Capítulo 5

Esperanza Condicional y Martingalas

5.1. Preliminares

Comenzamos recordando algunos conceptos fundamentales sobre Espacios de Hilbert.

Definición 5.1 Sea V un espacio vectorial real. Una función $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ es un *producto interno* si satisface

(i) $\langle x, x \rangle > 0, \forall x \neq 0$.

(ii) $\langle x, y \rangle = \langle y, x \rangle$ para todo $x, y \in V$.

(iii) $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$.

Si $\langle x, x \rangle \geq 0$ para todo x y (ii) y (iii) valen, decimos que $\langle \cdot, \cdot \rangle$ es una forma bilineal simétrica semipositiva definida o un semi-producto interno.

El producto interno $\langle \cdot, \cdot \rangle$ en V define una norma $||x|| = \langle x, x \rangle^{1/2}$.

Definición 5.2 Si $(V, \langle \cdot, \cdot \rangle)$ es completo respecto a la norma definida por el producto interno, decimos que es un *Espacio de Hilbert*.

Ejemplo 5.1
Para $f, g \in \mathcal{L}^2(\mu)$ definimos

$$\langle f, g \rangle = \int f g \, d\mu.$$

Si consideramos el espacio cociente $L^2(\mu)$ y f, g se toman como representantes de sus clases de equivalencia $\bar{f}, \bar{g}$ respectivamente, definimos $\langle \bar{f}, \bar{g} \rangle = \langle f, g \rangle$. Entonces $\langle \cdot, \cdot \rangle$ es un producto interno en L^2 y un semi-producto interno en $\mathcal{L}^2$. Sabemos que $(L^2, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert.

Definición 5.3 Sea V un espacio vectorial real con producto interno $\langle \cdot, \cdot \rangle$. Si $W \subset V$, el complemento ortogonal de W se define como

$$W^{\perp} = \{v \in V : \langle v, w \rangle = 0, \forall w \in W\}$$

Teorema 5.1 (Descomposición Ortogonal) *Sea $(V, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $W \subset V$ un subespacio lineal cerrado. Para todo $x \in V$ hay una representación única $x = y + z$ con $y \in W$ y $z \in W^{\perp}$.*

Demostración. Sea $x \in V$, $c = \text{ínf}\{||x-w|| : w \in W\}$ y sea $(w_n)_{n\geq 1}$ una sucesión en W con $||x-w_n|| \to c$ cuando $n \to \infty$. Usando la ley del paralelogramo obtenemos

$$||w_m - w_n||^2 = 2||w_m - x||^2 + 2||w_n - x||^2 - 4\left\|\frac{1}{2}(w_m + w_n) - x\right\|^2$$

Como W es un subespacio lineal, $(w_m + w_n)/2 \in W$ y en consecuencia $||\frac{1}{2}(w_m + w_n) - x|| \geq c$. Por lo tanto $(w_n)_{n\geq 1}$ es una sucesión de Cauchy: $||w_m - w_n|| \to 0$ si $m, n \to \infty$. Como V es completo y W es cerrado, también es completo, de modo que existe $y \in W$ con $w_n \to y$ cuando $n \to \infty$. Sea ahora $z = x - y$, entonces $||z|| = \text{lím}\, ||w_n - x|| = c$ por la continuidad de la norma.

Consideremos un $w \in W \setminus \{0\}$ cualquiera. Definimos $\rho = -\langle z, w\rangle/||w||^2$ y tenemos $y + \rho w \in W$. En consecuencia

$$c^2 \leq ||x - (y + \rho w)||^2 = ||z||^2 + \rho^2||w||^2 - 2\rho\langle z, w\rangle = c^2 - \rho^2||w||^2$$

y en conclusión $\langle z, w\rangle = 0$ para todo $w \in W$ y por lo tanto $z \in W^\perp$.

Falta ver la unicidad en la descomposición. Si $x = y' + z'$ es otra descomposición ortogonal entonces $y - y' \in W$ y $z - z' \in W^\perp$ y además $y - y' + z - z' = 0$. En consecuencia

$$\begin{aligned} 0 &= ||y - y' + z - z'||^2 = ||y - y'||^2 + ||z - z'||^2 + 2\langle y - y', z - z'\rangle \\ &= ||y - y'||^2 + ||z - z'||^2, \end{aligned}$$

de donde obtenemos que $y = y'$ y $z = z'$. ■

Teorema 5.2 (de Representación de Riesz-Fréchet) *Sea $(V, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $F : V \to R$ una función. Las siguientes afirmaciones son equivalentes:*

(i) F es lineal y continua.

(ii) Existe $f \in V$ con $F(x) = \langle x, f\rangle$ para todo $x \in V$.

El elemento $f \in V$ en (ii) es único.

Demostración. $(ii) \Rightarrow (i)$Para cualquier $f \in V$, la función $x \mapsto \langle x, f\rangle$ es lineal y continua por la definición del producto interno y la norma.

$(i) \Rightarrow (ii)$ Para ver el recíproco, si $F \equiv 0$ escogemos $f = 0$. Si F no es idénticamente cero, como es una función continua, el núcleo $W = F^{-1}(\{0\})$ es un subespacio lineal cerrado y propio de V. Sea $v \in V \setminus W$ y sea $v = y + z$ para $y \in W$ y $z \in W^\perp$ la descomposición ortogonal de v. Entonces $z \neq 0$ y $F(z) = F(v) - F(y) = F(v) \neq 0$. Por lo tanto podemos definir $u = z/F(z) \in W^\perp$. Claramente, $F(u) = 1$ y para cualquier $x \in V$, tenemos

$$F(x - F(x)u) = F(x) - F(x)F(u) = 0.$$

Por lo tanto $x - F(x)u \in W$ y $\langle x - F(x)u, u\rangle = 0$. En consecuencia

$$F(x) = \frac{1}{||u||^2}\langle x, u\rangle.$$

Definimos ahora $f = u/||u||^2$, entonces $F(x) = \langle x, f\rangle$ para todo $x \in V$.

Finalmente, para ver la unicidad, sea $\langle x, f\rangle = \langle x, g\rangle$ para todo $x \in V$. Si ponemos $x = f - g$ tenemos $0 = \langle f - g, f - g\rangle$ de donde obtenemos que $f = g$. ■

5.2. El Teorema de Radon-Nikodym

Sea μ, ν medidas sobre $(\Omega, \mathcal{F})$. Decimos que la función medible $f : \Omega \to [0, \infty)$ es una densidad de ν respecto de μ si para todo $A \in \mathcal{F}$,

$$\nu(A) = \int_A f \, d\mu \tag{5.1}$$

Por otro lado, dada cualquier función medible $f : \Omega \to [0, \infty)$, la ecuación (5.1) define una medida en $(\Omega, \mathcal{F})$. En este caso, si $g : \Omega \to [0, \infty]$ es medible entonces

$$\int g \, d\nu = \int gf \, d\mu, \tag{5.2}$$

de modo que $g \in \mathcal{L}(\nu)$ si y sólo si $gf \in \mathcal{L}(\mu)$, y en este caso (5.2) vale.

Teorema 5.3 *Sea ν una medida σ-finita. Si f_1 y f_2 son densidades de ν respecto de μ, entonces $f_1 = f_2$ c.s.(μ). Por lo tanto la densidad es única salvo por diferencias en conjuntos μ-nulos.*

Demostración. Sea $\Omega = \cup_{n\geq 1} E_n$ una descomposición del espacio Ω en subconjuntos medibles de medida finita. Como la descomposición es numerable, basta con demostrar que $f_1 = f_2$ c.s.(μ) en los conjuntos E_n. Sea E cualquiera de estos conjuntos y sea $A = E \cap \{f_1 > f_2\}$, entonces $\nu(A) < \infty$. Por lo tanto

$$\int_A (f_1 - f_2) \, d\mu = \nu(A) - \nu(A) = 0$$

y como en el conjunto A el integrando no es nulo, necesariamente $\mu(A) = 0$. De manera similar se demuestra que $\mu(\{f_1 < f_2\}) = 0$ y en consecuencia $f_1 = f_2$ c.s.(μ). ■

Usaremos la notación $d\nu/d\mu$ para la densidad de ν respecto de μ.

Definición 5.4 Sea μ y ν dos medidas sobre $(\Omega, \mathcal{F})$. Decimos que ν es *absolutamente continua* respecto de μ ($\nu \ll \mu$) si $\nu(A) = 0$ para todo $A \in \mathcal{F}$ con $\mu(A) = 0$.

Decimos que las medidas μ y ν son equivalentes ($\mu \approx \nu$) si $\nu \ll \mu$ y $\mu \ll \nu$.

Decimos que ν es singular respecto de ν ($\nu \perp \mu$) si existe un conjunto medible A tal que $\mu(A) = 0$ y $\nu(A^c) = 0$.

Ejemplos 5.2

1. Sea μ una medida sobre $(\mathbb{R}, \mathcal{B})$ con densidad f respecto de la medida de Lebesgue m. Entonces $\mu(A) = \int_A f \, dm = 0$ para todo $A \in \mathcal{B}$ con $m(A) = 0$, de modo que $\mu \ll m$. Por otro lado si $f > 0$ c.s. respecto de la medida de Lebesgue, entonces $\mu(A) = \int_A f \, dm > 0$ si $m(A) > 0$, y en consecuencia $\mu \approx m$. En cambio, si f se anula en un conjunto de medida de Lebesgue positiva, como $\mu(\{f = 0\}) = 0$ pero $m(\{f = 0\}) > 0$, la medida de Lebesgue no es absolutamente continua respecto de la medida μ.

2. La distribución de Poisson está concentrada en el conjunto $\{0, 1, 2, \dots\}$ que tiene medida 0, de modo que es singular respecto de m.

Teorema 5.4 (de Descomposición de Lebesgue) *Sea μ y ν medidas σ-finitas en $(\Omega, \mathcal{F})$. Entonces ν puede descomponerse de manera única en una parte absolutamente continua ν_a y una parte singular ν_s respecto de μ:*

$$\nu = \nu_a + \nu_s, \quad \text{donde } \nu_a \ll \mu \text{ y } \nu_s \perp \mu.$$

ν_a tiene densidad $\frac{d\nu_a}{d\mu}$ que es medible y finita c.s.(μ).

Demostración. Basta con hacer la demostración en el caso en el cual ambas medidas son finitas. Consideremos el funcional lineal $T : \mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu) \to \mathbb{R}$ definido por

$$T(h) = \int h\, d\nu.$$

Este funcional es acotado ya que, usando la desigualdad de Cauchy-Schwarz,

$$\begin{aligned} |T(h)| &= \Big| \int h\, d\nu \Big| \leq \Big(\int h^2\, d\nu \Big)^{1/2} \Big(\int 1\, d\nu \Big)^{1/2} \\ &\leq \nu^{1/2}(\Omega) ||h||_{L^2(\nu)} \leq \nu^{1/2}(\Omega) ||h||_{L^2(\mu+\nu)} \end{aligned}$$

y por lo tanto es continuo. Por el teorema de representación de Riesz-Fréchet, existe $g \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu)$ tal que para todo $h \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu)$,

$$\int h\, d\nu = \int hg\, d(\mu + \nu) \tag{5.3}$$

o equivalentemente, para todo $f \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu)$,

$$\int f(1 - g)\, d(\mu + \nu) = \int f\, d\mu \tag{5.4}$$

Si en (5.3) ponemos $h = \mathbf{1}_{\{g<0\}}$, entonces $g \geq 0$ c.s. $(\mu + \nu)$. Si en cambio ponemos $f = \mathbf{1}_{\{g>1\}}$ en (5.4), obtenemos que $g \leq 1$ c.s. $(\mu + \nu)$ y en consecuencia $0 \leq g \leq 1$.

Sea ahora $f \geq 0$ medible y sea $(f_n)_{n\geq 1}$ una sucesión de funciones medibles no-negativas en $\mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu)$ con $f_n \uparrow f$. Por el teorema de convergencia monótona aplicado a la medida $(1 - g)(\mu + \nu)$, obtenemos que (5.4) vale para toda $f \geq 0$ medible. De manera similar se demuestra que (5.3) vale para toda $h \geq 0$ medible.

Sea $E = g^{-1}(\{1\})$. Si ponemos $f = \mathbf{1}_E$ en (5.4) entonces obtenemos $\mu(E) = 0$. Definimos las medidas ν_a y ν_s para $A \in \mathcal{F}$ por

$$\nu_a(A) = \nu(A \setminus E) \quad \text{y} \quad \nu_s(A) = \nu(A \cap E)$$

Entonces $\nu = \nu_A + \nu_s$ y $\nu_s(\Omega \setminus E) = 0$, de modo que $\nu_s \perp \mu$. Si tenemos que $A \cap E = \varnothing$ y $\mu(A) = 0$ entonces $\int \mathbf{1}_A\, d\mu = 0$. Por (5.4) también se tiene que

$$\int_A (1 - g)\, d(\mu + \nu) = 0$$

Por otro lado tenemos que $1 - g > 0$ en A, de modo que $\mu(A) + \nu(A) = 0$ y $\mu(A) = \nu(A) = 0$. Más generalmente, si B es medible con $\mu(B) = 0$ entonces $\mu(B \backslash E) = 0$ y en consecuencia $\nu_A(B) = \nu_A(B \backslash E) = 0$. En consecuencia $\nu_a \ll \mu$ y $\nu = \nu_q + \nu_s$ es la descomposición que buscábamos.

Para obtener la densidad de ν_a respecto de μ, definimos

$$f = \frac{g}{1 - g} \mathbf{1}_{\Omega \setminus E}.$$

Para cualquier $A \in \mathcal{F}$, por (5.3) y (5.4) con $h = \mathbf{1}_{A \backslash E}$

$$\int_A f\, d\mu = \int_{A \cap E^c} g\, d(\mu + \nu) = \nu(A \setminus E) = \nu_a(A)$$

de modo que $f = \frac{d\nu_a}{d\mu}$. ■

Corolario 5.1 (Teorema de Radon-Nikodym) *Si μ y ν son medidas σ-finitas en $(\Omega, \mathcal{F})$, existe f medible tal que*

$$\nu(A) = \int_A f\, d\mu, \qquad \forall A \in \mathcal{F},$$

si y sólo si $\nu \ll \mu$.

Escribimos $f = \frac{d\nu}{dP}$ y también $d\nu = f dP$.

Corolario 5.2 *Sean P y Q medidas de probabilidad en $(\Omega, \mathcal{F})$ tales que $Q \ll P$. Sea $\mathcal{G} \subset \mathcal{F}$ una sub-σ-álgebra. Sean $Q\big|_{\mathcal{G}}$ y $P\big|_{\mathcal{G}}$ las restricciones de Q y P a $\mathcal{G}$. Entonces en $(\Omega, \mathcal{G})$*

$$Q\big|_{\mathcal{G}} \ll P\big|_{\mathcal{G}}$$

y

$$\frac{dQ\big|_{\mathcal{G}}}{dP\big|_{\mathcal{G}}} \quad \text{es } \mathcal{G}\text{-medible.}$$

5.3. Esperanza Condicional

Definición 5.5 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, X una v.a. integrable y $\mathcal{G} \subset \mathcal{F}$ una sub-σ-álgebra. La *esperanza condicional* de X dada la σ-álgebra $\mathcal{G}$ es cualquier variable aleatoria Z medible respecto a $\mathcal{G}$ e integrable tal que

$$\int_A Z\, dP = \int_A X\, dP \qquad \forall A \in \mathcal{G} \tag{5.5}$$

Observación 5.1 (a) Usamos la notación $\mathrm{E}(X|\mathcal{G})$ para la esperanza condicional de X dada $\mathcal{G}$.

(b) La esperanza condicional no es única: Cualquier variable aleatoria Y que sea equivalente a Z en el sentido de ser igual a ella casi seguramente tiene las mismas propiedades.

(c) Observamos que las integrales de X y Z sobre los conjuntos $A \in \mathcal{G}$ coinciden. Sin embargo $X \in \mathcal{F}$ mientras que $Z \in \mathcal{G}$.

(d) La esperanza condicional no es un número, es una variable aleatoria.

Veamos cuál es el sentido matemático de esta definición. Supongamos inicialmente que $X \geq 0$ es integrable y definamos

$$\nu(A) = \int_A X\, dP, \quad A \in \mathcal{F}.$$

Entonces ν es finita y absolutamente continua respecto a P. Por lo tanto

$$\nu\big|_{\mathcal{G}} \ll P\big|_{\mathcal{G}}.$$

Por el teorema de Radon-Nikodym sabemos que existe la derivada y ponemos

$$\mathrm{E}(X|\mathcal{G}) = \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}}$$

la cual es $\mathcal{G}$-medible por el corolario 5.2. Para cualquier $G \in \mathcal{G}$ se tiene

$$\begin{aligned}\int_G X\, dP = \nu|_{\mathcal{G}}(G) = \nu(G) &= \int_G \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}} dP|_{\mathcal{G}} \\ &= \int_G \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}} dP \\ &= \int_G \mathrm{E}(X|\mathcal{G})\, dP\end{aligned}$$

que es la ecuación (5.5).

Si X no es positiva entonces

$$\mathrm{E}(X|\mathcal{G}) = \mathrm{E}(X^+|\mathcal{G}) - \mathrm{E}(X^-|\mathcal{G})$$

satisface las condiciones de la definición.

Definición 5.6 Definimos la probabilidad condicional dada la (sub-)σ-álgebra $\mathcal{G}$ por

$$P(A|\mathcal{G}) = \mathrm{E}(\mathbf{1}_A|\mathcal{G})$$

para todo $A \in \mathcal{F}$. Por lo tanto $P(A|\mathcal{G})$ es una variable aleatoria tal que
(a) $P(A|\mathcal{G})$ es $\mathcal{G}$-medible e integrable.
(b) $P(A|\mathcal{G})$ satisface

$$\int_G P(A|\mathcal{G})\, dP = P(A \cap G), \tag{5.6}$$

para todo $G \in \mathcal{G}$.

Definición 5.7 Sea $\{X_t, t \in T\}$ una colección de v.a. definidas en el espacio $(\Omega, \mathcal{F}, P)$, donde T es algún conjunto de índices. Sea

$$\mathcal{G} = \sigma(X_t,\ t \in T)$$

la σ-álgebra generada por $\{X_t,\ t \in T\}$. Definimos

$$\mathrm{E}(X|X_t,\ t \in T) = \mathrm{E}(X|\mathcal{G}).$$

Ejemplo 5.3
Sea $\{\Lambda_i,\ i \geq 1\}$ una partición de Ω, de modo que $\Lambda_i \cap \Lambda_j = \varnothing$ si $i \neq j$ y $\cup_1^\infty \Lambda_n = \Omega$. Definimos $\mathcal{G} = \sigma(\Lambda_i,\ i \geq 1)$ de modo que podemos describir a esta σ-álgebra como

$$\mathcal{G} = \Big\{ \bigcup_{i \in J} \Lambda_i : J \subset \mathbb{N} \Big\}$$

Para $X \in L^1(\Omega)$ definimos

$$\alpha_n = \alpha_n(X) = \mathrm{E}(X|\Lambda_n) = \int X\, P(d\omega|\Lambda_n) = \frac{1}{P(\Lambda_n)} \int_{\Lambda_n} X\, dP$$

si $P(\Lambda_n) > 0$ y $\alpha_n = 0$ si $P(\Lambda_n) = 0$. Veamos que

(a) $\mathrm{E}(X|\mathcal{G}) \stackrel{c.s.}{=} \sum_{n=1}^{\infty} \alpha_n(X) \mathbf{1}_{\Lambda_n}$.

(b) Para cualquier $A \in \mathcal{F}$, $P(A|\mathcal{G}) \stackrel{c.s.}{=} \sum_{n=1}^{\infty} P(A|\Lambda_n) \mathbf{1}_{\Lambda_n}$.

Para verificar (a) observamos que

$$\sum_{n=1}^{\infty} \alpha_n(X) \mathbf{1}_{\Lambda_n} \in \mathcal{G}.$$

Ahora escogemos $\Lambda \in \mathcal{G}$ y basta mostrar que

$$\int_\Lambda \mathrm{E}(X|\mathcal{G})\, dP = \int_\Lambda \Big(\sum_{n=1}^{\infty} \alpha_n(X) \mathbf{1}_{\Lambda_n} \Big)\, dP = \int_\Lambda X\, dP. \tag{5.7}$$

Como $\Lambda \in \mathcal{G}$, Λ es de la forma $\Lambda = \cup_{i \in J} \Lambda_i$ para algún $J \subset \mathbb{N}$. Veamos que la forma propuesta en (a) satisface (5.7).

$$\begin{aligned}
\int_\Lambda \sum_{n=1}^\infty \alpha_n(X)\mathbf{1}_{\Lambda_n}\, dP &= \sum_{n\geq 1}\sum_{i\in J}\int_{\Lambda_i}\alpha_n(X)\mathbf{1}_{\Lambda_n}\, dP \\
&= \sum_{n\geq 1}\sum_{i\in J}\alpha_n(X)P(\Lambda_i\cap\Lambda_n) \\
&= \sum_{i\in J}\alpha_n(X)P(\Lambda_i) \\
&= \sum_{i\in J}\frac{1}{P(\Lambda_i)}\int_{\Lambda_i} X\, dP\ P(\Lambda_i) \\
&= \sum_{i\in J}\int_{\Lambda_i} X\, dP = \int_{\cup_{i\in J}\Lambda_i} X\, dP \\
&= \int_\Lambda X\, dP.
\end{aligned}$$

Esto demuestra (a). Para obtener (b) basta poner $X = \mathbf{1}_A$.

Ejemplo 5.4
Sea X una v.a. discreta con valores posibles $x_1, x_2, \dots$. Entonces para $A\in\mathcal{F}$

$$\begin{aligned}
P(A|X) = P(A|\sigma(X)) &= P(A|\sigma(\{X=x_i\},\ i\geq 1)) \\
&= \sum_{i=1}^\infty P(A|X=x_i)\mathbf{1}_{\{X=x_i\}}
\end{aligned}$$

Ejemplo 5.5
Supongamos que X e Y son v.a. cuya distribución conjunta es absolutamente continua con densidad $f(x,y)$, de modo que para $A\in\mathcal{B}^2$

$$P[(X,Y)\in A] = \iint_A f(x,y)\, dx\, dy.$$

¿Quién es $P(Y\in C|X)$ para $C\in\mathcal{B}$? Sea $\mathcal{G}=\sigma(X)$ y

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\, dy$$

la densidad marginal de X y definimos

$$\phi(x) = \begin{cases} \frac{1}{f_X(x)}\int_C f(x,y)\, dy, & \text{si } f_X(x)>0, \\ 0 & \text{si } f_X(x)=0. \end{cases}$$

Veamos que $P(Y\in C|X) = \phi(X)$. En primer lugar observamos que $\int_C f(X,y)\, dy$ es $\sigma(X)$-medible y por lo tanto $\phi(X)$ es $\sigma(X)$-medible. Falta demostrar que para cualquier $\Lambda\in\sigma(X)$,

$$\int_\Lambda \phi(X)\, dP = P[(Y\in C)\cap\Lambda].$$

Como $\Lambda\in\sigma(X)$, podemos escribir $\Lambda=\{X\in A\}$ para algún $A\in\mathcal{B}$. Por el teorema de transformación tenemos

$$\int_\Lambda \phi(X)\, dP = \int_{X^{-1}(A)}\phi(X)\, dP = \int_A \phi(x)\, dP_X(x)$$

y como existe la densidad conjunta de (X, Y) obtenemos que esta expresión es igual a

$$\begin{aligned}
&= \int_A \phi(x)\big(\int_{\mathbb{R}} f(x,y)\,dy\big)\,dx \\
&= \int_{A\cap\{x: f_X(x)>0\}} \phi(x) f_X(x)\,dx + \int_{A\cap\{x: f_X(x)=0\}} \phi(x) f_X(x)\,dx \\
&= \int_{A\cap\{x: f_X(x)>0\}} \phi(x) f_X(x)\,dx \\
&= \int_{A\cap\{x: f_X(x)>0\}} \frac{1}{f_X(x)} \int_C f(x,y)\,dy\ f_X(x)\,dx \\
&= \int_{A\cap\{x: f_X(x)>0\}} \big(\int_C f(x,y)\,dy\big)\,dx \\
&= \int_A \big(\int_C f(x,y)\,dy\big)\,dx = P(X\in A, Y\in C) \\
&= P((Y\in C)\cap\Lambda)
\end{aligned}$$

5.3.1. Propiedades de la Esperanza Condicional

Proposición 5.1 *Sean X e Y v.a. integrables, $\mathcal{G}\subset\mathcal{F}$ y a, b, c números reales. Entonces*

(a) $\mathrm{E}(\mathrm{E}(X|\mathcal{G})) = \mathrm{E}(X)$.
(b) $\mathrm{E}(aX+bY|\mathcal{G}) \overset{c.s.}{=} a\,\mathrm{E}(X|\mathcal{G}) + b\,\mathrm{E}(Y|\mathcal{G})$.
(c) Si $X\in\mathcal{G}$ *entonces* $\mathrm{E}(X|\mathcal{G}) \overset{c.s.}{=} X$.
(d) $\mathrm{E}(c|\mathcal{G}) \overset{c.s.}{=} c$.
(e) $\mathrm{E}(X|\{\varnothing,\Omega\}) \overset{c.s.}{=} \mathrm{E}(X)$.
(f) Si $X\geq 0$ *c.s. entonces* $\mathrm{E}(X|\mathcal{G})\geq 0$ *c.s.*
(g) Si $X\leq Y$ *c.s. entonces* $\mathrm{E}(X|\mathcal{G})\leq \mathrm{E}(Y|\mathcal{G})$ *c.s.*
(h) $|\,\mathrm{E}(X|\mathcal{G})|\leq \mathrm{E}(|X||\mathcal{G})$.
(i) Si X *es independiente de* $\mathcal{G}$ *entonces* $\mathrm{E}(X|\mathcal{G}) = \mathrm{E}(X)$ *c.s.*

Demostración. La primera propiedad se obtiene haciendo $\Lambda = \Omega$ en (5.5). (b) es

$$\begin{aligned}
\int_\Lambda \mathrm{E}(aX+bY|\mathcal{G})\,dP &= \int_\Lambda (aX+bY)\,dP = a\int_\Lambda X\,dP + b\int_\Lambda Y\,dP \\
&= a\int_\Lambda \mathrm{E}(X|\mathcal{G})\,dP + b\int_\Lambda \mathrm{E}(Y|\mathcal{G})\,dP \\
&= \int_\Lambda \big[a\,\mathrm{E}(X|\mathcal{G}) + b\,\mathrm{E}(Y|\mathcal{G})\big]\,dP.
\end{aligned}$$

(c) es consecuencia de la identidad trivial

$$\int_\Lambda X\,dP = \int_\Lambda X\,dP$$

y como X es $\mathcal{G}$-medible satisface las condiciones de la definición. Como toda constante es $\mathcal{G}$-medible, (d) es inmediata de (c).

Para (e)

$$\int_\Lambda X\,dP = \begin{cases} 0, & \text{para } \Lambda = \varnothing, \\ \mathrm{E}(X), & \text{para } \Lambda = \Omega, \end{cases}$$

es decir

$$\int_\Lambda X\,dP = \int_\Lambda \mathrm{E}(X)\,dP, \quad \text{para todo } \Lambda\in\{\varnothing,\Omega\},$$

de modo que podemos reemplazar $\mathrm{E}(X|\mathcal{G})$ por $\mathrm{E}(X)$. (f) se obtiene por

$$\int_\Lambda \mathrm{E}(X|\mathcal{G})\, dP = \int_\Lambda X\, dP \geq 0.$$

Usando esta propiedad y (b) aplicada a $Y - X$ obtenemos (g). Para (h) tenemos

$$|\,\mathrm{E}(X|\mathcal{G})| = |\,\mathrm{E}(X^+|\mathcal{G}) - \mathrm{E}(X^-|\mathcal{G})| \leq \mathrm{E}(X^+|\mathcal{G}) + \mathrm{E}(X^-|\mathcal{G}) = \mathrm{E}(|X||\mathcal{G}).$$

Para probar (i) observamos que $\mathrm{E}(X) \in \mathcal{G}$ y que para cualquier $\Lambda \in \mathcal{G}$,

$$\int_\Lambda \mathrm{E}(X)\, dP = \mathrm{E}(X)P(\Lambda) = \mathrm{E}(X\mathbf{1}_\Lambda) = \int_\Omega X\mathbf{1}_\Lambda\, dP = \int_\Lambda X\, dP.$$

Esta relación dice que $\mathrm{E}(X)$ satisface (5.5). ■

La siguiente proposición presenta las versiones condicionales del TCM, el lema de Fatou y el TCD.

Proposición 5.2 *(a) Si* $0 \leq X_n \uparrow X \in L^1(\Omega)$ *cuando* $n \to \infty$ *entonces* $\mathrm{E}(X_n|\mathcal{G}) \uparrow \mathrm{E}(X|\mathcal{G})$ *cuando* $n \to \infty$.

(b) Si $X_n \downarrow X$ *cuando* $n \to \infty$ *y* $X_1, X \in L^1(\Omega)$ *entonces* $\mathrm{E}(X_n|\mathcal{G}) \downarrow \mathrm{E}(X|\mathcal{G})$ *cuando* $n \to \infty$.

(c) Si $\{X_n,\ n \geq 1\}$ *son no-negativas e integrables, entonces*

$$\mathrm{E}(\liminf_{n\to\infty} X_n|\mathcal{G}) \leq \liminf_{n\to\infty} \mathrm{E}(X_n|\mathcal{G}) \quad c.s.$$

(d) Si $X_n \leq Y \in L^1$ *para todo* n*, entonces*

$$\mathrm{E}(\limsup_{n\to\infty} X_n|\mathcal{G}) \geq \limsup_{n\to\infty} \mathrm{E}(X_n|\mathcal{G}) \quad c.s.$$

(e) Si $|X_n| \leq Y \in L^1$ *y* $X_n \stackrel{c.s.}{\to} X$ *cuando* $n \to \infty$*, entonces*

$$\mathrm{E}(X_n|\mathcal{G}) \stackrel{c.s.}{\to} \mathrm{E}(X|\mathcal{G}) \quad \textit{cuando } n \to \infty.$$

Demostración. Sabemos por la proposición anterior que $\mathrm{E}(X_n|\mathcal{G})$ es monótona creciente y no-negativa, de modo que el límite $Z = \lim_{n\to\infty} \mathrm{E}(X_n|\mathcal{G})$ existe. Usando la relación (5.5) y el TCM obtenemos para $\Lambda \in \mathcal{G}$

$$\begin{aligned}\int_\Lambda Z\, dP &= \int_\Lambda \lim_{n\to\infty} \mathrm{E}(X_n|\mathcal{G})\, dP = \lim_{n\to\infty} \int_\Lambda \mathrm{E}(X_n|\mathcal{G})\, dP \\ &= \lim_{n\to\infty} \int_\Lambda X_n\, dP = \int_\Lambda \lim_{n\to\infty} X_n\, dP = \int_\Lambda X\, dP.\end{aligned}$$

Esto demuestra (a) y (b) sigue cambiando signos. Para demostrar (c) ponemos $Z_n = \inf_{k\geq n} X_k \leq X_n$ y observamos que $Z_n \stackrel{c.s.}{\to} \liminf X_n$ monotónicamente. Por la proposición anterior

$$\mathrm{E}(X_n|\mathcal{G}) \geq \mathrm{E}(Z_n|\mathcal{G}) \uparrow \mathrm{E}(\liminf_{n\to\infty} X_n|\mathcal{G})$$

de donde (c) sigue. Para (d) observamos que $Y - X_n \geq 0$ es integrable y podemos aplicar (c). Finalmente, para (e) usamos (c) y (d). ■

Proposición 5.3 *Si* X *y* XY *son integrables,* $Y \in \mathcal{G}$*, entonces*

$$\mathrm{E}(XY|\mathcal{G}) \stackrel{c.s.}{=} Y\,\mathrm{E}(X|\mathcal{G}).$$

Demostración. Supongamos inicialmente que X e Y son no-negativas. Para $Y = \mathbf{1}_A$, con $A \in \mathcal{G}$, también se tiene que $A \cap \Lambda \in \mathcal{G}$, para todo $\Lambda \in \mathcal{G}$ y usando (5.5)

$$\int_\Lambda Y\,\mathrm{E}(X|\mathcal{G})\,dP = \int_{\Lambda\cap A} \mathrm{E}(X|\mathcal{G})\,dP = \int_{\Lambda\cap A} X\,dP = \int_\Lambda XY\,dP$$

y esto demuestra la relación para indicadores, y por linealidad para variables simples. Si $\{Y_n,\ n \geq 1\}$ son v.a. simples tales que $Y_n \uparrow Y$ c.s. cuando $n \to \infty$, se tiene que $XY_n \uparrow XY$ y $Y_n\,\mathrm{E}(X|\mathcal{G}) \uparrow Y\,\mathrm{E}(X|\mathcal{G})$ c.s. cuando $n \to \infty$, y la conclusión sigue por convergencia monótona. El caso general sigue usando la descomposición $X = X^+ - X^-$ y $Y = Y^+ - Y^-$. ■

Muchas propiedades de martingalas se demuestran condicionando sucesivamente. El siguiente lema es fundamental para este procedimiento

Lema 5.1 (Suavizamiento) *Sea $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, entonces*

$$\mathrm{E}(\mathrm{E}(X|\mathcal{G})|\mathcal{H}) = \mathrm{E}(X|\mathcal{H}) = \mathrm{E}(\mathrm{E}(X|\mathcal{H})|\mathcal{G}) \quad c.s.$$

Demostración. Como $\mathrm{E}(X|\mathcal{H}) \in \mathcal{G}$ la segunda igualdad es consecuencia de la proposición 5.1 (c). Para demostrar la primera sea $\Lambda \in \mathcal{H} \subset \mathcal{G}$. Usando la relación (5.5) obtenemos

$$\int_\Lambda \mathrm{E}(\mathrm{E}(X|\mathcal{G})|\mathcal{H})\,dP = \int_\Lambda \mathrm{E}(X|\mathcal{G})\,dP = \int_\Lambda X\,dP = \int_\Lambda \mathrm{E}(X|\mathcal{H})\,dP$$

■

Para entender por qué el nombre del lema, recordemos el ejemplo 5.3 en el cual $\mathcal{G} = \sigma(\Lambda_n,\ n \geq 1)$, donde $\{\Lambda_n,\ n \geq 1\}$ es una partición numerable de Ω. Entonces

$$\mathrm{E}(X|\mathcal{G}) = \sum_{n=1}^{\infty} \alpha_n(X)\mathbf{1}_{\Lambda_n},$$

de modo que $\mathrm{E}(X|\mathcal{G})$ es constante en cada conjunto Λ_n.

Si $\mathcal{G}_1 \subset \mathcal{G}_2$ y ambas son generadas por particiones numerables $\{\Lambda_n^{(1)},\ n \geq 1\}$ y $\{\Lambda_n^{(2)},\ n \geq 1\}$, entonces $\Lambda_n^{(1)} \in \mathcal{G}_2$, de modo que existe un conjunto de índices $J \subset \mathbb{N}$ tal que $\Lambda_n^{(1)} = \sum_{j\in J} \Lambda_j^{(2)}$. Por lo tanto $\mathrm{E}(X|\mathcal{G}_1)$ es constante en $\Lambda_n^{(1)}$ pero $\mathrm{E}(X|\mathcal{G}_2)$ puede cambiar de valor a medida que ω se mueve entre los conjuntos $\Lambda_j^{(1)},\ j \in J$. Como función, $\mathrm{E}(X|\mathcal{G}_1)$ es más suave que $\mathrm{E}(X|\mathcal{G}_2)$.

Teorema 5.5 *Sea Y una v.a. con varianza finita y sea $\mathcal{G}$ una sub-σ-álgebra de $\mathcal{F}$. Entonces*

$$\mathrm{E}[(Y - \mathrm{E}(Y|\mathcal{G}))^2] = \mathrm{E}[Y^2] - \mathrm{E}[(\mathrm{E}(Y|\mathcal{G}))^2]$$

Demostración. Usando la proposición 5.3 y el lema 5.1,

$$\mathrm{E}(Y\,\mathrm{E}(Y|\mathcal{G})) = \mathrm{E}\left[\,\mathrm{E}(Y\,\mathrm{E}(Y|\mathcal{G}))|\mathcal{G}\right] = \mathrm{E}\left[\,\mathrm{E}(Y|\mathcal{G})\,\mathrm{E}(Y|\mathcal{G})\right] = \mathrm{E}\left[\,\mathrm{E}(Y|\mathcal{G})^2\right]$$

de modo que

$$\mathrm{E}[(Y - \mathrm{E}(Y|\mathcal{G}))^2] = \mathrm{E}[Y^2] + \mathrm{E}[(\mathrm{E}(Y|\mathcal{G}))^2] - 2\,\mathrm{E}[Y\,\mathrm{E}(Y|\mathcal{G})] = \mathrm{E}[Y^2] - \mathrm{E}[(\mathrm{E}(Y|\mathcal{G}))^2]$$

■

Proyecciones

Sea $\mathcal{G}$ una sub-σ-álgebra de $\mathcal{F}$. Sea $L^2(\mathcal{G})$ la clase de las variables aleatorias $\mathcal{G}$-medibles con segundo momento finito. Si $X \in L^2(\mathcal{F})$ entonces $\mathrm{E}(X|\mathcal{G})$ es la proyección de X a $L^2(\mathcal{G})$, un subespacio de $L^2(\mathcal{B})$.

La proyección de X en $L^2(\mathcal{G})$ es el (único) elemento de $L^2(\mathcal{G})$ en el cual se alcanza

$$\inf_{Z \in L^2(\mathcal{G})} ||X - Z||_2$$

Para alcanzar el ínfimo debe cumplirse que $X - Z$ sea ortogonal a todas las variables del subespacio $L^2(\mathcal{G})$:

$$\langle Y, X - Z \rangle = 0, \quad \forall\, Y \in L^2(\mathcal{G}).$$

Es decir que

$$\int Y(X - Z)\, dP = 0, \quad \forall\, Y \in L^2(\mathcal{G}).$$

Si probamos $Z = \mathrm{E}(X|\mathcal{G})$ obtenemos

$$\begin{aligned}\int Y(X - Z)\, dP &= \mathrm{E}(Y(X - \mathrm{E}(X|\mathcal{G}))) \\ &= \mathrm{E}(YX) - \mathrm{E}(Y\,\mathrm{E}(X|\mathcal{G})) \\ &= \mathrm{E}(YX) - \mathrm{E}(\mathrm{E}(YX|\mathcal{G})) \\ &= \mathrm{E}(YX) - \mathrm{E}(YX) = 0.\end{aligned}$$

5.3.2. Desigualdades de Momentos Condicionales.

Teorema 5.6 *Sea X e Y v.a. y supongamos que $\mathcal{G}$ es una sub-σ-álgebra $\mathcal{F}$. Las siguientes desigualdades de momentos valen c.s. siempre que los momentos correspondientes existan*

1. $$\mathrm{E}(|X + Y|^r|\mathcal{G}) \le c_r\big(\mathrm{E}(|X|^r|\mathcal{G}) + \mathrm{E}(|Y|^r|\mathcal{G})\big),$$
   donde $c_r = 1$ para $r \le 1$ y $c_r = 2^{r-1}$ para $r \ge 1$.

2. *Hölder. Si $1 < p < \infty$, $p^{-1} + q^{-1} = 1$,*
   $$|\,\mathrm{E}(XY|\mathcal{G})| \le \mathrm{E}(|XY||\mathcal{G}) \le (\mathrm{E}(|X|^p|\mathcal{G}))^{1/p} \cdot (\mathrm{E}(|Y|^q|\mathcal{G}))^{1/q}$$

3. *Minkowski*
   $$(\mathrm{E}(|X + Y|^p|\mathcal{G})^{1/p} \le (\mathrm{E}(|X|^p|\mathcal{G}))^{1/p} + (\mathrm{E}(|Y|^p|\mathcal{G})^{1/p}$$

4. *Jensen. Si g es una función convexa y $g(X) \in L^1$,*
   $$g(\mathrm{E}(X|\mathcal{G})) \le \mathrm{E}(g(X)|\mathcal{G}).$$

Demostración. Veamos la demostración de la desigualdad de Jensen. Consideremos la recta de soporte en x_0; por convexidad debe estar por debajo de la gráfica de g de modo que

$$g(x_0) + \lambda(x_0)(x - x_0) \le g(x) \tag{5.8}$$

donde $\lambda(x_0)$ es la pendiente de la recta de soporte que pasa por $(x_0, g(x_0))$. Reemplazamos x_0 por $\mathrm{E}(X|\mathcal{G})$ y x por X para obtener

$$g(\mathrm{E}(X|\mathcal{G})) + \lambda(\mathrm{E}(X|\mathcal{G}))(X - \mathrm{E}(X|\mathcal{G})) \le g(X) \tag{5.9}$$

Si no hay problemas de integrabilidad podemos tomar esperanza condicional respecto a $\mathcal{G}$ en ambos lados de (5.9). Para el lado izquierdo obtenemos

$$\begin{aligned} g(\mathrm{E}(X|\mathcal{G})) + \mathrm{E}[\lambda(\mathrm{E}(X|\mathcal{G}))(X - \mathrm{E}(X|\mathcal{G}))|\mathcal{G}] &= g(\mathrm{E}(X|\mathcal{G})) + \lambda(\mathrm{E}(X|\mathcal{G}))\,\mathrm{E}(X - \mathrm{E}(X|\mathcal{G}))|\mathcal{G}) \\ &= g(\mathrm{E}(X|\mathcal{G})) \end{aligned}$$

donde usamos que $\mathrm{E}(X - \mathrm{E}(X|\mathcal{G})|\mathcal{G}) = 0$. Por otro lado, al calcular la esperanza condicional del lado derecho de (5.9) obtenemos $\mathrm{E}(g(X)|\mathcal{G})$, y por lo tanto

$$g(\mathrm{E}(X|\mathcal{G})) \leq \mathrm{E}(g(X)|\mathcal{G}).$$

Para concluir veamos que no hay problemas de integrabilidad en (5.9). Observemos primero que podemos tomar $\lambda(x)$ como la derivada por la derecha

$$\lim_{h\downarrow 0} \frac{g(x+h) - g(x)}{h}$$

que siempre existe y por convexidad es no-decreciente en x. Si podemos demostrar que $\mathrm{E}(X|\mathcal{G})(\omega)$ es acotada como función de ω, entonces también lo serían $g(\mathrm{E}(X|\mathcal{G}))$ y $\lambda(\mathrm{E}(X|\mathcal{G}))$, de modo que todos los términos de (5.9) serían integrables. Sea ahora

$$X' = X\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}},$$

observamos que

$$\mathrm{E}(X'|\mathcal{G}) = \mathrm{E}(X\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}}|\mathcal{G}) = \mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}}\,\mathrm{E}(X|\mathcal{G})$$

es acotado y podemos usar el resultado de la desigualdad de Jensen para esperanzas condicionales acotadas. Obtenemos

$$g(\mathrm{E}(X'|\mathcal{G})) \leq \mathrm{E}(g(X')|\mathcal{G}).$$

Por lo tanto, cuando $n \to \infty$

$$\begin{aligned} \mathrm{E}\left(g(X')|\mathcal{G}\right) &= \mathrm{E}\left(g(X\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}})|\mathcal{G}\right) \\ &= \mathrm{E}\left(g(X\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}}) + g(0)\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|>n\}}|G\right) \\ &= \mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|\leq n\}}\,\mathrm{E}(g(X)|\mathcal{G}) + g(0)\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})|>n\}} \\ &\to \mathrm{E}(g(X)|\mathcal{G}). \end{aligned}$$

Además, cuando $n \to \infty$

$$g\big(\,\mathrm{E}(X'|\mathcal{G})\big) = g\big(\mathbf{1}_{\{|\,\mathrm{E}(X|\mathcal{G})\leq n\}}\,\mathrm{E}(X|\mathcal{G})\big) \to g(\mathrm{E}(X|\mathcal{G}))$$

ya que g es continua. ■

Proposición 5.4 *Si $X \in L^p$, definimos $||X||_p = (\mathrm{E}\,|X|^p)^{1/p}$ y supongamos que $p \geq 1$. Entonces*

$$||\,\mathrm{E}(X|\mathcal{G})||_p \leq ||X||_p. \tag{5.10}$$

Además, si $X_n \to X$ en L^p entonces $\mathrm{E}(X_n|\mathcal{G}) \to \mathrm{E}(X|\mathcal{G})$ en L^p.

Demostración. La desigualdad (5.10) vale si

$$(\mathrm{E}(|\,\mathrm{E}(X|\mathcal{G})|^p))^{1/p} \leq (\mathrm{E}(|X|^p))^{1/p}$$

o equivalentemente

$$\mathrm{E}(|\,\mathrm{E}(X|\mathcal{G})|^p) \leq \mathrm{E}(|X|^p).$$

A partir de la desigualdad de Jensen, si g es convexa, $g(\mathrm{E}(X|\mathcal{G})) \leq \mathrm{E}(g(X)|\mathcal{G})$. Como $g(x) = |x|^p$ es convexa para $p \geq 1$ obtenemos

$$\mathrm{E}(|\,\mathrm{E}(X|\mathcal{G})|^p) \leq \mathrm{E}(|X|^p).$$

Para ver la convergencia observamos que

$$||\,\mathrm{E}(X_n|\mathcal{G}) - \mathrm{E}(X|\mathcal{G})||_p = ||\,\mathrm{E}((X_n - X)|\mathcal{G})||_p \leq ||X_n - X||_p \to 0.$$

■

5.4. Martingalas.

Comenzamos con un espacio de probabilidad $(\Omega, \mathcal{F}, P)$.

Definición 5.8 Una sucesión $\{\mathcal{F}_n,\ n \geq 0\}$ de sub-σ-álgebras es una *filtración* si es una sucesión creciente, es decir,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots \subset \mathcal{F}.$$

Si interpretamos n como el tiempo, $\mathcal{F}_n$ contiene la información disponible al tiempo n. También consideramos $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

Definición 5.9 Una sucesión $\{X_n,\ n \geq 0\}$ de v.a. es *adaptada* a $\{\mathcal{F}_n,\ n \geq 1\}$ si $X_n \in \mathcal{F}_n$ para todo n. Si $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ decimos simplemente que la sucesión es *adaptada* y llamamos a $\{\mathcal{F}_n,\ n \geq 1\}$ la *filtración natural*.

Definición 5.10 Una sucesión $\{X_n,\ n \geq 0\}$ de v.a. es $\mathcal{F}_n$-*predecible* si $X_n \in \mathcal{F}_{n-1}$ para todo n.

Definición 5.11 Una sucesión $\{A_n,\ n \geq 0\}$ de v.a. es un *proceso creciente* si $A_0 = 0$, $A_n \nearrow$ y $\{A_n\}$ es predecible

Definición 5.12 Una sucesión $\{X_n,\ \mathcal{F}_n,\ n \geq 1\}$ donde X_n son v.a. integrables y $\{\mathcal{F}_n\}$ es una filtración, es una *martingala* si X_n es adaptada a $\mathcal{F}_n$ y se cumple que

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n \quad c.s. \quad \forall n \geq 0. \tag{5.11}$$

Decimos que la sucesión es una *submartingala* si

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) \geq X_n \quad c.s. \quad \forall n \geq 0. \tag{5.12}$$

y una supermartingala si

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) \leq X_n \quad c.s. \quad \forall n \geq 0. \tag{5.13}$$

Decimos que es una L^p-martingala si además $\mathrm{E}\,|X_n|^p < \infty$ para todo n. Decimos que es L^p-acotada si además $\sup_n \mathrm{E}\,|X_n|^p < \infty$.

Observación 5.2

1. $\{X_n\}$ es una martingala si es a la vez una submartingala y una supermartingala. $\{X_n\}$ es una supermartingala sii $\{-X_n\}$ es una submartingala.

2. La relación (5.11) vale sii
$$\int_\Lambda X_{n+1} = \int_\Lambda X_n, \quad \forall \Lambda \in \mathcal{F}_n.$$
Un comentario similar vale para sub y supermartingalas.

3. Podemos reemplazar la condición (5.11) por
$$\mathrm{E}(X_n|\mathcal{F}_m) = X_m \quad c.s. \quad \text{para cualesquiera } 0 \leq m < n. \tag{5.14}$$
Para ver esto basta usar repetidamente el lema de suavizamiento:
$$\begin{aligned}\mathrm{E}(X_n|\mathcal{F}_m) &= \mathrm{E}(\mathrm{E}(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m)\\ &= \mathrm{E}(X_{n-1}|\mathcal{F}_m) = \mathrm{E}(\mathrm{E}(X_{n-1}|\mathcal{F}_{n-2})|\mathcal{F}_m)\\ &= \mathrm{E}(X_{n-2}|\mathcal{F}_m) = \cdots = \mathrm{E}(X_{m+1}|\mathcal{F}_m) = X_m\end{aligned}$$

4. Si (X_n) es una martingala entonces $\mathrm{E}(X_n)$ es constante. En el caso de una submartingala, la media crece mientras que para una submartingala, decrece. Por ejemplo, para martingalas,
$$\mathrm{E}(X_m) = \mathrm{E}(\mathrm{E}(X_n|\mathcal{F}_m)) = \mathrm{E}(X_n).$$
5. Sea $\{X_n, \mathcal{F}_n,\ n \geq 0\}$ una (sub, super) martingala, y sea $\mathcal{G}_n = \sigma(X_0, X_1, \dots, X_n)$. Entonces $\{X_n, \mathcal{G}_n,\ n \geq 0\}$ también es una (sub, super) martingala.

   Para ver por qué es cierto esto observamos que como $X_n \in \mathcal{F}_n$, se tiene que $\mathcal{G}_n = \sigma(X_0, \dots, X_n) \subset \mathcal{F}_n$, y por suavizamiento
$$\mathrm{E}(X_{n+1}|\mathcal{G}_n) = \mathrm{E}(\mathrm{E}(X_{n+1}|\mathcal{F}_n)|\mathcal{G}_n) = \mathrm{E}(X_n|\mathcal{G}_n) = X_n.$$

Definición 5.13 Una sucesión integrable $\{U_n\}$ adaptada a $\{\mathcal{F}_n\}$ es una *sucesión de diferencias de martingala* si
$$\mathrm{E}(U_{n+1}|\mathcal{F}_n) = 0 \quad \forall\, n \geq 0. \tag{5.15}$$
$\{U_n\}$ es una *sucesión de diferencias de submartingala (supermartingala)* si
$$\mathrm{E}(U_{n+1}|\mathcal{F}_n) \geq (\leq) 0 \quad \forall\, n \geq 0. \tag{5.16}$$

Teorema 5.7 *Sea $\{U_n\}$ integrable y adaptada a$\{\mathcal{F}_n\}$ y sea $X_n = \sum_{k=0}^{n} U_k,\ n \geq 0$.*

(i) $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ es una martingala sii $\{(U_n, \mathcal{F}_n),\ n \geq 0\}$ es una sucesión de diferencias de martingala, una submartingala sii $\{(U_n, \mathcal{F}_n),\ n \geq 0\}$ es una sucesión de diferencias de submartingala, y una supermartingala sii $\{(U_n, \mathcal{F}_n),\ n \geq 0\}$ es una sucesión de diferencias de supermartingala.

(ii) Una sucesión de diferencias de martingala tiene esperanza constante 0, una sucesión de diferencias de submartingala tiene esperanza no-negativa, y una sucesión de diferencias de martingala tiene esperanza no-positiva.

5.4.1. Ejemplos.

Ejemplo 5.6
Sea $\{Y_n,\ n \geq 1\}$ una sucesión de v.a.i. con media 0 y sea $X_n = \sum_{k=1}^{n} Y_k$, $n \geq 1$ con $Y_0 = X_0 = 0$. Sea
$$\mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n) = \sigma(X_0, X_1, \dots, X_n), \quad n \geq 0.$$
Entonces $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ es una martingala y $\{(Y_n, \mathcal{F}_n),\ n \geq 1\}$ es una sucesión de diferencias de martingala:
$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = \mathrm{E}(X_n + Y_{n+1}|\mathcal{F}_n) = X_n + \mathrm{E}(Y_{n+1}|\mathcal{F}_n) = X_n \quad c.s.$$

Ejemplo 5.7
Sean $X \in L^1$ y $\{\mathcal{F}_n,\ n \geq 0\}$ una filtración. Para $n \geq 0$ definimos
$$X_n = \mathrm{E}(X|\mathcal{F}_n)$$
Entonces $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ es una martingala:
$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = \mathrm{E}(\mathrm{E}(X|\mathcal{F}_{n+1})|\mathcal{F}_n) = \mathrm{E}(X|\mathcal{F}_n) = X_n.$$

Ejemplo 5.8
Sea $\{Y_n,\ n \geq 1\}$ v.a.i. con $\mathrm{E}\, Y_k = \mu_k$, $\mathrm{Var}(Y_k) = \sigma_k^2$ y ponemos $s_n^2 = \sum_{k=1}^{n} \sigma_k^2$ para $n \geq 1$. Tomamos $\{\mathcal{F}_n,\ n \geq 1\}$ la filtración natural. Finalmente ponemos
$$X_n = \Big(\sum_{k=1}^{n}(Y_k - \mu_k)\Big)^2 - s_n^2, \quad n \geq 1.$$

Entonces $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ es una martingala.

Para ver esto podemos suponer, sin pérdida de generalidad, que todas las medias valen 0. tenemos

$$\begin{aligned}
\mathrm{E}(X_{n+1}|\mathcal{F}_n) &= \mathrm{E}\Big(\big(\sum_{k=1}^{n} Y_k + Y_{n+1}\big)^2 - s_{n+1}^2|\mathcal{F}_n\Big) \\
&= \mathrm{E}\Big(\big(\sum_{k=1}^{n} Y_k\big)^2|\mathcal{F}_n\Big) + \mathrm{E}(Y_{n+1}^2|\mathcal{F}_n) + 2\,\mathrm{E}\Big(\big(\sum_{k=1}^{n} Y_k\big)Y_{n+1}|\mathcal{F}_n\Big) - s_{n+1}^2 \\
&= \big(\sum_{k=1}^{n} Y_k\big)^2 + \sigma_{n+1}^2 + 2\big(\sum_{k=1}^{n} Y_k\big)\,\mathrm{E}(Y_{n+1}|\mathcal{F}_n) - s_n^2 - \sigma_{n+1}^2 \\
&= X_n + 2\big(\sum_{k=1}^{n} Y_k\big)\cdot 0 = X_n.
\end{aligned}$$

En particular, si $Y_1, Y_2, \dots$ son i.i.d. centradas, entonces $\{X_n = (\sum_{k=1}^{n} Y_k)^2 - n\sigma_1^2,\ n \geq 1\}$ es una martingala.

Ejemplo 5.9
Sea $\{Y_n,\ n \geq 1\}$ una sucesión de v.a.i. con media 1 y definimos $X_n = \prod_{k=1}^{n} Y_k,\ n \geq 1$, con $Y_0 = X_0 = 1$ y sea $\{\mathcal{F}_n,\ n \geq 1\}$ la filtración natural. Entonces $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ es una martingala porque

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = \mathrm{E}(X_n \cdot Y_{n+1}|\mathcal{F}_n) = X_n \cdot \mathrm{E}(Y_{n+1}|\mathcal{F}_n) = X_n \cdot 1 = X_n.$$

Ejemplo 5.10
Sea $X_0 = 1$ y para $n \geq 1$ definimos recursivamente

$$X_{n+1} = \begin{cases} 2X_n, & \text{con probabilidad } 1/2, \\ 0, & \text{con probabilidad } 1/2, \end{cases}$$

o equivalentemente

$$P(X_n = 2^n) = \frac{1}{2^n}, \qquad P(X_n = 0) = 1 - \frac{1}{2^n}.$$

Como

$$X_n = \prod_{k=1}^{n} Y_k,$$

donde $Y_1, Y_2, \dots$ son i.i.d. que valen 0 ó 2 con probabilidad 1/2, X_n es el producto de variables i.i.d. con media 1, y por el ejemplo anterior es una martingala.

Ejemplo 5.11
Sea $\{Y_n,\ n \geq 1\}$ una sucesión de v.a.i.i.d. con función generadora de momentos ψ finita, y sea $S_n = \sum_{k=1}^{n} Y_k,\ n \geq 1$. Entonces

$$X_n = \frac{e^{tS_n}}{\big(\psi(t)\big)^n} = \prod_{k=1}^{n} \frac{e^{tY_k}}{\psi(t)}, \quad n \geq 1,$$

para t dentro del rango de convergencia de la f.g.m., es una martingala que se conoce como la martingala exponencial. Esto es consecuencia del ejemplo 5.9, porque X_n es el producto de n factores independientes con media 1.

Ejemplo 5.12
Si $\{Y_n,\ n \geq 1\}$ son variables independientes con densidad común f, la sucesión de cocientes de verosimilitudes es

$$L_n = \prod_{k=1}^{n} \frac{f(Y_k;\theta_1)}{f(Y_k;\theta_0)}, \quad n \geq 0$$

donde θ_0 y θ_1 son los valores de algún parámetro bajo las hipótesis nula y alternativa, respectivamente. Esta sucesión es una martingala como la del ejemplo 5.9 bajo la hipótesis nula:

$$\mathrm{E}\Big(\frac{f(Y_k;\theta_1)}{f(Y_k;\theta_0)}\Big) = \int_{-\infty}^{\infty} \frac{f(y;\theta_1)}{f(y;\theta_0)} f(y;\theta_0)\,dy = \int_{-\infty}^{\infty} f(y;\theta_1)\,dy = 1.$$

Ejemplo 5.13
Sea $\{(X_n,\mathcal{F}_n),\ n\geq 0\}$ una martingala y $\{U_n,\ n\geq 0\}$ la sucesión de diferencias de martingala asociada. Sea $\{v_k,\ k\geq 0\}$ una sucesión predecible, ponemos $X_0=0$ y

$$X_n = \sum_{k=1}^{n} U_k v_k, \quad n\geq 1.$$

Una sucesión de este tipo se conoce como una transformada de martingala y es a su vez una martingala:

$$\begin{aligned}\mathrm{E}(X_{n+1}|\mathcal{F}_n) &= \sum_{k=1}^{n} \mathrm{E}(U_k v_k|\mathcal{F}_n) + \mathrm{E}(U_{n+1}v_{n+1}|\mathcal{F}_n)\\ &= X_n + v_{n+1}\,\mathrm{E}(U_{n+1}|\mathcal{F}_n) = X_n.\end{aligned}$$

Ejemplo 5.14
Cualquier sucesión integrable y adaptada puede 'ajustarse' para transformarla en una martingala. Sea $\{Y_n,\ n\geq 0\}$ una sucesión adaptada a $\{\mathcal{F}_n\}$ y pongamos $X_0=Y_0$ y

$$X_n = \sum_{k=1}^{n}(Y_k - \mathrm{E}(Y_k|\mathcal{F}_{k-1})), \qquad n\geq 1.$$

Por el lema de suavizamiento y el hecho de que $Y_k\in\mathcal{F}_n$ para $1\leq k\leq n$,

$$\begin{aligned}\mathrm{E}(X_{n+1}|\mathcal{F}_n) &= \mathrm{E}\,\Big(\sum_{k=1}^{n+1}(Y_k - \mathrm{E}(Y_k|\mathcal{F}_{k-1}))|\mathcal{F}_n\Big)\\ &= \sum_{k=1}^{n}\mathrm{E}\,\big(Y_k - \mathrm{E}(Y_k|\mathcal{F}_{k-1})|\mathcal{F}_n\big) + \mathrm{E}\,\big(Y_{n+1} - \mathrm{E}(Y_{n+1}|\mathcal{F}_n)|\mathcal{F}_n\big)\\ &= \sum_{k=1}^{n}(Y_k - \mathrm{E}(Y_k|\mathcal{F}_{k-1})) + \mathrm{E}(Y_{n+1}|\mathcal{F}_n) - \mathrm{E}(Y_{n+1}|\mathcal{F}_n)\\ &= X_n\end{aligned}$$

de modo que $\{(X_n,\mathcal{F}_n),\ n\geq 1\}$ es una martingala.

Como corolario obtenemos que las sumas parciales de cualquier sucesión integrable y adaptada puede descomponerse como la suma de una martingala mas la suma de las esperanzas condicionales:

$$\sum_{k=1}^{n} Y_k = X_n + \sum_{k=1}^{n}\mathrm{E}(Y_k|\mathcal{F}_{k-1}), \quad n\geq 1.$$

Si las variables $\{Y_n,\ n\geq 1\}$ son independientes con $\mathrm{E}\,Y_k=\mu_k$, las esperanzas condicionales se convierten en esperanzas ordinarias y la descomposición se reduce a

$$\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n}(Y_k-\mu_k) + \sum_{k=1}^{n}\mu_k, \quad n\geq 1.$$

Ejemplo 5.15
Sea $\{Y_n, n \geq 0\}$ una cadena de Markov con los enteros como espacio de estados y matriz de transición $P = (P_{ij})$. Sea f un eigenvector con eigenvalor asociado λ es decir,

$$Pf = \lambda f$$

o en términos de las componentes

$$\sum_j P_{ij} f(j) = \lambda f(i).$$

En términos de esperanzas tenemos

$$\mathrm{E}(f(Y_{n+1})|Y_n = i) = \lambda f(i)$$

o

$$\mathrm{E}(f(Y_{n+1})|Y_n) = \lambda f(Y_n)$$

y por la propiedad de Markov esto es

$$\mathrm{E}(f(Y_{n+1})|Y_n) = \mathrm{E}(f(Y_{n+1})|Y_0, Y_1, \dots, Y_n) = \lambda f(Y_n)$$

por lo tanto $(\lambda^{-n} f(Y_n), \sigma(Y_0, \dots, Y_n))$ para $n \geq 1$ es una martingala.

Un caso particular es el proceso de ramificación simple. Supongamos que $\{p_k,\ k \geq 0\}$ es la distribución de la descendencia, de modo que p_k es la probabilidad de que un individuo de la población tenga k descendientes. Sea $m = \sum_k k p_k$ el promedio de descendientes por individuo. Sea $\{Z^{(n)}(i), n \geq 0, i \geq 1\}$ una sucesión i.i.d. con función de probabilidad común $\{p_k\}$ y definimos de manera recursiva $Z_0 = 1$ y

$$Z_{n+1} = \begin{cases} Z^{(n)}(1) + \cdots Z^{(n)}(Z_n), & \text{si } Z_n > 0, \\ 0, & \text{si } Z_n = 0, \end{cases}$$

que representa el número de individuos en la generación $n+1$. Entonces Z_n es una cadena de Markov y

$$P_{ij} = P(Z_{n+1} = j | Z_n = i) = \begin{cases} \delta_{0j}, & \text{si } i = 0, \\ p_j^{*i}, & \text{si } i \geq 1, \end{cases}$$

donde, para $i \geq 1$, p_j^{*i} es la j-ésima componente de la i-ésima convolución de la sucesión $\{p_k\}$. Observamos que para $i \geq 1$,

$$\sum_{j=0}^{\infty} P_{ij} j = \sum_{j=1}^{\infty} p_j^{*i} j = im$$

mientras que para $i = 0$

$$\sum_{j=0}^{\infty} P_{ij} j = P_{00} \cdot 0 + 0 = 0 = im$$

Con $f(j) = j$ tenemos que $Pf = mf$ y por lo tanto el proceso $(m^{-n} Z_n, \sigma(Z_0, \dots, Z_n))$ para $n \geq 0$ es una martingala.

La siguiente proposición da algunos mecanismos para obtener nuevas martingalas a partir de ejemplos conocidos.

Proposición 5.5 *Sean $a, b \in \mathbb{R}$ y $\{(X_n^{(i)}, \mathcal{F}_n),\ n \geq 0\}$ para $i = 1, 2$ martingalas. Entonces*

(1) $\{(aX_n^{(1)} + bX_n^{(2)}, \mathcal{F}_n),\ n \geq 0\}$ es una martingala.

(2) $\{(\max\{X_n^{(1)}, X_n^{(2)}\}, \mathcal{F}_n),\ n \geq 0\}$ es una submartingala.

(2) $\{(\min\{X_n^{(1)}, X_n^{(2)}\}, \mathcal{F}_n),\ n \geq 0\}$ es una supermartingala.

Demostración. (1) es consecuencia de la linealidad de las esperanzas condicionales. Para ver (2), teniendo en cuenta que $\max\{X_n^{(1)}, X_n^{(2)}\} \geq X_n^{(1)}$ y $\max\{X_n^{(1)}, X_n^{(2)}\} \geq X_n^{(2)}$ tenemos

$$\mathrm{E}(\max\{X_{n+1}^{(1)}, X_{n+1}^{(2)}\}|\mathcal{F}_n) \geq \max\{\mathrm{E}(X_{n+1}^{(1)}|\mathcal{F}_n), \mathrm{E}(X_{n+1}^{(2)}|\mathcal{F}_n)\} = \max\{X_n^{(1)}, X_n^{(2)}\}$$

lo cual demuestra (2). Para ver (3) basta cambiar signos. ■

Proposición 5.6 *Si* $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ *es*
(a) una martingala y g *es una función convexa o*
(b) una submartingala y g *es convexa no-decreciente,*
y además $\mathrm{E}\,|g(X_n)| < \infty$ *para todo* n*, entonces* $\{(g(X_n), \mathcal{F}_n),\ n \geq 0\}$ *es una submartingala.*

Demostración. Sea $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ una martingala. Por convexidad

$$\mathrm{E}(g(X_{n+1}|\mathcal{F}_n) \geq g(\mathrm{E}(X_{n+1}|\mathcal{F}_n)) = g(X_n).$$

Para submartingalas la primera desigualdad es igual pero como $\mathrm{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ la segunda igualdad se convierte en una desigualdad $\geq$ si g es no-decreciente. ■

Teorema 5.8 *(a) Si* $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ *es una martingala, entonces* $\{(X_n^+, \mathcal{F}_n),\ n \geq 1\}$, $\{(X_n^-, \mathcal{F}_n),\ n \geq 1\}$ *y* $\{(|X_n|, \mathcal{F}_n),\ n \geq 1\}$ *son submartingalas.*
(b) Si $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ *es una martingala y* $\mathrm{E}\,|X_n|^p < \infty$ *para todo* n *y algún* $p > 1$*, entonces* $\{(|X_n|^p, \mathcal{F}_n),\ n \geq 1\}$ *es una submartingala.*
(c) Si $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ *es una submartingala, también lo es* $\{(X_n^+, \mathcal{F}_n),\ n \geq 1\}$
(d) Si $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ *es una submartingala no-negativa y* $\mathrm{E}\,|X_n|^p < \infty$ *para todo* n *y algún* $p \geq 1$ *entonces* $\{(|X_n|^p, \mathcal{F}_n),\ n \geq 1\}$ *es una submartingala.*

5.5. Ortogonalidad

Lema 5.2 *Sea* $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ *una martingala en* L^2 *con sucesión de diferencias de martingala* $\{U_n\}$.
(a) Se tiene

$$\mathrm{E}(U_m U_n) = \begin{cases} \mathrm{E}(U_m^2), & \text{para } n = m, \\ 0, & \text{en otro caso.} \end{cases}$$

(b) Para $m < n$

$$\begin{aligned} \mathrm{E}(U_n X_m) &= \mathrm{E}(U_n\,\mathrm{E}(X_n|\mathcal{F}_m)) = 0, \\ \mathrm{E}(X_n X_m) &= \mathrm{E}(X_m\,\mathrm{E}(X_n|\mathcal{F}_m)) = \mathrm{E}(X_m^2), \\ \mathrm{E}(X_n - X_m)^2 &= \mathrm{E}(X_n^2) - \mathrm{E}(X_m^2), \\ \mathrm{E}\Big(\sum_{k=m+1}^{n} U_k\Big)^2 &= \sum_{k=m+1}^{n} \mathrm{E}(U_k^2). \end{aligned}$$

(c) Si $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ *es una submartingala (super martingala) en* L^2*, las mismas relaciones valen con* $=$ *reemplazado por* $\geq$ $(\leq)$.

Demostración. La herramienta fundamental es el lema de suavizamiento.
(a) El resultado es inmediato en el caso $m = n$. Si $m < n$,

$$\mathrm{E}(U_n U_m) = \mathrm{E}(\mathrm{E}(U_n U_m|\mathcal{F}_m)) = \mathrm{E}(U_m\,\mathrm{E}(U_n|\mathcal{F}_m)) = \mathrm{E}(X_m \cdot 0) = 0.$$

(b) De manera similar,

$$\begin{aligned} \mathrm{E}(U_n X_m) &= \mathrm{E}(\mathrm{E}(U_n X_m|\mathcal{F}_m)) = \mathrm{E}(X_m\,\mathrm{E}(U_n|\mathcal{F}_m)) = \mathrm{E}(X_m \cdot 0) = 0, \\ \mathrm{E}(X_n X_m) &= \mathrm{E}(\mathrm{E}(X_n X_m|\mathcal{F}_m)) = \mathrm{E}(X_m\,\mathrm{E}(X_n|\mathcal{F}_m)) = \mathrm{E}(X_m^2). \end{aligned}$$

Usando este último resultado y

$$\mathrm{E}(X_n - X_m)^2 = \mathrm{E}(X_n^2) - 2\,\mathrm{E}(X_n X_m) + \mathrm{E}(X_m^2) = \mathrm{E}(X_n^2) - \mathrm{E}(X_m^2)$$

La última igualdad es una consecuencia de este resultado.

(c) Sigue de los resultados anteriores. ■

Observación 5.3 Para martingalas se puede reescribir la tercera relación de (b) como

$$\mathrm{E}(X_n^2) = \mathrm{E}(X_m^2) + \mathrm{E}(X_n - X_m)^2$$

que muestra que las martingalas tienen incrementos ortogonales, y además que $\mathrm{E}(X_n^2) \geq \mathrm{E}(X_m^2)$.

5.6. Descomposiciones

Teorema 5.9 (Desomposición de Doob) *Toda submartingala $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ puede descomponerse de manera única en la suma de una martingala $\{(M_n, \mathcal{F}_n),\ n \geq 0\}$ y un proceso creciente $\{(A_n, \mathcal{F}_n),\ n \geq 0\}$:*

$$X_n = M_n + A_n, \quad n \geq 0.$$

Demostración. Recordemos el ejemplo 5.14. Ponemos $M_0 = X_0$ de modo que $A_0 = X_0 - M_0 = 0$ y

$$M_n = \sum_{k=1}^{n} (X_k - \mathrm{E}(X_k|\mathcal{F}_{k-1})), \quad \text{y} \quad A_n = X_n - M_n,$$

de modo que $\{(M_n, \mathcal{F}_n),\ n \geq 0\}$ es una martingala.Falta ver que $\{(A_n, \mathcal{F}_n),\ n \geq 0\}$ es un proceso creciente, y ya sabemos que $A_0 = 0$. Además, el proceso A_n es predecible:

$$A_n = \sum_{k=1}^{n} \mathrm{E}(X_k|\mathcal{F}_{k-1}) - \sum_{k=1}^{n-1} X_k \in \mathcal{F}_{n-1}.$$

Finalmente,

$$\begin{aligned} A_{n+1} - A_n &= X_{n+1} - M_{n+1} - X_n + M_n = (X_{n+1} - X_n) - (M_{n+1} - M_n) \\ &= X_{n+1} - X_n - (X_{n+1} - \mathrm{E}(X_{n+1}|\mathcal{F}_n)) \\ &= \mathrm{E}(X_{n+1}|\mathcal{F}_n) - X_n \geq 0, \end{aligned}$$

por la definición de submartingala. Esto establece la existencia de la descomposición y falta ver que es única.

Supongamos que $X_n = M'_n + A'_n$ es otra descomposición. Como A_n es predecible,

$$\begin{aligned} A'_{n+1} - A'_n &= \mathrm{E}(A'_{n+1} - A'_n|\mathcal{F}_n) = \mathrm{E}\left((X_{n+1} - X_n) - (M'_{n+1} - M'_n)|\mathcal{F}_n\right) \\ &= \mathrm{E}(X_{n+1}|\mathcal{F}_n) - X_n - (M'_n - M'_n) = \mathrm{E}(X_{n+1}|\mathcal{F}_n) - X_n \\ &= A_{n+1} - A_n, \end{aligned}$$

y como $A_0 = A'_0 = 0$ esto demuestra la unicidad del proceso creciente. Pero

$$M_n = X_n - A_n = X_n - A'_n = M'_n,$$

y por lo tanto la martingala también es única. ■

Corolario 5.3 *Sea $\{X_n,\ n \geq 0\}$ una supermartingala. Existen una martingala $\{M_n,\ n \geq 0\}$ y un proceso decreciente $\{A_n,\ n \geq 0\}$ tales que*

$$X_n = M_n + A_n,$$

con $A_0 = 0$. Esta descomposición es única.

Ejemplo 5.16
Sea $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ una martingala en L^2, entonces $\{(X_n^2, \mathcal{F}_n),\ n \geq 0\}$ es una submartingala. Para ver esto basta tener en cuenta que $f(x) = x^2$ es una función convexa y la desigualdad de Jensen. Veamos cuál es su descomposición de Doob.

Ponemos $M_0 = X_0^2$ y

$$M_n = \sum_{k=1}^{n} (X_k^2 - \mathrm{E}(X_k^2|\mathcal{F}_{k-1})), \quad \text{y} \quad A_n = X_n^2 - M_n,$$

Sea $\{U_n,\ n \geq 0\}$ la sucesión de diferencias de martingala asociada a la martingala $\{X_n\}$. Sabemos que esta sucesión es centrada. Tenemos

$$X_k^2 = (X_{k-1} + U_k)^2 = X_{k-1}^2 + 2X_{k-1}U_k + U_k^2$$

y en consecuencia

$$\mathrm{E}(X_k^2|\mathcal{F}_{k-1}) = X_{k-1}^2 + 2X_{k-1}\,\mathrm{E}(U_k|\mathcal{F}_{k-1}) + \mathrm{E}(U_k^2|\mathcal{F}_{k-1}),$$

pero $\mathrm{E}(U_k|\mathcal{F}_{k-1}) = 0$ y entonces

$$\mathrm{E}(X_k^2|\mathcal{F}_{k-1}) = X_{k-1}^2 + \mathrm{E}(U_k^2|\mathcal{F}_{k-1}).$$

Ahora

$$\begin{aligned} A_{n+1} - A_n &= X_{n+1}^2 - X_n^2 - M_{n+1} + M_n = \mathrm{E}(X_{n+1}^2|\mathcal{F}_n) - X_n^2 \\ &= X_n^2 + \mathrm{E}(U_{m+1}^2|\mathcal{F}_n) - X_n^2 \\ &= \mathrm{E}(U_{m+1}^2|\mathcal{F}_n), \end{aligned}$$

y por lo tanto la descomposición de Doob para X_n^2 es

$$M_n = X_n^2 - \sum_{k=1}^{n} \mathrm{E}(U_k^2|\mathcal{F}_{k-1}), \qquad A_n = \sum_{k=1}^{n} \mathrm{E}(U_k^2|\mathcal{F}_{k-1})$$

El siguiente resultado, que presenta la descomposición de Krickeberg, lo enunciamos sin demostración, que puede verse en el libro *Probability. A Graduate Course* de A. Gut, p. 490.

Teorema 5.10 (Descomposición de Krickeberg) *(a) Para cualquier martingala* $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ *con* $\sup_n \mathrm{E}(X_n^+) < \infty$*, existen dos martingalas no-negativas* $\{(M_n^{(i)}, \mathcal{F}_n),\ n \geq 0\},\ i = 1, 2$ *tales que*

$$X_n = M_n^{(1)} - M_n^{(2)}.$$

(b) Para cualquier submartingala $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ *con* $\sup_n \mathrm{E}(X_n^+) < \infty$*, existen dos martingalas no-negativas* $\{(M_n^{(i)}, \mathcal{F}_n),\ n \geq 0\},\ i = 1, 2$ *y un proceso creciente* $\{(A - n, \mathcal{F}_n),\ n \geq 0\}$ *tales que*

$$X_n = M_n^{(1)} - M_n^{(2)} + A_n.$$

5.7. Tiempos de Paro

Definición 5.14 Una v.a. $T : \Omega \to \overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ es un *tiempo de paro* si $\{T \leq n\} \in \mathcal{F}_n$ para todo n.

Una v.a. constante con valor entero o $+\infty$ es un tiempo de paro. Podemos pensar que los tiempos de paro son el instante en el cual ocurre un evento aleatorio, con la convención de que toma el valor $+\infty$ si el evento nunca ocurre. Por ejemplo, supongamos que $(X_n)_{n\geq 0}$ es una martingala y nos interesa el primer instante en el cual vale al menos 12. Este instante es aleatorio y lo podemos describir como

$$T = \begin{cases} \text{ínf}_{n\geq 0}\{n : X_n \geq 12\} & \text{si } X_n \geq 12 \text{ para algún } n \in \mathbb{N}, \\ +\infty & \text{si no.} \end{cases}$$

Es decir,

$$T(\omega) = \operatorname*{ínf}_{n\geq 0}\{n : X_n(\omega) \geq 12\}$$

si $X_n(\omega) \geq 12$ para algún $n \in \mathbb{N}$ y $T(\omega) = +\infty$ si no. Observamos que el evento $\{\omega : T(\omega) \leq n\}$ se puede expresar como

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{X_k \geq 12\} \in \mathcal{F}_n$$

porque $\{X_k \geq 12\} \in \mathcal{F}_k \subset \mathcal{F}_n$ si $k \leq n$.

Un tiempo de paro es *acotado* si existe una constante c tal que $P(T \leq c) = 1$. Si T es un tiempo de paro finito denotamos por X_T a la variable

$$X_T(\omega) = X_{T(\omega)}(\omega),$$

es decir, toma el valor X_n siempre que $T = n$.

Teorema 5.11 *Sea T un tiempo de paro acotado por c y sea $\{X_n,\ n \geq 0\}$ una martingala. Entonces* $\mathrm{E}(X_T) = \mathrm{E}(X_0)$.

Demostración. Partimos de $X_T(\omega) = \sum_{n=0}^{\infty} X_n(\omega)\mathbf{1}_{\{T(\omega)=n\}}$. Por lo tanto, suponiendo sin pérdida de generalidad que c es entero,

$$\mathrm{E}(X_T) = \mathrm{E}\Big[\sum_{k=0}^{\infty} X_n \mathbf{1}_{\{T=n\}}\Big] = \mathrm{E}\Big[\sum_{k=0}^{c} X_n \mathbf{1}_{\{T=n\}}\Big] = \sum_{k=0}^{c} \mathrm{E}\big[X_n \mathbf{1}_{\{T=n\}}\big].$$

Como $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$ vemos que $\{T = n\} \in \mathcal{F}_n$, y obtenemos que la expresión anterior es

$$\begin{aligned} &= \sum_{k=0}^{c} \mathrm{E}\big[\mathrm{E}[X_c|\mathcal{F}_n]\mathbf{1}_{\{T=n\}}\big] = \sum_{k=0}^{c} \mathrm{E}\big[X_c \mathbf{1}_{\{T=n\}}\big] \\ &= \mathrm{E}\Big[X_c \sum_{n=0}^{c} \mathbf{1}_{\{T=n\}}\Big] = \mathrm{E}(X_c) = \mathrm{E}(X_0). \end{aligned}$$

■

Teorema 5.12 *Sea T un tiempo de paro acotado por $c \in \mathbb{N}$ y sea $\{X_n,\ n \geq 0\}$ una submartingala. Entonces* $\mathrm{E}(X_T) \leq \mathrm{E}(X_c)$.

Demostración. Es similar a la del teorema 5.11 y queda como ejercicio. ■

Definición 5.15 Sea T un tiempo de paro. Definimos la σ-álgebra $\mathcal{F}_T$ como

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n,\ \text{para todo } n\}$$

Teorema 5.13 *Si T es un tiempo de paro, $\mathcal{F}_T$ es una σ-álgebra.*

Demostración. Claramente $\varnothing$ y Ω están en $\mathcal{F}_T$. Si $A \in \mathcal{F}_T$ entonces

$$A^c \cap \{T \leq n\} = \{T \leq n\} \setminus (A \cap \{T \leq n\}),$$

y por lo tanto $A^c \in \mathcal{F}_T$. Además si $(A_i)_{i\geq 1}$ están en $\mathcal{F}_T$, entonces

$$\big(\cup_{i\geq 1} A_i\big) \cap \{T \leq n\} = \cup_{i\geq 1}\big(A_i \cap \{T \leq n\}\big) \in \mathcal{F}_n,$$

de modo que $\mathcal{F}_T$ es cerrada bajo complementos y uniones numerables, y por lo tanto es una σ-álgebra. ∎

Teorema 5.14 *Sean S, T tiempos de paro, con $S \leq T$. Entonces $\mathcal{F}_S \subset \mathcal{F}_T$.*

Demostración. Como $S \leq T$ tenemos $\{T \leq n\} \subset \{S \leq n\}$. Por lo tanto si $A \in \mathcal{F}_S$ se tiene

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\}$$

pero $A \cap \{S \leq n\} \in \mathcal{F}_n$ y $\{T \leq n\} \in \mathcal{F}_n$ de modo que $A \cap \{T \leq n\} \in \mathcal{F}_n$ y en consecuencia $A \in \mathcal{F}_T$. ∎

Sea ahora $\{X_n,\ n \geq 1\}$ una sucesión de v.a. adaptada a la filtración $\{\mathcal{F}_n\}$. Sea T un tiempo de paro con $P(T < \infty) = 1$. Entonces $X_T = \sum_{n\geq 0} X_n \mathbf{1}_{\{T=n\}}$, y tenemos el siguiente resultado:

Teorema 5.15 *X_T es $\mathcal{F}_T$-medible.*

Demostración. Sea $\Lambda \in \mathcal{B}$, queremos mostrar que $\{X_T \in \Lambda\} \in \mathcal{F}_T$, es decir, tenemos que ver que $\{X_T \in \Lambda\} \cap \{T \leq n\} \in \mathcal{F}_n$. Pero

$$\{X_T \in \Lambda\} \cap \{T \leq n\} = \bigcup_{k=0}^{n} \{X_T \in \Lambda\} \cap \{T = k\} = \bigcup_{k=0}^{n} \{X_k \in \Lambda\} \cap \{T = k\},$$

y $\{X_k \in \Lambda\} \cap \{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ para $k \leq n$. ∎

Teorema 5.16 (Teorema de Muestreo Opcional de Doob) *Sea $\{X_n,\ n \geq 0\}$ una martingala y sean S, T tiempos de paro acotados por una constante c, con $S \leq T$ c.s. Entonces*

$$\mathrm{E}(X_T|\mathcal{F}_S) = X_S \quad c.s.$$

Demostración. En primer lugar $|X_T| \leq \sum_{n=0}^{c} |X_n|$ es integrable, y lo mismo ocurre para X_S, y además X_S es $\mathcal{F}_S$-medible por el teorema anterior. Falta demostrar que para todo $A \in \mathcal{F}_S$

$$\mathrm{E}(X_S \mathbf{1}_A) = \int_A X_T\, dP = \int_A X_S\, dP = \mathrm{E}(X_T \mathbf{1}_A)$$

Definimos una nueva variable aleatoria R por

$$R(\omega) = S(\omega)\mathbf{1}_A(\omega) + T(\omega)\mathbf{1}_{A^c}(\omega).$$

Entonces R también es un tiempo de paro:

$$\{R \leq n\} = \big(A \cap \{S \leq n\}\big) \cup \big(A^c \cap \{T \leq n\}\big),$$

y $A \cap \{S \leq n\} \in \mathcal{F}_n$ porque $A \in \mathcal{F}_S$. Como $A \in \mathcal{F}_S$ tenemos $A^c \in \mathcal{F}_S \subset \mathcal{F}_T$. En consecuencia $A^c \cap \{T \leq n\} \in \mathcal{F}_n$ y concluimos que $\{R \leq n\} \in \mathcal{F}_n$ y R es un tiempo de paro. Por lo tanto $\mathrm{E}(X_R) = \mathrm{E}(X_T) = \mathrm{E}(X_0)$. Pero

$$\mathrm{E}(X_R) = \mathrm{E}(X_S \mathbf{1}_A + X_T \mathbf{1}_{A^c}),$$
$$\mathrm{E}(X_T) = \mathrm{E}(X_T \mathbf{1}_A + X_T \mathbf{1}_{A^c}),$$

restando obtenemos

$$\mathrm{E}(X_S \mathbf{1}_A) - \mathrm{E}(X_T \mathbf{1}_A) = 0.$$

∎

Teorema 5.17 *Sea $\{X_n,\ n \geq 0\}$ una sucesión de v.a. adaptada a la filtración $\{\mathcal{F}_n,\ n \geq 0\}$. Supongamos que $\mathrm{E}\,|X_n| < \infty$ para todo n y $\mathrm{E}(X_T) = \mathrm{E}(X_0)$ para todo tiempo de paro acotado T. Entonces X es una martingala.*

Demostración. Sea $0 \leq m < n < \infty$, y sea $A \in \mathcal{F}_m$ Definimos un tiempo aleatorio por

$$T(\omega) = \begin{cases} m & \text{si } \omega \in A^c, \\ n & \text{si } \omega \in A. \end{cases}$$

Entonces T es un tiempo de paro, de modo que

$$\mathrm{E}(X_0) = \mathrm{E}(X_T) = \mathrm{E}(X_m \mathbf{1}_{A^c} + X_n \mathbf{1}_A).$$

Pero también $\mathrm{E}(X_0) = \mathrm{E}(X_m \mathbf{1}_{A^c} + X_m \mathbf{1}_A)$. Restando obtenemos $\mathrm{E}(X_n \mathbf{1}_A) = \mathrm{E}(X_m \mathbf{1}_A)$ o equivalentemente $\mathrm{E}(X_n | \mathcal{F}_m) = X_m$ c.s. ■

5.8. Desigualdades.

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\{\mathcal{F}_n,\ n \geq 0\}$ una filtración. Sea $\{X_n,\ n \geq 0\}$ una sucesión de v.a. integrables con $X_n \in \mathcal{F}_n$. Sea $W_n = \sup_{j \leq n} |X_j|$. Observamos que $W_n \leq W_{n+1}$ y si X_n es una martingala, W_n es una submartingala ya que

$$\mathrm{E}(W_n) \leq \mathrm{E}\,\Big[\sum_{j=1}^{n} |X_j|\Big] < \infty.$$

La desigualdad de Markov dice que, para $\alpha > 0$,

$$P(W_n \geq \alpha) = \mathrm{E}[\mathbf{1}_{\{W_n \geq \alpha\}}] \leq \frac{1}{\alpha}\,\mathrm{E}(W_n).$$

En el caso de una martingala podemos reemplazar W_n por $|X_n|$ en el lado derecho.

Teorema 5.18 (Primera Desigualdad de Doob) *Sea $\{X_n,\ n \geq 1\}$ una martingala o una submartingala positiva. Entonces*

$$P(W_n \geq \alpha) \leq \frac{1}{\alpha}\,\mathrm{E}(|X_n|).$$

Demostración. Sea $T = \min\{j : |X_j| \geq \alpha\}$ con la convención de que el mínimo de un conjunto vacío es $+\infty$. Como $g(x) = |x|$ es convexa y creciente en $\mathbb{R}^+$, tenemos que $|X_n|$ es una submartingala. Teniendo en cuenta que

$$\{T \leq n\} = \{W_n \geq \alpha\}$$

tenemos

$$P(W_n \geq \alpha) \leq P(T \leq n) \leq \mathrm{E}(\mathbf{1}_{\{T \leq n\}}) \leq \mathrm{E}\,\Big[\frac{|X_T|}{\alpha}\mathbf{1}_{\{T \leq n\}}\Big],$$

y como $X_T = X_{T \wedge n}$ en $\{T \leq n\}$,

$$P(W_n \geq \alpha) \leq \frac{1}{\alpha}\,\mathrm{E}[|X_{T\wedge n}|\mathbf{1}_{\{T \leq n\}}] \leq \frac{1}{\alpha}\,\mathrm{E}[|X_{T\wedge n}|] \leq \frac{1}{\alpha}\,\mathrm{E}[|X_n|]$$

por el teorema 5.12. ■

Lema 5.3 *Sea $X \geq 0$ una v.a., $p > 0$ y $\mathrm{E}[X^p] < \infty$. Entonces*

$$\mathrm{E}[X^p] = \int_0^\infty p\lambda^{p-1} P(X > \lambda)\, d\lambda.$$

Demostración. Tenemos

$$\int_0^\infty p\lambda^{p-1} P(X > \lambda)\, d\lambda = \int_0^\infty p\lambda^{p-1}\, \mathrm{E}(\mathbf{1}_{\{X>\lambda\}})\, d\lambda,$$

y por el teorema de Fubini

$$= \mathrm{E}\Big[\int_0^\infty p\lambda^{p-1} \mathbf{1}_{\{X>\lambda\}}\, d\lambda\Big] = \mathrm{E}\Big[\int_0^X p\lambda^{p-1}\, d\lambda\Big] = \mathrm{E}[X^p].$$

■

Teorema 5.19 (Segunda Desigualdad de Doob) *Sea $\{X_n,\ n \geq 1\}$ una martingala o una submartingala positiva. Sea $1 < p < \infty$. Existe una constante c_p que depende únicamente de p tal que*

$$\mathrm{E}[(W_n)^p] \leq c_p\, \mathrm{E}[|X_n|^p].$$

Demostración. Daremos la demostración para el caso de una martingala. Como $g(x) = |x|$ es convexa, $|X_n|$ es una submartingala. Sea $\alpha > 0$ y $Y_n = X_n \mathbf{1}_{\{|X_n|>\alpha/2\}}$. Para n fijo definimos

$$Z_j = \mathrm{E}[Y_n|\mathcal{F}_j], \quad 0 \leq j \leq n.$$

Observamos que $\{Z_j, 0 \leq j \leq n\}$ es una martingala (ver ejemplo 4.7) y además que $W_n \leq Z_n^* + \frac{\alpha}{2}$ con $Z_n^* = \max_{1\leq j\leq n} |Z_j|$ ya que

$$\begin{aligned}
|X_j| &= |\,\mathrm{E}(X_n|\mathcal{F}_j)| = |\,\mathrm{E}(X_n\mathbf{1}_{\{|X_n|>\alpha/2\}} + X_n\mathbf{1}_{\{|X_n|\leq\alpha/2\}}|\mathcal{F}_j)| \\
&= |\,\mathrm{E}(Y_n + X_n\mathbf{1}_{\{|X_n|\leq\alpha/2\}}|\mathcal{F}_j)| \\
&\leq |\,\mathrm{E}(Y_n|\mathcal{F}_j)| + \frac{\alpha}{2} = |Z_j| + \frac{\alpha}{2}.
\end{aligned}$$

Por la primera desigualdad de Doob tenemos

$$P(W_n > \alpha) \leq P\big(Z_n^* > \frac{\alpha}{2}\big) \leq \frac{2}{\alpha}\,\mathrm{E}(|Z_n|) \leq \frac{2}{\alpha}\,\mathrm{E}(|Y_n|) = \frac{2}{\alpha}\,\mathrm{E}[|X_n|\mathbf{1}_{\{|X_n|>\alpha/2\}}]$$

Por el lema 5.3 y usando el teorema de Fubini tenemos

$$\begin{aligned}
\mathrm{E}[W_n^p] &= \int_0^\infty p\lambda^{p-1} P(M_n > \lambda)\, d\lambda \\
&\leq \int_0^\infty 2p\lambda^{p-2}\, \mathrm{E}[|X_n|\mathbf{1}_{\{|X_n|>\lambda/2\}}]\, d\lambda \\
&= \mathrm{E}\Big[\int_0^{2|X_n|} 2p\lambda^{p-2}\, d\lambda\, |X_n|\Big] \\
&= \frac{2^p p}{p-1}\, \mathrm{E}[|X_n|^p].
\end{aligned}$$

■

En la demostración vimos que $c_p \leq 2^p p/(p-1)$. Es posible demostrar que $c_p^{1/p} = p/(p-1)$, lo que permite reescribir el teorema de la siguiente manera

Teorema 5.20 (Doob) *Sea $\{X_n,\ n \geq 1\}$ una martingala o una submartingala positiva. Sea $1 < p < \infty$. Entonces*

$$(\mathrm{E}[(W_n)^p])^{1/p} \leq \frac{p}{p-1}(\mathrm{E}[|X_n|^p])^{1/p}.$$

Para la última desigualdad de esta sección introducimos la noción de cruces hacia arriba (*upcrossings*). Sea $\{X_n,\ n \geq 0\}$ una submartingala y sea $a < b$. El número de cruces hacia arriba del intervalo $[a, b]$ es el número de veces que el proceso pasa de estar por debajo de a a estar por encima de b en un tiempo posterior. Esta idea se puede expresar de manera simple con tiempos de paro. Definimos $T_0 = 0$ e inductivamente para $j \geq 0$

$$S_{j+1} = \text{mín}\{k > T_j : X_k \leq a\}, \qquad T_{j+1} = \text{mín}\{k > S_{j+1} : X_k \geq b\} \tag{5.17}$$

con la convención usual de que el mínimo de un conjunto vacío es $+\infty$. Tomando como convención que el máximo de un conjunto vacío es 0, definimos

$$U_n = \text{máx}\{j : T_j \leq n\}, \tag{5.18}$$

U_n es el número de cruces hacia arriba de $[a, b]$ al tiempo n

Teorema 5.21 (Desigualdad de Doob para cruces hacia arriba) *Sea $\{X_n,\ n \geq 0\}$ una submartingala, sean $a < b$ y sea U_n el número de cruces hacia arriba de $[a, b]$ al tiempo n según la definición (5.18). Entonces*

$$\mathrm{E}[U_n] \leq \frac{1}{b-a}\,\mathrm{E}[(X_n - a)^+]$$

donde $(X_n - a)^+ = \text{máx}\{X_n - a, 0\}$.

Demostración. Sea $Y_n = (X_n - a)^+$. Como la función $\varphi(x) = (x - a)^+$ es convexa y no-decreciente, tenemos que (Y_n) es una submartingala. Como $S_{n+1} > n$ obtenemos

$$Y_n = Y_{S_1 \wedge n} + \sum_{i=1}^{n}(Y_{T_i \wedge n} - Y_{S_i \wedge n}) + \sum_{i=1}^{n}(Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}) \tag{5.19}$$

Cada cruce hacia arriba de (X_n) entre los tiempos 0 y n corresponde a un entero i tal que $S_i < T_i \leq n$, con $Y_{S_i} = 0$ y $Y_{T_i} = Y_{T_i \wedge n} \geq b - a$ mientras que $Y_{T_i \wedge n} - Y_{S_i \wedge n} \geq 0$ por construcción para todo i. Por lo tanto

$$\sum_{i=1}^{n}(Y_{T_i \wedge n} - Y_{S_i \wedge n}) \geq (b - a)U_n.$$

Por (5.19) obtenemos

$$(b - a)U_n \leq Y_n - Y_{S_i \wedge n} - \sum_{i=1}^{n}(Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}),$$

y como $Y_{S_1 \wedge n} \geq 0$ obtenemos

$$(b - a)U_n \leq Y_n - \sum_{i=1}^{n}(Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}),$$

Tomamos esperanzas en ambos lados: Como (Y_n) es una submartingala, los tiempos de paro $T_i \wedge n$ y $S_{i+1} \wedge n$ están acotados y $T_i \wedge n \leq S_{i+1} \wedge n$ tenemos $\mathrm{E}[Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}] \geq 0$ y por lo tanto

$$(b - a)\,\mathrm{E}[U_n] \leq \mathrm{E}[Y_n].$$

■

5.9. Teoremas de Convergencia.

Teorema 5.22 (Teorema de Convergencia de Martingalas) *Sea $\{X_n,\ n \geq 1\}$ una submartingala tal que $\sup_n \mathrm{E}(X_n^+) < \infty$. Entonces $\lim_{n\to\infty} X_n = X$ existe c.s. y es finito c.s. Además, $X \in L^1$.*

Observación 5.4 El teorema no dice que hay convergencia en L^1. Esto no es cierto en general.

Demostración. Sea U_n el número de cruces hacia arriba de $[a,b]$ antes de n, entonces U_n es no-decreciente y por lo tanto $U(a,b) = \lim_{n\to\infty} U_n$ existe. Por el teorema de convergencia monótona

$$\begin{aligned}
\mathrm{E}[U(a,b)] &= \lim_{n\to\infty} \mathrm{E}[U_n] \\
&\leq \frac{1}{b-a} \sup_n \mathrm{E}[(X_n - a)^+] \\
&\leq \frac{1}{b-a} \big(\sup_n \mathrm{E}[X_n^+] + |a| \big) \leq \frac{c}{b-a} < \infty
\end{aligned}$$

para alguna constante c que por hipótesis satisface $c < \infty$. La primera desigualdad viene del teorema 5.21 y la segunda de $(x-a)^+ \leq x^+ + |a|$ para $a, x \in \mathbb{R}$. Como $\mathrm{E}[U(a,b)] < \infty$, tenemos $P(U(a,b) < \infty) = 1$. En consecuencia X_n cruza $[a,b]$ hacia arriba sólo un número finito de veces c.s. y si ponemos

$$\Lambda_{a,b} = \{\limsup_{n\to\infty} X_n \geq b;\ \liminf_{n\to\infty} X_n \leq a\},$$

entonces $P(\Lambda_{a,b}) = 0$. Sea

$$\Lambda = \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \Lambda_{a,b},$$

entonces $P(\Lambda) = 0$ ya que los pares de racionales son numerables. Pero

$$\Lambda = \{\limsup_n X_n > \liminf_n X_n\},$$

y concluimos que $\lim_{n\to\infty} X_n$ existe c.s.

Veamos que el límite es finito. Como X_n es una submartingala, $\mathrm{E}(X_n) \geq \mathrm{E}(X_0)$, y en consecuencia,

$$\mathrm{E}(|X_n|) = \mathrm{E}(X_n^+) + \mathrm{E}(X_n^-) = 2\,\mathrm{E}(X_n^+) - \mathrm{E}(X_n) \leq 2\,\mathrm{E}(X_n^+) - \mathrm{E}(X_0), \tag{5.20}$$

y por lo tanto

$$\mathrm{E}(\lim_n |X_n|) \leq \liminf_{n\to\infty} \mathrm{E}(|X_n|) \leq 2 \sup_n \mathrm{E}(X_n^+) - \mathrm{E}(X_0) < \infty,$$

donde hemos usado el lema de Fatou, (5.20) y la hipótesis de que $\sup_n \mathrm{E}(X_n^+) < \infty$. Por lo tanto X_n converge c.s. a un límite finito X. Hemos mostrado además que $\mathrm{E}(|X|) = \mathrm{E}(\lim_n |X_n|) < \infty$, de modo que $X \in L^1$. ∎

Corolario 5.4 *Si X_n es una supermartingala no-negativa, o una martingala acotada superior o inferiormente, entonces $\lim_{n\to\infty} X_n = X$ existe c.s. y $X \in L^1$.*

Demostración. Si X_n es una supermartingala no-negativa, $(-X_n)$ es una submartingala acotada superiormente por 0 y podemos usar el teorema anterior.

Si X_n es una martingala acotada inferiormente, entonces $X_n \geq -c$ c.s. para todo n para alguna constante $c > 0$. Sea $Y_n = X_n + c$, entonces Y_n es una martingala no-negativa y por lo tanto también una supermartingala no-negativa y podemos aplicar lo que hemos demostrado del corolario. Si X_n es una martingala acotada superiormente entonces $-X_n$ es una martingala acotada inferiormente. ∎

Ejemplo 5.17
Consideremos un paseo al azar simple: $X_n = \sum_1^n Y_k$ donde las variables $Y_k,\ k \geq 1$ son independientes con distribución de Bernoulli de parámetro 1/2. No es difícil demostrar que que el paseo al azar $\{X_n,\ n \geq 1\}$ no converge porque oscila con excursiones que se alejan cada vez más del origen, pero regresa al origen infinitas veces.

Por el TCL sabemos que $X_n/\sqrt{n} \to \mathcal{N}(0,1)$ en distribución y en L^1 y por lo tanto $\mathrm{E}\,|X_n| \sim (2n/\pi)^{1/2}$ cuando $n \to \infty$.

Para obtener convergencia en L^1 necesitamos la hipótesis de integrabilidad uniforme. Recordemos la definición de este concepto.

Definición 5.16 Una colección de variables aleatorias $\mathcal{H} \subset L^1$ es *uniformemente integrable* si

$$\lim_{a\to\infty} \sup_{X\in\mathcal{H}} \mathrm{E}[|X|\mathbf{1}_{\{|X|\geq a\}}] = 0.$$

Recordamos también las siguientes condiciones suficientes para integrabilidad uniforme, que están contenidas en la sección 1.6.

Teorema 5.23 *Sea $\mathcal{H}$ una clase de variables aleatorias*

a) Si $\sup_{X\in\mathcal{H}} \mathrm{E}(|X|^p) < \infty$ para algún $p > 1$, entonces $\mathcal{H}$ es uniformemente integrable.

b) Si existe una v.a. Y tal que $|X| \leq Y$ c.s. para toda $X \in \mathcal{H}$ y $\mathrm{E}(Y) < \infty$, entonces $\mathcal{H}$ es uniformemente integrable.

Definición 5.17 Decimos que una martingala $\{(X_n, \mathcal{F}_n),\ n \geq 1\}$ es *cerrada, regular o completa* si existe una variable aleatoria Y con $\mathrm{E}(|Y|) < \infty$ y $X_n = \mathrm{E}(Y|\mathcal{F}_n)$ para todo n.

Teorema 5.24 (Teorema de Convergencia de Martingalas) *a) Sea $\{X_n,\ n \geq 1\}$ una martingala uniformemente integrable. Entonces*

$$\lim_{n\to\infty} X_n = X_\infty \quad \textit{existe c.s.}$$

$X_\infty \in L^1$ y X_n converge a X_∞ en L^1. Además $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$.

b) Recíprocamente sea $Y \in L^1$ y consideremos la martingala $X_n = \mathrm{E}(Y|\mathcal{F}_n)$. Entonces $\{X_n,\ n \geq 1\}$ es uniformemente integrable.

En otras palabras, la martingala es regular si y sólo si es uniformemente integrable.

Demostración.

a) Como (X_n) es uniformemente integrable, para $\varepsilon > 0$ existe c tal que $\sup_n \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|\geq c\}}) \leq \varepsilon$. Por lo tanto

$$\mathrm{E}(|X_n|) = \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|\geq c\}}) + \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|<c\}}) \leq \varepsilon + c.$$

Así $(X_n)_{n\geq 1}$ está acotada en L^1 y en consecuencia $\sup_n \mathrm{E}(X_n^+) < \infty$. Por el teorema 5.22 tenemos que

$$\lim_{n\to\infty} X_n = X_\infty$$

existe c.s. y $X_\infty \in L^1$.

Para ver que X_n converge a X_∞ en L^1, definimos

$$f_c(x) = \begin{cases} c & \text{si } x > c, \\ x & \text{si } |x| \leq c, \\ -c & \text{si } x < -c. \end{cases}$$

Entonces f_c es Lipschitz. Por integrabilidad uniforme existe c suficientemente grande tal que para $\varepsilon > 0$ dado

$$\mathrm{E}(|f_c(X_n) - X_n|) < \frac{\varepsilon}{3} \quad \text{para todo } n, \tag{5.21}$$

$$\mathrm{E}(|f_c(X_\infty) - X_\infty|) < \frac{\varepsilon}{3}. \tag{5.22}$$

Como $\lim X_n = X_\infty$ c.s. tenemos $\lim_{n\to\infty} f_c(X_n) = f_c(X_\infty)$ y por el TCD tenemos para $n \geq N$, N suficientemente grande

$$\mathrm{E}(|f_c(X_n) - f_c(X_\infty)|) < \frac{\varepsilon}{3} \tag{5.23}$$

Usando (5.21), (5.22) y (5.23) tenemos

$$\mathrm{E}(|X_n - X_\infty|) < \varepsilon, \quad \text{para } n \geq N.$$

Por lo tanto $X_n \to X_\infty$ en L^1. Falta demostrar que $\mathrm{E}(X_\infty|\mathcal{F}_n) = X_n$. Sea $\Lambda \in \mathcal{F}_m$ y $n \geq m$, entonces

$$\mathrm{E}(X_n \mathbf{1}_\Lambda) = \mathrm{E}(X_m \mathbf{1}_\Lambda)$$

por la propiedad de martingala. Sin embargo,

$$|\,\mathrm{E}(X_n \mathbf{1}_\Lambda) - \mathrm{E}(X_\infty \mathbf{1}_\Lambda)| \leq \mathrm{E}(|X_n - X_\infty|\mathbf{1}_\Lambda) \leq \mathrm{E}(|X_n - X_\infty|)$$

que tiende a 0 cuando $n \to \infty$. Por lo tanto $\mathrm{E}(X_m \mathbf{1}_\Lambda) = \mathrm{E}(X_\infty \mathbf{1}_\Lambda)$ y $\mathrm{E}(X_\infty|\mathcal{F}_n) = X_n$ c.s.

b) Ya sabemos que $\{X_n,\ n \geq 1\}$ es una martingala. Si $c > 0$ tenemos

$$X_n \mathbf{1}_{\{|X_n| \geq c\}} = E(Y|\mathcal{F}_n)\mathbf{1}_{\{|X_n| \geq c\}} = \mathrm{E}(Y \mathbf{1}_{\{|X_n| \geq c\}}|\mathcal{F}_n),$$

porque $\{|X_n| \geq c\} \in \mathcal{F}_n$. para cualquier $d > 0$ tenemos

$$\begin{aligned}
\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n| \geq c\}}) &\leq \mathrm{E}(|Y|\mathbf{1}_{\{|X_n| \geq c\}}) \\
&\leq \mathrm{E}(|Y|\mathbf{1}_{\{|Y| > d\}}) + dP(|X_n| \geq c) \\
&\leq \mathrm{E}(|Y|\mathbf{1}_{\{|Y| > d\}}) + \frac{d}{c}\,\mathrm{E}(|X_n|) \\
&\leq \mathrm{E}(|Y|\mathbf{1}_{\{|Y| > d\}}) + \frac{d}{c}\,\mathrm{E}(|Y|).
\end{aligned} \tag{5.24}$$

Tomamos $\varepsilon > 0$ y escogemos d de modo que el primer término de (5.24) sea menor que $\varepsilon/2$, y luego escogemos c de modo que el segundo término también sea menor que $\varepsilon/2$. Hemos demostrado que $\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n| > c\}}) \leq \varepsilon$. ■

Corolario 5.5 *Sea $\{\mathcal{F}_n\ n \geq 0\}$ una filtración y $\mathcal{F}_\infty = \sigma(\cup_{n\geq 0}\mathcal{F}_n)$ la σ-álgebra generada por la filtración. Si $Y \in L^1(\mathcal{F}_\infty)$ entonces*

$$\lim_{n\to\infty} \mathrm{E}[Y|\mathcal{F}_n] = Y$$

donde el límite es c.s. y en L^1.

Demostración.

Sea $X_n = \mathrm{E}(Y|\mathcal{F}_n)$, entonces X_n es una martingala uniformemente integrable. Observemos que si $\Lambda \in \mathcal{F}_m$

$$\int_\Lambda Y\,dP = \int_\Lambda X_n\,dP \to \int_\Lambda X_\infty\,dP$$

cuando $n \to \infty$. Esto permite demostrar que

$$\int_\Lambda Y\,dP = \int_\Lambda X_\infty\,dP \quad \forall \Lambda \in \mathcal{F}_\infty$$

y en consecuencia

$$X_\infty = \mathrm{E}(Y|\mathcal{F}_\infty) = Y.$$

porque $Y \in \mathcal{F}_\infty$ ■

Ejemplo 5.18
Recordemos el ejemplo 4.10 en el cual se duplica la apuesta cuando se pierde: $X_0 = 0$ y

$$X_{n+1} = \begin{cases} 2X_n, & \text{con probabilidad } 1/2, \\ 0 & \text{con probabilidad } 1/2 \end{cases}$$

Vimos que $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ es una martingala multiplicativa con media 1.

Como $P(X_n = 2^n) = 1 - P(X_n = 0) = 2^{-n}$, una aplicación del lema de Borel-Cantelli muestra que $X_n \to 0$ c.s. cuando $n \to \infty$. Pero $\mathrm{E}(X_n = 1)$ para todo n, el valor esperado no converge a 0, de modo que la martingala no es uniformemente integrable y no es regular.

Una aplicación del teorema de convergencia de martingalas es la siguiente.

Teorema 5.25 (Kolmogorov) *Sea $(Y_n,\ n \geq 1)$ v.a.i. centradas con $\mathrm{E}(Y_n^2) < \infty$ para todo n. Supongamos que $\sum_{n=1}^\infty \mathrm{E}(Y_n^2) < \infty$. Sea $S_n = \sum_{j=1}^n Y_j$, entonces*

$$\lim_{n\to\infty} S_n = \sum_{j=1}^\infty Y_j$$

existe c.s. y es finito c.s.

Demostración. Sea $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ y observemos que

$$\mathrm{E}(S_{n+1} - S_n|\mathcal{F}_n) = \mathrm{E}(Y_{n+1}|\mathcal{F}_n) = \mathrm{E}(Y_{n+1}) = 0$$

de modo que $(S_n,\ n \geq 1)$ es una $\mathcal{F}_n$-martingala. Observamos además que

$$\sup_n \mathrm{E}(S_n^+) \leq \sup_n (\mathrm{E}(S_n^2) + 1) \leq \sum_{k=1}^\infty \mathrm{E}(Y_k^2) + 1 < \infty.$$

El resultado sigue del teorema de convergencia de martingalas. ■

Definición 5.18 Consideremos ahora una sucesión creciente de σ-álgebras con índices negativos $\{\mathcal{F}_{-n},\ n \geq 0\}$, es decir que $\mathcal{F}_{-(n+1)} \subset \mathcal{F}_{-n}$. Una *martingala invertida* es una sucesión $\{X_{-n},\ n \geq 0\}$ de v.a. integrables con $X_{-n} \in \mathcal{F}_{-n}$ y satisface

$$\mathrm{E}[X_{-n}|\mathcal{F}_{-m}] = X_{-m} \tag{5.25}$$

con $0 \leq n < m$.

Una diferencia fundamental entre una martingala y una martingala invertida es que esta última tiene último elemento pero no tiene primer elemento, mientras que para una martingala es lo contrario.

Por la definición tenemos que

$$\mathrm{E}(X_0|\mathcal{F}_{-n}) = X_{-n}, \quad \text{para todo } n \geq 0$$

lo que implica que una martingala invertida es regular y por lo tanto uniformemente integrable.

Teorema 5.26 (Teorema de Convergencia para Martingalas Invertidas) *Sea* $\{(X_{-n}, \mathcal{F}_{-n}),\ n \geq 0\}$ *una martingala invertida, y sea* $\mathcal{F}_{-\infty} = \cap_{n=0}^{\infty} \mathcal{F}_{-n}$. *Entonces la sucesión* (X_{-n}) *converge c.s. y en* L^1 *a un límite* X *cuando* $n \to +\infty$. *En particular* X *es c.s. finita e integrable.*

Demostración. Sea U_{-n} el número de cruces hacia arriba del intervalo $[a, b]$ por $(X_{-n})_{n\geq 0}$ entre el instante $-n$ y 0. Entonces U_{-n} crece con n y sea $U(a, b) = \lim_{n\to\infty} U_{-n}$. Por el TCM

$$\mathrm{E}[U(a, b)] = \lim_{n\to\infty} \mathrm{E}[U_{-n}] \leq \frac{1}{b - a} \mathrm{E}[(X_0 - a)^+] < \infty,$$

por lo tanto $P(U(a, b) < \infty) = 1$. El mismo argumento usado en la demostración del teorema 5.22 implica que $X = \lim_{n\to\infty} X_{-n}$ existe c.s.

Sea $\varphi(x) = x^+ = (x \wedge 0)$, que es convexa y creciente. Además $\varphi(X_{-n})$ es integrable para todo n. La desigualdad de Jensen y (5.25) implican que $X_{-n}^+ \leq \mathrm{E}(X_0^+|\mathcal{F}_{-n})$, de modo que $\mathrm{E}(X_{-n}^+) \leq \mathrm{E}(X_0^+)$. Por el lema de Fatou y como $X_{-n}^+ \geq 0$ y $X_{-n}^+ \to X^+$ c.s. obtenemos

$$\mathrm{E}(X^+) \leq \liminf_n \mathrm{E}(X_{-n}^+) \leq \mathrm{E}(X_0^+) < \infty.$$

Por lo tanto $X^+ \in L^1$ y el mismo argumento aplicado a la martingala $(-X_{-n})$ muestra que $X^- \in L^1$. En consecuencia $X \in L^1$.

Falta demostrar que la convergencia también es en L^1. Observamos que en la demostración del teorema 5.24 vimos que si $X_{-n} \to X$ c.s., si $X \in L^1$ y si la sucesión (X_{-n}) es uniformemente integrable entonces $X_{-n} \to X$ en L^1. Todas estas condiciones se satisfacen en este caso. ■

Como aplicación demostramos la LFGN de Kolmogorov.

Teorema 5.27 (LFGN) *Sea* $(X_n)_{n\geq 1}$ *una sucesión i.i.d. con* $\mathrm{E}(|X_1|) < \infty$ *y sea* $S_n = X_1 + \cdots + X_n$. *Entonces*

$$\lim_{n\to\infty} \frac{1}{n} S_n = \mathrm{E}(X_1) \quad c.s.$$

Demostración. Sea $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \dots)$. Entonces $\mathcal{F}_{-n} \subset \mathcal{F}_{-m}$ si $n \geq m$, y el proceso

$$M_{-n} = \mathrm{E}(X_1|\mathcal{F}_{-n})$$

es una martingala invertida. Observemos que $\mathrm{E}(M_{-n}) = \mathrm{E}(X_1)$, para todo n, y además, como la sucesión es i.i.d., para $1 \leq j \leq n$

$$\mathrm{E}(X_1|\mathcal{F}_{-n}) = \mathrm{E}(X_j|\mathcal{F}_{-n}) \quad c.s. \tag{5.26}$$

Por lo tanto,

$$M_{-n} = \mathrm{E}(X_1|\mathcal{F}_{-n}) = \mathrm{E}(X_2|\mathcal{F}_{-n}) = \cdots = \mathrm{E}(X_n|\mathcal{F}_{-n}).$$

En consecuencia

$$M_{-n} = \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}(X_j|\mathcal{F}_{-n}) = \frac{1}{n} \mathrm{E}(S_n|\mathcal{F}_{-n}) = \frac{1}{n} S_n \quad c.s.$$

Por el teorema 5.26

$$\lim_{n\to\infty} \frac{1}{n} S_n = X \quad c.s.$$

con $\mathrm{E}(X) = \mathrm{E}(X_1)$. Además X es medible respecto a la σ-álgebra cola, y por la ley 0-1 de Kolmogorov, tenemos que X es constante c.s. En consecuencia debe ser igual a su valor esperado. ■