right) - \bar{\alpha}_n\left(\frac{i}{R}\right) \right|.\end{aligned}$$

en efecto, recordemos que

$$\bar{\omega}_n(a) = \sup_{|t-s|\leq a} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)|.$$

Sean s, t tales que $|t-s| \leq a$ y $s \leq t$. Sea

$$s \in \left[\frac{i}{R}, \frac{i+1}{R}\right].$$

Entonces $s = \frac{i}{R} + \tau$, donde $0 \leq \tau \leq \frac{1}{R}$. Sea

$$t \in \left[\frac{j}{R}, \frac{j+1}{R}\right] \quad \text{p.a. } j.$$

Entonces

$$\begin{aligned} t &= s + \beta, \quad \text{donde } 0 < \beta \leq a \\ &= \frac{i}{R} + s_1 + \beta \quad \text{donde } 0 \leq s_1 \leq \frac{1}{R} \text{ y } 0 < \beta \leq a. \end{aligned}$$

Así que

$$t \in \left[\frac{i}{R}, \frac{i}{R} + a\right] \quad \text{o} \quad t \in \left[\frac{i+1}{R}, \frac{i+1}{R} + a\right].$$

Entonces,

$$t = \frac{i}{R} + \gamma, \quad 0 \leq \gamma \leq a \quad \text{o} \quad t = \frac{i+1}{R} + \gamma, \quad 0 \leq \gamma \leq a.$$

Luego, si $t = \frac{i}{R} + \gamma$, $0 \leq \gamma \leq a$, se tiene

$$\begin{aligned} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)| &\leq \left|\bar{\alpha}_n(t) - \bar{\alpha}_n(\frac{i}{R})\right| + \left|\bar{\alpha}_n(s) - \bar{\alpha}_n(\frac{i}{R})\right| \\ &\leq \left|\bar{\alpha}_n(t) - \bar{\alpha}_n(\frac{i}{R})\right| + 2\left|\bar{\alpha}_n(s) - \bar{\alpha}_n(\frac{i}{R})\right| \\ &= \left|\bar{\alpha}_n(\frac{i}{R} + \gamma) - \bar{\alpha}_n(\frac{i}{R})\right| + 2\left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right|. \end{aligned}$$

De aquí se sigue que

$$\begin{aligned} \sup_{|t-s| \leq a} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)| &\leq \max_{0 \leq i \leq R-1} \sup_{0 \leq \gamma \leq a} \left|\bar{\alpha}_n(\frac{i}{R} + \gamma) - \bar{\alpha}_n(\frac{i}{R})\right| \\ &+ 2 \max_{0 \leq i \leq R-1} \sup_{0 \leq \tau \leq \frac{1}{R}} \left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right|. \end{aligned}$$

Ahora, si $t = \frac{i+1}{R} + \gamma$, $0 \leq \gamma \leq a$, se tiene

$$\begin{aligned} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)| &\leq \left|\bar{\alpha}_n(t) - \bar{\alpha}_n(\frac{i+1}{R})\right| + \left|\bar{\alpha}_n(\frac{i+1}{R}) - \bar{\alpha}_n(s)\right| \\ &\leq \left|\bar{\alpha}_n(t) - \bar{\alpha}_n(\frac{i+1}{R})\right| + \left|\bar{\alpha}_n(\frac{i+1}{R}) - \bar{\alpha}_n(\frac{i}{R})\right| \end{aligned}$$

$$
\begin{aligned}
&+ \left|\bar{\alpha}_n(s) - \bar{\alpha}_n(\frac{i}{R})\right| \\
&= \left|\bar{\alpha}_n(\frac{i+1}{R} + \gamma) - \bar{\alpha}_n(\frac{i+1}{R})\right| + \left|\bar{\alpha}_n(\frac{i}{R} + \frac{1}{R}) - \bar{\alpha}_n(\frac{i}{R})\right| \\
&+ \left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right|.
\end{aligned}
$$

Se sigue entonces que

$$
\begin{aligned}
\sup_{|t-s|\leq a} |\bar{\alpha}_n(t) - \bar{\alpha}_n(s)| \;\leq\; & \max_{0\leq i\leq R-1} \sup_{0\leq\gamma\leq a} \left|\bar{\alpha}_n(\frac{i}{R} + \gamma) - \bar{\alpha}_n(\frac{i}{R})\right| \\
& + \; 2 \max_{0\leq i\leq R-1} \sup_{0\leq\tau\leq\frac{1}{R}} \left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right|.
\end{aligned}
$$

Esto último es porque

$$
\max_{0\leq i\leq R-1} \sup_{0\leq\tau\leq\frac{1}{R}} \left|\bar{\alpha}_n(\frac{i}{R} + \frac{1}{R}) - \bar{\alpha}_n(\frac{i}{R})\right| \leq \max_{0\leq i\leq R-1} \sup_{0\leq\tau\leq\frac{1}{R}} \left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right|.
$$

Finalmente, de los dos casos con $\gamma = t$, se tiene lo deseado.

Como

$$
\{Y + Z \leq x\} \supset \left\{Y \leq \frac{x}{1+\delta}\right\} \cap \left\{Z \leq \frac{\delta x}{1+\delta}\right\},
$$

se sigue que

$$
\{Y + Z > x\} \subset \left\{Y > \frac{x}{1+\delta}\right\} \cup \left\{Z > \frac{\delta x}{1+\delta}\right\}.
$$

Entonces

$$
\begin{aligned}
\mathbb{P}(\bar{\omega}_n(a) > x) \leq \mathbb{P}&\left(\max_{0\leq i\leq R-1} \sup_{0\leq t\leq a} \left|\bar{\alpha}_n(\frac{i}{R} + t) - \bar{\alpha}_n(\frac{i}{R})\right| > \frac{x}{1+\delta}\right) \\
+ \;\; \mathbb{P}&\left(2 \max_{0\leq i\leq R-1} \sup_{0\leq\tau\leq\frac{1}{R}} \left|\bar{\alpha}_n(\frac{i}{R} + \tau) - \bar{\alpha}_n(\frac{i}{R})\right| > \frac{\delta x}{(1+\delta)}\right) \\
\leq \;\; & R\mathbb{P}\left(\sup_{0\leq t\leq a} |\bar{\alpha}_n(t)| > \frac{x}{1+\delta}\right) + R\mathbb{P}\left(\sup_{0\leq t\leq\frac{1}{R}} |\bar{\alpha}_n(t)| > \frac{x\delta}{2(1+\delta)}\right).
\end{aligned}
$$

La última desigualdad, se sigue del hecho de que $\bar{\alpha}_n$ tiene incrementos estacionarios, así que se toma $t = 0$, $\tau = 0$ y tomando en cuenta que hay R

intervalos.

Observemos que

$$\mathbb{P}\left(\sup_{0\leq t\leq a}|\bar{\alpha}_n(t)| > \frac{x}{1+\delta}\right) \leq 2\mathbb{P}\left(|\bar{\alpha}_n(a)| > \frac{x(1-\delta)}{1+\delta}\right).$$

en efecto: se tienen las condiciones

(i) $a < \delta/2$ por hipótesis

(ii) $8 \leq [s\delta/(1+\delta)]^2$ implica $8a \leq [s\sqrt{a}\delta/(1+\delta)]^2 = \left(\frac{x}{1+\delta}\right)^2\delta^2$.

y se aplica el lema 2.9.1. También se tiene

$$\mathbb{P}\left(\sup_{0\leq t\leq \frac{1}{R}}|\bar{\alpha}_n(t)| > \frac{\delta x}{2(1+\delta)}\right) \leq 2\mathbb{P}\left(\left|\bar{\alpha}_n(\frac{1}{R})\right| > \frac{x\delta(1-\delta)}{2(1+\delta)}\right).$$

en efecto: se tienen las condiciones

(i) $a < \delta/2$ y $\frac{1}{\sqrt{R}} \leq \delta\sqrt{a}/2$ por hipótesis. Entonces $\frac{1}{R} \leq \frac{\delta^2 a}{4} \leq \frac{\delta^2}{2}\frac{\delta}{2} < \frac{\delta}{2}$

(ii) $8 \leq [s\delta/(1+\delta)]^2$ implica $\frac{8}{R} \leq [s\delta/(1+\delta)]^2\frac{1}{R} \leq [s\delta/(1+\delta)]^2\frac{\delta^2 a}{4} = [s\sqrt{a}\delta/2(1+\delta)]^2\delta^2 = \left[\frac{x\delta}{2(1+\delta)}\right]^2\delta^2$,

y se aplica el lema 2.9.1.

Por lo tanto, se tiene

$$\mathbb{P}(\bar{\omega}_n(a) > x) \leq 2R\left[\mathbb{P}\left(|\bar{\alpha}_n(a)| > \frac{x(1-\delta)}{1+\delta}\right) + \mathbb{P}\left(\left|\bar{\alpha}_n(\frac{1}{R})\right| > \frac{x\delta(1-\delta)}{2(1+\delta)}\right)\right].$$

Por otra parte

$$\begin{aligned}\mathbb{P}\left(|\bar{\alpha}_n(a)| > \frac{x(1-\delta)}{1+\delta}\right) &= \mathbb{P}\left(\left|n^{1/2}(\bar{F}_n(a)-a)\right| > \frac{s\sqrt{a}(1-\delta)}{1+\delta}\right)\\ &= \mathbb{P}\left(|n\bar{F}_n(a)-na| > \frac{s\sqrt{na}(1-\delta)}{1+\delta}\right).\end{aligned}$$

Por otro lado, observemos que

$$\frac{s\sqrt{na}(1-\delta)}{1+\delta} \leq nax_\delta,$$

en efecto: por la hipótesis (iii) $s \leq \delta x_\delta \sqrt{na/4} = \delta x_\delta \sqrt{na}/2$. Multiplicando por $2\sqrt{na}/\delta$ se tiene $2s\sqrt{na}/\delta \leq nax_\delta$. Luego, $(1-\delta)2s\sqrt{na}/(1+\delta) \leq 2s\sqrt{na}/(1+\delta) \leq 2s\sqrt{na}/\delta$.

Esta es la condición para usar (2.26) y obtener

$$\begin{aligned}
&\mathbb{P}\left(\left|n\bar{F}_n(a) - na\right| > \tfrac{s\sqrt{na}(1-\delta)}{1+\delta}\right) \\
&\qquad\leq\; 2\exp[-(1-\delta)s^2na(1-\delta)^2/(1+\delta)^2 2na] \\
&\qquad=\; 2\exp[-s^2(1-\delta)^3/2(1+\delta)^2].
\end{aligned}$$

Para la otra probabilidad

$$\begin{aligned}
\mathbb{P}\left(\mid \bar{\alpha}_n(\frac{1}{R}) \mid > \frac{x\delta(1-\delta)}{2(1+\delta)}\right) &= \mathbb{P}\left(\mid n\bar{F}_n\left(\frac{1}{R}\right) - \frac{n}{R} \mid > \frac{s\sqrt{na}\delta(1-\delta)}{2(1+\delta)}\right) \\
&= \mathbb{P}\left(\mid n\bar{F}_n\left(\frac{1}{R}\right) - \frac{n}{R} \mid > \frac{s\sqrt{n}(1-\delta)}{1+\delta}\frac{\delta\sqrt{a}}{2}\right) \\
&\leq \mathbb{P}\left(\mid n\bar{F}_n\left(\frac{1}{R}\right) - \frac{n}{R} \mid > \frac{s\sqrt{n}(1-\delta)}{1+\delta}\frac{1}{\sqrt{R}}\right).
\end{aligned}$$

La última desigualdad, se debe a que $\frac{\delta\sqrt{a}}{2} > \frac{1}{\sqrt{R}}$. Luego,

$$\frac{s\sqrt{n}(1-\delta)}{(1+\delta)\sqrt{R}} \leq \frac{n}{R}x_\delta,$$

en efecto: por la hipótesis (iii), y por la elección de R,

$$s \leq \delta x_\delta \sqrt{na}/4 = \frac{x_\delta\sqrt{n}}{2}\frac{\delta\sqrt{a}}{2} < \frac{x_\delta\sqrt{n}}{2}\frac{1}{\sqrt{R-1}} < \frac{x_\delta\sqrt{n}}{\sqrt{R}}.$$

Se sigue que $s\sqrt{n}/\sqrt{R} \leq \frac{n}{R}x_\delta$. Finalmente, $\frac{1-\delta}{1+\delta} < 1$, y esto lleva al resultado.

Esta es la condición para usar (2.26) y obtener

$$\begin{aligned}
\mathbb{P}\left(\mid n\bar{F}_n\left(\frac{1}{R}\right) - \frac{n}{R} \mid > \frac{s\sqrt{n}(1-\delta)}{(1+\delta)\sqrt{R}}\right) &\leq 2\exp\left[-\frac{(1-\delta)s^2n(1-\delta)^2}{(1+\delta)^2R}\frac{1}{2n/R}\right] \\
&= 2\exp[-(1-\delta)^3s^2/2(1+\delta)^2].
\end{aligned}$$

Entonces,

$$\mathbb{P}(\bar{\omega}_n(a) > x) \;\leq\; 2R\Big\{2\exp\left[-\frac{s^2(1-\delta)^3}{2(1+\delta)^2}\right] + 2\exp\left[-\frac{s^2(1-\delta)^3}{2(1+\delta)^2}\right]\Big\}$$

$$
\begin{aligned}
&= 8R\exp\left[-\frac{s^2(1-\delta)^3}{2(1+\delta)^2}\right] = 8R\exp\left[-\frac{s^2(1-\delta)^5}{2(1-\delta^2)^2}\right] \\
&\leq 8R\exp\left[-\frac{s^2(1-\delta)^5}{2}\right] \leq 64\delta^{-2}a^{-1}\exp[-s^2(1-\delta)^5/2].
\end{aligned}
$$

La última desigualdad, se debe a que $\delta\sqrt{a}/2 < \frac{1}{\sqrt{R-1}}$, elevando al cuadrado y despejando se tiene $R-1 < 4/\delta^2 a$. Entonces,

$$
R < \frac{4}{\delta^2 a} + 1 < \frac{4}{\delta^2 a} + \frac{4}{\delta^2 a} = 8\delta^{-2}a^{-1}.
$$

□

2.9.4. Teorema. *Para todo $\varepsilon > 0$ y para todo $\eta > 0$, existe $a > 0$, tal que para toda $n \geq n_0(\varepsilon, \eta)$,*

$$
\mathbb{P}(\bar{\omega}_n(a) \geq \varepsilon) \leq \eta.
$$

Demostración: Sean $\delta = 1/2$ y $s = a^{-1/4}$. Entonces, las condiciones (i)-(iii) del lema 2.9.3 se satisfacen para una $a > 0$ suficientemente chica y n suficientemente grande. Esto es

(i) $a < \delta/2 = \frac{1}{4}$

(ii) $8 \leq [a^{-1/4}(1/2)/(3/2)]^2 = [a^{-1/4}/3]^2 = 1/9\sqrt{a}$

(iii) $a^{1/4} \leq \frac{1}{2}x_\delta\sqrt{na}/4$.

Entonces

$$
\begin{aligned}
\mathbb{P}(\bar{\omega}_n(a) > s\sqrt{a}) &\leq c_{1/2}a^{-1}\exp[-a^{-1/2}(1/2)^5/2] \\
&= c_{1/2}a^{-1}\exp[-a^{-1/2}/64].
\end{aligned}
$$

Es claro que para toda $\eta > 0$ y toda $\varepsilon > 0$, podemos elegir una $a > 0$ suficientemente chica, con $\varepsilon \geq a^{1/4}$, de tal manera que

$$
c_{1/2}a^{-1}\exp[-a^{-1/2}/64] \leq \eta.
$$

Por otro lado, observemos que

$$
\{\bar{\omega}_n(a) > \varepsilon\} \subset \{\bar{\omega}_n(a) > s\sqrt{a}\},
$$

en efecto: si $\omega \in \{\bar{\omega}_n(a) > \varepsilon\}$ entonces $\bar{\omega}_n(a) > \varepsilon \geq a^{1/4} = s\sqrt{a}$. Por lo tanto $\omega \in \{\bar{\omega}_n(a) > s\sqrt{a}\}$ y

$$\mathbb{P}(\bar{\omega}_n(a) > \varepsilon) \leq \mathbb{P}(\bar{\omega}_n(a) > s\sqrt{a}).$$

Se sigue entonces que

$$\mathbb{P}(\bar{\omega}_n(a) > \varepsilon) \leq \eta.$$

□

2.9.5. Teorema. *Ver* [14]. *Sea* (a_n) *una sucesión que converge a cero cuando* $n \to \infty$. *Supongamos que se cumplen*

(i) $\ln a_n^{-1} = o(na_n)$

(ii) $\sum_{n\geq 1} a_n^r < \infty$ *para alguna* $r > 0$.

Entonces

$$\bar{\omega}_n(a_n) = O(\sqrt{a_n \ln a_n^{-1}}) \quad \textit{con probabilidad uno.}$$

2.10. Cotas exponenciales

En esta sección daremos algunas cotas de probabilidad para sumas de variables aleatorias independientes, ver Hoeffding [6].

Sean $X_1, \ldots, X_n$ v.a.i., no necesariamente provenientes de la misma distribución. Definamos

$$S := \sum_{i=1}^{n} X_i, \quad \bar{S} := S/n$$

y

$$\mu := \mathbb{E}(\bar{S}) = n^{-1}\sum_{i=1}^{n} \mathbb{E}(X_i).$$

2.10.1. Lema. *Para toda* $\varepsilon > 0$ *y toda* $h > 0$,

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq e^{-h(n\varepsilon + \mathbb{E}(S))} \prod_{i=1}^{n} \mathbb{E}e^{hX_i}.$$

Demostración:

$$\begin{aligned}
\mathbb{P}(\bar{S} - \mu \geq \varepsilon) &= \mathbb{P}(S - \mathbb{E}(S) \geq n\varepsilon) \\
&= \mathbb{P}(S \geq n\varepsilon + \mathbb{E}(S)) \\
&\leq e^{-h(n\varepsilon + \mathbb{E}(S))}\mathbb{E}e^{hS} \\
&= e^{-h(n\varepsilon + \mathbb{E}(S))}\prod_{i=1}^{n}\mathbb{E}e^{hX_i}.
\end{aligned}$$

La desigualdad, se debe a la desigualdad de Chernoff, ver apéndice B.2.2, y la última igualdad, es por la hipótesis de independencia. Esto completa la prueba. □

2.10.2. Lema. *Sea $a \leq X \leq b$ y $h \in \mathbb{R}$. Entonces,*

$$\mathbb{E}e^{hX} \leq \frac{b - \mathbb{E}(X)}{b-a}e^{ha} + \frac{\mathbb{E}(X) - a}{b-a}e^{hb}.$$

Demostración: Observemos que

$$e^{hX} = e^{h\frac{b-X}{b-a}a + h\frac{X-a}{b-a}b}.$$

Sea

$$l(X) = h\frac{b-X}{b-a}a + h\frac{X-a}{b-a}b,$$

entonces $l(a) = ha$ y $l(b) = hb$. Ahora, si definimos $f(x) = e^x$ entonces f es convexa y por eso

$$\begin{aligned}
f(h\frac{b-X}{b-a}a + h\frac{X-a}{b-a}b) &= e^{h\frac{b-X}{b-a}a + h\frac{X-a}{b-a}b} \\
&\leq \frac{b-X}{b-a}f(ha) + \frac{X-a}{b-a}f(hb) \\
&= \frac{b-X}{b-a}e^{ha} + \frac{X-a}{b-a}e^{hb}.
\end{aligned}$$

Por lo tanto,

$$e^{h\frac{b-X}{b-a}a + h\frac{X-a}{b-a}b} \leq \frac{b-X}{b-a}e^{ha} + \frac{X-a}{b-a}e^{hb}.$$

Finalmente, aplicando esperanza obtenemos

$$\mathbb{E}e^{hX} \leq \frac{b - \mathbb{E}(X)}{b-a}e^{ha} + \frac{\mathbb{E}(X) - a}{b-a}e^{hb}.$$

□

2.10.3. Lema. ***(Hoeffding).*** *Sean $X_1, \ldots, X_n$ independientes con $0 \leq X_i \leq 1$. Entonces, si $0 < \varepsilon < 1 - \mu$,*

$$\begin{aligned}\mathbb{P}(\bar{S} - \mu \geq \varepsilon) &\leq \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^n \\ &\leq e^{-n\varepsilon^2 g(\mu)} \leq e^{-2n\varepsilon^2},\end{aligned}$$

donde

$$g(\mu) = \begin{cases} \frac{1}{1-2\mu}\ln\frac{1-\mu}{\mu}, & si \quad 0 \leq \mu < 1/2 \\ \frac{1}{2\mu(1-\mu)}), & si \quad 1/2 \leq \mu \leq 1. \end{cases}$$

Demostración: Sea $\mu_i = \mathbb{E}X_i$ para todo $i = 1, \ldots, n$. Aplicando el lema 2.10.2, con $a = 0$ y $b = 1$, se obtiene

$$\mathbb{E}e^{hX_i} \leq 1 - \mu_i + \mu_i e^h.$$

Entonces,

$$\prod_{i=1}^n \mathbb{E}e^{hX_i} \leq \prod_{i=1}^n (1 - \mu_i + \mu_i e^h).$$

Luego, la media geométrica es menor o igual que la media aritmética, ver [3], esto es,

$$\left[\prod_{i=1}^n (1 - \mu_i + \mu_i e^h)\right]^{1/n} \leq \frac{1}{n}\sum_{i=1}^n (1 - \mu_i + \mu_i e^h).$$

Se sigue que

$$\prod_{i=1}^n (1 - \mu_i + \mu_i e^h) \leq \left[\frac{1}{n}\sum_{i=1}^n (1 - \mu_i + \mu_i e^h)\right]^n = [1 - \mu + \mu e^h]^n.$$

Aplicando el lema 2.10.1, se tiene

$$\begin{aligned}\mathbb{P}(\bar{S} - \mu \geq \varepsilon) &\leq e^{-h(n\varepsilon + \mathbb{E}(S))}\prod_{i=1}^n \mathbb{E}e^{hX_i} \\ &\leq e^{-hn(\varepsilon+\mu)}[(1 - \mu + \mu e^h)]^n \\ &= \left[e^{-h(\varepsilon+\mu)}(1 - \mu + \mu e^h)\right]^n.\end{aligned}$$

Por lo tanto,

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq \left[e^{-h(\varepsilon+\mu)}(1 - \mu + \mu e^h)\right]^n.$$

Sea
$$f(h) := \left[e^{-h(\varepsilon+\mu)}(1-\mu+\mu e^{h})\right]^{n}.$$
Derivando esta función encontramos que
$$h = \ln\frac{(\varepsilon+\mu)(1-\mu)}{(1-\mu-\varepsilon)\mu},$$
es un mínimo de f. Sustituyendo este valor en f obtenemos
$$\begin{aligned}
f(h) &= \left[\left(\frac{(1-\mu-\varepsilon)\mu}{(1-\mu)(\mu+\varepsilon)}\right)^{\varepsilon+\mu}\left(1-\mu+\mu\frac{(1-\mu)(\varepsilon+\mu)}{(1-\mu-\varepsilon)\mu}\right)\right]^{n} \\
&= \left[\left(\frac{(1-\mu-\varepsilon)\mu}{(1-\mu)(\mu+\varepsilon)}\right)^{\varepsilon+\mu}\left(1-\mu+\frac{(1-\mu)(\varepsilon+\mu)}{1-\mu-\varepsilon}\right)\right]^{n} \\
&= \left[\mu^{\varepsilon+\mu}\left(\frac{1-\mu-\varepsilon}{(1-\mu)(\mu+\varepsilon)}\right)^{\varepsilon+\mu}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)\right]^{n} \\
&= \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\varepsilon+\mu}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^{n}.
\end{aligned}$$
Por lo tanto,
$$\mathbb{P}(\bar{S}-\mu\geq\varepsilon) \leq \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\varepsilon+\mu}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^{n}.$$
Definamos
$$G(\varepsilon,\mu) := \frac{\mu+\varepsilon}{\varepsilon^{2}}\ln\left(\frac{\mu+\varepsilon}{\mu}\right)+\frac{1-\mu-\varepsilon}{\varepsilon^{2}}\ln\left(\frac{1-\mu-\varepsilon}{1-\mu}\right).$$
Entonces,
$$\begin{aligned}
-n\varepsilon^{2}G(\varepsilon,\mu) &= -n(\mu+\varepsilon)\ln\left(\frac{\mu+\varepsilon}{\mu}\right)-n(1-\mu-\varepsilon)\ln\left(\frac{1-\mu-\varepsilon}{1-\mu}\right) \\
&= \ln\left(\frac{\mu}{\mu+\varepsilon}\right)^{n(\mu+\varepsilon)}+\ln\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{n(1-\mu-\varepsilon)} \\
&= \ln\left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^{n}.
\end{aligned}$$
por lo tanto
$$e^{-n\varepsilon^{2}G(\varepsilon,\mu)} = \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^{n},$$

y

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq e^{-n\varepsilon^2 G(\varepsilon,\mu)}.$$

Luego, $\frac{\partial G(\varepsilon,\mu)}{\partial \varepsilon}$

$$= \frac{1}{\varepsilon^2}\left[\left(1 - \frac{2(1-\mu)}{\varepsilon}\right)\ln\left(1 - \frac{\varepsilon}{1-\mu}\right) - \left(1 - \frac{2(\mu+\varepsilon)}{\varepsilon}\right)\ln\left(1 - \frac{\varepsilon}{\mu+\varepsilon}\right)\right].$$

Entonces, $\frac{\varepsilon^2 \partial(\varepsilon,\mu)}{\partial \varepsilon}$

$$\begin{aligned} &= \left(1 - \frac{2(1-\mu)}{\varepsilon}\right)\ln\left(1 - \frac{\varepsilon}{1-\mu}\right) - \left(1 - \frac{2(\mu+\varepsilon)}{\varepsilon}\right)\ln\left(1 - \frac{\varepsilon}{\mu+\varepsilon}\right) \\ &\equiv H\left(\frac{\varepsilon}{1-\mu}\right) - H\left(\frac{\varepsilon}{\mu+\varepsilon}\right), \end{aligned}$$

donde

$$H(s) = \left(1 - \frac{2}{s}\right)\ln(1-s); \quad 0 < s < 1.$$

En nuestro caso: $0 < \frac{\varepsilon}{1-\mu} < 1$ y $0 < \frac{\varepsilon}{\mu+\varepsilon} < 1$. En general, expandiendo $H(s)$ en su serie de Taylor sobre $|s| < 1$, se obtiene

$$H(s) = 2 + \left(\frac{2}{3} - \frac{1}{2}\right)s^2 + \left(\frac{2}{4} - \frac{1}{3}\right)s^3 + \left(\frac{2}{5} - \frac{1}{4}\right)s^4 + \dots$$

$H(s)$ es creciente en $0 < s < 1$, debido a que todos los coeficientes de la serie son positivos. Así

$$\frac{\partial G(\varepsilon,\mu)}{\partial \varepsilon} > 0 \text{ si, y sólo si } H\left(\frac{\varepsilon}{1-\mu}\right) - H\left(\frac{\varepsilon}{\mu+\varepsilon}\right) > 0.$$

Debido a que H es creciente, se sigue que

$$\frac{\varepsilon}{1-\mu} > \frac{\varepsilon}{\mu+\varepsilon},$$

o lo que es lo mismo

$$\varepsilon > 1 - 2\mu.$$

Por lo tanto, si $1 - 2\mu > 0$, i.e., $\mu < 1/2$, entonces $G(\varepsilon,\mu)$ alcanza su mínimo en $\varepsilon = 1 - 2\mu$. Entonces $G(\varepsilon,\mu)$ es igual a

$$g(\mu) = \frac{\mu - 1 - 2\mu}{(1-2\mu)^2}\ln\frac{\mu + 1 - 2\mu}{\mu} + \frac{1 - \mu - (1-2\mu)}{(1-2\mu)^2}\ln\frac{1 - \mu - (1-2\mu)}{1-\mu}$$

$$\begin{aligned}
&= \frac{1-\mu}{(1-2\mu)^2}\ln\frac{1-\mu}{\mu} + \frac{\mu}{(1-2\mu)^2}\ln\frac{\mu}{1-\mu} \\
&= \frac{1-\mu}{(1-2\mu)^2}\ln\frac{1-\mu}{\mu} - \frac{\mu}{(1-2\mu)^2}\ln\frac{1-\mu}{\mu} \\
&= \frac{1}{1-2\mu}\ln\frac{1-\mu}{\mu}.
\end{aligned}$$

Por lo tanto,

$$g(\mu) = \frac{1}{1-2\mu}\ln\frac{1-\mu}{\mu} \quad \text{si } 0<\mu<\frac{1}{2}.$$

Ahora, si $1-2\mu \leq 0$, i.e., $1/2 \leq \mu$, entonces $G(\varepsilon,\mu)$ alcanza su mínimo en $\varepsilon = 0$. Por otro lado, expandiendo $G(\varepsilon,\mu)$ en su serie de Taylor, se tiene que $G(\varepsilon,\mu)$ es igual a

$$\frac{1}{2}\left(\frac{1}{1-\mu}+\frac{1}{\mu}\right) + \frac{1}{2\cdot 3}\left(\frac{1}{(1-\mu)^2}-\frac{1}{\mu^2}\right)\varepsilon + \frac{1}{3\cdot 4}\left(\frac{1}{(1-\mu)^3}+\frac{1}{\mu^3}\right)\varepsilon^3 + \ldots$$

Entonces,

$$g(\mu) = G(0,\mu) = \frac{1}{2}\left(\frac{1}{1-\mu}+\frac{1}{\mu}\right) = \frac{1}{2\mu(1-\mu)} \quad \text{si } \frac{1}{2}\leq\mu<1.$$

Por lo tanto,

$$g(\mu) = \inf_{0\leq\varepsilon<1-\mu} G(\varepsilon,\mu) = \begin{cases} \frac{1}{1-2\mu}\ln\frac{1-\mu}{\mu}, & \text{si } \ 0<\mu<1/2 \\ \frac{1}{2\mu(1-\mu)}, & \text{si } \ 1/2\leq\mu<1. \end{cases}$$

Se sigue que $g(\mu)\leq G(\varepsilon,\mu)$. Multiplicando por $-n\varepsilon^2$ se tiene $-n\varepsilon^2 g(\mu) \geq -n\varepsilon^2 G(\varepsilon,\mu)$. Entonces,

$$\begin{aligned}
\mathbb{P}(\bar{S}-\mu\geq\varepsilon) &\leq \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^n \\
&= e^{-n\varepsilon^2 G(\varepsilon,\mu)} \leq e^{-n\varepsilon^2 g(\mu)}.
\end{aligned}$$

Por último, veamos que

$$g(\mu)\geq 2.$$

Si $1/2\leq\mu<1$, $g(\mu)=\frac{1}{2\mu(1-\mu)}$. Derivando $g(\mu)$ encontramos que $\mu=\frac{1}{2}$ es un mínimo para g. Luego, $g(1/2)=2$. Por lo tanto,

$$g(\mu)\geq 2 \quad \text{para } \frac{1}{2}\leq\mu\leq 1.$$

Si $0 < \mu < 1/2$, $g(\mu) = \frac{1}{1-2\mu} \ln \frac{1-\mu}{\mu}$. Luego, $g(\mu) \geq 2$ si, y sólo si $\frac{1}{1-2\mu} \ln \frac{1-\mu}{\mu} \geq 2$. Despejando, encontramos que

$$\frac{1}{\mu} - 1 \geq e^{2(1-2\mu)}$$

lo cual es cierto para $0 < \mu < 1/2$. Por lo tanto,

$$\begin{aligned} \mathbb{P}(\bar{S} - \mu \geq \varepsilon) &\leq \left[\left(\frac{\mu}{\mu + \varepsilon} \right)^{\mu+\varepsilon} \left(\frac{1-\mu}{1-\mu-\varepsilon} \right)^{1-\mu-\varepsilon} \right]^n \\ &= e^{-n\varepsilon^2 G(\varepsilon,\mu)} \leq e^{-n\varepsilon^2 g(\mu)} \leq e^{-2n\varepsilon^2}. \end{aligned}$$

□

La desigualdad de Hoeffding, se puede extender al caso en que las cotas para las $X_i's$ varían.

2.10.4. Lema. ***(Hoeffding).*** *Sean $X_1, \ldots, X_n$ v.a.i. con $a_i \leq X_i \leq b_i$. Entonces, para toda $\varepsilon > 0$,*

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq \exp\left[-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2} \right].$$

Demostración: Sea $\mu_i = \mathbb{E}X_i$ para todo $i = 1, \ldots, n$. Aplicando el lema 1.10.1, y por la hipótesis de independencia se tiene

$$\begin{aligned} \mathbb{P}(\bar{S} - \mu \geq \varepsilon) &\leq e^{-hn\varepsilon - h\mathbb{E}(S)} \prod_{i=1}^n \mathbb{E}e^{hX_i} \\ &= e^{-hn\varepsilon} e^{-h\mathbb{E}(S)} \prod_{i=1}^n \mathbb{E}e^{hX_i} \\ &= e^{-hn\varepsilon} \prod_{i=1}^n e^{-h\mathbb{E}X_i} \prod_{i=1}^n \mathbb{E}e^{hX_i} \\ &= e^{-hn\varepsilon} \prod \mathbb{E}e^{h(X_i - \mathbb{E}X_i)}. \end{aligned}$$

Ahora, por el lema 2.10.2,

$$\begin{aligned} \mathbb{E}e^{h(X_i - \mu_i)} &= e^{-h\mu_i} \mathbb{E}e^{hX_i} \\ &\leq e^{-h\mu_i} \left(\frac{b_i - \mu_i}{b_i - a_i} e^{ha_i} + \frac{\mu_i - a_i}{b_i - a_i} e^{hb_i} \right) \quad \text{para todo } i = 1, \ldots, n. \end{aligned}$$

Definamos

$$L(h_i) := -h_i p_i + \ln(1 - p_i + p_i e^{h_i}).$$

donde

$$h_i = h(b_i - a_i) \quad \text{y} \quad p_i = \frac{\mu_i - a_i}{b_i - a_i}.$$

Luego,

$$\begin{aligned}
e^{L(h_i)} &= e^{-h_i p_i}(1 - p_i + p_i e^{h_i}) \\
&= e^{-h(b_i - a_i)\frac{\mu_i - a_i}{b_i - a_i}} \left(1 - \frac{\mu_i - a_i}{b_i - a_i} + \frac{\mu_i - a_i}{b_i - a_i} e^{h(b_i - a_i)}\right) \\
&= e^{-h(\mu_i - a_i)} \left(\frac{b_i - a_i - \mu_i + a_i}{b_i - a_i} + \frac{\mu_i - a_i}{b_i - a_i} e^{h(b_i - a_i)}\right) \\
&= e^{-h\mu_i} \left(\frac{b_i - \mu_i}{b_i - a_i} e^{ha_i} + \frac{\mu_i - a_i}{b_i - a_i} e^{hb_i}\right).
\end{aligned}$$

Así que

$$\mathbb{E} e^{h(X_i - \mu_i)} \leq e^{L(h_i)}.$$

Observemos que

$$\begin{aligned}
L'(h_i) &= -p_i + \frac{p_i e^{h_i}}{1 - p_i + p_i e^{h_i}} = -p_i + \frac{p_i e^{h_i}}{[(1 - p_i)e^{-h_i} + p_i]e^{h_i}} \\
&= -p_i + \frac{p_i}{(1 - p_i)e^{-h_i} + p_i},
\end{aligned}$$

y

$$L''(h_i) = \frac{p_i(1 - p_i)e^{-h_i}}{[(1 - p_i)e^{-h_i} + p_i]^2}.$$

Luego, $L''(h_i)$ es de la forma $u(1 - u)$ con $0 < u < 1$, en efecto,

Sea

$$u = \frac{p_i}{(1 - p_i)e^{-h_i} + p_i}.$$

Entonces, $0 < u < 1$ y

$$\begin{aligned}
u(1 - u) &= \frac{p_i}{(1 - p_i)e^{-h_i} + p_i} \left(1 - \frac{p_i}{(1 - p_i)e^{-h_i} + p_i}\right) \\
&= \frac{p_i}{(1 - p_i)e^{-h_i} + p_i} \left(\frac{(1 - p_i)e^{-h_i}}{(1 - p_i)e^{-h_i} + p_i}\right) \\
&= \frac{p_i(1 - p_i)e^{-h_i}}{[(1 - p_i)e^{-h_i} + p_i]^2}.
\end{aligned}$$

Luego, si $f(u) = u(1-u)$; $0 < u < 1$, entonces $f'(u) = (1-u) + u(-1) = 1 - 2u = 0$ si, y sólo si $u = 1/2$. Además $f''(u) = -2 < 0$. Por lo tanto, $f(u)$ alcanza su máximo en $u = 1/2$. Como $f(1/2) = 1/4$, se sigue que

$$L''(h_i) \leq \frac{1}{4}.$$

Por otra parte, $L(0) = 0$, $L'(0) = 0$ y $L''(\xi_i) \leq \frac{1}{4}$ p.a. $0 < \xi_i < h_i$. Luego entonces

$$\begin{aligned} L(h_i) &= L(0) + L'(0)h_i + \frac{L''(\xi_i)h_i^2}{2!} \quad \text{donde } 0 < \xi_i < h_i \\ &\leq \frac{1}{4}\frac{h_i^2}{2} = \frac{1}{8}h^2(b_i - a_i)^2. \end{aligned}$$

Por lo tanto,

$$\mathbb{E}e^{h(X_i - \mu_i)} \leq e^{L(h_i)} \leq e^{\frac{1}{8}h^2(b_i - a_i)^2},$$

y

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq \exp\left[-hn\varepsilon + \frac{1}{8}h^2 \sum_{i=1}^{n}(b_i - a_i)^2\right].$$

Por último, si definimos $g(h) = -hn\varepsilon + \frac{1}{8}h^2 \sum_{i=1}^{n}(b_i - a_i)^2$, encontramos que

$$h_0 = \frac{4n\varepsilon}{\sum_{i=1}^{n}(b_i - a_i)^2},$$

es un mínimo para $g(h)$. Evaluando este valor en g, obtenemos

$$\begin{aligned} g(h_0) &= \exp\left[-\frac{4n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2} + \frac{2n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right] \\ &= \exp\left[-2n^2\varepsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2\right]. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(\bar{S} - \mu \geq \varepsilon) \leq \exp\left[-2n^2\varepsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2\right].$$

□

2.11. La función característica empírica

Sean $X_1, X_2, \ldots$ v.a.i.i.d. con función de distribución F. Denotemos con

$$\psi(t) \equiv \int e^{itx} dF(x), \quad t \in \mathbb{R},$$

la **función característica** de F. La función característica empírica está definida por

$$\psi_n(t) \equiv \int e^{itx} dF_n(x), \quad t \in \mathbb{R}.$$

El siguiente resultado, muestra la convergencia uniforme entre ψ_n y ψ.

2.11.1. Teorema. *Sea $\mathbb{E}\,|X_1| < \infty$. Entonces, para cada $a > 0$ finito,*

$$\sup_{|t|\leq a} |\psi_n(t) - \psi(t)| \to 0 \quad c.s.\ [\mathbb{P}].$$

Demostración: Sea t de la forma $\pm k/n$, donde $k \leq [na]$ y $\varepsilon > 0$. Como $\mathbb{E}[\psi_n(t)] = \psi(t)$ entonces, por la desigualdad de Hoeffding.

$$\mathbb{P}(|\psi_n(t) - \psi(t)| \geq \varepsilon) \leq c_1 \exp[-c_2\varepsilon^2 n].$$

Por lo que

$$\sum_{n=1}^{\infty} \mathbb{P}(|\psi_n(t) - \psi(t)| \geq \varepsilon) \leq \sum_{n=1}^{\infty} c_1 \exp[-c_2\varepsilon^2 n].$$

Observemos que

$$\sum_{n=1}^{\infty} e^{-c_2\varepsilon^2 n} = \sum_{n=1}^{\infty} \left(e^{-c_2\varepsilon^2}\right)^n < \infty \quad \text{puesto que } 0 < e^{-c_2\varepsilon^2} < 1.$$

Por lo tanto,

$$\sum_{n=1}^{\infty} \mathbb{P}(|\psi_n(t) - \psi(t)| \geq \varepsilon) < \infty.$$

Aplicando Borel-Cantelli, obtenemos

$$\mathbb{P}(\limsup_{n\to\infty} |\psi_n(t) - \psi(t)| \geq \varepsilon) = 0 \quad \text{para toda } \varepsilon > 0.$$

Se sigue que

$$\mathbb{P}(\limsup_{n\to\infty} |\psi_n(\pm k/n) - \psi(\pm k/n)| \geq \varepsilon) = 0 \quad \text{para toda } \varepsilon > 0.$$

Por lo tanto,

$$\sup_{0\le k\le [na]} |\psi_n(\pm k/n) - \psi(\pm k/n)| \to 0 \quad \text{c.s. } [\mathbb{P}].$$

Ahora, para una t arbitraria (digamos $t > 0$), existe $k \le [na]$ tal que $|t - k/n| < \frac{1}{n}$. Observemos que

$$\begin{aligned} |\psi(t) - \psi(k/n)| &= \left|\int (e^{itx} - e^{i\frac{k}{n}x}) dF(x)\right| \\ &\le \int \left|e^{itx} - e^{i\frac{k}{n}x}\right| dF(x). \end{aligned}$$

Por otro lado,

$$e^{itx} = \sum_{j=0}^{\infty} \frac{(itx)^j}{j!} = 1 + itx + R_1$$

y

$$e^{i\frac{k}{n}x} = \sum_{j=0}^{\infty} \frac{(i\frac{k}{n}x)^j}{j!} = 1 + i\frac{k}{n}x + R_2.$$

Entonces,

$$\begin{aligned} \left|e^{itx} - e^{i\frac{k}{n}x}\right| &= \left|itx - i\frac{k}{n}x\right| + (R_1 - R_2) \\ &= |x|\,|t - k/n| + 0(1/n). \end{aligned}$$

Por lo tanto,

$$|\psi(t) - \psi(k/n)| \le c \int |x|\, dF(x)/n = 0(1/n).$$

Analogamente

$$|\psi_n(t) - \psi_n(k/n)| \le c \int |x|\, dF_n(x)/n = 0(1/n).$$

Luego,

$$\begin{aligned} |\psi_n(t) - \psi(t)| &\le |\psi_n(t) - \psi_n(k/n)| + |\psi_n(k/n) - \psi(k/n)| + |\psi(k/n) - \psi(t)| \\ &= 0(1/n) + |\psi_n(k/n) - \psi(k/n)| + 0(1/n) \\ &= 0(1/n) + |\psi_n(k/n) - \psi(k/n)|. \end{aligned}$$

Entonces

$$\sup_{|t|\le a} |\psi_n(t) - \psi(t)| \to 0 \quad \text{cuando } n \to \infty \text{ c.s. } [\mathbb{P}].$$

□

Capítulo 3

Estimación no paramétrica

3.1. Estimación de la densidad

En esta sección daremos una breve introducción a un estimador empírico f_n de la verdadera densidad f. Este estimador fue propuesto primeramente por Rosenblatt (1956) y Parzen (1962), ver [10] y [9] respectivamente. Sean $X_1, \ldots, X_n$ una muestra aleatoria con función de distribución F y densidad $f = F'$. Notemos que

$$\lim_{a \downarrow 0} \frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a}$$

$$\begin{aligned}
&= \lim_{a \downarrow 0} \frac{F(x+\frac{a}{2}) - F(x) - (F(x-\frac{a}{2}) - F(x))}{a} \\
&= \lim_{a \downarrow 0} \frac{\frac{1}{2}(F(x+\frac{a}{2}) - F(x))}{a/2} + \frac{\frac{1}{2}(F(x-\frac{a}{2}) - F(x))}{-a/2} \\
&= \frac{1}{2}\left[\lim_{a \downarrow 0} \frac{F(x+\frac{a}{2}) - F(x)}{a/2} + \lim_{a \downarrow 0} \frac{F(x-\frac{a}{2}) - F(x)}{-a/2}\right] \\
&= \frac{1}{2}[F'(x+) + F'(x-)] \\
&= \frac{1}{2} 2F'(x) \\
&= F'(x).
\end{aligned}$$

Por lo tanto,

$$f(x) = F'(x) = \lim_{a \downarrow 0} \frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a}.$$

Podemos considerar $a > 0$, como una banda fija y estimar el cociente del lado derecho por
$$f_n(x) = \frac{F_n(x+\frac{a}{2}) - F_n(x-\frac{a}{2})}{a}.$$
En primera instancia, debido a que F_n no es diferenciable en los datos, no es posible tomar el límite de este cociente cuando $a \to 0$. También observemos que
$$\mathbb{E}f_n(x) = \frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a}$$
y, $Varf_n(x)$
$$= \frac{1}{a^2}\left[VarF_n(x+\frac{a}{2}) + VarF_n(x-\frac{a}{2}) - 2Cov(F_n(x+\frac{a}{2}), F_n(x-\frac{a}{2}))\right].$$
Recordemos que
$$VarF_n(x+\frac{a}{2}) = \frac{1}{n}F(x+\frac{a}{2})(1-F(x+\frac{a}{2})),$$
$$VarF_n(x-\frac{a}{2}) = \frac{1}{n}F(x-\frac{a}{2})(1-F(x-\frac{a}{2})),$$
y
$$Cov(F_n(x+\frac{a}{2}), F_n(x-\frac{a}{2})) = \frac{1}{n}F(x-\frac{a}{2}) - \frac{1}{n}F(x+\frac{a}{2})F(x-\frac{a}{2}).$$
Sustituyendo estos valores en $Varf_n(x)$, se tiene
$$\begin{aligned} Varf_n(x) &= \frac{1}{na^2}\left(F(x+\frac{a}{2}) - F(x-\frac{a}{2})\right)\left(1 - F(x+\frac{a}{2}) + F(x-\frac{a}{2})\right) \\ &= \frac{1}{na}\left(\frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a}\right)\left(1 - F(x+\frac{a}{2}) + F(x-\frac{a}{2})\right) \\ &\sim \frac{f(x)}{na}. \end{aligned}$$
De la expresión para $\mathbb{E}f_n(x)$, vemos que $f_n(x)$ resulta ser un **estimador sesgado** de $f(x)$. Definamos
$$Sesgof_n(x) = \mathbb{E}f_n(x) - f(x).$$
Si f es dos veces continuamente diferenciable, la fórmula de Taylor dá:
$$Sesgof_n(x) = \frac{a^2}{24}f''(x) + 0(a^4),$$

en efecto, expandiendo $F(y)$ alrededor de x se tiene

$$F(y) = F(x) + F^{'}(x)(y-x) + \frac{F^{''}(x)(y-x)^2}{2!} + \frac{F^{'''}(x)(y-x)^3}{3!} + \dots$$

Entonces,

$$F(x+\frac{a}{2}) = F(x) + f(x)(\frac{a}{2}) + \frac{f^{'}(x)(\frac{a}{2})^2}{2!} + \frac{f^{''}(x)(\frac{a}{2})^3}{3!} + \frac{f^{'''}(x)(\frac{a}{2})^4}{4!} + \dots$$

y

$$F(x-\frac{a}{2}) = F(x) - f(x)(\frac{a}{2}) + \frac{f^{'}(x)(\frac{a}{2})^2}{2!} - \frac{f^{''}(x)(\frac{a}{2})^3}{3!} + \frac{f^{'''}(x)(\frac{a}{2})^4}{4!} + \dots$$

Entonces

$$F(x+\frac{a}{2}) - F(x-\frac{a}{2}) = af(x) + \frac{2f^{''}(x)a^3}{2^3 3!} + ca^5.$$

Por lo tanto

$$\frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a} = f(x) + \frac{f^{''}(x)a^2}{24} + ca^4.$$

Finalmente,

$$\mathbb{E}f_n(x) - f(x) = \frac{F(x+\frac{a}{2}) - F(x-\frac{a}{2})}{a} - f(x) = \frac{a^2}{24}f^{''}(x) + 0(a^4).$$

Una medida global entre la desviación entre f_n y f en x, es el **error cuadrático medio**

$$ECMf_n(x) = Sesgo^2 f_n(x) + Var f_n(x).$$

Ignorando los terminos del error, obtenemos

$$ECMf_n(x) = \frac{a^4}{(24)^2}[f^{''}(x)]^2 + \frac{f(x)}{na} := c_1 a^4 + \frac{c_2}{na} := h(a).$$

Luego,

$$h^{'}(a) = 4c_1 a^3 - \frac{c_2}{na^2} = 0 \text{ si y sólo si } \frac{4c_1 na^5 - c_2}{na^2} = 0.$$

Despejando a en esta igualdad obtenemos

$$a = \left(\frac{c_2}{4nc_1}\right)^{1/5} := cn^{-1/5}.$$

Como
$$h''(a) = 12c_1a^2 + \frac{2c_2}{na^3} > 0.$$
se sigue que $ECMf_n(x)$ se minimiza en $a = cn^{-1/5}$.

El problema de esto, es que la constante c depende de f. Por eso, el valor de a es desconocido.

Una posible solución a esto, es parametrizar la longitud de banda como $tn^{-1/5}$ y estudiar a f_n como un proceso estocástico indexado por t. En esta tesis no se discutirá este hecho.

Veamos que al menos hay una cota para el error estocástico. Para cualquier longitud de banda $a = a_n$ que tiende a cero,
$$\sqrt{na_n}[f_n(x) - \mathbb{E}f_n(x)] = a_n^{-1/2}[\alpha_n(x + \frac{a_n}{2}) - \alpha_n(x - \frac{a_n}{2})].$$
Asi que el valor absoluto, está acotado uniformemente en x por
$$a_n^{-1/2}\omega_n(a_n) = 0\left(\sqrt{\ln a_n^{-1}}\right).$$
Luego, para $a = tn^{-1/5}$,
$$\sup_x |f_n(x) - \mathbb{E}f_n(x)| = 0\left(\frac{\sqrt{\ln n}}{n^{2/5}}\right).$$
f_n se puede extender a una clase de estimadores más generales. Definamos
$$K(y) = 1_{\{-\frac{1}{2}\leq y<\frac{1}{2}\}}, \quad y \in \mathbb{R}.$$
Entonces se tiene,
$$f_n(x) = \frac{1}{a}\int K\left(\frac{x-y}{a}\right)dF_n(y) \tag{3.1}$$
en efecto:
$$K\left(\frac{x-y}{a}\right) = 1_{\{x-\frac{a}{2}<y\leq x+\frac{a}{2}\}}.$$
Luego,
$$\frac{1}{a}\int K\left(\frac{x-y}{a}\right)dF_n(y)$$

$$\begin{aligned}
&= \frac{1}{a}\int 1_{\{x-\frac{a}{2}<y\leq x+\frac{a}{2}\}}dF_n(y)\\
&= \frac{1}{a}\int \left(1_{(-\infty,x+\frac{a}{2}]}(y) - 1_{(-\infty,x-\frac{a}{2}]}(y)\right) dF_n(y)\\
&= \frac{1}{a}\int 1_{(-\infty,x+\frac{a}{2}]}(y)dF_n(y) - \frac{1}{a}\int 1_{(-\infty,x-\frac{a}{2}]}(y)dF_n(y)\\
&= \frac{1}{a}F_n(x+\frac{a}{2}) - \frac{1}{a}F_n(x-\frac{a}{2})\\
&= f_n(x).
\end{aligned}$$

La función K es no negativa y tiene la integral de Lebesgue

$$\int K(y)dy = 1.$$

La ecuación (3.1) también se puede aplicar a otras densidades K. La función K se llama el **Kernell** y el resultante, f_n, es un **estimador del Kernell**.

3.2. Regresión no paramétrica

Sea (X, Y) un vector aleatorio en $\mathbb{R}^{d+1}$. El modelo clásico que mide el grado de dependencia entre X y Y es

$$Y = X^T\beta + \varepsilon,\ \mathbb{E}(\varepsilon \mid X) = 0,\ \ \text{Var}\varepsilon = \sigma^2 < \infty \tag{3.2}$$

donde β es el parámetro de interés. Cuando se asume que ε sigue una distribución normal, se deducen muchas cantidades de interés. Por ejemplo, se puede calcular explícitamente β por mínimos cuadrados.

Observemos que

$$\begin{aligned}
\mathbb{E}[Y \mid X] &= \mathbb{E}[X^T\beta \mid X] + \mathbb{E}[\varepsilon \mid X]\\
&= X^T\beta,
\end{aligned}$$

es decir,

$$\mathbb{E}[Y \mid X = x] = x^T\beta.$$

Entonces, para conocer β, es equivalente a conocer la **función de regresión**

$$m(x) = \mathbb{E}[Y \mid X = x].$$

Se sabe de la existencia de m, debido al Teorema de Radon-Nikodym. En regresión no paramétrica, la estimación de m, es basada en una muestra i.i.d.,

sin hacer suposición alguna, como en el modelo (3.2). Así, sean (X_i, Y_i), $1 \leq i \leq n$ una muestra i.i.d. provenientes de la misma distribución que (X, Y).
Para la estimación de m, supongamos que x es un X-átomo, i.e., $\mathbb{P}(X = x) > 0$. Para este caso, $m(x)$ está dada por

$$m(x) = \frac{1}{\mathbb{P}(X = x)} \int_{\{X=x\}} Y d\mathbb{P} \tag{3.3}$$

Oservemos que

$$\mathbb{E}[Y_1 1_{\{X_1=x\}}] = \int_{\{X_1=x\}} Y_1 d\mathbb{P} \quad \text{y} \quad \mathbb{E}[1_{\{X_1=x\}}] = \int_{\{X_1=x\}} d\mathbb{P}.$$

Entonces, la LFGN implica que con probabilidad uno

$$n^{-1} \sum_{i=1}^{n} Y_i 1_{\{X_i=x\}} \to \int_{\{X=x\}} Y d\mathbb{P}$$

y

$$n^{-1} \sum_{i=1}^{n} 1_{\{X_i=x\}} \to \mathbb{P}(X = x).$$

Así, definamos

$$m_n(x) = \frac{\sum_{i=1}^n Y_i 1_{\{X_i=x\}}}{\sum_{i=1}^n 1_{\{X_i=x\}}} = \frac{\sum_{i=1}^n Y_i 1_{\{x\}}(X_i)}{\sum_{i=1}^n 1_{\{x\}}(X_i)}.$$

Por el Teorema de Slutsky, ver [12], obtenemos

$$m_n(x) \to m(x) \quad \text{con probabilidad uno.}$$

Si x no es un X-átomo, no se puede aplicar (3.3). Para ello, cuando $d = 1$, se propone reemplazar al conjunto con un solo punto $\{x\}$ por un intervalo $(x - a_n, x + a_n]$. En este caso,

$$m_n(x) = \frac{\sum_{i=1}^n Y_i 1_{\{x-a_n<X_i\leq x+a_n\}}}{\sum_{i=1}^n 1_{\{x-a_n<X_i\leq x+a_n\}}} \tag{3.4}$$

$$(= 0 \quad \text{si el denominador es cero}).$$

$a_n > 0$ se considera nuevamente una longitud de banda, que tiende a cero conforme n crece. Definamos

$$K(y) = 1_{[-1,1)}(y).$$

Entonces,

$$K\left(\frac{x-X_i}{a_n}\right) = 1_{\{x-a_n<X_i\leq x+a_n\}}.$$

Entonces, la expresión de m_n de (3.4), se puede reescribir como

$$m_n = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{a_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right)} \tag{3.5}$$

Notemos que que en este caso, no se requiere que $\int K(u)du = 1$, como en el caso de la estimación de la densidad. La expresión (3.5) se puede extender a Kernels K, más generales.

Observemos que del numerador y denominador de (3.5) se tiene:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{na_n}\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{a_n}\right)\right] &= \mathbb{E}\left[\frac{1}{a_n}\frac{1}{n}\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{a_n}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{a_n} Y K\left(\frac{x-X}{a_n}\right)\right] \\
&= \mathbb{E}\Big\{\mathbb{E}\left[\frac{1}{a_n} Y K\left(\frac{x-X}{a_n}\right) \mid X\right]\Big\} \\
&= \mathbb{E}\Big\{\frac{1}{a_n} K\left(\frac{x-X}{a_n}\right)\mathbb{E}[Y \mid X]\Big\} \\
&= \mathbb{E}\Big\{\frac{1}{a_n} m(X) K\left(\frac{x-X}{a_n}\right)\Big\}.
\end{aligned}$$

La segunda igualdad se debe a que X, Y tiene la misma distribución que X_i y Y_i, para toda $i = 1, \ldots, n$. La tercera igualdad es por que $\frac{1}{a_n} Y K\left(\frac{x-X}{a_n}\right)$ es integrable, esto es:

$$\int \left|\frac{1}{a_n} Y K\left(\frac{x-X}{a_n}\right)\right| dP \leq \frac{1}{a_n}\int \mid Y \mid dP < \infty.$$

La cuarta igualda, se debe a que $\frac{1}{a_n} K\left(\frac{x-X}{a_n}\right)$ es X-medible.

Asumiendo que X tiene una densidad f, la última expresión anterior se escribe como

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{a_n} m(y) K\left(\frac{x-y}{a_n}\right) f(y)dy &= a_n^{-1}\int_{-\infty}^{\infty} m(y) f(y) K\left(\frac{x-y}{a_n}\right) dy \\
&= -\int_{-\infty}^{\infty} m(x-a_n u) f(x-a_n u) K(u) du.
\end{aligned}$$

Cuando $a_n \to 0$, y bajo condiciones de regularidad para f y m, se tiene que la última integral converge a

$$m(x)f(x)\int K(u)du.$$

Por otro lado,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{na_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{a_n}\right)\right] &= \mathbb{E}\left[\frac{1}{a_n}K\left(\frac{x-X}{a_n}\right)\right] \\
&= \int_{-\infty}^{\infty}\frac{1}{a_n}K\left(\frac{x-y}{a_n}\right)f(y)dy \\
&= a_n^{-1}\int_{-\infty}^{\infty} f(y)K\left(\frac{x-y}{a_n}\right)dy \\
&= \int_{-\infty}^{\infty} f(x-a_n u)k(u)du.
\end{aligned}$$

Bajo las mismas condiciones que antes, esta última integral converge a

$$1.f(x)\int K(u)du.$$

Luego, si $f(x) > 0$, el cociente de las dos expresiones anteriores, converge a $m(x)$ como se deseaba.

Desde otro punto de vista, m_n, es motivada por el hecho de suponer que (X, Y) tiene una densidad bivariada g, como sigue

$$\mathbb{E}[Y \mid X] = m(x) = \frac{\int yg(x,y)dy}{f(x)} \tag{3.6}$$

donde

$$f(x) = \int g(x,y)dy$$

es la densidad marginal de X. Denotemos

$$H_n(u,v) = n^{-1}\sum_{i=1}^{n} 1_{\{X_i\le u, Y_i\le v\}}, \quad u, v \in \mathbb{R}.$$

a la función de distribución empírica bivariada de (X_i, Y_i), $1 \le i \le n$. Observemos que

$$\begin{aligned}
g_n(x,y) &= a_n^{-2}\int\int K\left(\frac{x-u}{a_n}\right)K\left(\frac{y-v}{a_n}\right)dH_n(u,v) \\
&= a_n^{-2}\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{a_n}\right)K\left(\frac{y-Y_i}{a_n}\right).
\end{aligned}$$

g_n constituye la extención para el estimador del Kernel de la densidad al caso bivariado, con Kernel

$$K_0(u,v) = K(u)K(v).$$

Ahora,

$$\begin{aligned}
\int y g_n(x,y)dy &= a_n^{-2}\int\int\int yK\left(\frac{x-u}{a_n}\right)K\left(\frac{y-v}{a_n}\right)dH_n(u,v)dy \\
&= a_n^{-2}\int\int\int yK\left(\frac{x-u}{a_n}\right)K\left(\frac{y-v}{a_n}\right)dydH_n(u,v) \\
&= a_n^{-1}\int\int\int (a_n w + v)K\left(\frac{x-u}{a_n}\right)K(w)dwdH_n(u,v) \\
&= a_n^{-1}\int\int K\left(\frac{x-u}{a_n}\right)\int (a_n w + v)K(w)dwdH_n(u,v) \\
&= a_n^{-1}\int\int K\left(\frac{x-u}{a_n}\right)(0+v)dH_n(u,v) \\
&= a_n^{-1}\int\int vK\left(\frac{x-u}{a_n}\right)dH_n(u,v) \\
&= a_n^{-1}\frac{1}{n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{a_n}\right).
\end{aligned}$$

En la segunda igualdad se aplica Fubini, y en la quinta igualdad se considera el hecho de que $\int K(w)dw = 1$ y $\int wK(w)dw = 0$. Luego, si se sustituye en (3.6) a g y f, por g_n y

$$f_n(x) = a_n^{-1}\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{a_n}\right)$$

respectivamente, entonces se obtiene

$$m_n \equiv \frac{\int y g_n(x,y)dy}{f_n(x)},$$

la expresión de (3,5), como se esperaba.

Definamos

$$W_{in}(x) := \frac{K\left(\frac{x-X_i}{a_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{a_n}\right)}.$$

Entonces

$$m_n(x) = \sum_{i=1}^{n} W_{in}(x)Y_i \tag{3.7}$$

3.3. Consistencia en regresión no paramétrica

Sea $(X, Y) \in \mathbb{R}^{d+1}$, donde $d \geq 1$ y $\|\cdot\|$ denota alguna norma sobre $\mathbb{R}^d$. Para x y y en $\mathbb{R}$, $x \vee y$ y $x \wedge y$ denota el máximo y mínimo de x y y respectivamente.

$$L_r = L_r(\Omega, \mathcal{F}, P) = \{f : \Omega \to \mathbb{R} \mid f \text{ es medible y } \int \mid f \mid^r dP < \infty\}.$$

Todas las variables aleatorias que se consideren se asumen estar definidas sobre el espacio de probabilidad $(\Omega, \mathcal{F}, P)$.

3.3.1. Definición. *Se dice que una sucesión $\{m_n\}$ es consistente en L_r, $r \geq 1$, si*

$$m_n(X) \to m(X) \quad \textit{en } L_r,$$

i.e.,

$$\int \mid m_n(x) - m(x) \mid^r d\mu(x) \to 0.$$

Aquí μ es la distribución de X y $m(x) = \mathbb{E}[Y \mid X = x]$ es la veradera función de regresión. Asumiremos que $\mathbb{E} \mid Y \mid^r < \infty$.

En esta sección se probará la consistencia de $m_n(x) = \sum_{i=1}^{n} W_{in}(x)Y_i$, donde la función de peso W_n, es de la forma

$$W_{in}(x) = W_{in}(x, X_1, \ldots, X_n), \quad 1 \leq i \leq n.$$

La función de peso W_n, se llama función de peso de probabilidad, si es no negativa y $\sum_{i=1}^{n} W_{in}(x) = 1$. Consideremos las siguientes condiciones

(i)

$$\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid f(X_i)\right] \leq C\mathbb{E}f(X),$$

para toda $f \geq 0$, $n \geq 1$ y p.a. $C \geq 1$.

(ii) Para alguna $D \geq 1$,

$$\mathbb{P}\left(\sum_{i=1}^{n} \mid W_{in}(X) \mid \leq D\right) = 1.$$

(iii)

$$\sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{\|X_i - X\| > \varepsilon\}} \to 0 \quad \text{en probabilidad, para cada } \varepsilon > 0.$$

(iv)

$$\sum_{i=1}^{n} W_{in}(X) \to 1 \quad \text{en probabilidad.}$$

(v)

$$\max_{1 \leq i \leq n} \mid W_{in}(X) \mid \to 0 \quad \text{en probabilidad.}$$

3.3.2. Teorema. ***(Stone)***, *ver* [13]. *Bajo las condiciones (i)-(v),*

$$m_n(X) \to m(X) \quad en\ L_r.$$

Para probar el presente teorema, necesitaremos una serie de lemas.

3.3.3. Lema. *Bajo las condiciones (i)-(iii), y asumiendo que* $\mathbb{E} \mid f(X) \mid^r < \infty$,

$$\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid^r\right] \to 0.$$

Demostración: Sea $\varepsilon > 0$ y h, una función continua sobre $\mathbb{R}^d$ con soporte compacto, tal que

$$\mathbb{E} \mid f(X) - h(X) \mid^r \leq \varepsilon.$$

Esto es posible, debido a la densidad entre las funciones r-integrables y las funciones continuas con soporte compacto, ver [7].

Como $\mid f(X_i) - h(X_i) \mid^r \geq 0$, aplicando la condición (i), obtenemos

$$\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - h(X_i) \mid^r\right] \leq C\mathbb{E} \mid f(X) - h(X) \mid^r \leq C\varepsilon.$$

Por (ii),

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X) - h(X) \mid^r\right] &\leq \mathbb{E}[D \mid f(X) - h(X) \mid^r] \\ &= D\mathbb{E} \mid f(X) - h(X) \mid^r \\ &= D\varepsilon.\end{aligned}$$

Esto muestra que sólo necesitamos probar el lema para funciones continuas f, con soporte compacto. Así, sea $M = \|f\|_\infty$. Puesto que f es uniformemente continua, dada $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\|x - x_1\| \leq \delta \quad \text{implica} \quad | f(x) - f(x_1) |^r \leq \varepsilon.$$

Ahora, como

$$\begin{aligned}
&\sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r \\
= & \sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r 1_{\{\|X_i - X\| > \delta\}} \\
+ & \sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r 1_{\{\|X_i - X\| \leq \delta\}}.
\end{aligned}$$

Aplicando esperanza, la condición (ii) y debido a la independencia, se tiene

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r\right] \\
= & \ \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r 1_{\{\|X_i - X\| > \delta\}}\right] \\
+ & \ \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r 1_{\{\|X_i - X\| \leq \delta\}}\right] \\
\leq & \ \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | (| f(X_i) | + | f(X) |)^r 1_{\{\|X_i - X\| > \delta\}}\right] \\
+ & \ \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | \varepsilon\right] \\
\leq & \ \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | (2M)^r 1_{\{\|X_i - X\| > \delta\}}\right] + \varepsilon D \\
= & \ (2M)^r \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\|X_i - X\| > \delta\}}\right] + \varepsilon D.
\end{aligned}$$

Luego,

$$\mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\|X_i - X\| > \delta\}}\right] \to 0 \quad \text{cuando } n \to \infty.$$

en efecto:

(1) Por (iii),

$$\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\|X_i - X\| > \delta\}} \to 0 \quad \text{en probabilidad.}$$

(2) Por (ii),

$$\sum_{i=1}^{n} | W_{in}(X) | \leq D \quad \text{p.a. } D \geq 1 \text{ con probabilidad } 1,$$

se sigue que

$$\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\|X_i - X\| > \delta\}} \leq D \quad \text{con probabilidad 1.}$$

Aplicando el Teorema de convergencia de Lebesgue a (1) y (2):

$$\mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\|X_i - X\| > \delta\}}\right] \to 0 \quad \text{cuando } n \to \infty.$$

Finalmente, haciendo tender ε a cero, se concluye que

$$\mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) || f(X_i) - f(X) |^r\right] \to 0 \quad \text{cuando } n \to \infty.$$

□

3.3.4. Lema. *Bajo las condiciones (i)-(iii), sea $\{W_n\}$ una sucesión de pesos no negativos, tal que para constantes no negativos M_n y N_n,*

$$\mathbb{P}\left(M_n \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n\right) \to 1.$$

Sea f una función Borel medible no negativa sobre $\mathbb{R}^d$ tal que $\mathbb{E}f(X) < \infty$. Entonces

$$\liminf_{n\to\infty} \mathbb{E} \sum_{i=1}^{n} W_{in}(X) f(X_i) \geq (\liminf_{n\to\infty} M_n) \mathbb{E}f(X),$$

y

$$\limsup_{n\to\infty} \mathbb{E} \sum_{i=1}^{n} W_{in}(X) f(X_i) \leq (\limsup_{n\to\infty} N_n) \mathbb{E}f(X).$$

Demostración: Definamos

$$A_n = \Big\{ M_n \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n \Big\}.$$

Supongamos s.p.g. que $M_n \leq D$. Entonces

$$M_n - D1_{A_n^c} \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n + D1_{A_n^c},$$

en efecto: Sea $\omega \in \Omega$, entonces $\omega \in A_n$ o $\omega \in A_n^c$.

Si $\omega \in A_n$ entonces $D1_{A_n^c} = 0$ y,

$$M_n - D1_{A_n^c} \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n + D1_{A_n^c}.$$

Si $\omega \in A_n^c$ entonces $D1_{A_n^c} = D \geq M_n$ y,

$$M_n - D1_{A_n^c} \leq 0 \leq \sum_{i=1}^{n} W_{in}(X).$$

Por lo tanto,

$$M_n - D1_{A_n^c} \leq \sum_{i=1}^{n} W_{in}(X).$$

Por otro lado, Como $\omega \in A_n^c$ entonces

$$\sum_{i=1}^{n} W_{in}(X) < M_n \quad \text{o} \quad \sum_{i=1}^{n} W_{in}(X) > N_n.$$

Si $\sum_{i=1}^{n} W_{in}(X) < M_n$ entonces

$$\sum_{i=1}^{n} W_{in}(X) - M_n < 0 \leq N_n + D1_{A_n^c}.$$

Por lo tanto,

$$\sum_{i=1}^{n} W_{in}(X) < N_n + D1_{A_n^c}.$$

Anolagamente, si $\sum_{i=1}^{n} W_{in}(X) > N_n$ entonces

$$\sum_{i=1}^{n} W_{in}(X) < N_n + D1_{A_n^c}.$$

Por lo tanto,

$$M_n - D1_{A_n^c} \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n + D1_{A_n^c}.$$

Entonces,

$$M_n f(X) - D1_{A_n^c} f(X) \leq \sum_{i=1}^{n} W_{in}(X) f(X) \leq N_n f(X) + D1_{A_n^c} f(X).$$

Aplicando esperanza, se obtiene

$$\begin{aligned} M_n \mathbb{E} f(X) - D\mathbb{E}1_{A_n^c} f(X) &\leq \mathbb{E}\left[\sum_{i=1}^{n} W_{in}(X) f(X)\right] \\ &\leq N_n \mathbb{E} f(X) + D\mathbb{E}1_{A_n^c} f(X). \end{aligned}$$

Como

$$\mathbb{P}(M_n \leq \sum_{i=1}^{n} W_{in}(X) \leq N_n) = \mathbb{P}(M_n f(X) \leq \sum_{i=1}^{n} W_{in}(X) f(X) \leq N_n f(X)).$$

entonces

$$\mathbb{P}(M_n f(X) \leq \sum_{i=1}^{n} W_{in}(X) f(X) \leq N_n f(X)) \to 1 \quad \text{cuando } n \to \infty.$$

Por lo tanto

$$\mathbb{E}1_{A_n^c} f(X) \to 0 \quad \text{cuando } n \to \infty.$$

Entonces,

$$\liminf_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} W_{in}(X) f(X)\right] \geq (\liminf_{n\to\infty} M_n) \mathbb{E} f(X)$$

y,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} W_{in}(X) f(X)\right] \leq (\limsup_{n\to\infty} N_n) \mathbb{E} f(X).$$

Luego, como $\mathbb{E} \mid f(X) \mid = \mathbb{E}f(X) < \infty$, aplicando el lema 3.3.3, para $r = 1$ se tiene

$$\mathbb{E} \mid \sum_{i=1}^{n} W_{in}(X)(f(X_i) - f(X)) \mid \leq \mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid \mid f(X_i) - f(X) \mid\right] \to 0.$$

Por lo tanto,

$$\lim_{n\to\infty} \mathbb{E}\sum_{i=1}^{n} W_{in}(X)f(X_i) = \lim_{n\to\infty} \mathbb{E}\sum_{i=1}^{n} W_{in}(X)f(X).$$

Entonces

$$\liminf_{n\to\infty} \mathbb{E}\sum_{i=1}^{n} W_{in}(X)f(X_i) \geq (\liminf_{n\to\infty} M_n)\mathbb{E}f(X),$$

y

$$\limsup_{n\to\infty} \mathbb{E}\sum_{i=1}^{n} W_{in}(X)f(X_i) \leq (\limsup_{n\to\infty} N_n)\mathbb{E}f(X).$$

□

3.3.5. Lema. *Bajo la condiciones (i)-(iii), supongamos que para algunas constantes M_n y N_n,*

$$\mathbb{P}\left(M_n \leq \sum_{i=1}^{n} W_{in}^2(X) \leq N_n\right) \to 1.$$

Entonces, para toda función f μ-integrable no negativa,

$$\liminf_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} W_{in}^2(X)f(X_i)\right] \geq (\liminf_{n\to\infty} M_n)\mathbb{E}f(X)$$

y,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} W_{in}^2(X)f(X_i)\right] \leq (\limsup_{n\to\infty} N_n)\mathbb{E}f(X).$$

Demostración:

(1)

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}^2(X) \mid f(X_i)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid (\mid W_{in}(X) \mid f(X_i))\right] \\
&\leq C\mathbb{E}\left[\mid W_{in}(X) \mid f(X)\right] \\
&\leq C\mathbb{E}[Df(X)] = CD\mathbb{E}f(X).
\end{aligned}
$$

(2) Sabemos que

$$\sum_{i=1}^{n} \mid W_{in}(X) \mid \leq D \quad \text{con probabilidad 1.}$$

Entonces,

$$\sum_{i=1}^{n} \mid W_{in}^2(X) \mid + \sum_{i \neq j} \mid W_{in}(X) W_{in}(X) \mid \leq D^2 \quad \text{con probabilidad uno.}$$

Entonces,

$$\sum_{i=1}^{n} \mid W_{in}^2(X) \mid \leq D^2 \quad \text{con probabilidad uno}$$

y,

$$\mathbb{P}\left(\sum_{i=1}^{n} \mid W_{in}^2(X) \mid \leq D^2\right) = 1.$$

(3) Como los pesos son ≤ 1, entonces

$$\sum_{i=1}^{n} \mid W_{in}^2(X) \mid 1_{\{\|X_i - X\| > \varepsilon\}} \leq \sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{\|X_i - X\| > \varepsilon\}},$$

y esto último converge a cero en probabilidad por la condición (iii), para cada $\varepsilon > 0$.

Así, se cumplen las condiciones (i)-(iii) reemplazando C, D por CD y D^2 respectivamente. Entonces, aplicando el lema 3.3.4 a $\{W_n^2\}$, se concluye la prueba. $\square$

3.3.6. Lema. *Bajo las condiciones (i)-(iii), para toda función Borel medible* f,

$$\sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{f(X_i)-f(X)\mid>\varepsilon\}} \to 0 \quad \textit{en probabilidad, para toda } \varepsilon > 0.$$

Demostración: Sea $\varepsilon > 0$. Supongamos que f está acotada. Entonces existe $M > 0$ tal que $\mid f(X) \mid \leq M$. Se sigue entonces que $\mathbb{E} \mid f(X) \mid < \infty$. Aplicando el lema 3.3.3, se sigue que

$$\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid\right] \to 0,$$

i.e.,

$$\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid\right] = \int_{\Omega} \sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid dP$$

$$\begin{aligned} &= \int_{\{\mid f(X_i)-f(X)\mid>\varepsilon\}} \sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid dP \\ &+ \int_{\{\mid f(X_i)-f(X)\mid\leq\varepsilon\}} \sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid dP \to 0. \end{aligned}$$

Entonces,

$$\int_{\{\mid f(X_i)-f(X)\mid>\varepsilon\}} \sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid dP \to 0.$$

Por otra parte,

$$\varepsilon \int_{\Omega} \sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{f(X_i)-f(X)\mid>\varepsilon\}} dP$$

$$\begin{aligned} &= \varepsilon \int_{\{f(X_i)-f(X)\mid>\varepsilon\}} \sum_{i=1}^{n} \mid W_{in}(X) \mid dP \\ &\leq \int_{\{\mid f(X_i)-f(X)\mid>\varepsilon\}} \sum_{i=1}^{n} \mid W_{in}(X) \mid\mid f(X_i) - f(X) \mid dP. \end{aligned}$$

Por lo tanto,

$$\varepsilon \int_{\Omega} \sum_{i=1}^{n} | W_{in}(X) | 1_{\{f(X_i)-f(X)|>\varepsilon\}} dP \to 0.$$

Se sigue que

$$\sum_{i=1}^{n} | W_{in}(X) | 1_{\{f(X_i)-f(X)|>\varepsilon\}} \to 0 \quad \text{en probabilidad.}$$

Para una f arbitraria, sea $M > 0$ tal que $\mathbb{P}(| f(X) |> M) \leq \delta$. Definamos

$$\bar{f} = (f \wedge M) \vee (-M).$$

Si $| f(X_i) |\leq M$ entonces $\bar{f} = f$. Entonces $\bar{f}(X_i) = f(X_i)$. Se sigue que

$$\{\bar{f}(X_i) \neq f(X_i)\} \subset \{| f(X_i) |> M\}.$$

Luego,

$$\sum_{i=1}^{n} | W_{in}(X) | 1_{\{|f(X_i)-f(X)|>\varepsilon\}}$$

$$\begin{aligned} \leq & \sum_{i=1}^{n} | W_{in}(X) | 1_{\{\bar{f}(X_i)-\bar{f}(X)|>\varepsilon\}} + \sum_{i=1}^{n} | W_{in}(X) | 1_{\{\bar{f}(X_i)\neq f(X_i)\}} \\ + & \sum_{i=1}^{n} | W_{in}(X) | 1_{\{\bar{f}(X)\neq f(X)\}}. \end{aligned}$$

Por un lado,

$$\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\bar{f}(X_i)-\bar{f}(X)|>\varepsilon\}} \to 0 \quad \text{en probabilidad porque } \bar{f} \text{ es acotada.}$$

Por la condición (i)

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{n} | W_{in}(X) | 1_{\{\bar{f}(X_i)\neq f(X_i)\}}\right] &\leq C\mathbb{E}\left[1_{\{\bar{f}(X)\neq f(X)\}}\right] \\ &= C\mathbb{P}(\bar{f}(X) \neq f(X)) \\ &\leq C\mathbb{P}(| f(X) |> M) \leq C\delta. \end{aligned}$$

Por la condición (ii)

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{\bar{f}(X)\neq f(X)\}}\right] &\leq \mathbb{E}\left[D1_{\{\bar{f}(X)\neq f(X)\}}\right] \\
&= D\mathbb{P}(\bar{f}(X)\neq f(X)) \\
&\leq D\mathbb{P}(\mid f(X)\mid > M) \leq D\delta.
\end{aligned}
$$

Haciendo tender δ a cero se cocluye que

$$
\sum_{i=1}^{n} \mid W_{in}(X) \mid 1_{\{\mid f(X_i)-f(X)\mid > \varepsilon\}} \to 0 \quad \text{en probabilidad.}
$$

□

3.3.7. Lema. *Bajo las condiciones (i)-(iv), supongamos que $f(X)\in L_r$, para alguna $r\geq 1$. Entonces,*

$$
\sum_{i=1}^{n} W_{in}(X)f(X_i) \to f(X) \quad \textit{en } L_r.
$$

Demostración: Por la condición (ii),

$$
\begin{aligned}
\mid \sum_{i=1}^{n} W_{in}(X) - 1 \mid^r &\leq \left(\mid \sum_{i=1}^{n} W_{in}(X) \mid +1\right)^r \leq \left(\sum_{i=1}^{n} \mid W_{in}(X) \mid +1\right)^r \\
&\leq (D+1)^r \quad \text{con probalididad uno.}
\end{aligned}
$$

Por otro lado, sabemos de la condición (iv)

$$
\sum_{i=1}^{n} W_{in}(X) \to 1 \quad \text{en probabilidad,}
$$

i.e.,

$$
\mathbb{P}\left(\mid \sum_{i=1}^{n} W_{in}(X) - 1 \mid^r > \delta\right) \to 0.
$$

Ahora veamos que

$$
\mathbb{E}\left|\left[\left(\sum_{i=1}^{n} W_{in}(X)\right) - 1\right] f(X)\right|^r \to 0.
$$

en efecto, sea

$$A = \Big\{ \mid \sum_{i=1}^{n} W_{in}(X) - 1 \mid^r > \delta \Big\}.$$

Entonces

$$0 \leq \mathbb{E}\left| \left[\left(\sum_{i=1}^{n} W_{in}(X) \right) - 1 \right] f(X) \right|^r = \int_{\Omega} \left| \left[\left(\sum_{i=1}^{n} W_{in}(X) \right) - 1 \right] f(X) \right|^r dP$$

$$\begin{aligned}
&= \int_A \left| \left[\left(\sum_{i=1}^{n} W_{in}(X) \right) - 1 \right] f(X) \right|^r dP \\
&+ \int_{A^c} \left| \left[\left(\sum_{i=1}^{n} W_{in}(X) \right) - 1 \right] f(X) \right|^r dP \\
&\leq \int_A (D+1)^r \mid f(x) \mid^r dP + \delta \int_{\Omega} \mid f(X) \mid^r dP \\
&= (D+1)^r \int_A \mid f(x) \mid^r dP + \delta \mathbb{E} \mid f(X) \mid^r \\
&= (D+1)^r \mathbb{E}[\mid f(X) \mid^r 1_A] + \delta \mathbb{E} \mid f(X) \mid^r .
\end{aligned}$$

El primer sumando converge a cero porque $\mathbb{P}(\sum_{i=1}^{n} W_{in}(X) - 1 \mid^r > \delta) \to 0$. Luego, haciendo tender δ a cero, se tiene

$$\mathbb{E}\left| \left[\left(\sum_{i=1}^{n} W_{in}(X) \right) - 1 \right] f(X) \right|^r \to 0 \tag{3.8}$$

Así, aplicando la condición (ii), el lema 3.3.3 y la desigualdad de Hölder, se tiene

$$\mathbb{E}\left| \sum_{i=1}^{n} W_{in}(X)[f(X_i) - f(X)] \right|^r \to 0 \tag{3.9}$$

Es claro que de (3.8) y (3.9), se sigue el resultado. □

Ahora sí se tienen todas las condiciones para probar el teorema 3.3.2

Supongamos que se cumplen las condiciones (i)-(iv), y sea $r \geq 1$. Puesto que $\mathbb{E} \mid Y \mid^r < \infty$ por hipótesis y por la desigualdad de Jensen,

$$\mathbb{E} \mid m(X) \mid^r = \mathbb{E} \mid \mathbb{E}(Y \mid X) \mid^r \leq \mathbb{E}(\mathbb{E} \mid Y \mid^r \mid X) = \mathbb{E} \mid Y \mid^r < \infty.$$

Por lo tanto

$$m(X) \in L_r.$$

Sea

$$Z_i = Y_i - \mathbb{E}[Y_i \mid X_i] = Y_i - m(X_i).$$

Entonces

$$\begin{aligned}\sum_{i=1}^{n} W_{in}(X)Y_i - m(X) &= \left[\sum_{i=1}^{n} W_{in}(X)m(X_i) - m(X)\right] \\ &+ \sum_{i=1}^{n} W_{in}(X)Z_i.\end{aligned}$$

El primer sumando converge a cero en L_r por el lema 3.3.7. Para la segunda suma consideremos el caso $r = 2$. Por definición de m, y por la hipótesis de independencia, se tiene

$$\mathbb{E}\left[\sum_{i=1}^{n} W_{in}(x)Z_i\right]^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[W_{in}(x)W_{jn}(x)Z_iZ_j]$$

$$\begin{aligned} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[\mathbb{E}(W_{in}(x)W_{jn}(x)Z_iZ_j \mid X_1, \ldots, X_n)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[W_{in}(x)W_{jn}(x)\mathbb{E}(Z_iZ_j \mid X_1, \ldots, X_n)] \\ &= \sum_{i=1}^{n} \mathbb{E}[W_{in}^2(x)\mathbb{E}(Z_i^2 \mid X_i)] := \sum_{i=1}^{n} \mathbb{E}(W_{in}^2(x)h(X_i)].\end{aligned}$$

Por lo tanto,

$$\mathbb{E}\left[\sum_{i=1}^{n} W_{in}(X)Z_i\right]^2 = \mathbb{E}\left[\sum_{i=1}^{n} W_{in}^2(X)h(X_i)\right].$$

Por el lema 3.3.5,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} W_{in}^2(X)h(X_i)\right] \leq (\limsup_{n\to\infty} N_n)\mathbb{E}h(X) \tag{3.10}$$

donde N_n es tal que

$$\mathbb{P}\left(\sum_{i=1}^{n} W_{in}^2(X) \leq N_n\right) \to 1.$$

Ahora, por (ii) y (iv),

$$\sum_{i=1}^{n} W_{in}^2(X) \to 0 \quad \text{en probabilidad},$$

i.e., N_n la podemos tomar como $N_n \equiv \delta$, donde $\delta > 0$ es cualquier número real positivo. La expresión (3.10) del lado derecho es ahora $\delta\mathbb{E}h(X)$. Por último, haciendo tender δ a cero, la segunda suma converge a cero y el teorema queda probado. □

Observemos que

$$W_{in}(X) = W_{in}(X; X_1, \ldots X_n)$$

es una función medible para los vectores aleatorios i.i.d. $X, X_1, \ldots, X_n$. Entonces se tiene

$$\mathbb{E}[\mid W_{in}(X) \mid f(X_i)] = \mathbb{E}[\mid W_{in}(X_i; X_1, \ldots, X, \ldots, X_n) \mid f(X)]$$

es por eso que la condición (1) llega a ser

$$\mathbb{E}\left[\left\{\sum_{i=1}^{n} \mid W_{in}(X_i; X_1, \ldots, X, \ldots, X_n) \mid \right\} f(X)\right] \leq C\mathbb{E}f(X),$$

para toda $f \geq 0$. En particular, la condición (i) se satisface siempre que

$$(i^*) \quad \sum_{i=1}^{n} \mid W_{in}(X_i; X_1, \ldots, x, \ldots, X_n) \mid \leq C, \text{ para cada } x \in \mathbb{R}^d$$

se cumpla.

Apéndice A

Conceptos básicos de probabilidad

A.1. Propiedades

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. El conjunto Ω es el evento seguro, también llamado espacio muestral y los elementos de $\mathcal{F}$ se llaman eventos. El simbolo $\emptyset$ denota al conjunto vacío. $A, B, C, \ldots$, con o sin subíndices denotan a los eventos.
Notemos que si $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, entonces A_n^c, $\bigcup_{n=1}^{\infty} A_n$, $\bigcap_{n=1}^{\infty} A_n$, $\liminf_{n\to\infty} A_n$, $\limsup_{n\to\infty} A_n$ (si es que existen), son eventos. La medida de probabilidad P, está definida sobre $\mathcal{F}$, y debe satisfacer para cualesquiera eventos A, A_n,

$$P(A) \geq 0, \quad P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n), \quad A_i \cap A_j = \emptyset \quad \text{para todo } i \neq j.$$

y $P(\Omega) = 1$.

Se siguen inmediatamente la siguientes propiedades:

$$P(\emptyset) = 0, \quad P(A) \leq P(B) \text{ si } A \subset B, \quad P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

El siguiente resultado se conoce como la propiedad de continuidad de P.

A.1.1. Proposición. *Si* $\{E_n\}_{n\geq 1}$ *es una sucesión de eventos creciente, i.e.,* $E_1 \subset E_2 \subset \ldots \subset E = \bigcup_{n=1}^{\infty} E_n$ *o decreciente, i.e.,* $E_1 \supset E_2 \supset \ldots \supset E = \bigcap_{n=1}^{\infty} E_n$*, entonces*

$$\lim_{n\to\infty} P(E_n) = P(E).$$

A.1.2. Definición. *Sean* $A, B \in \mathcal{F}$. *Se dice que* A, B *son independientes si* $P(A \cap B) = P(A)P(B)$.

A.1.3. Definición. *Sea* $\{A_i\}_{i\in I} \subset \mathcal{F}$. *Se dice que la colección* $\{A_i\}_{i\in I}$ *son independientes si para toda colección finita* $\{i_1, \ldots, i_k\} \subset I$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k}).$$

A.1.4. Proposición. ***(Lema de Borel-Cantelli).*** *Sea* $\{E_n\}_{n\geq 1} \subset \mathcal{F}$. *Entonces,*

(i) *Si* $\sum_{n=1}^{\infty} P(E_n) < \infty$ *entonces* $P(\limsup_{n\to\infty} E_n) = 0$.

(ii) *Si* $\{E_n\}_{n\geq 1} \subset \mathcal{F}$ *son independientes y si* $\sum_{n=1}^{\infty} P(E_n) = \infty$ *entonces* $P(\limsup_{n\to\infty} E_n) = 1$.

Apéndice B

Esperanza matemática

B.1. Definiciones

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, y X una variable aleatoria definida sobre él. Sea g una función Borel medible sobre $\mathbb{R}$. Se sigue que $g(X)$ es también una variable aleatoria.

B.1.1. Definición. *Decimos que la esperanza matemática (o simplemente la esperanza) de $g(X)$ existe, si $g(X)$ es integrable sobre Ω con respecto a P. En este caso se define la esperanza $\mathbb{E}g(X)$ de la variable aleatoria $g(X)$ por*

$$\mathbb{E}g(X) = \int_\Omega g(X(\omega))dP(\omega) = \int_\Omega g(X)dP.$$

Es claro que $\mathbb{E}g(X)$ existe si, y sólo si $\mathbb{E} \mid g(X) \mid$ existe.

Si F, la función de distribución de la variable aleatoria X es absolutamente continua sobre $\mathbb{R}$ con función de densidad $f(x) = F^{'}(x)$, entonces la $\mathbb{E}g(X)$ existe si, y sólo si $\int_{-\infty}^{\infty} \mid g(x) \mid f(x)dx < \infty$, y en este caso

$$\mathbb{E}g(X) := \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Sea $\mathcal{L}_1 = \mathcal{L}_1(\Omega, \mathcal{F}, P)$ el conjuntos de variables aleatorias con esperanza finita. Entonces se tienen las siguientes propiedades:

(i) Si $X, Y \in \mathcal{L}_1$ y $\alpha, \beta \in \mathbb{R}$ entonces $\alpha X + \beta Y \in \mathcal{L}_1$ y $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$.

(ii) Si $X \in \mathcal{L}_1$ entonces $|\mathbb{E}X| \leq \mathbb{E}|X|$.

(iii) Si $X \in \mathcal{L}_1$ y $E \in \mathcal{F}$ entonces $X1_E \in \mathcal{L}_1$, y

$$\int_E X dP = \mathbb{E}(X1_E).$$

Sea X una variable aleatoria sobre $(\Omega, \mathcal{F}, P)$ y $g(X)$ integrable sobre Ω con respecto a P.

B.1.2. Definición.

(i) *Sea $g(x) = (x-\gamma)^n$, donde γ es un número real y n un entero positivo. Entonces $\mathbb{E}(X-\gamma)^n$, si es que existe, se llama el* ***momento*** *de X de orden n al rededor de γ. En particular, si $\gamma = \mathbb{E}X$, entonces $\mu_n := \mathbb{E}(X-\mathbb{E}X)^n$ se llama el* ***momento central*** *de X de orden n. Cuando $n=2$,*

$$\mu_2 = \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

se llama la ***varianza*** *de X, y se denota por $VarX$.*

(ii) *Sea $g(x) = e^{tx}$, $t \in (-\delta, \delta)$ para algún número positivo δ. Entonces $M(t) := \mathbb{E}e^{tX}$, si es que existe, se llama la* ***función generadora de momentos*** *de la variable aleatoria X.*

(iii) *Sea $g(x) = e^{itx} = \cos tx + i \operatorname{sen} tx$, donde $t \in \mathbb{R}$ e $i = \sqrt{-1}$. Entonces*

$$\varphi(t) := \mathbb{E}e^{itX} = \mathbb{E}\cos tX + i\mathbb{E}\operatorname{sen} tX.$$

se llama la ***función característica*** *de X.*

B.2. Algunas desigualdades importantes

B.2.1. Proposición. ***(Desigualdad de Markov).*** *Si X es una variable aleatoria sobre $(\Omega, \mathcal{F}, P)$, $\lambda > 0$ y $\mathbb{E}|X|^\lambda < \infty$ entonces para todo $\varepsilon > 0$,*

$$P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|^\lambda}{\varepsilon^\lambda}.$$

Demostración: Si $\mathbb{E} \mid X \mid^\lambda < \infty$ entonces

$$\begin{aligned}\mathbb{E} \mid X \mid^\lambda &= \int_{|X|\geq\varepsilon} \mid X \mid^\lambda dP + \int_{|X|<\varepsilon} \mid X \mid^\lambda dP \\ &\geq \varepsilon^\lambda \int_{|X|\geq\varepsilon} dP = \varepsilon^\lambda P(\mid X \mid\geq \varepsilon).\end{aligned}$$

□

En particular, si $\lambda = 2$, se obtiene

$$P(\mid X \mid\geq \varepsilon) \leq \frac{\mathbb{E}X^2}{\varepsilon^2}.$$

Entonces si $\mu = \mathbb{E}X$,

$$P(\mid X - \mu \mid\geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

Esta desigualdad se conoce como la **desigualdad de Chevichev**.

B.2.2. Proposición. ***(Desigualdad de Chernoff).*** *Ver [11]. Sea X una variable aleatoria con función generadora de momentos $M(t) = \mathbb{E}[e^{tX}]$. Entonces para todo $\varepsilon > 0$,*

$$\mathbb{P}(X \geq \varepsilon) \leq e^{-t\varepsilon} M(t) \quad \textit{para todo } t > 0$$

$$\mathbb{P}(X \leq \varepsilon) \leq e^{-t\varepsilon} M(t) \quad \textit{para todo } t < 0.$$

Demostración: Sea $t > 0$. Entonces

$$P(X \geq \varepsilon) = P(e^{tX} \geq e^{t\varepsilon}) \leq \mathbb{E}[e^{tX}]e^{-t\varepsilon}.$$

La última desigualdad se debe a la desigualdad de Markov con $\lambda = 1$. La prueba para $t < 0$ es similar. □

B.2.3. Proposición. ***(Desigualdad de Hölder).*** *Ver [3]. Sean p y q números reales positivos tales que $1 < p,\ q < \infty$ y $p^{-1} + q^{-1} = 1$. Sean X y Y variables aleatorias tales que $\mathbb{E} \mid X \mid^p < \infty$ y $\mathbb{E} \mid Y \mid^q < \infty$. Entonces $\mathbb{E} \mid XY \mid < \infty$, y*

$$\mathbb{E} \mid XY \mid \leq [\mathbb{E} \mid X \mid^p]^{1/p} [\mathbb{E} \mid Y \mid^q]^{1/q}.$$

Apéndice C

Conceptos de convergencia

Los resultados que se presentan en este apéndice, se pueden consultar en [8].

C.1. Convergencia casi segura

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias definidas sobre él.

C.1.1. Definición. *Se dice que la sucesión de variables aleatorias $\{X_n\}$ converge c.s. a la variable aleatoria X, si existe $E \in \mathcal{F}$ con $P(E) = 0$ tal que para todo $\omega \notin E$, $| X_n(\omega) - X(\omega) | \to 0$ cuando $n \to \infty$. En este caso escribimos $X_n \overset{c.s.}{\to} X$.*

C.1.2. Definición. *Se dice que $\{X_n\}$ es de Cauchy c.s., si existe $E \in \mathcal{F}$ con $P(E) = 0$ tal que para todo $\omega \notin E$, $| X_n(\omega) - X_m(\omega) | \to 0$ cuando $n, m \to \infty$.*

C.1.3. Proposición. *Sea $\{X_n\}$ una sucesión de variables aleatorias definidas sobre $(\Omega, \mathcal{F}, P)$. Entonces*

$$X_n \overset{c.s.}{\to} X \ \text{ si, y sólo si } \{X_n\} \text{ es de Cauchy c.s.}$$

Demostración: Si $X_n \overset{c.s.}{\to} X$, existe un conjunto nulo $E \in \mathcal{F}$ tal que para todo $\omega \notin E$, $| X_n(\omega) - X(\omega) | \to 0$, cuando $n \to \infty$. Entonces, para todo $\omega \notin E$ y n, m enteros positivos arbitrarios,

$$| X_m(\omega) - X_n(\omega) | \leq | X_m(\omega) - X(\omega) | + | X_n(\omega) - X(\omega) | \to 0 \text{ si } n, m \to \infty.$$

Por lo tanto $\{X_n\}$ es de Cauchy c.s.
Inversamente, si $\{X_n\}$ es de Cauchy c.s., existe un conjunto nulo $E \in \mathcal{F}$ tal que para todo $\omega \notin E$, $\mid X_n(\omega) - X_m(\omega) \mid \to 0$ cuando $n, m \to \infty$. Entonces, $\{X_n(\omega)\}$ es una sucesión de Cauchy de numeros reales. Por lo tanto, existe número real $X(\omega)$, tal que $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ para todo $\omega \notin E$, esto es $X_n \overset{c.s.}{\to} X$. $\square$

C.1.4. Proposición. *$X_n \overset{c.s.}{\to} X$ si, y sólo si*

$$\lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} (\mid X_m - X \mid \geq \varepsilon)\right) = 0 \quad \textit{para todo } \varepsilon > 0.$$

El concepto de convergencia casi segura, en estadística tiene una interpretación especial. Cuando $X_n \overset{c.s.}{\to} X$, se dice que X_n es un estimador *fuertemente consistente* c.s. de X.

C.2. Convergencia en probabilidad

Sean $\{X_n\}_{n\geq 1}$, X variables aleatorias sobre $(\Omega, \mathcal{F}, P)$.

C.2.1. Definición. *Se dice que $\{X_n\}$ converge en probabilidad a X, si para todo $\varepsilon > 0$,*

$$P(\mid X_n(\omega) - X(\omega) \mid \geq \varepsilon) \to 0 \quad \textit{cuando } n \to \infty.$$

En este caso se escribe $X_n \overset{P}{\to} X$.

Se dice también que $\{X_n\}$ es de Cauchy en probabilidad si, y sólo si para todo $\varepsilon > 0$, $P(\mid X_m - X_n \mid \geq \varepsilon) \to 0$ cuando $n, m \to \infty$.

La relación que existe entre convergencia casi segura y convergencia en probabilidad, es la siguiente:

C.2.2. Proposición. *Si $X_n \overset{c.s.}{\to} X$ entonces $X_n \overset{P}{\to} X$.*

Demostración: Por la proposición C.1.4

$$X_n \overset{c.s.}{\to} X \text{ si, y sólo si para todo } \varepsilon > 0 \ \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} \mid X_m - X \mid \geq \varepsilon\right) = 0.$$

Pero los eventos

$$\{| X_n - X |\geq \varepsilon\} \subset \bigcup_{m=n}^{\infty} \{| X_m - X |\geq \varepsilon\}.$$

Entonces,

$$P(| X_n - X |\geq \varepsilon) \leq P\left(\bigcup_{m=n}^{\infty} | X_m - X |\geq \varepsilon\right).$$

Por lo tanto,

$$\lim_{n\to\infty} P(| X_n - X |\geq \varepsilon) \leq \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} | X_m - X |\geq \varepsilon\right) = 0,$$

y $X_n \xrightarrow{P} X$. □

El regreso de esta proposición no se cumple, por lo que la convergencia casi segura, es más fuerte que la convergencia en probabilidad.

C.3. Convergencia en L_1

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. $L_1 = L_1(\Omega, \mathcal{F}, P)$ denota el conjunto de variables aleatorias X sobre Ω tales que $\mathbb{E} \mid X \mid < \infty$.

C.3.1. Definición. *Sea $\{X_n\}_{n\geq 1} \subset L_1$. Se dice que la sucesión $\{X_n\}$ converge en media a la variable aleatoria $X \in L_1$, si $\mathbb{E} \mid X_n - X \mid \to 0$ cuando $n \to \infty$. En este caso se escribe $X_n \xrightarrow{L_1} X$.*

C.3.2. Definición. *Se dice que la sucesión $\{X_n\}$ es de Cauchy en media si $\mathbb{E} \mid X_n - X_m \mid \to 0$ cuando $n, m \to \infty$.*

C.3.3. Proposición. *Sea $\{X_n\}$ una sucesión de variables aleatorias sobre Ω. $X_n \xrightarrow{L_1} X$ implica $X_n \xrightarrow{P} X$.*

Demostración: Supongamos que $X_n \xrightarrow{L_1} X$. Sea $\lambda = 1$ en la desigualdad de Markov, esto es

$$P(| X_n - X |\geq \varepsilon) \leq \frac{\mathbb{E}(| X_n - X |)}{\varepsilon} \quad \text{para todo } \varepsilon > 0.$$

Esto último converge a cero cuando $n \to \infty$, por hipótesis. Por lo tanto,

$$X_n \xrightarrow{P} X.$$

□

C.3.4. Teorema. ***(Teorema de convergencia de Lebesgue).*** *Sea $\{X_n\}$ una sucesión de variables aleatorias, tal que $X_n \xrightarrow{P} X$ o $X_n \xrightarrow{c.s.} X$. Supongamos que existe $Y \in L_1$ tal que $| X_n | \leq Y$ c.s. Entonces $X_n, X \in L_1$ para todo $n \geq 1$, y $X_n \xrightarrow{L_1} X$. En particular $\mathbb{E}X_n \to \mathbb{E}X$ cuando $n \to \infty$.*

La definición de convergencia en media se puede generalizar para una $1 \leq r < \infty$ como sigue:
$L_r = L_r(\Omega, \mathcal{F}, P)$ denota el conjunto de variables aleatorias X sobre Ω tales que $\mathbb{E} | X |^r < \infty$.

C.3.5. Definición. *Se dice que la sucesión $\{X_n\}$ converge en p-media a la variable aleatoria $X \in L_r$ si $\mathbb{E} | X_n - X |^r \to 0$ cuando $n \to \infty$. En este caso se escribe $X_n \xrightarrow{L_r} X$.*

C.4. LFGN y TLC

(1) **Ley Fuerte de los Grandes Números (LFGN).** Sean $X_1, X_2, \ldots,$ variables aleatorias independientes, con esperanza finita m, y $S_n = X_1 + \ldots + X_n$, entonces

$$S_n/n \to m \quad \text{c.s.}$$

(2) **Teorema del Límite Central (TLC).** Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e identicamente distribuidas, tales que $P(X_1 = 1) = p, P(X_1 = 0) = 1 - p$, $0 < p < 1$. Sea $S_n = \sum_{i=1}^{n} X_i$. Entonces,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{cuando } n \to \infty.$$

Apéndice D

Esperanza condicional

D.1. Propiedades

Recordemos que si X es un conjunto, $\mathcal{F} \in 2^X$ es un σ-álgebra si, y sólo si $X, \emptyset \in \mathcal{F}$ y $\mathcal{F}$ es cerrado bajo un número a lo más numerable de operaciones de conjuntos $\{A_i\}_{i\geq 1} \subset \mathcal{F}$. La pareja $(X, \mathcal{F})$ se le llama *espacio medible*. Una *medida* μ es una función de $\mathcal{F}$ en $[0, \infty]$ tal que $\mu(\emptyset) = 0$. Si además $A_1, A_2, \ldots, \in \mathcal{F}$ son tales que $A_i \cap A_j = \emptyset$ para todo $i \neq j$, entonces $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$. La terna $(X, \mathcal{F}, \mu)$ se le llama *espacio de medida*. Si $(X, \mathcal{F})$ y $(Y, \mathcal{A})$ son dos espacios de medida, $f : X \to Y$ es *medible* si, y sólo si para todo $A \in \mathcal{A}$, $f^{-1}(A) = \{x \in X \mid f(x) \in \mathcal{A}\} \in \mathcal{F}$.

Daremos los resultados principales de esperanza condicional. Para quién desee conocer mas de este tema, ver [1].

D.1.1. Definición. *Sean μ, λ medidas sobre el espacio medible $(\Omega, \mathcal{F})$. Se dice que μ es absolutamente continua con respecto a λ ($\mu \prec\prec \lambda$) si siempre que $\lambda(A) = 0$, se sigue que $\mu(A) = 0$ para todo $A \in \mathcal{F}$.*

D.1.2. Teorema. ***(Radon-Nikodyn).*** *Sean μ medida σ-finita o finita y λ, medidas sobre $(\Omega, \mathcal{F})$. Si $\mu \prec\prec \lambda$ entonces existe $g : \Omega \to \bar{\mathbb{R}}$ medible, tal que*

$$\mu(A) = \int_A g d\lambda \quad \textit{para todo } A \in \mathcal{F}.$$

Si $h : \Omega \to \bar{\mathbb{R}}$ es otra función medible con $\mu(A) = \int_A g d\lambda$, entonces $g = h$ c.s. $[\lambda]$.

El siguiente teorema garantiza la existencia de densidades.

D.1.3. Teorema. *Sea* $X : (\Omega, \mathcal{F}, P) \to (\Omega', \mathcal{F}')$ *medible y* $B \in \mathcal{F}$ *fijo. Entonces, existe* $g : (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B})$ *medible, tal que para todo* $A \in \mathcal{F}'$,

$$P(\{X \in A\} \cap B) = \int_A g(x) dP_X(x).$$

Si h *es otra función medible tal que* $P(\{X \in A\} \cap B) = \int_A h(x) dP_X(x)$, *entonces* $h = g$ *c.s.*

En este caso, se define $P(B \mid X = x) := g(x)$.

D.1.4. Teorema. *Sean* $Y : (\Omega, \mathcal{F}, P) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ *y* $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ *medibles. Si* $\mathbb{E}[Y]$ *existe, entonces existe* $g : (\Omega', \mathcal{F}') \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ *medible, tal que para todo* $A \in \mathcal{F}'$,

$$\int_{\{X \in A\}} Y dP = \int_A g(x) dP_X(x).$$

Si h *es otra función medible que cumple la igualdad anterior, entonces* $g = h$ *c.s.* $[P_X]$.

En este caso se define $\mathbb{E}[Y \mid X = x] := g(x)$, llamada la esperanza condicional de Y, dado que $X = x$.

D.1.5. Ejemplo. Sea X variable aleatoria con valores $x_1, x_2, \dots$, tal que $P(X = x_i) > 0$ para todo $i = 1, 2, \dots$,. Entonces

$$\mathbb{E}[Y \mid X = x_i] = \frac{1}{P[X = x_i]} \int_{\{X = x_i\}} Y dP.$$

Demostración: Aquí $\Omega' = \{x_1, x_2, \dots\}$ y $\mathcal{F}' = 2^{\Omega'}$. Sea

$$g(x_i) = \frac{1}{P(X = x_i)} \int_{\{X = x_i\}} Y dP \quad \text{para todo } i = 1, 2, \dots,$$

Entonces

$$\int_{\{X \in A\}} Y dP = \sum_{x_i \in A} P(X = x_i) \frac{1}{P(X = x_i)} \int_{\{X = x_i\}} Y dP = *$$

Esta igualdad se debe a que $\{X \in A\} = \bigcup_{x_i \in A}\{X = x_i\}$, es una unión ajena. Luego

$$* = \sum_{x_i \in A} P(X = x_i)g(x_i) = \int_A g(x)dP_X(x).$$

Por lo tanto,

$$\mathbb{E}[Y \mid X = x_i] = g(x_i) = \frac{1}{P(X = x_i)} \int_{\{X = x_i\}} Y dP.$$

■

Si en el teorema D.1.4 definimos $h(\omega) = g(X(\omega))$, entonces $h : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ es medible. Además

$$\int_{\{X \in A\}} Y dP = \int_{\{X \in A\}} h dP \quad \text{para todo } A \in \mathcal{F}',$$

en efecto:

$$\begin{aligned}
\int_{\{X \in A\}} h dP &= \int_{\Omega} g(X(\omega))1_A(X(\omega))dP(\omega) \\
&= \int_{\Omega'} g(x)1_A(x)dP_X(x) \\
&= \int_A g(x)dP_X(x) \\
&= \int_{\{X \in A\}} Y dP.
\end{aligned}$$

Así que,

$$\int_C h dP = \int_C Y dP \quad \text{para todo } C \in X^{-1}(\mathcal{F}') := \mathcal{F}(X).$$

$\mathcal{F}(X)$ es un sub σ-álgebra de $\mathcal{F}$ y es el más chico que hace medible a X. Entonces,

$$h : (\Omega, \mathcal{F}(X)) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}})) \quad \text{es medible.}$$

Esto caracteriza a h como medible y dá la definición de esperanza condicional dada un σ-álgebra.

D.1.6. Teorema. *Sea $Y : (\Omega, \mathcal{F}, P) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ medible y $\mathcal{G} \subset \mathcal{F}$ sub σ-álgebra. Si $\mathbb{E}[Y]$ existe, entonces existe $\mathbb{E}[Y \mid \mathcal{G}] : (\Omega, \mathcal{G}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$*

medible, llamada la esperanza condicional de Y dada $\mathcal{G}$, tal que para todo $C \in \mathcal{G}$,

$$\int_C Y dP = \int_C \mathbb{E}[Y \mid \mathcal{G}]dP.$$

Si h es otra función que satisface esta igualdad, entonces $h = \mathbb{E}[Y \mid \mathcal{G}]$ c.s. $[P]$.

D.1.7. Teorema. *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\mathcal{G} \subset \mathcal{F}$ sub σ-álgebra. Sea $B \in \mathcal{F}$ fijo, entonces existe $P(B \mid \mathcal{G}) : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ medible, llamado la probabilidad condicional de B dado $\mathcal{G}$, tal que*

$$P(C \cap B) = \int_C P(B \mid \mathcal{G})dP \quad \textit{para todo } C \in \mathcal{G}.$$

Si h es otra función que satisface la igualdad anterior, entonces $h = P(B \mid \mathcal{G})$ c.s. $[P]$, y se cumple $P(B \mid \mathcal{G}) = \mathbb{E}[1_B \mid \mathcal{G}]$.

La demostración de los últimos cuatro teoremas, se siguen del teorema Radon-Nikodyn.

Ahora mencionaremos algunas propiedades que se usan en este trabajo.

D.1.8. Teorema. *Sea $X : (\Omega, \mathcal{F}, P) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ medible, y $Y, Y_1, Y_2, \ldots,$: $(\Omega, \mathcal{F}, P) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ medibles, con $\mathbb{E}[Y] < \infty$ y $\mathbb{E}[Y_n] < \infty$.*

(i) *Si $Y_1 \leq Y_2$ c.s. $[P]$, entonces $\mathbb{E}[Y_1 \mid \mathcal{G}] \leq \mathbb{E}[Y_2 \mid \mathcal{G}]$ c.s. $[P]$.*

(ii) $\mid \mathbb{E}(Y \mid \mathcal{G}) \mid \leq \mathbb{E}[\mid Y \mid\mid \mathcal{G}]$ *c.s.* $[P]$.

(iii) $\mathbb{E}\left[\sum_{m=1}^{\infty} Y_m \mid \mathcal{G}\right] = \sum_{m=1}^{\infty} \mathbb{E}[Y_m \mid \mathcal{G}]$ *c.s.* $[P]$.

(iv) $\mathbb{E}[\mathbb{E}(Y \mid \mathcal{G})] = \mathbb{E}[Y]$ *si Y es integrable.*

(v) $\mathbb{E}[Y \mid \mathcal{F}] = Y$ *c.s.* $[P]$.

(vi) *Si $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, entonces $\mathbb{E}[\mathbb{E}(Y \mid \mathcal{G}_2) \mid \mathcal{G}_1] = \mathbb{E}[Y \mid \mathcal{G}_1]$ c.s. $[P]$.*

(vii) *Si Z es $\mathcal{G}$-medible y Y, YZ son integrables, entonces $\mathbb{E}[YZ \mid \mathcal{G}] = Z\mathbb{E}[Y \mid \mathcal{G}]$ c.s. $[P]$.*

D.1.9. Proposición. ***(Desigualdad de Jensen).*** *Sea* $Y : (\Omega, \mathcal{F}, P) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ *tal que* $\mathbb{E}[Y]$ *existe. Sea* g *continua y convexa sobre* $\mathbb{R}$*, tal que* $\mathbb{E}[g(y)]$ *existe. Sea* $\mathcal{G} \subset \mathcal{F}$ *sub* σ*-álgebra. Entonces,*

$$g[\mathbb{E}(Y \mid \mathcal{G})] \leq \mathbb{E}[g(Y) \mid \mathcal{G}] \quad c.s. \ [P].$$

En particular $\mathbb{E}[g(Y)] \geq g[\mathbb{E}(Y)]$.

D.2. Martingalas

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $(T, \prec)$ un conjunto con orden parcial. Sea $\{\mathcal{F}_t \mid t \in T\}$ una familia de sub σ-álgebras de $\mathcal{F}$ tal que si

$$s \prec t \quad \text{entonces} \quad \mathcal{F}_s \subset \mathcal{F}_t.$$

Por ejemplo, si $\{\mathcal{F}_n \mid n \in \mathbb{N}\}$ es una familia de sub σ-álgebras de $\mathcal{F}$, con $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, entonces es claro que si $n < m$, $\sigma(X_1, \ldots, X_n) \subset \sigma(X_1, \ldots, X_m)$.

Sea $\mathcal{X} = \{X_t : t \in T\}$ una familia de variables aleatorias definidas en Ω, con $\mathbb{E}[X_t]$ finita.

D.2.1. Definición. *La clase* $\mathcal{X}$ *es una martingala con respecto a* $\{\mathcal{F}_t \mid t \in T\}$ *si, y solo si*

(i) X_t *es* $\mathcal{F}_t$*-medible para todo* $t \in T$.

(ii) *Para todo* $s, t \in T$*, con* $s < t$, $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ *c.s.* $[P]$.

$\mathcal{X}$ *es una submartingala si para todo* $s, t \in T$*, con* $s < t$, $\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s$ *c.s.* $[P]$.
$\mathcal{X}$ *es una supermartingala si para todo* $s, t \in T$*, con* $s < t$, $\mathbb{E}[X_t \mid \mathcal{F}_s] \leq X_s$ *c.s.* $[P]$.
Por martingala se denota $\{X_t, \mathcal{F}_t\}$, y cuando $T = \{1, 2, \ldots,\}$ y $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, se denota por $\{X_n, \mathcal{F}_n\}$

D.2.2. Proposición. *Sea* $\{X_n, \mathcal{F}_n\}$ *una martingala (sub o super). Las siguientes afirmaciones son equivalentes.*

(i) $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}) = (\geq \ o \ \leq) \ X_{n-1}$ *para todo* $n > 1$ *c.s.* $[P]$.

(ii) $\mathbb{E}[X_n \mid \mathcal{F}_m] = (\geq \ o \ \leq) \ X_m$ *para todo* $m < n$ *c.s.* $[P]$.

Bibliografía

[1] Ash, R. B. (2000). *Probability and Measure Theory.* Second Edition. Ed. Academic Press, San Diego.

[2] Birnbaum, Z. W. and Tingey, F. H. (1951). One-side confidence for probability distribution function. *Ann. Math. Statist.* Vol. **22**, pp. 592-596.

[3] Dudley, R. M. (1989). *Real Analysis and Probability.* Ed. Wadsworth & Brooks/Cole Mathematics Series, Pacific Grove, California.

[4] Durbin, J. (1973). *Distribution theory of test based on the sample distribution function.* Regional Conference Series in Appl. Math., No. **9**, SIAM, Philadelphia.

[5] Dvoretzky A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, pp. 642-669.

[6] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, pp. 13-30.

[7] Inder, K. R. (2002). *An introduction to measure and integration.* Second Edition. Ed. American Mathematical Society, Providence, Rhode Island.

[8] Laha, R. G. and Rohatgi, V. K. (1979). *Probability theory.* Ed. John Wiley & Sons, New York.

[9] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **35**, pp. 1065-1076.

[10] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, pp. 832-837.

[11] Ross, S. M. (1996). *Stochastic Processes.* Second Edition. Ed. Wiley series in probability and Mathematical Statistic, New York.

[12] Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Ed. Wiley, New York.

[13] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statist.* Vol. **5**, No. **4**, pp. 595-645.

[14] Stute, W. (1982). The oscillation behavior of empirical processes. *Annals of probability.* Vol. **10**, No. **1**, pp. 86-107.

[15] Stute, W. (2001). *Lecture notes on empirical processes.* Preprint.