



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO
Grado en Matemáticas

LAS LEYES DE LOS GRANDES NÚMEROS EN LA TEORÍA DE LA PROBABILIDAD

Autor: Diego Alonso de la Maza
Tutor: Carlos Matrán Bea
Año 2025

Índice general

1. Introducción	5
2. Conceptos y resultados preliminares	7
2.1. Definiciones básicas y lemas de Borel-Cantelli	7
2.2. Variables aleatorias y convergencia	9
2.3. Desigualdades clásicas	11
2.4. Desigualdades exponenciales	12
2.5. Ensayos de Bernoulli	13
2.6. Distribución binomial	14
3. La probabilidad como límite de frecuencias	16
3.1. La ley débil de los grandes números	16
3.2. La ley fuerte de los grandes números	18
4. Distribuciones más generales	23
4.1. El método de truncamiento	24
4.2. Ley débil de los grandes números	25
4.3. Ley fuerte de los grandes números	28
4.4. Comparación entre demostraciones	39
5. Teorema de Etemadi	43
5.1. Extensiones teorema de Etemadi	47
5.2. Argumentos de truncación	55
6. Otros resultados	59
6.1. Momentos de orden 4	59
6.2. Momentos de orden par	62

ÍNDICE GENERAL

A. Desigualdades avanzadas en teoría de la probabilidad	66
A.1. Desigualdad de Marcinkiewicz–Zygmund	66
A.2. Desigualdades integrales básicas	67

Capítulo 1

Introducción

La teoría de la probabilidad es una rama de las matemáticas que proporciona principios y herramientas necesarias para estudiar y analizar el comportamiento de fenómenos aleatorios y estocásticos a lo largo del tiempo. Uno de los resultados más importantes de esta teoría es la **Ley de los Grandes Números** (LGN). Este resultado formaliza la intuición según la cual *"si repetimos un experimento aleatorio muchas veces, el promedio de los resultados tiende hacia un valor esperable determinado"*; es decir, aunque el resultado de un experimento aleatorio individual es incierto, el promedio tras muchas repeticiones muestra un comportamiento estable y predecible.

La forma más primitiva de esta idea aparece en el trabajo del matemático italiano Gerolamo Cardano, quien observó que, cuanto mayor fuera el número de ensayos mayor sería la precisión de los resultados. Expresó esta idea en términos de que el número de veces que ocurre un suceso con probabilidad p es aproximadamente igual a $\mu = np$, aunque nunca tuvo una formulación matemática rigurosa ni se conoce ningún intento de demostrarlo. Este enunciado propuesto por Cardano no representa una ley de los grandes números, pero es el primer rastro que existe de este pensamiento.

Tuvieron que pasar 200 años hasta que el matemático suizo Jacob Bernoulli publicara la primera demostración de lo que Cardano observó siglos antes. Bernoulli no solo habla sobre lo intuitiva que es esta idea, sino también de su gran importancia. Más de 100 años después, fue el matemático francés Poisson quien acuñó el nombre de Ley de los Grandes Números para referirse a este

tipo de resultados.

Desde entonces, matemáticos como Markov, Chebyshev, Khinchin o Kolmogorov han contribuido a la evolución de esta ley tratando de refinar las hipótesis, generalizar la idea e incluso haciendo pequeñas variaciones del enunciado. Llegando a resultados cuya importancia a día de hoy es difícil de exagerar. Las leyes de los grandes números se utilizan en diversos campos como la estadística, la teoría de la probabilidad e incluso en el ámbito económico.

Actualmente se distinguen principalmente dos Leyes de los Grandes Números.

- La **ley débil de los grandes números**, que establece la convergencia en probabilidad del promedio de una sucesión de variables aleatorias.
- La **ley fuerte de los grandes números**, que afirma la convergencia casi segura del mismo promedio.

Aunque ambas tengan enunciados similares, tienen propiedades diferentes. Así como las técnicas utilizadas para sus demostraciones e incluso sus hipótesis también varían.

El objetivo de este trabajo es estudiar las algunas versiones destacadas de las leyes de los grandes números, analizando su evolución desde los casos más simples hacia formulaciones más generales. Para ello comenzaremos por el caso más intuitivo, ensayos de Bernoulli, que podemos asociarlo al experimento de lanzar una moneda al aire. A partir de ahí, presentaremos generalizaciones hacia sucesiones de variables aleatorias independientes e idénticamente distribuidas, donde veremos diversas formas de abordar el problema, tratando de aprender y entender distintas vías metodológicas. Esto nos ayudará a poner en valor distintos enunciados y demostraciones que tratan este importante resultado conocido como Ley de los Grandes Números. Finalmente trataremos casos más generales donde las variables aleatorias sean arbitrariamente distribuidas, o no necesariamente independientes.

Capítulo 2

Conceptos y resultados preliminares

Para el desarrollo riguroso de este trabajo partimos de los conocimientos obtenidos en la asignatura Teoría de la Probabilidad y Estadística Matemática de tercer curso. En particular aquellos dentro del enfoque axiomático debido a Kolmogorov, en el que la probabilidad se interpreta como una medida.

No obstante, en este capítulo recordaremos algunos de los resultados fundamentales que utilizaremos con frecuencia a lo largo del trabajo. Además, repasaremos los conceptos más básicos relacionados con los ensayos de Bernoulli y la distribución binomial.

2.1. Definiciones básicas y lemas de Borel-Cantelli

Comenzaremos introduciendo las nociones matemáticas mas fundamentales para el estudio de la teoría probabilidad.

Definición 2.1. Un **espacio muestral** Ω es el conjunto de todos los posibles resultados de un experimento aleatorio. Además, si el espacio es finito o numerable, diremos que el espacio es discreto.

Definición 2.2. Sea Ω un conjunto no vacío. Llamamos **σ -álgebra** sobre Ω a una familia $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ no vacía de subconjuntos de Ω que satisface:

1. $\Omega \in \mathcal{F}$, es decir, contiene al total.
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$, es cerrada bajo complementarios.

CAPÍTULO 2. CONCEPTOS Y RESULTADOS PRELIMINARES

3. Si $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, es cerrada bajo uniones numerables.

Al par $(\Omega, \mathcal{F})$ se le denomina espacio medible. En un contexto probabilístico, a los elementos de $\mathcal{F}$ se les llaman **sucesos**.

Típicamente, en un espacio discreto la σ -álgebra que se utiliza es la de todos los subconjuntos de Ω , que en este caso coincide con la mínima que contiene a todos los conjuntos unipuntuales, también llamados sucesos elementales.

Definición 2.3. Sea $\mathcal{F}$ una σ -álgebra sobre Ω , decimos que $P : \mathcal{F} \rightarrow \mathbb{R}$ es una probabilidad sobre el espacio medible $(\Omega, \mathcal{F})$ si

1. $P(A) \geq 0, \forall A \in \mathcal{F}$
2. $P(\Omega) = 1$
3. (σ -aditividad) Sea $\{A_n\}_{i \in I}^{\infty}$ una familia numerable de eventos de $\mathcal{F}$ disjuntos dos a dos, entonces se tiene que

$$P\left(\bigcup_{i \in I} A_n\right) = \sum_{i \in I} P(A_n).$$

Nota 2.1. En un espacio discreto, cada conjunto es un suceso, y tiene una probabilidad asociada, que coincide con la suma de las probabilidades de los sucesos elementales que lo componen. En ese caso, la probabilidad queda determinada conociendo las probabilidades de todos los sucesos elementales: la función de masa de probabilidad.

Teorema 2.1 (Desigualdad de Boole). Sea $A_1, A_2, \dots$ una familia finita o numerable de sucesos, entonces

$$P\left(\bigcup_n A_n\right) \leq \sum_n P(A_n)$$

Definición 2.4. Un **espacio de probabilidad** es una terna $(\Omega, \mathcal{F}, P)$, donde:

1. Ω es un espacio muestral,
2. $\mathcal{F}$ es una σ -álgebra sobre Ω ,
3. $P : \mathcal{F} \rightarrow [0, 1]$ es una probabilidad.

2.2. VARIABLES ALEATORIAS Y CONVERGENCIA

Lemas de Borel-Cantelli: Sea $\{A_n\}$ una sucesión de sucesos de (Ω, σ, P) .

1. Primer Lema de Borel-Cantelli:

Si $\sum_{n=1}^{\infty} P(A_n) < \infty$ entonces $P(\limsup A_n) = 0$

2. Segundo Lema de Borel-Cantelli:

Si los sucesos $\{A_n\}$ son independientes y $\sum_{n=1}^{\infty} P(A_n) = \infty$ entonces $P(\limsup A_n) = 1$.

Nota 2.2. *En un espacio probabilístico discreto, el primer Lema es absolutamente trivial: si A_n es una sucesión cualquiera de sucesos, y $\sum_n P(A_n) < \infty$, ningún suceso elemental ω con probabilidad $P(\omega) = \delta > 0$ puede pertenecer a infinitos de esos conjuntos, porque si estuviera en los conjuntos $A_{n_1}, A_{n_2}, \dots$, tendríamos que $\sum_n P(A_n) \geq \sum_{n_k} P(A_{n_k}) \geq \sum_{n_k} \delta = \infty$.*

Nota 2.3. *El segundo Lema merece también cierta atención. Podemos suponer que todos los sucesos elementales tienen probabilidad estrictamente positiva (podemos quitar del espacio todos los que tengan probabilidad 0). Por tanto, el único suceso que tiene probabilidad 1 es el espacio total, por lo que si (A_n) es una sucesión de sucesos independientes y algún ω está en infinitos de ellos ($\omega \in \limsup A_n$), entonces $P(\limsup A_n) = 1$ y todos los ω 's estarían en infinitos de ellos. Además, como es bien conocido, en ningún espacio probabilístico discreto puede existir una sucesión (A_n) de sucesos independientes con probabilidades acotadas inferior y superiormente: $0 < \delta < P(A_n) < 1 - \delta < 1$. En otras palabras, el interés de los lemas de Borel-Cantelli en espacios probabilísticos discretos es muy limitado, y su aparición en el libro de Feller se debe a la “asignación razonable” de probabilidades a sucesos que no están en el espacio, pero que pueden considerarse como límites de otros (ver las secciones VIII.1 a VIII.3 en el libro de Feller [8]).*

2.2. Variables aleatorias y convergencia

Definición 2.5. Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad. Una **variable aleatoria** (v.a.) sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$ es una función $X : (\Omega, \mathcal{F}, \mathcal{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ medible, es decir,

$$X^{-1}(B) \in \mathcal{F} \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}),$$

CAPÍTULO 2. CONCEPTOS Y RESULTADOS PRELIMINARES

donde $\mathcal{B}(\mathbb{R})$ denota la σ -álgebra de Borel en $\mathbb{R}$.

Definición 2.6. Una variable aleatoria X se dice **indicadora** de un suceso A si

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

Se denota por $\mathbf{1}_A$.

Definición 2.7. Dos variables aleatorias X e Y se dice incorreladas si

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = 0$$

Definición 2.8. Se dice que una familia de variables aleatorias $X_1, \dots, X_n$ es incorrelada por pares si

$$\text{Cov}(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]$$

Definición 2.9. Sean $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria. Se dice que:

1. $X_n \rightarrow X$ converge **casi seguro (c.s.)** si

$$P\left(\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1.$$

2. $X_n \rightarrow X$ converge **en probabilidad** si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Nota 2.4. Es importante saber que la convergencia casi seguro implica la convergencia en probabilidad. Además, en el caso discreto son equivalentes.

Teorema 2.2. Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria, entonces

$$X_n \rightarrow X \text{ c.s.} \iff P\left(\sup_{k \geq n} |X_k - X| \geq \varepsilon\right) \rightarrow 0, \text{ cuando } n \rightarrow \infty$$

para todo $\varepsilon > 0$

2.3. DESIGUALDADES CLÁSICAS

Corolario 2.1. *Dado que*

$$P\left(\sup_{k \geq n} |X_k - X| \geq \varepsilon\right) = P\left(\bigcup_{k \geq n} (|X_k - X| \geq \varepsilon)\right) \leq \sum_{k \geq n} P(|X_k - X| \geq \varepsilon)$$

una condición suficiente para que $X_n \rightarrow X$ casi seguro es que

$$\sum_{k=1}^{\infty} P(|X_k - X| \geq \varepsilon) < \infty$$

se cumpla para todo $\varepsilon > 0$

2.3. Desigualdades clásicas

Teorema 2.3 (Desigualdad de Markov). *Sea X una variable aleatoria no negativa tal que $\exists E(X)$. Entonces, para todo $t > 0$,*

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

A partir de la desigualdad de Markov se tiene que si ϕ es una función estrictamente monótona creciente y no negativa, entonces para cualquier variable aleatoria X y número real t se tiene que

$$P(X \geq t) = P(\phi(X) \geq \phi(t)) \leq \frac{E(\phi(X))}{\phi(t)}$$

Aplicando esto a la función $\phi(x) = x^2$ obtenemos la conocida desigualdad de Chebyshev.

Teorema 2.4 (Desigualdad de Chebyshev). *Sea X una variable aleatoria con varianza finita. Entonces, para todo $t > 0$,*

$$P(|X| \geq t) \leq t^{-2} E(X^2).$$

En particular, si $EX = \mu$ entonces

$$P(|X - \mu| \geq t) \leq t^{-2} \text{Var}(X)$$

2.4. Desigualdades exponenciales

Ahora introduciremos desigualdades de tipo exponencial, estas en muchos casos nos facilitaran en gran medida las demostraciones y obtendremos resultados muy interesantes a partir de ellas. La primera de ellas se obtiene aplicando la desigualdad de Markov a la función $\phi(x) = e^{sx}$ donde s es un número positivo y arbitrario.

Teorema 2.5 (Desigualdad de Chernoff). *Sea X una variable aleatoria y $t > 0$, se tiene que*

$$P(X \geq t) = P(e^{sX} \geq e^{st}) \leq \frac{E(e^{sX})}{e^{st}}$$

Ahora podemos mencionar ciertas desigualdades para sumas de variables aleatorias independientes. Nuestro objetivo es poder acotar superiormente lo siguiente $P(S_n - E(S_n) \geq t)$, con $S_n = \sum_{i=1}^n X_i$, donde $X_1, \dots, X_n$ son variables aleatorias reales e independientes. La desigualdad de Chebyshev y la independencia implican de manera inmediata lo siguiente:

$$P(|S_n - E(S_n)| \geq t) \leq \frac{\text{Var}(S_n)}{t^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{t^2}$$

Veamos ahora el método de Chernoff, el cual es muy útil para acotar colas de probabilidades de sumas de variables aleatorias independientes. Esto no es más que aplicar la desigualdad de Chernoff de la siguiente forma:

$$\begin{aligned} P(S_n - E(S_n) \geq t) &\leq e^{-st} E \left[\exp \left(s \sum_{i=1}^n (X_i - E(X_i)) \right) \right] \\ &= e^{-st} \prod_{i=1}^n E \left[e^{s(X_i - E(X_i))} \right] \quad \text{por independencia} \end{aligned} \quad (2.1)$$

Ahora el hecho de acotar lo anterior de manera más precisa recae en acotar la función generadora de momentos de las variables aleatorias $X_i - E(X_i)$. Para ello existen muchos resultados, uno de los más elegantes es mediante la desigualdad de Hoeffding .

Lema 2.1 (Desigualdad de Hoeffding). *Sea X una variable aleatoria con $E(X) = 0$, $a \leq X \leq b$. Entonces para $s > 0$,*

$$E(e^{sX}) \leq e^{s^2(b-a)^2/8}$$

Si aplicamos este lema a (2.1) llegamos al siguiente resultado.

2.5. ENSAYOS DE BERNOULLI

Teorema 2.6. *Sea $X_1, \dots, X_n$ variables aleatorias independientes y acotadas, de modo que X_i toma valores en el intervalo $[a_i, b_i]$ con probabilidad uno. Entonces, para cualquier $t > 0$*

$$P(S_n - E(S_n) \geq t) \leq e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2}$$

y

$$P(S_n - E(S_n) \leq -t) \leq e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2}$$

Uno de los principales problemas de la desigualdad de Hoeffding es que no tiene en cuenta la varianza de las variables aleatorias implicadas. Con el objetivo de presentar desigualdades más precisas tenemos la desigualdad de Bennett y la desigualdad de Bernstein, que tendrán en cuenta dicha varianza.

Teorema 2.7 (Desigualdad de Bennett). *Sea $X_1, \dots, X_n$ una sucesión de variables aleatorias independientes, de media cero y tales que $X_i \leq 1$ con probabilidad 1. Si definimos*

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \text{Var}(X_i),$$

entonces, para cualquier $t > 0$,

$$P\left(\sum_{i=1}^n X_i > t\right) \leq \exp\left(-n\sigma^2 h\left(\frac{t}{n\sigma^2}\right)\right),$$

donde $h(u) = (1 + u) \log(1 + u) - u$, para $u \geq 0$.

Este resultado puede simplificarse aún más si se usa la cota elemental $h(u) \geq \frac{u^2}{2(2+u/3)}$, lo cual nos lleva a la siguiente desigualdad más clásica.

Teorema 2.8 (Desigualdad de Bernstein). *Bajo las mismas hipótesis que el teorema anterior, para todo $\epsilon > 0$, se tiene:*

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i > \epsilon\right) \leq \exp\left(-\frac{n\epsilon^2}{2(\sigma^2 + \epsilon/3)}\right).$$

2.5. Ensayos de Bernoulli

Un caso fundamental en el estudio de experimentos aleatorios repetidos son los ensayos de Bernoulli, que son sucesos independientes repetidos que solo

CAPÍTULO 2. CONCEPTOS Y RESULTADOS PRELIMINARES

tienen dos posibles resultados con probabilidad fija durante todos los ensayos.

Definición 2.10. Un **ensayo de Bernoulli** es un experimento aleatorio que solo puede tener dos resultados. Normalmente se denotan las dos probabilidades como p y q , y nos referimos al resultado con probabilidad p como éxito, y al otro resultado como fracaso con probabilidad $q = 1 - p$.

El ejemplo más conocido de ensayos de Bernoulli es el de lanzar una moneda simétrica al aire, en este caso $p = q = 1/2$. Las sucesivas tiradas de una moneda son sucesos independientes entre sí que forman una sucesión de ensayos de Bernoulli.

Definición 2.11. Una variable aleatoria X se dice que tiene **distribución de Bernoulli** con parámetro p si

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Notamos $X \sim \text{Bernoulli}(p)$.

Proposición 2.1. Si $X \sim \text{Bernoulli}(p)$, entonces:

$$E(X) = p, \quad \text{Var}(X) = p(1 - p).$$

Proposición 2.2. Toda variable indicadora tiene distribución de Bernoulli con parámetro $P(A)$:

$$\mathbf{1}_A \sim \text{Bernoulli}(P(A)).$$

2.6. Distribución binomial

La distribución binomial describe el número de éxitos en un número fijo de ensayos de Bernoulli independientes.

Definición 2.12. Sea $n \in \mathbb{N}$ y $p \in (0, 1)$. Una variable aleatoria X tiene distribución **binomial** con parámetros n y p , denotado $X \sim \text{Binomial}(n, p)$, si representa el número de éxitos en n ensayos de Bernoulli independientes con probabilidad de éxito p .

Su función de probabilidad es:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

2.6. DISTRIBUCIÓN BINOMIAL

Además, denotamos $P(X = k)$ por $b(k; n, p)$.

Proposición 2.3. *Si $X \sim \text{Binomial}(n, p)$, entonces:*

$$E(X) = np, \quad \text{Var}(X) = np(1 - p).$$

Proposición 2.4. *Sea $X_i \sim \text{Bernoulli}(p)$ una sucesión de variables aleatorias independientes. Entonces*

$$S_n := \sum_{i=1}^n X_i \sim \text{Binomial}(n, p).$$

Capítulo 3

La probabilidad como límite de frecuencias

Naturalmente, la idea que tenemos de *probabilidad* está fundada en ejemplos como el que sigue. Si dos amigos, Pedro y Juan, lanzan una moneda homogénea dos veces, obteniendo cara en ambas tiradas, es claro que nadie pensaría que la probabilidad de que salga cara es 1, este resultado es simple azar en una muestra de tamaño muy pequeño. En cambio, si estos dos amigos deciden lanzar esta moneda durante una tarde entera y apuntar el resultado de todas las tiradas, cabe esperar que el número de caras obtenidas, dividido entre el número total de tiradas, será un valor cercano a la probabilidad de que salga cara.

3.1. La ley débil de los grandes números

Generalizando este experimento, consideremos una secuencia de n ensayos idénticos, donde n es suficientemente grande, y un suceso A que ocurre v veces, entonces v/n se aproximará a la probabilidad de que ocurra A . Sin embargo, una idea intuitiva no constituye una definición matemática formal. Para formalizarla, modelamos la sucesión de ensayos idénticos a través de ensayos de Bernoulli con probabilidad p de éxito. Sea S_n el número de éxitos en n ensayos, se tiene que S_n/n es el promedio de éxitos y debe aproximarse a p .

Para demostrar esta afirmación observamos que S_n sigue una distribución binomial. Estudiemos la probabilidad de que haya al menos r éxitos en n

3.1. LA LEY DÉBIL DE LOS GRANDES NÚMEROS

ensayos de Bernoulli. Esto es:

$$P(S_n \geq r) = \sum_{v=0}^{\infty} b(r+v; n, p) \quad (3.1)$$

Estudiemos los términos de la serie (3.1), para ello recordemos que $b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$ y se tiene que:

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{(n-k+1)p}{kq} = \frac{(n+1)p - k + kq}{kq} = 1 + \frac{(n+1)p - k}{kq} \quad (3.2)$$

Por lo tanto los términos $b(k; n, p)$ crecen si $k < (n+1)p$ y decrecen si $k > (n+1)p$.

Supongamos que $r > np$, es obvio por (3.2) que los términos de la serie (3.1) decrecen más rápido que los términos de una serie geométrica de razón $\delta = 1 - (r - np)/rq$. Así que podemos comparar la serie de probabilidades (3.1) con una serie geométrica de la siguiente manera:

$$b(r; n, p) + b(r+1; n, p) + b(r+2; n, p) + \dots \leq b(r; n, p) + b(r; n, p)\delta + b(r; n, p)\delta^2 + \dots$$

Por lo que $P(S_n \geq r) \leq b(r; n, p) \cdot \frac{rq}{r-np}$ si $r > np$.

Por otro lado, hay al menos $r - np$ enteros k tales que $(n+1) \leq k \leq r$. Entonces $b(k; n, p) \geq b(r; n, p)$ por (3.2), y $\sum_k b(k; n, p) \leq 1$ por definición de probabilidad.

Por lo tanto $(r - np) \cdot b(r; n, p) \leq \sum_k b(k; n, p) \leq 1$, por lo que $b(r; n, p) \leq (r - np)^{-1}$.

Concluyendo así que

$$P(S_n \geq r) \leq \frac{rq}{(r - np)^2} \quad \text{si } r > np. \quad (3.3)$$

Podemos ahora estimar la probabilidad de que haya como mucho r éxitos como la probabilidad de que haya al menos $n - r$ fracasos y teniendo en cuenta que el número de fracasos en n ensayos es $n - S_n \sim \text{Binomial}(n, q)$. Entonces:

$$P(S_n \leq r) = P(n - S_n \geq n - r)$$

CAPÍTULO 3. LA PROBABILIDAD COMO LÍMITE DE FRECUENCIAS

Y por (3.3) tenemos:

$$P(n - S_n \geq n - r) \leq \frac{(n - r)p}{((n - r) - np)^2} = \frac{(n - r)p}{(np - r)^2} \quad \text{si } n - r > np \Rightarrow r < np.$$

Resumiendo, hemos visto que

$$P(S_n \leq r) \leq \frac{(n - r)p}{(np - r)^2} \quad \text{si } r < np. \quad (3.4)$$

Con esto podemos estudiar la probabilidad de que S_n/n sea mayor que $p + \varepsilon$, con $\varepsilon > 0$ arbitrariamente pequeño. Notemos que $P(S_n/n > p + \varepsilon) = P(S_n > n(p + \varepsilon))$ y por (3.3) se da la siguiente desigualdad:

$$P(S_n > n(p + \varepsilon)) < \frac{1}{n\varepsilon^2} \rightarrow 0 \quad \text{cuando } n \text{ se incrementa}$$

Probando así que la probabilidad de que S_n/n sea mayor que $p + \varepsilon$ tiende a cero para cualquier $\varepsilon > 0$ fijo. Lo mismo ocurre con la probabilidad de que S_n/n sea menor que $p - \varepsilon$. Se prueba de manera análoga con (3.4) que $P(S_n < n(p - \varepsilon)) \rightarrow 0$. Luego

$$P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) \rightarrow 1$$

Hemos probado que cuando n crece, la probabilidad de que S_n/n se aleje de p en más que una cantidad $\varepsilon > 0$ arbitrariamente pequeña tiende a cero. Este resultado es conocido como la ley débil de los grandes números y es la base de la idea intuitiva de *probabilidad* a la que habíamos hecho referencia previamente. Sin embargo podemos tratar de afinar aún más este resultado, para ello conviene que introduzcamos la Ley Fuerte de los Grandes Números la cual proporciona una caracterización más precisa de la convergencia de S_n/n hacia p .

3.2. La ley fuerte de los grandes números

La ley débil de los grandes números dice que para cualquier n suficientemente grande y fijo, S_n/n se acerca mucho a p , pero no asegura que en n nuevos ensayos idénticos S_n/n siga siendo próximo a p . Por ello es natural preguntarse

3.2. LA LEY FUERTE DE LOS GRANDES NÚMEROS

si podemos asegurar la siguiente convergencia

$$\frac{S_n}{n} \rightarrow p \quad (3.5)$$

siendo S_n el número de éxitos en los primeros n ensayos de una sucesión de ensayos de Bernoulli. En la teoría nuestro espacio muestral considera el caso de una sucesión infinita de éxitos, en la cual no podría darse dicha convergencia. El objetivo de la ley fuerte de los grandes números es el de probar que (3.5) se cumple con probabilidad 1, pudiendo desestimar así los casos donde no se da la convergencia. Probemos que $S_n/n - p$ se hace pequeño y hay probabilidad uno de que se mantenga así. Debemos recordar que la existencia de sucesiones de variables aleatorias independientes e igualmente distribuidas (que no sean constantes) es imposible en cualquier espacio probabilístico discreto, por lo que la equivalencia entre las convergencias casi seguro y en probabilidad no puede utilizarse en este contexto.

Teorema 3.1 (Ley fuerte de los grandes números). *Para todo $\varepsilon > 0$, hay probabilidad uno de que ocurra solo un número finito de eventos en los que*

$$\left| \frac{S_n}{n} - p \right| > \varepsilon \quad (3.6)$$

Esto prueba que, en términos de espacios muestrales finitos, para todo $\varepsilon > 0$ y $\delta > 0$, existe un r tal que para todo n la probabilidad de que se den las n siguientes desigualdades

$$\left| \frac{S_{r+k}}{r+k} - p \right| < \varepsilon, \quad k = 1, 2, \dots, n$$

es mayor que $1 - \delta$. Y por tanto (3.5) se da con probabilidad uno.

Demostración. Por el primer lema de Borel-Cantelli, si vemos que

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) < \infty$$

entonces $P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon \text{ infinitas veces}\right) = 0$, lo cual probaría (3.6).

Lo primero en lo que uno piensa al intentar resolver este problema es en acotar el término $P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right)$ de la serie. Sin embargo para acotarlo no nos sirve con desigualdades clásicas como la desigualdad de Markov o Chebyshev,

CAPÍTULO 3. LA PROBABILIDAD COMO LÍMITE DE FRECUENCIAS

ya que obtendríamos cotas no lo suficientemente finas para asegurar la convergencia de la serie.

Por ejemplo aplicando la desigualdad de Chebyshev de la siguiente forma:

$$P(|S_n - np| \geq n\varepsilon) \leq \frac{npq}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2}$$

Que no es una serie convergente.

Este inconveniente Feller en [8] lo resuelve probando una afirmación más fuerte, considera el evento A_k

$$|S_k^*| = \left| \frac{S_k - kp}{\sqrt{kpq}} \right| \geq \sqrt{2a \log k}, \quad \text{donde } a > 1$$

donde S_k^* es la variable aleatoria normalizada, es decir, mide la desviación típica de S_k respecto a su media.

Para posteriormente acotar $P(A_k)$ con el fin de garantizar la convergencia de la serie $\sum P(A_k)$. Para ello, en el libro de Feller se establece la siguiente desigualdad:

$$P(A_k) < e^{-a \log k} = \frac{1}{k^a}$$

Dicha desigualdad se presenta como una consecuencia del teorema de De Moivre, es decir, el teorema central del límite aplicado a variables de Bernoulli. No obstante, la teoría previa necesaria para poder justificar este resultado tal y como se expone en [8] no es en absoluto trivial. Sin embargo, en la actualidad existen herramientas alternativas que permiten obtener este tipo de cotas de forma más directa y accesible. Estas son, en particular, las desigualdades exponenciales. Estas nacen gracias al Teorema Central del Límite, el cual sugiere que sumas de variables aleatorias independientes e igualmente distribuidas se comporten como variables aleatorias normalmente distribuidas; y entonces, como las colas de probabilidad de una variable aleatoria normal pueden aproximarse por funciones exponenciales, cabe esperar que lo mismo ocurra para sumas de variables aleatorias independientes e igualmente distribuidas. Este pensamiento es en el que se basan desigualdades exponenciales como la de Hoeffding, Bennett o Bernstein.

Esto nos va a permitir demostrar esencialmente lo mismo utilizando operaciones y teoría mucho más elemental. Por el Teorema 2.6, si $X_1, \dots, X_n$ son variables aleatorias e independientes y tales que X_i toma valores en el intervalo

3.2. LA LEY FUERTE DE LOS GRANDES NÚMEROS

$[a_i, b_i]$ con probabilidad uno. Entonces para cualquier $t > 0$

$$\begin{aligned} P(|S_n - E(S_n)| \geq t) &= P(S_n - E(S_n) \geq t) + P(S_n - E(S_n) \leq -t) \\ &\leq e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2} + e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2} = 2e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2} \end{aligned}$$

En nuestro caso, tenemos ensayos de Bernoulli, que son variables aleatorias e independientes y tales que cada una de ellas toma valores en el intervalo $[0, 1]$ con probabilidad uno. Podemos por tanto aplicar la desigualdad de Hoeffding para cualquier $t > 0$. Nosotros queremos acotar lo siguiente

$$P\left(\left|\frac{S_n - np}{n}\right| \geq \varepsilon\right) = P(|S_n - np| \geq n\varepsilon)$$

Ahora estamos en condiciones de aplicar la desigualdad de Hoeffding

$$P(|S_n - np| \geq t) \leq 2e^{-2t^2/k}$$

con $t = n\varepsilon$. Entonces:

$$P(|S_n - np| \geq n\varepsilon) \leq 2e^{-2\varepsilon^2 n}$$

Ahora bien, la serie $\sum 2e^{-2\varepsilon^2 n}$ es convergente por el Criterio del Cociente, pues

$$\frac{a_{n+1}}{a_n} = \frac{1}{e^{2\varepsilon^2}} < 1$$

Luego $\sum P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) < \infty$, y por el primer lema de Borel-Cantelli podemos asegurar que con probabilidad 1 solo ocurrirán un número finito de eventos como $\left|\frac{S_n}{n} - p\right| > \varepsilon$. Como queríamos demostrar. $\square$

Acabamos de presentar una demostración alternativa a la de Feller [8]. Sin embargo, en el Capítulo 6, introduciremos un resultado que nos permitirá, aprovechando que los ensayos de Bernoulli son variables aleatorias acotadas (y, por tanto, poseen momentos de todos los órdenes finitos), establecer una nueva vía para demostrar este teorema. Veremos que esta no es la única alternativa existente y obtendremos una demostración especialmente sencilla y elegante.

Los teoremas presentados hasta ahora son casos particulares de resultados límite más generales. A continuación, estudiaremos la ley de los grandes números en un contexto más general, considerando variables aleatorias con cualquier

CAPÍTULO 3. LA PROBABILIDAD COMO LÍMITE DE FRECUENCIAS

distribución en lugar de ensayos de Bernoulli.

Capítulo 4

Distribuciones más generales

En el capítulo anterior, presentamos la Ley Débil y la Ley Fuerte de los Grandes Números en el contexto particular de los ensayos de Bernoulli, que constituyen el caso más intuitivo y primitivo dentro de las leyes de los grandes números. Ahora extenderemos estos resultados al caso más general de sucesiones de variables aleatorias independientes e idénticamente distribuidas (i.i.d.). Es importante destacar que los ensayos de Bernoulli no dejan de ser un caso particular de variables i.i.d., por lo que este nuevo marco constituye una generalización natural de los resultados previamente estudiados. Para ello nos basaremos, como en el primer capítulo, en el libro de Feller [8], donde todos los desarrollos se realizan en el caso discreto y sin recurrir a la teoría de la medida. Recordemos que en el caso discreto, la convergencia en probabilidad y la convergencia casi segura son equivalentes. Lo primero que uno pensaría es que esto facilitaría las cosas ya que la demostración de ambas leyes no exige herramientas distintas, con demostrar la Ley Débil de los Grandes Números quedaría probada también la Ley Fuerte, sin embargo, como ya se ha dicho, el modelo de variables independientes e igualmente distribuidas es imposible en los espacios discretos, por lo que los argumentos deberán justificarse siempre a través de la Teoría de la Probabilidad sobre espacios generales o utilizando aproximaciones que garanticen los resultados. En particular, uno de los objetivos de este capítulo será comparar la demostración de la LFGN dada por Feller [8] con una demostración alternativa proveniente del libro de Chung [4], en el cual sí se hace uso de los espacios de probabilidad y de la teoría de la medida. Esta comparación nos permitirá identificar con claridad las diferencias entre ambos enfoques y comprender la estrategia de Feller para formular una de-

mostración con las mismas ideas clave pero prescindiendo de las herramientas formales que proporciona la teoría de la medida.

4.1. El método de truncamiento

Antes de introducir la Ley Débil de los Grandes Números para variables aleatorias independientes e idénticamente distribuidas (i.i.d.), presentaremos el método de truncamiento. Este se emplea cuando trabajamos con variables aleatorias que, aunque tienen esperanza finita, no están acotadas o no tienen varianza finita. En estos casos, los valores extremos pueden dificultar el análisis de la convergencia de los promedios.

La idea central del truncamiento consiste en descomponer cada variable aleatoria X_i en la suma de dos partes: una parte truncada, que consiste en recortar los valores extremos de la variable, y un resto que recoge las desviaciones “grandes” respecto a un umbral previamente fijado. Para ello, dado un umbral $a > 0$, definimos la versión truncada de X_i como

$$U_i := X_i \cdot \mathbf{1}_{\{|X_i| \leq a\}},$$

y el resto como

$$V_i := X_i \cdot \mathbf{1}_{\{|X_i| > a\}}.$$

De este modo, se tiene la descomposición

$$X_i = U_i + V_i$$

El objetivo es analizar por separado el comportamiento de los promedios

$$\frac{1}{n} \sum_{i=1}^n U_i \quad \text{y} \quad \frac{1}{n} \sum_{i=1}^n V_i,$$

y demostrar que el segundo término puede hacerse arbitrariamente pequeño (en probabilidad) al tomar a suficientemente grande, gracias a la existencia de la esperanza de X_i . Una vez controlado el resto, el análisis de la parte truncada se puede llevar a cabo usando técnicas más elementales, ya que las variables truncadas son acotadas y por tanto más manejables analíticamente.

Este procedimiento permite extender la demostración de la Ley de los Grandes Números más allá del caso de variables acotadas, sin requerir supuestos

4.2. LEY DÉBIL DE LOS GRANDES NÚMEROS

adicionales sobre la varianza o momentos superiores. En la siguiente demostración utilizaremos esta técnica adaptándola al caso de variables i.i.d. con esperanza finita.

4.2. Ley débil de los grandes números

Comenzaremos viendo la demostración presentada por Feller ya que esta no necesita ningún concepto ni ningún resultado previo.

Ley débil de los grandes números. *Sea $\{X_k\}$ una sucesión de variables aleatorias mutuamente independientes e igualmente distribuidas. Si la esperanza $\mu = E(X_k)$ existe, entonces para todo $\varepsilon > 0$ cuando $n \rightarrow \infty$*

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) \rightarrow 0 \quad (4.1)$$

es decir, la probabilidad de que S_n/n difiera de la esperanza μ menos de un ε arbitrario tiende a uno.

Demostración. Podemos, sin pérdida de generalidad, suponer que $\mu = E(X_k) = 0$, en caso contrario podríamos tomar la variable aleatoria $X_k - \mu$ y proceder de igual manera. Antiguas demostraciones de este teorema imponían la condición innecesaria de que la varianza $Var(X_k)$ fuese finita. En este caso la ley de los grandes números es una consecuencia trivial de la desigualdad de Chebyshev

$$P(|S_n| > t) \leq \frac{n\sigma^2}{t^2}$$

Para $t = \varepsilon n$ el lado derecho tiende hacia 0, cumpliéndose así (4.1). El caso en el que el segundo momento no existe es más complicado. Lo demostraremos utilizando el método de truncamiento. Sea $\delta > 0$ una constante que definiremos más adelante. Para cada n , definimos las variables aleatorias truncadas de la siguiente manera:

$$\begin{aligned} U_k &= X_k, & V_k &= 0 & \text{si } |X_k| \leq \delta n, \\ U_k &= 0, & V_k &= X_k & \text{si } |X_k| > \delta n. \end{aligned}$$

Donde $k = 1, \dots, n$, además hemos de tener en cuenta la dependencia de U_k y V_k con respecto a n . Ahora, por definición, podemos escribir la variable

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

aleatoria X_k de la siguiente manera:

$$X_k = U_k + V_k$$

y para demostrar la ley de los grandes números, basta con probar que, dado $\varepsilon > 0$, podemos elegir una constante δ de tal forma que cuando $n \rightarrow \infty$

$$P\left(|U_1 + \cdots + U_n| > \frac{1}{2}\varepsilon n\right) \rightarrow 0 \quad (4.2)$$

y

$$P\left(|V_1 + \cdots + V_n| > \frac{1}{2}\varepsilon n\right) \rightarrow 0 \quad (4.3)$$

Denotamos los valores que puede tomar cada variable X_j por $x_1, x_2, \dots$ y sus respectivas probabilidades por $f(x_j)$ y definimos

$$a = E(|X_j|) = \sum_j |x_j| f(x_j)$$

La variable U_1 está acotada por δn , luego es claro que:

$$E(U_1^2) \leq a\delta n$$

Las variables $U_1, \dots, U_n$ tienen la misma distribución y son mutuamente independientes. Por lo tanto,

$$\text{Var}(U_1 + \cdots + U_n) = n\text{Var}(U_1) \leq nE(U_1^2) \leq a\delta n^2.$$

Ahora, por definición de U_n cuando $n \rightarrow \infty$

$$E(U_n) \rightarrow E(X_1) = 0.$$

Con esto intentaremos acotar la esperanza del cuadrado de la suma de las variables truncadas $U_1, \dots, U_n$. Como las variables truncadas no están necesariamente centradas, desarrollaremos directamente el cuadrado de manera explícita:

$$E\left((U_1 + \cdots + U_n)^2\right) = \sum_{k=1}^n E(U_k^2) + 2 \sum_{1 \leq i < j \leq n} E(U_i U_j).$$

Como las variables $U_1, \dots, U_n$ son independientes por definición, se tiene que

4.2. LEY DÉBIL DE LOS GRANDES NÚMEROS

$E(U_i U_j) = E(U_i)E(U_j)$. Por lo tanto,

$$E\left((U_1 + \cdots + U_n)^2\right) = \sum_{k=1}^n E(U_k^2) + 2 \sum_{1 \leq i < j \leq n} E(U_i)E(U_j).$$

Debido a que las U_k son independientes e idénticamente distribuidas, se deduce que $E(U_k^2) = E(U_1^2)$ y $E(U_k) = E(U_1)$, con lo cual:

$$E\left((U_1 + \cdots + U_n)^2\right) = nE(U_1^2) + n(n-1)(E(U_1))^2.$$

Ahora bien, ya sabemos que

$$nE(U_1^2) \leq a\delta n^2.$$

Además, aunque no tenemos control exacto sobre $E(U_1)^2$, sí sabemos que $E(U_n) \rightarrow E(U_1) = 0$ cuando $n \rightarrow \infty$. Por tanto, para n suficientemente grande, $E(U_n)$ es suficientemente pequeño como para que podemos garantizar que

$$n(n-1)(E(U_n))^2 \leq n^2 (E(U_n))^2 \leq a\delta n^2,$$

y obtenemos en conjunto, para un n suficientemente grande, la cota

$$E\left((U_1 + \cdots + U_n)^2\right) \leq 2a\delta n^2$$

Ahora (4.2) es consecuencia de la desigualdad de Chebyshev por la cual:

$$P\left(|U_1 + \cdots + U_n| > \frac{1}{2}\varepsilon n\right) \leq \frac{8a\delta}{\varepsilon^2}$$

Escogiendo δ lo suficientemente pequeño, podemos hacer que el lado derecho sea tan pequeño como queramos, cumpliéndose así (4.2). Para probar (4.3) vemos que

$$\{V_1 + \cdots + V_n \neq 0\} \subseteq \bigcup_{k=1}^n \{V_k \neq 0\}$$

Luego

$$P(V_1 + \cdots + V_n \neq 0) \leq P\left(\bigcup_{k=1}^n (V_k \neq 0)\right) \leq \sum_{k=1}^n P(V_k \neq 0)$$

Y como $V_k = X_k$ si $|X_k| > \delta n$ y $V_k = 0$ en el resto de casos, y como las X_k son

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

igualmente distribuidas, se tiene que $P(V_k \neq 0) = P(V_1 \neq 0)$ para todo k . De lo que se deduce:

$$P(V_1 + \cdots + V_n \neq 0) \leq nP(V_1 \neq 0) \quad (4.4)$$

Ahora para $\delta > 0$ arbitrario tenemos que

$$P(V_1 \neq 0) = P(|X_1| > \delta n) = \sum_{|x_j| > \delta n} f(x_j) \leq \frac{1}{\delta n} \sum_{|x_j| > \delta n} |x_j| f(x_j).$$

Nota 4.1. *Esta última desigualdad; que es elemental, es esencialmente la idea que proporciona la desigualdad de Markov. Este tipo de herramientas son las que camufla Feller para poder demostrar resultados de este tipo sin la necesidad de la teoría de la medida.*

El último sumatorio tiende hacia cero cuando $n \rightarrow \infty$. Luego el lado izquierdo de (4.4) también tiende hacia 0. Lo que implica que (4.3) también tienda hacia 0, terminando así la demostración. $\square$

4.3. Ley fuerte de los grandes números

La Ley Débil de los Grandes Números establece que, para cada $\varepsilon > 0$, la probabilidad de que la desviación $|S_n - m_n|$ sea mayor que εn tiende a cero cuando $n \rightarrow \infty$. Es decir, garantiza que la sucesión de promedios S_n/n converge en probabilidad hacia el valor medio esperado m_n/n . Sin embargo, esta forma de convergencia no implica que $|S_n - m_n|/n$ permanezca pequeño para todos los valores grandes de n ; únicamente asegura que la probabilidad de grandes desviaciones decrece. Al trabajar con ensayos de Bernoulli, mencionamos el caso de que incluso bajo la validez de la LDGN, pueden existir casos remotos en los que la desviación relativa siga siendo grande en ciertos momentos, aunque poco probables.

La Ley Fuerte de los Grandes Números, por el contrario, afirma que estos comportamientos remotos no ocurren salvo en un conjunto de probabilidad nula. El objetivo será, por tanto, demostrar que los casos en los que $|S_n - m_n|/n$ se mantiene grande son tan raros que pueden ser ignorados con certeza.

Decimos que la sucesión X_k cumple con la ley fuerte de los grandes números

4.3. LEY FUERTE DE LOS GRANDES NÚMEROS

si para cada par $\epsilon > 0$, $\delta > 0$, existe un N tal que con probabilidad $1 - \delta$ o mejor de que para todo $r > 0$ se satisfacen las $r + 1$ desigualdades

$$\frac{|S_n - m_n|}{n} < \epsilon, \quad n = N, N + 1, \dots, N + r \quad (4.5)$$

Esto significa que, con probabilidad arbitrariamente alta, $|S_n - m_n|/n$ permanecerá pequeño para todos los valores de n suficientemente grandes. Así, la LFGN no se limita a controlar el comportamiento promedio de las desviaciones, sino que garantiza que las desviaciones son pequeñas para todo n suficientemente grande.

En el enfoque seguido por Feller, se trabaja exclusivamente dentro del caso discreto, sin utilizar herramientas de teoría de la medida. Por ello, presentaremos dos demostraciones aparentemente diferentes de la Ley Fuerte de los Grandes Números: la primera, basada en el tratamiento clásico de Feller [8]; sin teoría de la medida, y la segunda, basada en el enfoque más general de Chung [4], donde si se trabaja con la teoría de la medida. Esta distinción será clave para entender no solo las diferencias entre ambas demostraciones de la LFGN; aprendiendo de ambos enfoques metodológicos, sino también las herramientas utilizadas por Feller para construir una demostración de la LFGN independiente.

Antes de presentar la demostración de la Ley Fuerte de los Grandes Números, necesitamos una herramienta fundamental: la desigualdad de Kolmogorov.

Teorema 4.1 (Desigualdad de Kolmogorov). *Sean $X_1, \dots, X_n$ variables aleatorias mutuamente independientes con $E(X_k) = \mu_k$ y varianzas $\text{Var}(X_k) = \sigma_k^2$. Definimos:*

$$\begin{aligned} S_k &= X_1 + \dots + X_k, \\ m_k &= E(S_k) = \mu_1 + \dots + \mu_k \\ s_k^2 &= \text{Var}(S_k) = \sigma_1^2 + \dots + \sigma_k^2 \end{aligned}$$

Entonces, para todo $t > 0$ la probabilidad de que se cumplan simultáneamente las n desigualdades

$$|S_k - m_k| < ts_n \quad k = 1, 2, \dots, n. \quad (4.6)$$

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

es al menos $1 - t^{-2}$, es decir que

$$P(\exists k \in \{1, \dots, n\} : |S_k - m_k| \geq ts_n) \leq t^{-2}.$$

Para $n = 1$ este teorema no es más que la desigualdad de Chebyshev. Sin embargo, para $n > 1$, la desigualdad de Chebyshev da la misma cota para la probabilidad de que se cumpla la desigualdad $|S_n - m_n| < ts_n$, por lo tanto la desigualdad de Kolmogorov es mucho más fuerte.

Demostración. Queremos aproximar la probabilidad x de que al menos una de las desigualdades (4.6) no se cumpla. Definimos variables indicadoras Y_v de la siguiente forma:

$$Y_v = 1 \quad \text{si} \quad |S_v - m_v| \geq ts_n \quad \text{y} \quad |S_k - m_k| < ts_n \quad \text{para} \quad k = 1, 2, \dots, v-1,$$

$$Y_k = 0 \quad \text{en el resto de casos.}$$

En palabras, $Y_v = 1$ si el instante v es el primero en el que no se cumple la desigualdad. Entonces, como mucho uno de los Y_k puede ser igual a 1, por lo tanto la suma $Y_1 + Y_2 + \dots + Y_n$ solo puede tomar los valores 0 o 1, donde es 1 si y solo si al menos una de las desigualdades (4.6) no se cumple. Por lo tanto

$$x = P(Y_1 + \dots + Y_n = 1)$$

Como $Y_1 + \dots + Y_n$ es igual a 0 o a 1, se tiene que $\sum Y_k \leq 1$. Multiplicando ambos lados por $(S_n - m_n)^2$ y tomando esperanzas:

$$E \left[(S_n - m_n)^2 \sum_{k=1}^n Y_k \right] = \sum_{k=1}^n E \left[Y_k (S_n - m_n)^2 \right] \leq E[(S_n - m_n)^2] = s_n^2.$$

Luego

$$\sum_{k=1}^n E \left[Y_k (S_n - m_n)^2 \right] \leq s_n^2 \tag{4.7}$$

Ahora vamos a intentar evaluar el término de la izquierda de (4.7), para ello definimos:

$$U_k := (S_n - m_n) - (S_k - m_k) = \sum_{v=k+1}^n (X_v - \mu_v)$$

Nótese que $E(X_v - \mu_v) = 0$ para cada v , es decir, $E(U_k) = 0$

4.3. LEY FUERTE DE LOS GRANDES NÚMEROS

Ahora reescribimos como sigue:

$$(S_n - m_n)^2 = (S_k - m_k + U_k)^2 = (S_k - m_k)^2 + 2(S_k - m_k)U_k + U_k^2$$

Entonces, al tomar esperanzas, por linealidad:

$$E[Y_k(S_n - m_n)^2] = E[Y_k(S_k - m_k)^2] + 2E[Y_k(S_k - m_k)U_k] + E[Y_k U_k^2].$$

Observamos que:

1. U_k depende sólo de $X_{k+1}, \dots, X_n$,
2. Y_k y S_k dependen sólo de $X_1, \dots, X_k$,

Por lo tanto, $Y_k(S_k - m_k)$ es independiente de U_k , y como $E(U_k) = 0$, se tiene:

$$E[Y_k(S_k - m_k)U_k] = E[Y_k(S_k - m_k)] \cdot E(U_k) = 0.$$

Nota 4.2. *Feller en [8] define la independencia de variables aleatorias como: Sean $X, Y, \dots, W$ variables aleatorias definidas en el mismo espacio muestral. Se dice que son mutuamente independientes si para cualquier combinación de valores $(x, y, \dots, w)$ que puedan tomar se cumple que*

$$P(X = x, Y = y, \dots, W = w) = P(X = x)P(Y = y) \cdots P(W = w)$$

Esta definición no nos aporta nada sobre otras combinaciones de esas mismas variables aleatorias, entonces demostrar que $Y_k(S_k - m_k)$ es independiente de U_k con esta definición de independencia no es en absoluto trivial. Sin embargo, con la teoría de la medida a nuestro alcance sabemos que esto es elemental. Esta es una buena prueba de la oportunidad de la utilidad de los argumentos propios de la Teoría de la Medida en la Teoría de la Probabilidad, y del enriquecimiento de la primera por el desarrollo de la segunda.

Así, concluimos que:

$$E[Y_k(S_n - m_n)^2] \geq E[Y_k(S_k - m_k)^2]. \quad (4.8)$$

Además, si $Y_k = 1$, entonces por definición $|S_k - m_k| \geq ts_n$, y por tanto:

$$Y_k(S_k - m_k)^2 \geq t^2 s_n^2 Y_k.$$

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

Sustituyendo en (4.7) y usando (4.8), obtenemos:

$$s_n^2 \geq \sum_{k=1}^n E[Y_k(S_k - m_k)^2] \geq t^2 s_n^2 \sum_{k=1}^n E(Y_k) = t^2 s_n^2 \cdot E[Y_1 + \cdots + Y_n].$$

Finalmente, como $E(Y_1 + \cdots + Y_n) = P(Y_1 + \cdots + Y_n = 1) = x$, se concluye:

$$x \leq \frac{1}{t^2}.$$

□

Esta desigualdad refuerza el control probabilístico sobre desviaciones grandes en sucesiones de sumas parciales, siendo considerablemente más fuerte que la desigualdad de Chebyshev, ya que acota la probabilidad de que alguna suma parcial se desvíe más de lo permitido.

Una vez vista la desigualdad de Kolmogorov, disponemos de las herramientas necesarias para introducir y demostrar el Criterio de Kolmogorov, clave en la demostración de la LFGN, y en buena parte de lo que resta de trabajo.

Criterio de Kolmogorov. *La convergencia de la serie*

$$\sum \frac{\sigma_k^2}{k^2}$$

es una condición suficiente para que la sucesión de variables aleatorias mutuamente independientes X_k con varianzas σ_k^2 cumpla la ley fuerte de los grandes números, es decir:

$$\frac{S_n - E(S_n)}{n} \rightarrow 0 \text{ casi seguro}$$

Demostración. Sea A_ν el evento de que existe al menos un n con $2^{\nu-1} < n \leq 2^\nu$ tal que la desigualdad (4.5) no se cumple. Es obvio que, basta con probar que para todos los ν suficientemente grandes y todo r se cumple que

$$P\{A_\nu\} + P\{A_{\nu+1}\} + \cdots + P\{A_{\nu+r}\} < \delta,$$

es decir, que la serie $\sum P\{A_\nu\}$ converja. Ahora bien, el evento A_ν implica que para algún n con $2^{\nu-1} < n \leq 2^\nu$ se cumple

$$|S_n - m_n| \geq \varepsilon \cdot 2^{\nu-1}$$

4.3. LEY FUERTE DE LOS GRANDES NÚMEROS

y por la desigualdad de Kolmogorov:

$$P\{A_\nu\} \leq 4\varepsilon^2 \cdot \sigma_{2^\nu}^2 \cdot 2^{-2\nu}$$

Entonces:

$$\sum_{\nu=1}^{\infty} P\{A_\nu\} \leq 4\varepsilon^{-2} \sum_{\nu=1}^{\infty} 2^{-2\nu} \sum_{k=1}^{2^\nu} \sigma_k^2 = 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{2^\nu \geq k} 2^{-2\nu} \leq 8\varepsilon^{-2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

lo que, por el primer lema de Borel-Cantelli completa la demostración. $\square$

Ahora, aplicando el criterio de Kolmogorov podemos probar que

Teorema 4.2. *Si las variables aleatorias mutuamente independientes X_k tienen una distribución común dada por la función de masa de probabilidad $f(x_j)$ y si $\mu = E(X_k)$ existe, entonces la ley fuerte de los grandes números se aplica a la sucesión $\{X_k\}$.*

Este teorema es, evidentemente, más fuerte que la ley débil demostrada anteriormente. Se tratan por separado debido al interés metodológico de sus demostraciones.

Demostración. Usamos nuevamente el método de truncamiento. Introducimos dos nuevas sucesiones de variables aleatorias:

$$U_k = \begin{cases} X_k, & \text{si } |X_k| < k \\ 0, & \text{si } |X_k| \geq k \end{cases}, \quad V_k = \begin{cases} 0, & \text{si } |X_k| < k \\ X_k, & \text{si } |X_k| \geq k \end{cases}$$

Las U_k son mutuamente independientes; aunque ya no son igualmente distribuidas, y queremos ver que satisfacen el criterio de Kolmogorov. Para $\sigma_k^2 = \text{Var}(U_k)$ tenemos:

$$\sigma_k^2 \leq E(U_k^2) = \sum_{|x_j| < k} x_j^2 f(x_j). \quad (4.9)$$

Por comodidad, definimos:

$$a_\nu = \sum_{v-1 \leq |x_j| < v} |x_j| f(x_j).$$

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

Entonces la serie $\sum a_\nu$ converge, ya que $E(X_k)$ existe. Además, por (4.9),

$$\sigma_k^2 \leq a_1 + 2a_2 + 3a_3 + \cdots + ka_k,$$

y

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{\nu=1}^k \nu a_\nu = \sum_{\nu=1}^{\infty} \nu a_\nu \sum_{k=\nu}^{\infty} \frac{1}{k^2} < 2 \sum_{\nu=1}^{\infty} a_\nu < \infty. \quad (4.10)$$

Así, el criterio de Kolmogorov se cumple para $\{U_k\}$. Ahora:

$$E(U_k) = \mu_k = \sum_{|x_j| < k} x_j f(x_j),$$

por lo tanto, $\mu_k \rightarrow \mu$, y entonces $(\mu_1 + \mu_2 + \cdots + \mu_n)/n \rightarrow \mu$. De la ley fuerte de los grandes números para $\{U_k\}$ concluimos por tanto que, con probabilidad $1 - \delta$, o mayor:

$$\left| \frac{1}{n} \sum_{k=1}^n U_k - \mu \right| < \varepsilon,$$

para todo $n > N$, siempre que N se elija suficientemente grande. Queda por demostrar que la misma afirmación es cierta cuando cambiamos los U_k por los X_k . Para ello basta con ver que N puede ser elegido tan grande que con una probabilidad arbitrariamente cercana a 1 el evento $U_k = X_k$ ocurre para todo $k > N$. Esto también quiere decir que con probabilidad 1 solamente un número finito de variables V_k son distintas de cero. Por el primer lema de Borel-Cantelli este es el caso en el que la serie $\sum P\{V_k \neq 0\}$ converge, ahora podemos completar la demostración probando la convergencia de esta serie.

$$P\{V_k \neq 0\} = \sum_{|x_j| \geq n} f(x_j) \leq \frac{a_{n+1}}{n} + \frac{a_{n+2}}{n+1} + \frac{a_{n+3}}{n+2} + \cdots$$

por lo tanto

$$\sum P\{V_k \neq 0\} \leq \sum_{n=1}^{\infty} \sum_{\nu=n}^{\infty} \frac{a_{\nu+1}}{\nu} \sum_{n=1}^{\nu} 1 = \sum_{\nu} a_{\nu+1} < \infty$$

probando así la convergencia. □

Presentamos ahora una segunda demostración de la Ley Fuerte de los Grandes Números, tomada del libro de Chung, que se apoya en conceptos generales de probabilidad y convergencia. Antes de exponerla, introduciremos algunas

4.3. LEY FUERTE DE LOS GRANDES NÚMEROS

nociones previas necesarias

Definición 4.1. Sean $\{X_n\}$ e $\{Y_n\}$ dos sucesiones de variables aleatorias, se dice que son asintóticamente equivalentes si y solo si

$$\sum_n P(X_n \neq Y_n) < \infty \quad (4.11)$$

En la práctica obtendremos sucesiones equivalentes a partir del método de truncamiento que vimos anteriormente.

Teorema 4.3. *Se tiene que*

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$$

luego $E(|X|) < \infty$ si y solo si esa serie converge. En nuestro caso particular utilizaremos este teorema para ver que series de este tipo convergen como consecuencia de conocer que la esperanza es finita y viceversa.

Teorema 4.4. *Si $\{X_n\}$ e $\{Y_n\}$ son asintóticamente equivalentes, entonces*

$$\sum_n (X_n - Y_n) \quad \text{converge c.s.}$$

Además, si $a_n \uparrow \infty$, entonces

$$\frac{1}{a_n} \sum_{j=1}^n (X_j - Y_j) \rightarrow 0 \quad \text{c.s.}$$

Demostración. Por el lema de Borel–Cantelli y por ser $\{X_n\}$ e $\{Y_n\}$ equivalentes, (4.11) implica que

$$P(X_n \neq Y_n \text{ infinitas veces}) = 0.$$

Esto quiere decir que existe un conjunto de medida nula N tal que, si $\omega \in \Omega \setminus N$, existe un entero $n_0(\omega)$ de modo que si

$$n \geq n_0(\omega) \Rightarrow X_n(\omega) = Y_n(\omega).$$

Es decir, para cada $\omega \notin N$, las sucesiones $\{X_n(\omega)\}$ y $\{Y_n(\omega)\}$ difieren sólo en un número finito de términos (cuántos, depende de ω). En otras palabras, la

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

serie

$$\sum_n (X_n(\omega) - Y_n(\omega))$$

está compuesta por ceros a partir de cierto índice en adelante. Ambas afirmaciones del teorema se deducen trivialmente de este hecho. $\square$

Corolario 4.1. *Si $\{X_n\}$ e $\{Y_n\}$ son asintóticamente equivalentes, entonces con probabilidad 1 la expresión*

$$\sum_n X_n \quad o \quad \frac{1}{a_n} \sum_{j=1}^n X_j$$

converge, diverge a $+\infty$ o $-\infty$, o actúa de igual manera que

$$\sum_n Y_n \quad o \quad \frac{1}{a_n} \sum_{j=1}^n Y_j$$

respectivamente. En particular, si

$$\frac{1}{a_n} \sum_{j=1}^n X_j$$

converge a X en probabilidad, entonces lo hace de igual manera

$$\frac{1}{a_n} \sum_{j=1}^n Y_j$$

Lema 4.1 (Lema de Kronecker). *Sea $\{x_k\}$ una sucesión de números reales, $\{a_k\}$ una sucesión de números > 0 y tales que $\uparrow \infty$. Entonces*

$$\sum_n \frac{x_n}{a_n} < \infty \text{ (converge)} \Rightarrow \frac{1}{a_n} \sum_{j=1}^n x_j \rightarrow 0$$

Demostración. Para $1 \leq n \leq \infty$, definimos

$$b_n = \sum_{j=1}^n \frac{x_j}{a_j}$$

Si consideramos $a_0 = 0$, $b_0 = 0$ tenemos que $x_n = a_n(b_n - b_{n-1})$ y además, por

4.3. LEY FUERTE DE LOS GRANDES NÚMEROS

la fórmula de sumación de Abel:

$$\frac{1}{a_n} \sum_{j=1}^n x_j = \frac{1}{a_n} \sum_{j=1}^n a_j (b_j - b_{j-1}) = b_n - \frac{1}{a_n} \sum_{j=0}^{n-1} b_j (a_{j+1} - a_j)$$

Ahora como $a_{j+1} - a_j \geq 0$ es obvio que

$$\frac{1}{a_n} \sum_{j=0}^{n-1} (a_{j+1} - a_j) = 1$$

Y como $b_n \rightarrow b_\infty$, que es un número finito por definición, se tiene que

$$\frac{1}{a_n} \sum_{j=1}^n x_j \rightarrow b_\infty - b_\infty = 0$$

Con lo que se concluye la demostración □

Teorema 4.5. *Sea $\{X_n\}$ una sucesión de variables aleatorias independientes tales que $E(X_n) = 0$ para todo n , sea $\{a_n\}$ una sucesión tal que $0 < a_n \uparrow \infty$. Y sea φ una función positiva, par y continua en $\mathbb{R}^1$ tal que cuando $|x|$ crece se tiene que $\frac{\varphi(x)}{|x|} \uparrow$ y $\frac{\varphi(x)}{|x|} \downarrow$. Entonces si*

$$\sum_n \frac{E(\varphi(X_n))}{\varphi(a_n)} < \infty$$

se tiene que

$$\sum_n \frac{X_n}{a_n} \text{ converge casi seguro} \quad (4.12)$$

Nota 4.3. *Nótese que si tomamos $\varphi(x) = x^2$ y $a_n = n$, obtenemos el Criterio de Kolmogorov. Es decir, el Criterio de Kolmogorov es un caso particular del Teorema que acabamos de presentar.*

Aplicando el Lema de Kronecker a (4.12) a cada ω de un conjunto de probabilidad uno, obtenemos el siguiente resultado

Corolario 4.2. *Bajo las hipótesis del Teorema anterior, tenemos que*

$$\frac{1}{a_n} \sum_{j=1}^n X_j \rightarrow 0 \text{ c.s.}$$

Un caso particular y de nuestro interés de este corolario es el siguiente.

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

Sea $\varphi(x) = |x|^p$, con $1 \leq p \leq 2$ y donde $a_n = n$. Entonces tenemos que:

$$\sum_n \frac{1}{n^p} E(|X_n|^p) < \infty \Rightarrow \frac{1}{n} \sum_{j=1}^n X_j \rightarrow 0 \text{ c.s.} \quad (4.13)$$

Una vez vistos estos resultados previos, procedemos a dar una segunda demostración de la ley fuerte de los grandes números.

Teorema 4.6. *Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e igualmente distribuidas. Entonces se tiene que:*

$$E(|X_1|) < \infty \Rightarrow \frac{S_n}{n} \rightarrow E(X_1) \text{ c.s.} \quad (4.14)$$

Demostración. Para probar (4.14) lo primero que haremos será definir una sucesión de variables aleatorias $\{Y_n\}$ truncadas de la siguiente manera:

$$Y_n(\omega) = \begin{cases} X_n(\omega), & \text{si } |X_n(\omega)| \leq n, \\ 0, & \text{si } |X_n(\omega)| > n. \end{cases}$$

Según esta definición y debido a que todas las variables aleatorias X_n son igualmente distribuidas, la siguiente igualdad es evidente

$$\sum_n P(X_n \neq Y_n) = \sum_n P(|X_n| > n) = \sum_n P(|X_1| > n)$$

Además, por el Teorema 4.3, como $E(|X_1|) < \infty$, tenemos que $\sum_n P(|X_1| > n) < \infty$ y por tanto $\{X_n\}$ e $\{Y_n\}$ son asintóticamente equivalentes.

Apliquemos ahora (4.13) a $\{Y_n - E(Y_n)\}$, con $\varphi(x) = x^2$. Obtenemos lo siguiente:

$$\sum_n \frac{\sigma^2(Y_n)}{n^2} \leq \sum_n \frac{E(Y_n^2)}{n^2} = \sum_n \frac{1}{n^2} \int_{|x| \leq n} x^2 dF(x) \quad (4.15)$$

donde $F(x)$ es la función de distribución común a todas las variables aleatorias. Ahora bien, por definición, las Y_n están acotadas, y por tanto el segundo momento que aparece en la anterior desigualdad existe. Intentaremos ahora apro-

4.4. COMPARACIÓN ENTRE DEMOSTRACIONES

comparar ese valor en términos del primer momento:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^n \int_{j-1 < |x| < j} x^2 dF(x) &= \sum_{j=1}^{\infty} \int_{j-1 < |x| < j} x^2 dF(x) \sum_{n=j}^{\infty} \frac{1}{n^2} \\
&\leq \sum_{j=1}^{\infty} j \int_{j-1 < |x| < j} |x| dF(x) \cdot \frac{C}{j} \leq C \sum_{j=1}^{\infty} \int_{j-1 < |x| < j} |x| dF(x) \\
&= CE(|X_1|) < \infty
\end{aligned} \tag{4.16}$$

donde hemos utilizado que $\sum_{n=j}^{\infty} n^{-2} \leq Cj^{-1}$ para algún $C > 0$ y para todo $j \geq 1$. Con esto hemos probado que la serie $\sum_n \frac{\sigma^2(Y_n)}{n^2}$ converge y por (4.13) concluimos que

$$\frac{1}{n} \sum_{j=1}^n \{Y_j - E(Y_j)\} \rightarrow 0 \text{ c.s.}$$

Obviamente $E(Y_n) \rightarrow E(X_1)$ cuando $n \rightarrow \infty$, por lo tanto

$$\frac{1}{n} \sum_{j=1}^n E(Y_j) \rightarrow E(X_1)$$

Y como $\frac{1}{n} \sum_{j=1}^n \{Y_j - E(Y_j)\} \rightarrow 0$ c.s, entonces:

$$\frac{1}{n} \sum_{j=1}^n Y_j \rightarrow E(X_1) \text{ c.s.}$$

Ahora, por el Teorema 4.4, podemos sustituir $\frac{1}{n} \sum_{j=1}^n Y_j$ por $\frac{1}{n} \sum_{j=1}^n X_j$, llegando a que:

$$\frac{1}{n} \sum_{j=1}^n X_j \rightarrow E(X_1) \text{ c.s.}$$

Quedando así probado (4.14) □

Posteriormente veremos como existen generalizaciones de este Teorema en las que se consideran diferentes condiciones. Como por ejemplo el Teorema de Etemadi, en el cual se consideran variables aleatorias independientes por pares en lugar de mutuamente independientes.

4.4. Comparación entre demostraciones

Una vez presentadas ambas demostraciones de la Ley Fuerte de los Grandes Números, vamos a destacar los puntos más importantes de las mismas

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

para intentar encontrar las herramientas utilizadas por Feller para superar la condición autoimpuesta de trabajar en el ámbito discreto.

Ambas demostraciones tienen tres puntos clave

1. Utilizar el método de truncamiento
2. Demostrar que las sucesiones truncadas cumplen el Criterio de Kolmogorov
3. Con lo anterior, probar que si las variables aleatorias truncadas satisfacen la LFGN entonces las variables aleatorias originales también.

Repasaremos ahora estos tres puntos clave uno a uno:

1. Método de truncamiento

En ambos casos hemos utilizado el método de truncamiento. Sin embargo, salta a la vista la principal diferencia en los truncamientos. En el Teorema 4.2 hemos introducido dos sucesiones de variables aleatorias de la siguiente manera:

$$U_k = \begin{cases} X_k, & \text{si } |X_k| < k \\ 0, & \text{si } |X_k| \geq k \end{cases}, \quad V_k = \begin{cases} 0, & \text{si } |X_k| < k \\ X_k, & \text{si } |X_k| \geq k \end{cases}$$

mientras que en el Teorema 4.6 solo hemos necesitado una nueva sucesión de variables aleatorias truncadas:

$$Y_n(\omega) = \begin{cases} X_n(\omega), & \text{si } |X_n(\omega)| \leq n, \\ 0, & \text{si } |X_n(\omega)| > n. \end{cases}$$

En el punto (3) entenderemos el porqué de esta diferencia, y veremos que la existencia de las V_k es la herramienta auxiliar que en este caso utiliza Feller para poder avanzar en la demostración sin la teoría de la medida a su alcance.

2. Criterio de Kolmogorov

En ambas demostraciones se prueba el Criterio de Kolmogorov con el mismo objetivo, demostrar que las variables aleatorias truncadas U_k e Y_n satisfacen la LFGN. Bien es cierto que en el Teorema 4.6 no se menciona directamente el Criterio de Kolmogorov, esto es por que utilizamos el Teorema 4.5 que no es

4.4. COMPARACIÓN ENTRE DEMOSTRACIONES

más que una generalización del Criterio de Kolmogorov (tal y como se explica en la nota 4.3).

Lo importante aquí es ver que la manera de acotar la serie del Criterio de Kolmogorov es esencialmente la misma, pero utilizando en cada caso las herramientas de las que disponemos. En ambos casos utilizamos la desigualdad trivial por la cual $\text{Var}(X) \leq E(X^2)$. Ahora bien, por un lado en el Teorema 4.2 definimos $E(U_k^2) = \sum_{|x_j| < k} x_j^2 f(x_j)$ y acotamos $\sigma^2(U_k)$ de la siguiente manera:

$$\sigma^2(U_k) \leq \sum_{|x_j| < k} x_j^2 f(x_j) \leq \sum_{v=1}^k v \sum_{v-1 \leq |x_j| < v} |x_j| f(x_j)$$

Obteniendo así que

$$\sum_{k=1}^{\infty} \frac{\sigma^2(U_k)}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{|x_k| < k} x_k^2 f(x_k) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{v=1}^k v \sum_{v-1 \leq |x_j| < v} |x_j| f(x_j) \quad (4.17)$$

Por otro lado en el Teorema 4.6 se define la esperanza como una integral en vez de como un sumatorio gracias a la teoría de la medida y se acota de la siguiente manera:

$$\sum_{n=1}^{\infty} \frac{\sigma^2(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{|x| \leq n} x^2 dF(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^n \int_{j-1 < |x| < j} x^2 dF(x) \quad (4.18)$$

Las similitudes entre (4.17) y (4.18) son evidentes, su idea es exactamente. De hecho estas similitudes continúan durante la prueba de la convergencia de la serie del Criterio de Kolmogorov. En ambas se hace exactamente lo mismo, reordenar sumatorios y utilizar la cota por la cual $\sum_{n=j}^{\infty} n^{-2} \leq Cj^{-1}$ para algún $C > 0$. (Ver (4.10) y (4.16))

3. Equivalencia de las variables aleatorias

En la demostración del Teorema 4.6 lo primero que se ve es que $\{X_n\}$ e $\{Y_n\}$ son asintóticamente equivalentes, esto es muy sencillo de probar por la caracterización de esperanza finita. Con esto y por el Teorema 4.4 se concluye que $\{X_n\}$ satisface la LFGN.

Ahora bien; en el caso discreto (Teorema 4.2), una vez probado el Criterio de Kolmogorov es claro que las U_k satisfacen la LFGN, queremos ahora ver que las X_k también lo hacen, esto es equivalente a ver que $P(U_k = X_k) = 1$. Hasta este punto esto es exactamente lo mismo que lo que se busca en el Teorema

CAPÍTULO 4. DISTRIBUCIONES MÁS GENERALES

4.6, pero ahora no contamos con la caracterización de esperanza finita, y ahí es donde entra en juego la herramienta auxiliar de la que habíamos hablado en el punto número uno: las variables aleatorias V_k .

Nótese que por definición $\{U_k = X_k\} = \{V_k = 0\}$. Luego por el primer Lema de Borel-Cantelli si vemos que

$$\sum P(V_k \neq 0) < \infty$$

Entonces se tendría $P(V_k \neq 0 \text{ i.v.}) = P(U_k \neq X_k \text{ i.v.}) = 1$, concluyendo así en que las X_k satisfacen la LFGN.

Conclusión

Tras haber analizado en detalle ambas demostraciones, podemos concluir que, a pesar de desarrollarse en marcos teóricos distintos, ambas comparten una estructura esencialmente idéntica.

Las tres ideas clave que estructuran ambas demostraciones se mantienen inalteradas, variando únicamente las herramientas disponibles en cada marco. Así, lo que cambia no es el esquema lógico de la demostración, sino el lenguaje técnico y los recursos empleados para ejecutarla.

Capítulo 5

Teorema de Etemadi

Visto lo anterior tenemos sentadas las principales bases de las leyes de los grandes números. Primero hemos tratado el caso más primitivo e intuitivo considerando ensayos de Bernoulli, probando la convergencia en probabilidad y la convergencia casi seguro. Posteriormente hemos generalizado el concepto considerando sucesiones de variables aleatorias independientes e igualmente distribuidas, probando de nuevo ambos tipos de convergencia y analizando distintos enfoques metodológicos para la Ley Fuerte de los Grandes Números.

A partir de ahora nuestro objetivo será presentar versiones de la Ley de los Grandes Números para variables aleatorias idénticamente distribuidas bajo condiciones menos restrictivas que las que han sido necesarias hasta ahora, en particular, la independencia entre las variables aleatorias. Comenzaremos con un resultado presentado por Etemadi en [7], quien presentó una demostración relativamente sencilla de una versión de la Ley Fuerte de los Grandes Números bajo condiciones considerablemente más débiles y utilizando exclusivamente la desigualdad de Chebyshev, en lugar de la desigualdad de Kolmogorov.

Teorema de Etemadi. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes por pares e igualmente distribuidas con $E|X_1| < \infty$. Entonces*

$$\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = EX_1 \text{ casi seguro (c.s.)}$$

Es fácil ver que la principal mejora implementada por Etemadi con respecto a la ley de los grandes números presentada por Kolmogorov es que ya no sería necesario considerar variables aleatorias mutuamente independientes, sino que

CAPÍTULO 5. TEOREMA DE ETEMADI

bastaría con considerarlas independientes por pares. Mientras que la condición de idénticamente distribuidas permanece.

Demostración. Como toda variable aleatoria real X_i se puede descomponer en su parte positiva y negativa como $X_i = X_i^+ - X_i^-$, donde $X_i^+ = \max(X_i, 0)$ y $X_i^- = \max(-X_i, 0)$, y además

$$|X_i| = X_i^+ + X_i^-,$$

entonces si $E|X_1| < \infty$, también se cumple que $E(X_1^+) < \infty$ y $E(X_1^-) < \infty$. Es decir, tanto la parte positiva como la parte negativa de cada variable aleatoria X_i cumplen las hipótesis del teorema, por lo tanto podemos suponer sin pérdida de generalidad que $X_i \geq 0$. Sea $Y_i = X_i I_{\{X_i \leq i\}}$ con I la función indicadora y $S_n^* = \sum_{i=1}^n Y_i$. Para $\varepsilon > 0$, sea $k_n = [\alpha^n]$ con $\alpha > 1$. Utilizando la independencia para reescribir la varianza de la siguiente manera:

$$\text{Var}(S_{k_n}^*) = \sum_{i=1}^{k_n} \text{Var}(Y_i) \leq \sum_{i=1}^{k_n} E(Y_i^2) = \sum_{i=1}^{k_n} E(X_i^2 I_{\{X_i \leq i\}}) \leq k_n E(X_1^2 I_{\{X_1 \leq k_n\}})$$

Ahora por la desigualdad de Chebyshev y por lo anterior se tiene que:

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\left|\frac{S_{k_n}^* - E(S_{k_n}^*)}{k_n}\right| > \varepsilon\right) &\leq c \sum_{n=1}^{\infty} \frac{\text{Var}(S_{k_n}^*)}{k_n^2} \\ &\leq c \sum_{n=1}^{\infty} \frac{1}{k_n} E(X_1^2 I_{\{X_1 \leq k_n\}}) \leq cE\left(X_1^2 \sum_{n=1}^{\infty} \frac{1}{k_n} I_{\{X_1 \leq k_n\}}\right) \end{aligned}$$

donde $c = 1/\varepsilon^2$.

Sea $K = 2\alpha/(\alpha - 1)$ y $x > 0$. Si N es el primer n natural tal que $k_n \geq x$, entonces $\alpha^N \geq x$, y como $y \leq 2[y]$ para $y \geq 1$,

$$\sum_{k_n \geq x} \frac{1}{k_n} \leq 2 \sum_{n \geq N} \frac{1}{\alpha^n} = K \frac{1}{\alpha^N} \leq K \frac{1}{x}$$

Por lo tanto

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_{k_n}^* - E(S_{k_n}^*)}{k_n}\right| > \varepsilon\right) \leq cE\left(X_1^2 \sum_{n=1}^{\infty} \frac{1}{k_n} I_{\{X_1 \leq k_n\}}\right) \quad (5.1)$$

$$\leq cE(K X_1^2 X_1^{-1}) = KxE(X_1) < \infty \quad (5.2)$$

Además, por definición, tenemos que

$$E(X_1) = \lim_{n \rightarrow \infty} \int_0^n x dF(x) = \lim_{n \rightarrow \infty} E(Y_n) = \lim_{n \rightarrow \infty} \frac{E(S_{k_n}^*)}{k_n} \quad (5.3)$$

Ahora, por el lema de Borel-Cantelli tenemos que:

$$\lim_{n \rightarrow \infty} \frac{S_{k_n}^*}{k_n} = E(X_1) \quad \text{casi seguro.}$$

Además,

$$\begin{aligned} \sum_{n=1}^{\infty} P(Y_n \neq X_n) &= \sum_{n=1}^{\infty} P(X_n > n) = \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} \int_i^{i+1} dF(x) = \sum_{i=1}^{\infty} i \int_i^{i+1} dF(x) \leq \\ &\leq \sum_{i=1}^{\infty} \int_i^{i+1} x dF(x) \leq E(X_1) < \infty \end{aligned}$$

Por el lema de Borel-Cantelli, el evento $X_n \neq Y_n$ ocurre solo un número finito de veces, y por tanto

$$\lim_{n \rightarrow \infty} \frac{S_{k_n}}{k_n} = E(X_1) \quad \text{casi seguro.} \quad (5.4)$$

Ahora, como X_i son variables aleatorias positivas se tiene que los sumatorios S_n son monótonamente crecientes. Es obvio que para cualquier $n \in \mathbb{N}$ existe m tal que $k_m \leq n \leq k_{m+1}$. Por ser S_n monótona se cumple

$$S_{k_m} \leq S_n \leq S_{k_{m+1}} \quad (5.5)$$

dividiendo todo entre $n \in [k_m, k_{m+1}]$, se cumple que:

$$\frac{S_{k_m}}{k_{m+1}} \leq \frac{S_n}{n} \leq \frac{S_{k_{m+1}}}{k_m} \quad (5.6)$$

Tenemos ahora una cota superior e inferior de $\frac{S_n}{n}$ en función de los valores de S_{k_m} que sabemos que convergen. Tomando límite inferior y límite superior de los extremos de la desigualdad y teniendo en cuenta que $k_{m+1} \leq \alpha k_m$ obtenemos:

$$\frac{1}{\alpha} E(X_1) \leq \liminf_{n \rightarrow \infty} \frac{S_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \alpha E(X_1) \quad (5.7)$$

CAPÍTULO 5. TEOREMA DE ETEMADI

Finalmente, tomando $\alpha \rightarrow 1$, concluimos que:

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = E(X_1) \quad \text{casi seguro.}$$

□

Nota 5.1. *Nótese que gracias a la construcción de los k_n se puede demostrar la cota*

$$\sum_{n=1}^{\infty} k_n^{-1} I_{\{X_1 \leq k_n\}} \leq \frac{2\alpha}{\alpha - 1} X_1^{-1}$$

clave para probar que las variables truncadas Y_n satisfacen la LFGN a lo largo de la subsucesión $\{k_n\}$.

Con esta demostración, Etemadi establece una versión más general de la Ley Fuerte de los Grandes Números, al relajar el requisito clásico de independencia mutua por el de independencia por pares, manteniendo únicamente la hipótesis de idéntica distribución. Además, logra este resultado con una prueba sencilla y elegante, sin necesidad de utilizar ni la desigualdad de Kolmogorov ni su criterio de convergencia.

La clave de esta demostración; y su diferencia principal respecto a la de Kolmogorov, reside en la construcción de la sucesión $k_n = [\alpha^n]$, con $\alpha > 1$. Esta elección es fundamental para el desarrollo de la prueba. Desde el inicio, los k_n permiten aplicar el Lema de Borel-Cantelli gracias a una cota precisa sobre la suma $\sum_{n=1}^{\infty} k_n^{-1} I_{\{X_1 \leq k_n\}}$, como se ha discutido en la Nota 5.1.

Además esta construcción también es clave en la parte final de la demostración, al querer probar

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = E(X_1)$$

a partir del límite a lo largo de la subsucesión $\{k_n\}$. Gracias a $k_n = [\alpha^n]$, podemos asegurar que para todo $n \in \mathbb{N}$ existe un m tal que $k_m \leq n \leq k_{m+1}$. Esta propiedad permite comparar $\frac{S_n}{n}$ con los cocientes $\frac{S_{k_m}}{k_{m+1}}$ y $\frac{S_{k_{m+1}}}{k_m}$, los cuales sí sabemos que convergen.

Tal control no sería posible si simplemente usáramos estimaciones como $\frac{S_n}{n+1}$ o $\frac{S_{n+1}}{n}$, pues no tendríamos una relación clara entre los denominadores y los índices de los sumatorios. En cambio, al trabajar con los k_n , obtenemos cotas por arriba y por abajo que, al tomar límites y hacer tender $\alpha \rightarrow 1$, permiten concluir rigurosamente la convergencia de $\frac{S_n}{n}$ hacia $E(X_1)$ casi seguro.

5.1. Extensiones teorema de Etemadi

El teorema de Etemadi se ha extendido en diversas ocasiones, proporcionando condiciones suficientes para que una sucesión de variables aleatorias, no necesariamente idénticamente distribuidas, satisfaga la ley fuerte de los grandes números. Por ejemplo en [3]:

Teorema 5.1 (Chandra y Goswami). *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes por pares tal que*

$$\int_0^\infty \sup_{n \in \mathbb{N}} P(|X_n| > t) dt < \infty$$

Entonces

$$\lim_{n \rightarrow \infty} \frac{S_n - E(S_n)}{n} = 0 \quad \text{casi seguro}$$

Este teorema trata un caso muy general, sin embargo su demostración no será de nuestro interés, veremos una manera aún más sencilla de trabajar y de probar la ley de los grandes números incluso en el caso de que las variables no sean idénticamente distribuidas. Para ello, nos apoyaremos y seguiremos la estructura de [9], primero introduciremos los siguientes conceptos:

Definición 5.1. Una sucesión de variables aleatorias (arbitrariamente distribuidas) $(X_n)_{n \in \mathbb{N}}$ se dice que:

- (I) Satisface la condición de Kolmogorov si, y solo si, cada una de las X_n , $n \in \mathbb{N}$, tiene varianza finita $\text{Var}(X_n)$ y además se cumple que

$$\sum_{n \in \mathbb{N}} \frac{\text{Var}(X_n)}{n^2} < \infty;$$

- (II) Es cuasi-incorrelada si, y solo si, cada una de las X_n , $n \in \mathbb{N}$, tiene varianza finita, y existe una constante positiva c tal que $\text{Var}(S_n) \leq c \sum_{k=1}^n \text{Var}(X_k)$ para todo $n \in \mathbb{N}$. Hemos de tener en cuenta que si las X_n son incorreladas por pares, entonces tenemos la igualdad $\text{Var}(S_n) = \sum_{k=1}^n \text{Var}(X_k)$ (Bienaymé) y así las X_n son, en particular, cuasi-incorreladas.

- (III) Satisface la Ley Fuerte de los Grandes Números (LFGN) si, y solo si,

$$\lim_{n \rightarrow \infty} \frac{S_n - E(S_n)}{n} = 0 \quad \text{casi seguro};$$

CAPÍTULO 5. TEOREMA DE ETEMADI

donde $S_n = X_1 + X_2 + \cdots + X_n$.

Además, si existe el límite $a := \lim_{n \rightarrow \infty} E(X_n)$, entonces (por la sumación de Cesàro) se tiene que $\lim_{n \rightarrow \infty} n^{-1}S_n = a$ casi seguro.

Una vez introducidas estas nociones podemos ver el siguiente resultado:

Teorema 5.2 (LFGN bajo la condición de Kolmogorov). *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias positivas, cuasi-incorreladas y satisfaciendo la condición de Kolmogorov, tal que $A := \sup_{n \in \mathbb{N}} n^{-1}E(S_n) < \infty$. Entonces la sucesión $(X_n)_{n \in \mathbb{N}}$ cumple la LFGN.*

Se define el ínfimo esencial de una variable aleatoria X como el mayor número r tal que $X \geq r$ casi seguro, y lo denotamos como $\text{essinf} X$. Considerando la variable aleatoria $\widetilde{X}_n = X_n - \text{essinf} X_n$ y aplicando el teorema 5.2 podemos sustituir $\sup_{n \in \mathbb{N}} n^{-1}E(S_n) < \infty$ por una condición más débil como la que sigue:

$$\sup_{n \in \mathbb{N}} \frac{E(S_n) - \sum_{k=1}^n \text{essinf} X_k}{n} < \infty \quad (5.8)$$

Además, para cualquier sucesión de números $a_n \geq 0$, si se cumple que $\sup_{n \in \mathbb{N}} a_n = \infty$ entonces necesariamente $\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sup_{k \geq n} a_k = \infty$. Con esto la condición (5.8) es equivalente a

$$\limsup_{n \rightarrow \infty} \frac{E(S_n) - \sum_{k=1}^n \text{essinf} X_k}{n} < \infty$$

También la condición de no negatividad de las variables aleatorias puede ser sustituida por que cada una de ellas esté acotada uniformemente casi seguro inferior o superiormente. Por ejemplo, si cada una de las X_n está acotada inferiormente por un número real r , podemos aplicar el Teorema 5.2 a la transformación $\widetilde{X}_n := X_n - r$.

Para tratar la demostración de este teorema, hemos de introducir previamente algunas definiciones y probar dos lemas auxiliares que serán claves en el desarrollo de la prueba.

Definición 5.2. Fijemos, de ahora en adelante, dos números reales positivos $\alpha > 1$ y $\varepsilon > 0$ y consideremos variables aleatorias X_n que cumplan las condiciones del Teorema. Recordemos que, por hipótesis, $A := \sup_{n \in \mathbb{N}} n^{-1}E(S_n) < \infty$.

5.1. EXTENSIONES TEOREMA DE ETEMADI

Entonces podemos definir el número entero $L := \lfloor A/\varepsilon \rfloor$, donde $\lfloor x \rfloor$ representa la parte entera de x , es decir, el mayor entero no superior a x .

Para cada $n \in \mathbb{N}$, definimos $m(n) := \lfloor \log_\alpha n \rfloor$. Así, se cumple que

$$m(n) \rightarrow \infty \quad \text{cuando } n \rightarrow \infty \quad \text{y} \quad \alpha^{m(n)} \leq n < \alpha^{m(n)+1} \quad \text{para todo } n.$$

Dada una variable aleatoria positiva X_n , sea $s(n) \in \{0, 1, \dots, L\}$ números naturales tales que

$$\frac{E(S_n)}{n} \in [\varepsilon \cdot s(n), \varepsilon \cdot (s(n) + 1))$$

Los $s(n)$ siempre están bien definidos porque los intervalos $[\varepsilon \cdot s(n), \varepsilon \cdot (s(n) + 1))$ son disjuntos y su unión contiene al intervalo $[0, A]$.

Definición 5.3. Sean X_n variables aleatorias como en el Teorema. Para todo $n \in \mathbb{N}$ y $s \in \{0, 1, \dots, L\}$, definimos el conjunto

$$T_{n,s} := \left\{ k \in \mathbb{N} : \alpha^n \leq k < \alpha^{n+1} \text{ y } \varepsilon \cdot s \leq \frac{E(S_k)}{k} < \varepsilon \cdot (s + 1) \right\}.$$

Puede ocurrir que, para ciertos valores de n y s , el conjunto $T_{n,s}$ sea vacío. Sin embargo, el conjunto $T_{m(n),s(n)}$ no lo está, ya que por construcción $n \in T_{m(n),s(n)}$ para todo $n \in \mathbb{N}$.

Si $T_{n,s}$ no es vacío, definimos:

$$k(n, s)^+ := \max T_{n,s}, \quad k(n, s)^- := \min T_{n,s}.$$

Por otro lado, si $T_{n,s}$ es vacío, por conveniencia definimos:

$$k(n, s)^+ = k(n, s)^- := \lfloor \alpha^n \rfloor.$$

De ahora en adelante, utilizaremos la notación $k(n, s)^\pm$ cuando una afirmación sea válida tanto para $k(n, s)^+$ como para $k(n, s)^-$. Por ejemplo, se tiene que $k(n, s)^\pm \geq \lfloor \alpha^n \rfloor \rightarrow \infty$ cuando $n \rightarrow \infty$.

El siguiente resultado está relacionado con una condición tipo de Kolmogorov para las subsucesiones de la forma $S_{k(n,s)^\pm}$

Lema 5.1. Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias que cumplan las condiciones del teorema. Entonces las subsucesiones $S_{k(n,s)^\pm}$ cumplen la

CAPÍTULO 5. TEOREMA DE ETEMADI

condición de Kolmogorov, es decir,

$$\sum_{n=1}^{\infty} \frac{\text{Var}(S_{k(n,s)^{\pm}})}{(k(n,s)^{\pm})^2} < \infty \quad \text{para todo } s \in \{0, 1, \dots, L\}$$

La idea que introduce este lema es una consecuencia natural ya que las $(X_n)_{n \in \mathbb{N}}$ satisfacen la condición de Kolmogorov por hipótesis y los números $k(n,s)^{\pm}$ crecen exponencialmente respecto a n .

Demostración. Dado que las variables X_n son cuasi-incorreladas, tenemos que

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\text{Var} S_{k(n,s)^{\pm}}}{(k(n,s)^{\pm})^2} &\leq c \sum_{n=1}^{\infty} \left(\frac{1}{(k(n,s)^{\pm})^2} \sum_{j=1}^{k(n,s)^{\pm}} \text{Var} X_j \right) = \\ &= c \sum_{j=1}^{\infty} \left(\text{Var} X_j \sum_{\{n: k(n,s)^{\pm} \geq j\}} \frac{1}{(k(n,s)^{\pm})^2} \right) \end{aligned}$$

Para facilitar los cálculos definimos:

$$\kappa_j := \sum_{\{n: k(n,s)^{\pm} \geq j\}} \frac{1}{(k(n,s)^{\pm})^2}.$$

Ahora por definición tenemos que para todo n, s ,

$$\lfloor \alpha^n \rfloor \leq k(n,s)^{\pm} \leq \alpha^{n+1} \quad \text{de modo que } \{n : k(n,s)^{\pm} \geq j\} \subset \left\{n : \alpha^n \geq \frac{j}{\alpha}\right\}$$

Podemos por lo tanto acotar κ_j de modo que:

$$\kappa_j \leq \sum_{\{n: \alpha^n \geq j/\alpha\}} \frac{1}{\lfloor \alpha^n \rfloor} = \frac{\alpha^2}{j^2} \sum_{n=0}^{\infty} \frac{1}{\lfloor \alpha^n \rfloor^2}$$

Ahora teniendo en cuenta que $\sum_{n=0}^{\infty} \frac{1}{\lfloor \alpha^n \rfloor^2}$ se comporta como una serie geométrica podemos acotarlo de la siguiente manera

$$\frac{\alpha^2}{j^2} \sum_{n=0}^{\infty} \frac{1}{\lfloor \alpha^n \rfloor^2} \leq \frac{\alpha^2}{j^2} C \frac{\alpha^2}{\alpha^2 - 1}$$

Teniendo en cuenta que las X_n satisfacen la condición de Kolmogorov, llegamos

5.1. EXTENSIONES TEOREMA DE ETEMADI

a la siguiente desigualdad la cual completa la prueba:

$$\sum_{n=1}^{\infty} \frac{DS_{k(n,s)^{\pm}}}{(k(n,s)^{\pm})^2} \leq c \cdot C \frac{\alpha^4}{\alpha^2 - 1} \sum_{j=1}^{\infty} \frac{\text{Var}(X_j)}{j^2} < \infty$$

□

Lema 5.2 (LFGN para las subsucesiones $S_{k(n,s)^{\pm}}$). *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias que cumplan las condiciones del teorema. Entonces las subsucesiones $S_{k(n,s)^{\pm}}$ cumplen la ley fuerte de los grandes números, es decir:*

$$\lim_{n \rightarrow \infty} \frac{S_{k(n,s)^{\pm}} - E(S_{k(n,s)^{\pm}})}{k(n,s)^{\pm}} = 0 \quad \text{casi seguro para cualquier } s \in \{0, 1, \dots, L\}$$

Demostración. Fijamos $s \in \{0, 1, \dots, L\}$. Sea $\delta > 0$ definimos los eventos

$$\begin{aligned} A_{\delta}(n) &:= \left\{ \left| \frac{S_{k(n,s)^{\pm}}}{k(n,s)^{\pm}} - E \left(\frac{S_{k(n,s)^{\pm}}}{k(n,s)^{\pm}} \right) \right| > \delta \right\} = \left\{ |S_{k(n,s)^{\pm}} - E(S_{k(n,s)^{\pm}})| > k(n,s)^{\pm} \delta \right\} \\ &= \left\{ |S_{k(n,s)^{\pm}} - E(S_{k(n,s)^{\pm}})|^2 > (k(n,s)^{\pm})^2 \delta^2 \right\} \end{aligned}$$

Utilizando la desigualdad de Chebyshev y el lema previo obtenemos lo siguiente

$$\sum_{n=1}^{\infty} P(A_{\delta}(n)) \leq \frac{1}{\delta^2} \sum_{n=1}^{\infty} \frac{\text{Var} S_{k(n,s)^{\pm}}}{(k(n,s)^{\pm})^2} < \infty$$

Ahora estamos en condiciones de aplicar el lema de Borell-Cantelli y con ello obtener que $P(\limsup_{n \rightarrow \infty} A_{\delta}(n)) = 0$ para cualquier $\delta > 0$. Lo cual implica la convergencia casi seguro a 0 cuando $n \rightarrow \infty$, terminando así la prueba. □

Una vez introducidos y demostrados estos resultados estamos en condiciones de demostrar el Teorema 5.2.

Demostración Teorema 5.2. Durante esta demostración utilizaremos la notación abreviada $k^{\pm}$ para referirnos a $k(m(n), s(n))^{\pm}$. A partir de los resultados anteriores, nuestro objetivo es demostrar que si la LFGN se cumple para las subsucesiones $k^{\pm}$, entonces también se verifica para la sucesión $(X_n)_{n \in \mathbb{N}}$ utilizando la monotonía. Veremos, además, que la condición de positividad de las variables X_n es necesaria para que esto ocurra.

CAPÍTULO 5. TEOREMA DE ETEMADI

Por las definiciones 5.2 y 5.3, tenemos que, para cualquier $\varepsilon > 0$ y $n \in \mathbb{N}$:

$$k^- \leq n \leq k^+ \quad \text{y} \quad \left| \frac{ES_{k^\pm}}{k^\pm} - \frac{ES_n}{n} \right| \leq \varepsilon. \quad (5.9)$$

Como todas las variables X_n son positivas es evidente que $0 \leq S_n \leq S_m$ para todo $n \leq m$, y en particular, $0 \leq ES_n \leq ES_m$. También, por las definiciones 5.2 y 5.3, tenemos que $n \geq k^+/\alpha$. O lo que es lo mismo, $1/n \leq \alpha/k^+$

Ahora bien, como $(1/n - \alpha/k^+) \leq 0$, tomando B tal que $ES_{k^+}/k^+ \leq B$ llegamos a la siguiente desigualdad

$$\left(\frac{1}{n} - \frac{\alpha}{k^+} \right) S_{k^+} + (\alpha - 1) \frac{ES_{k^+}}{k^+} \leq (\alpha - 1)B \quad (5.10)$$

Operando y despejando vemos que esto es equivalente a

$$\frac{S_{k^+}}{n} - \frac{ES_{k^+}}{k^+} \leq \frac{\alpha}{k^+} (S_{k^+} - ES_{k^+}) + (\alpha - 1)B \quad (5.11)$$

De manera análoga $\alpha k^- \geq n$. Y por tanto $1/(\alpha k^-) \leq 1/n$. Ahora por la definición de B tenemos que:

$$\begin{aligned} & - \left(1 - \frac{1}{\alpha} \right) B + \frac{1}{\alpha k^-} (S_{k^-} - ES_{k^-}) \\ & \leq - \left(1 - \frac{1}{\alpha} \right) \frac{1}{k^-} ES_{k^-} + \frac{1}{\alpha k^-} (S_{k^-} - ES_{k^-}) \\ & = \frac{S_{k^-}}{\alpha k^-} - E \frac{S_{k^-}}{k^-} \leq \frac{S_{k^-}}{n} - E \frac{S_{k^-}}{k^-} \stackrel{(5.9)}{\leq} \frac{S_{k^-}}{n} - E \frac{S_n}{n} + \varepsilon. \end{aligned} \quad (5.12)$$

Despejando ε y utilizando (5.10), (5.11) y (5.12) tenemos que

$$\begin{aligned} & -\varepsilon - \left(1 - \frac{1}{\alpha} \right) B + \frac{1}{\alpha k^-} (S_{k^-} - ES_{k^-}) \\ & \leq \frac{S_{k^-}}{n} - \frac{ES_n}{n} \leq \frac{1}{n} (S_n - ES_n) \\ & \leq \frac{1}{n} S_{k^+} + \frac{1}{k^+} ES_{k^+} + \varepsilon \\ & \leq \frac{\alpha}{k^+} (S_{k^+} - ES_{k^+}) + (\alpha - 1)B + \varepsilon \end{aligned} \quad (5.13)$$

Por el Lema 5.2 tenemos que para todo $\alpha > 1$ y $\varepsilon > 0$ existe un suceso medible $\Omega_{\alpha, \varepsilon} \subset \Omega$ tal que

$$\mathbf{P}(\Omega_{\alpha, \varepsilon}) = 1 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{S_{k^\pm}(\omega) - E(S_{k^\pm})}{\alpha k^\pm} = 0 \quad \text{para todo} \quad \omega \in \Omega_{\alpha, \varepsilon}.$$

5.1. EXTENSIONES TEOREMA DE ETEMADI

Ahora es obvio que si $\lim_{n \rightarrow \infty} \frac{S_{k\pm}(\omega) - E(S_{k\pm})}{\alpha k^\pm} = 0$ entonces el límite inferior y superior coinciden, tomando límites en 5.13 llegamos a lo siguiente

$$\begin{aligned} & -\varepsilon - \left(1 - \frac{1}{\alpha}\right) B + \liminf_{n \rightarrow \infty} \frac{1}{\alpha k^-} (S_{k^-} - E(S_{k^-})) = -\varepsilon - \left(1 - \frac{1}{\alpha}\right) B \\ & \leq \liminf_{n \rightarrow \infty} \frac{1}{n} (S_n(\omega) - E(S_n)) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} (S_n(\omega) - E(S_n)) \\ & \leq \limsup_{n \rightarrow \infty} \frac{\alpha}{k^+} (S_{k^+} - E(S_{k^+})) + (\alpha - 1)B + \varepsilon = (\alpha - 1)B + \varepsilon \quad \text{para todo } \omega \in \Omega_{\alpha, \varepsilon} \end{aligned} \quad (5.14)$$

Esto se cumple para cualquier $\alpha > 1$ y $\varepsilon > 0$, podemos ahora tomar $\alpha = 1 + 1/m$ y $\varepsilon = 1/k$ con $m, k \in \mathbb{N}$. Donde $P\left(\bigcap_{k, m \in \mathbb{N}} \Omega_{1+1/m, 1/k}\right) = 1$ por la σ -aditividad de $\mathbf{P}$. Haciendo que $m, k \rightarrow \infty$ se tiene que $\alpha = 1 + 1/m \rightarrow 1$ y $\varepsilon = 1/k \rightarrow 0$, haciendo que en 5.14 el límite inferior y superior coincidan, concluyendo en que

$$\lim_{n \rightarrow \infty} \frac{1}{n} (S_n(\omega) - E(S_n)) = 0$$

para casi todo $\omega \in \Omega$, con lo que se concluye la prueba. $\square$

Al principio de la demostración hemos mencionado la necesidad de que las variables X_n sean positivas. Sin embargo, esta condición no es estrictamente necesaria cuando las variables aleatorias X_n son independientes por pares en lugar de simplemente incorreladas por pares. Este es el caso del siguiente teorema que aparece en [6].

Teorema 5.3. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes por pares tales que*

1. *Satisfacen la condición de Kolmogorov $\sum_{n \in \mathbb{N}} \frac{\text{Var } X_n}{n^2} < \infty$*
2. *El promedio de las desviaciones absolutas respecto a las esperanzas esta uniformemente acotado para todo n suficientemente grande, o lo que es lo mismo:*

$$\frac{1}{n} \sum_{k=1}^n E(|X_k - E(X_k)|) = \mathcal{O}(1)$$

Entonces:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) = 0 \quad \text{casi seguro}$$

La demostración de nuestro Teorema 5.2 se basa en la que presentaron Csörgő, Tandori y Totik en [6] para el Teorema 5.3, motivo por el cual no

CAPÍTULO 5. TEOREMA DE ETEMADI

desarrollaremos aquí dicha prueba. No obstante, a partir de este resultado, los autores presentan un teorema adicional en el que, mediante la construcción de un contraejemplo, muestran que en el caso de variables aleatorias incorreladas por pares, la condición de no negatividad sí es indispensable para que se cumpla la ley fuerte de los grandes números. Nuestro interés recae en el siguiente Teorema:

Teorema 5.4. *Existe una sucesión $(X_n)_{n \in \mathbb{N}}$ de variables aleatorias incorreladas por pares que satisfacen las condiciones (1) y (2) del Teorema 5.3 pero que*

$$\limsup_{n \rightarrow \infty} \frac{|S_n - E(S_n)|}{n} = \infty \quad \text{casi seguro}$$

De hecho la sucesión construida en [6] cumple que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E(|X_k - E(X_k)|) = 0$$

que es una condición incluso más fuerte que (2).

Con este teorema queda clara la importancia de la no negatividad. Como ya hemos mencionado anteriormente, en el caso de variables aleatorias independientes por pares, esta condición puede ser deseada.

Ahora bien, aunque el Teorema exija que las variables aleatorias sean positivas, descomponiendo las variables aleatorias en su parte positiva y su parte negativa podemos obtener resultados muy interesantes como el siguiente, el cual demuestra una versión de la Ley Fuerte de los Grandes Números (con variables aleatorias arbitrariamente distribuidas) utilizando el Teorema 5.2, obteniendo así una demostración muy sencilla.

Corolario 5.1. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes por pares que satisfacen la condición de Kolmogorov y cumplen que*

$$\limsup_{n \rightarrow \infty} \frac{\sum_{k=1}^n E|X_k - E(X_k)|}{n} < \infty \quad (5.15)$$

Entonces la sucesión $(X_n)_{n \in \mathbb{N}}$ satisface la ley fuerte de los grandes números.

Demostración. Podemos suponer sin pérdida de generalidad que $E(X_n) = 0$ para todo $n \in \mathbb{N}$, en cualquier otro caso podríamos considerar la transformación $\tilde{X}_n := X_n - E(X_n)$. Podemos ahora descomponer las variables aleatorias

5.2. ARGUMENTOS DE TRUNCACIÓN

de la siguiente manera, $X_n = X_n^+ - X_n^-$. Como las variables X_n son independientes por pares, podemos asegurar las sucesiones $(X_n^+)_{n \in \mathbb{N}}$ y $(X_n^-)_{n \in \mathbb{N}}$ son positivas e independientes por pares, y además ambas satisfacen la condición de Kolmogorov (ya que $Var X_n^+ \leq Var X_n$ y $Var X_n^- \leq Var X_n$). Además ,

$$\sup_{n \in \mathbb{N}} \frac{\sum_{k=1}^n E(X_k^+)}{n} \leq \sup_{n \in \mathbb{N}} \frac{\sum_{k=1}^n E(|X_k|)}{n} < \infty \quad \text{por (5.15)}$$

y análogamente ocurre con $(X_k^-)_{k \in \mathbb{N}}$. Como las sucesiones $(X_n^+)_{n \in \mathbb{N}}$ y $(X_n^-)_{n \in \mathbb{N}}$ cumplen las condiciones del Teorema 5.2 se tiene que:

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k^+ - E(X_k^+)}{n} = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k^- - E(X_k^-)}{n} = 0$$

Con lo que se concluye la demostración. $\square$

De nuevo, Csörgő, Tandori y Totik en [6] demostraron que existe una sucesión de variables aleatorias independientes por pares que satisfacen la condición de Kolmogorov, y sin embargo no satisface la LFGN. El resultado que presentan afirma que una condición auxiliar del tipo 5.15 es necesaria para que una sucesión de variables aleatorias independientes por pares que satisfacen la condición de Kolmogorov pueda satisfacer la LFGN.

5.2. Argumentos de truncación

Nuestro objetivo ahora será debilitar la condición de Kolmogorov, utilizando para ello argumentos de truncación. Una de las primeras cosas que debemos tener en cuenta es que ya no vamos a poder suponer que las variables aleatorias X_n sean incorreladas por pares. En su lugar, vamos a considerar que son independientes por pares. Esto se debe a que las variables truncadas de las X_n siguen siendo independientes por pares si lo son las X_n , mientras que esto no ocurre necesariamente si las variables son incorreladas por pares.

Las truncaciones de las X_n serán de la forma $Y_n := X_n \cdot \mathbf{1}_{\{X_n \leq n\}}$, donde $\mathbf{1}_{\{X_n \leq n\}}$ es la función indicadora del suceso $\{X_n \leq n\}$.

Teorema 5.5. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias no negativas e independientes por pares tal que la sucesión de variables truncadas $(Y_n)_{n \in \mathbb{N}}$, donde $Y_n := X_n \cdot \mathbf{1}_{\{X_n \leq n\}}$, cumple la condición de Kolmogorov y tal que $X_n - Y_n$*

CAPÍTULO 5. TEOREMA DE ETEMADI

tiende a cero en L^1 y casi seguro cuando $n \rightarrow \infty$. Entonces la sucesión $(X_n)_{n \in \mathbb{N}}$ satisface la ley fuerte de los grandes números.

Además, podemos modificar el enunciado para evitar la condición de que las variables sean no negativas. Para ello utilizaremos la siguiente notación $X_n^+ := \max\{X_n, 0\}$ y $X_n^- := -\min\{X_n, 0\}$ e introducimos el siguiente corolario.

Corolario 5.2. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tal que las sucesiones $(X_n^+)_{n \in \mathbb{N}}$ y $(X_n^-)_{n \in \mathbb{N}}$ satisfacen las condiciones del Teorema 5.5. Entonces $(X_n)_{n \in \mathbb{N}}$ satisface la ley fuerte de los grandes números.*

Previo a la demostración del Teorema 5.5 es necesario introducir dos lemas previos sobre truncación de variables aleatorias.

El primer resultado demuestra que la esperanza de las variables truncadas no difiere demasiado respecto a las variables originales.

Lema 5.3. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias no negativas tales que $X_n - Y_n \rightarrow 0$ cuando $n \rightarrow \infty$ en L^1 , donde $Y_n := X_n \cdot \mathbf{1}_{\{X_n \leq n\}}$. Entonces, definiendo $T_n = Y_1 + \dots + Y_n$, se tiene que:*

$$\lim_{n \rightarrow \infty} \frac{E(S_n) - E(T_n)}{n} = 0$$

Demostración. Esta demostración es sencilla. Como $X_n - Y_n \rightarrow 0$ en L^1 cuando $n \rightarrow \infty$, se sigue que $E(X_n - Y_n) \rightarrow 0$. Entonces, por la sumación de Césaro, si $E(X_n - Y_n) \rightarrow 0$, tenemos que

$$\frac{1}{n} \sum_{k=1}^n E(X_k - Y_k) \rightarrow 0$$

Con esto se da la siguiente desigualdad:

$$0 \leq \lim_{n \rightarrow \infty} \frac{E(S_n) - E(T_n)}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n E(X_k - Y_k)}{n} = 0$$

Y así se concluye que

$$\lim_{n \rightarrow \infty} \frac{E(S_n) - E(T_n)}{n} = 0$$

como queríamos demostrar. □

5.2. ARGUMENTOS DE TRUNCACIÓN

El segundo resultado demuestra que si la sucesión $(T_n)_{n \in \mathbb{N}}$ satisface la ley fuerte de los grandes números, entonces también lo hace la sucesión $(S_n)_{n \in \mathbb{N}}$.

Lema 5.4. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias no negativas tales que $X_n - Y_n \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$. Entonces:*

$$\lim_{n \rightarrow \infty} \frac{S_n - T_n}{n} = 0 \quad \text{casi seguro}$$

Demostración. Esta vuelve a ser una demostración sencilla. Como $X_n - Y_n \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$, utilizando la misma idea que en el lema 5.3, por la sumación de Césaro llegamos a que:

$$0 \leq \frac{S_n - T_n}{n} = \frac{\sum_{k=1}^n (X_k - Y_k)}{n} \xrightarrow{\text{c.s.}} 0 \quad \text{cuando } n \rightarrow \infty$$

Y así se concluye que

$$\lim_{n \rightarrow \infty} \frac{S_n - T_n}{n} = 0 \quad \text{casi seguro}$$

como queríamos demostrar. □

Demostración Teorema 5.5. Esta demostración no es más que una consecuencia de los resultados que acabamos de introducir y del Teorema 5.2.

Primero, por los Lemas 5.3 y 5.4 sabemos que para casi todo $\omega \in \Omega$ se tiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{S_n(\omega) - E(S_n)}{n} &= \lim_{n \rightarrow \infty} \left(\frac{T_n(\omega) - E(T_n)}{n} + \frac{E(T_n) - E(S_n)}{n} + \frac{S_n(\omega) - T_n(\omega)}{n} \right) \\ &= \lim_{n \rightarrow \infty} \frac{T_n(\omega) - E(T_n)}{n}. \end{aligned}$$

Ahora, por hipótesis la sucesión $(Y_n)_{n \in \mathbb{N}}$ satisface la condición de Kolmogorov. Además las variables Y_n son independientes por pares, podemos por tanto aplicar el Teorema 5.2 y llegar a que

$$\lim_{n \rightarrow \infty} \frac{T_n - E(T_n)}{n} \quad \text{casi seguro}$$

con lo que se daría la igualdad

$$\lim_{n \rightarrow \infty} \frac{T_n - E(T_n)}{n} = \lim_{n \rightarrow \infty} \frac{S_n - E(S_n)}{n} = 0$$

completándose así la prueba. $\square$

Demostración Corolario 5.2. Esta demostración es análoga a la demostración del Corolario 5.1. Utilizando la misma descomposición y aplicando el Teorema 5.5 a las sucesiones $(X_n^+)_{n \in \mathbb{N}}$ y $(X_n^-)_{n \in \mathbb{N}}$ se termina la prueba.

Estos resultados son especialmente útiles, ya que permiten debilitar las condiciones necesarias para que una sucesión de variables aleatorias satisfaga la Ley Fuerte de los Grandes Números (LFGN), lo cual amplía su aplicabilidad y simplifica muchas situaciones prácticas. Sin embargo, el objetivo principal de esta sección no es solo generalizar esta ley, sino también mostrar que el Teorema de Etemadi se puede deducir fácilmente como una consecuencia del Corolario 5.2. Para ello, primero necesitamos ver lo siguiente:

Teorema 5.6. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias medibles, no negativas e idénticamente distribuidas. Entonces la sucesión de variables aleatorias truncadas $(Y_n)_{n \in \mathbb{N}}$, donde $Y_n := X_n \cdot \mathbf{1}_{\{X_n \leq n\}}$ satisface la condición de Kolmogorov, y $X_n - Y_n \rightarrow 0$ en L^1 y casi seguro cuando $n \rightarrow \infty$.*

Con este resultado, es claro que si tenemos una sucesión de variables aleatorias independientes por pares e idénticamente distribuidas $X_1, X_2, \dots$ con $E|X_1| < \infty$, podemos aplicar el Teorema 5.6 por separado a las sucesiones $(X_n^+)_{n \in \mathbb{N}}$ y $(X_n^-)_{n \in \mathbb{N}}$. Ambas cumplen todas las condiciones del Corolario 5.2. Entonces la sucesión $(X_n)_{n \in \mathbb{N}}$ satisface la ley fuerte de los grandes números. Así obtenemos el Teorema de Etemadi como una consecuencia directa de varios resultados intermedios más específicos, cuyas demostraciones resultan, en conjunto, más sencillas que una demostración directa del propio teorema.

Capítulo 6

Otros resultados

Hasta este punto, hemos seguido de manera lineal la evolución de la Ley de los Grandes Números, avanzando desde sus formulaciones más primitivas hasta versiones modernas que logran generalizar considerablemente el enunciado, prescindiendo incluso de condiciones clásicas como la independencia o la idéntica distribución de las variables aleatorias. Este progreso representa un avance significativo. No obstante, existen formulaciones importantes de la ley que no hemos tratado hasta ahora, no porque carezcan de relevancia, sino porque no encajaban en la línea evolutiva que hemos seguido. En este capítulo nos detendremos en algunos de estos enunciados y hablaremos sobre la importancia del avance de la ley de los grandes números tanto en la Teoría de la probabilidad como en otras Teorías matemáticas. También veremos una pequeña sección en la que trataremos no directamente la ley de los grandes números, sino condiciones necesarias o suficientes para que se de la misma.

6.1. Momentos de orden 4

El primer resultado que vamos a presentar establece la convergencia casi seguro de una sucesión de variables aleatorias bajo la hipótesis de que tienen momentos de orden cuatro finitos. Aunque esta condición no es la más habitual en los enunciados de las leyes de los grandes números, su interés recae en que permite obtener la convergencia casi seguro mediante una demostración sencilla y elegante. Además este resultado que fue uno de los primeros en los que se logró probar una forma fuerte de la Ley de los Grandes Números.

CAPÍTULO 6. OTROS RESULTADOS

Teorema 6.1 (Cantelli). *Sean $X_1, X_2, \dots$ variables aleatorias independientes con momentos de orden cuatro finitos; esto quiere decir $E(X^k) < \infty$, y tales que:*

$$E(|X_n - E(X_n)|^4) \leq C, \quad n \geq 1 \quad (6.1)$$

para alguna constante C . Entonces cuando $n \rightarrow \infty$ se tiene que:

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

Demostración. Podemos suponer, sin pérdida de generalidad que $E(X_n) = 0$ para $n \geq 1$. Veamos que

$$\sum_n P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) < \infty$$

para todo $\varepsilon > 0$, lo cual implicaría la convergencia casi seguro. Además, por la desigualdad de Chebyshev, esto se cumpliría si vemos que

$$\sum_n E\left|\frac{S_n}{n}\right|^4 < \infty$$

Veamos entonces que, bajo las hipótesis del teorema, esta última condición se satisface.

Tenemos que

$$\begin{aligned} S_n^4 &= (X_1 + \dots + X_n)^4 = \sum_{i=1}^n X_i^4 + \sum_{i < j} \frac{4!}{2!2!} X_i^2 X_j^2 \\ &\quad + \sum_{\substack{i \neq j \\ j < k}} \frac{4!}{2!1!1!} X_i^2 X_j X_k + \sum_{i < j < k < l} 4! X_i X_j X_k X_l \\ &\quad + \sum_{i \neq j} \frac{4!}{3!1!} X_i^3 X_j. \end{aligned}$$

Tomando esperanzas y recordando que $E(X_k) = 0$, $k \leq n$, se tiene que:

$$\begin{aligned} E(S_n^4) &= \sum_{i=1}^n E(X_i^4) + 6 \sum_{i,j=1}^n E(X_i^2) E(X_j^2) \leq nC + 6 \sum_{\substack{i,j=1 \\ i < j}}^n \sqrt{E(X_i^4) \cdot E(X_j^4)} \\ &\leq nC + \frac{6n(n-1)}{2} C = (3n^2 - 2n)C < 3n^2 C \end{aligned}$$

6.1. MOMENTOS DE ORDEN 4

Deduciendo que

$$\sum_n E \left(\frac{S_n}{n} \right)^4 \leq 3C \sum_n \frac{1}{n^2} < \infty$$

Lo cual implica la convergencia casi seguro, con lo que se termina la demostración. $\square$

Concluimos así una demostración muy sencilla, aplicando conceptos básicos como la desigualdad de Chebyshev, de una Ley Fuerte de los Grandes Números que pese a tener una condición inusual como es la de la existencia de momentos de orden 4, demuestra la convergencia casi seguro para variables aleatorias arbitrariamente distribuidas con argumentos muy sencillos, relajando así condiciones clásicas como la igualdad de distribución o la existencia de cotas uniformes sobre la varianza, reemplazándolas por una condición más técnica pero manejable: el acotamiento uniforme del momento de orden 4 centrado, sin recurrir a herramientas como el Criterio de Kolmogorov. Este teorema resulta especialmente útil en contextos donde las variables aleatorias son independientes pero no necesariamente idénticamente distribuidas, y donde se dispone de un control sobre momentos de orden superior, como el cuarto momento centrado. Un ejemplo claro es en aplicaciones estadísticas o físicas donde los datos pueden tener colas pesadas pero aún así poseer momentos finitos de orden cuatro, como ocurre en algunos modelos de errores en mediciones o en teoría del ruido.

Nota 6.1. *Evidentemente, el caso de variables independientes e igualmente distribuidas con momento de orden 4 finito es un caso particular del Teorema 6.1.*

Nota 6.2. *Nótese que podríamos dar un enunciado igual para variables aleatorias acotadas, ya que si están acotadas tienen momentos de todos los órdenes finitos.*

Además, este enunciado sirve para variables de Bernoulli independientes e igualmente distribuidas, obteniendo así una demostración muy sencilla, alternativa a la que hemos obtenido en el Capítulo 3 a partir de la desigualdad de Hoeffding.

6.2. Momentos de orden par

Veremos ahora un resultado visto en [5], el cual presenta una LFGN con hipótesis bastante fuertes (la existencia del momento $E(|X_n|^{2r})$ para algún $r \geq 1$), pero con una demostración elegante y sencilla. Antes de presentar este resultado, necesitamos ver la siguiente proposición

Proposición 6.1. *Sea $\{X_n\}$, $n = 1, 2, \dots$ una sucesión de variables aleatorias independientes. Consideremos lo siguiente*

- (1) $\frac{S_n}{n} \rightarrow 0 \quad \text{c.s.};$
- (2) $\frac{S_n}{n} \rightarrow 0 \quad \text{en probabilidad};$
- (3) $\frac{S_{2^n}}{2^n} \rightarrow 0 \quad \text{c.s.};$
- (4) $\frac{S_{2^{n+1}} - S_{2^n}}{2^n} \rightarrow 0 \quad \text{c.s.};$
- (5) $\sum_n P(|S_{2^{n+1}} - S_{2^n}| > 2^n \varepsilon) < \infty;$
- (6) $\sum_n P(|S_{2^n}| > 2^n \varepsilon) < \infty.$

Entonces se tienen las siguientes relaciones:

I. (1) $\Rightarrow$ (2): Convergencia casi segura implica convergencia en probabilidad, resultado que ya conocemos.

II. (1) $\Rightarrow$ (3) $\Leftrightarrow$ (4) $\Leftrightarrow$ (5): La equivalencia entre (3), (4) y (5) es trivial; que (1) implique (3) y (5) se deduce del lema de Borel-Cantelli.

III. (5) $\Leftarrow$ (6) $\Rightarrow$ (3): Que (6) implique (3) es consecuencia del lema de Borel-Cantelli; que (6) implique (5) se deduce de la desigualdad de Boole.

IV. (2) y (3) $\Rightarrow$ (1).

Demostración. (3) es equivalente a

$$P(|S_{2^n}| > 2^n \varepsilon \text{ i.v.}) = 0.$$

Para cada entero positivo k definimos $n(k)$ tal que $2^{n(k)-1} \leq k < 2^{n(k)}$. Dado que

$$P(|S_{2^{n(k)}} - S_k| \leq 2^{n(k)} \varepsilon) \geq 1 - P(|S_{2^{n(k)}}| > 2^{n(k)-1} \varepsilon) - P(|S_k| > 2^{n(k)-1} \varepsilon),$$

6.2. MOMENTOS DE ORDEN PAR

por (2) tenemos que si $k > k_0 = 2^{n_0}$, entonces

$$P\left(|S_{2^{n(k)}} - S_k| \leq 2^{n(k)}\varepsilon\right) > \frac{1}{2},$$

entonces, ahora $|S_k| > 2^{n(k)+1}\varepsilon$ y $|S_{2^{n(k)}} - S_k| \leq 2^{n(k)}\varepsilon$ implican $|S_{2^{n(k)}}| > 2^{n(k)}\varepsilon$, y que los dos primeros eventos son independientes, se deduce que

$$P\left(|S_k| > 2^{n(k)+1}\varepsilon \text{ para algún } k > k_0\right) \leq 2P\left(|S_{2^{n(k)}}| > 2^{n(k)}\varepsilon \text{ para algún } k > k_0\right).$$

Tomando $k_0 \rightarrow \infty$, obtenemos:

$$P\left(|S_k| > 4^k\varepsilon \text{ i.v.}\right) \leq 2P\left(|S_{2^n}| > 2^n\varepsilon \text{ i.v.}\right) = 0.$$

V. (1) $\Rightarrow$ (2) y (3) $\Rightarrow$ (2) y (4) $\Rightarrow$ (2) y (5) $\Leftrightarrow$ (2) y (6) $\Rightarrow$ (1): de I-IV.

Podemos ahora ver el siguiente enunciado que da una condición suficiente para que se cumpla la LFGN.

Teorema 6.2. Sea $\{X_n\}$, $n = 1, 2, \dots$ una sucesión de variables aleatorias independientes tal que $E(X_n) = 0$ para todo n y $E(|X_n|^{2r}) < \infty$ para algún número real $r \geq 1$. Si

$$\sum_n \frac{E(|X_n|^{2r})}{n^{r+1}} < \infty \quad (6.2)$$

Entonces se cumple que

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0\right) = 1$$

es decir, cumple la LFGN.

Demostración. Primero hemos de ver la siguiente desigualdad:

$$E\left(\left|\sum_{k=1}^n X_k\right|^{2r}\right) \leq An^{r-1} \sum_{k=1}^n E(|X_k|^{2r}) \quad (6.3)$$

donde A depende de r . Para ver esta desigualdad utilizaremos la desigualdad de Marcinkiewicz-Zygmund, por la cual:

$$E\left(\left|\sum_{k=1}^n X_k\right|^{2r}\right) \leq AE\left[\left(\sum_{k=1}^n X_k^2\right)^r\right]$$

CAPÍTULO 6. OTROS RESULTADOS

Y ahora por la desigualdad de Hölder:

$$\left(\sum_{k=1}^n X_k^2 \right)^r \leq n^{r-1} \sum_{k=1}^n |X_k|^{2r};$$

con lo que se demuestra (6.3).

Ahora, para $\varepsilon > 0$, por la desigualdad de Chebyshev y por (6.3) obtenemos lo siguiente:

$$\begin{aligned} P(|S_{2^{n+1}} - S_{2^n}| > 2^n \varepsilon) &\leq \frac{1}{(2^n \varepsilon)^{2r}} E \left(\left| \sum_{k=2^n+1}^{2^{n+1}} X_k \right|^{2r} \right) \\ &\leq \frac{A 2^{n(r-1)}}{(2^n \varepsilon)^{2r}} \sum_{k=2^n+1}^{2^{n+1}} k^{r+1} \frac{E(|X_k|^{2r})}{k^{r+1}} \leq \frac{A 2^{n(r-1)} 2^{(n+1)(r+1)}}{(2^n \varepsilon)^{2r}} \sum_{k=2^n+1}^{2^{n+1}} \frac{E(|X_k|^{2r})}{k^{r+1}} \\ &= \frac{A 2^{r+1}}{\varepsilon^{2r}} \sum_{k=2^n+1}^{2^{n+1}} \frac{E(|X_k|^{2r})}{k^{r+1}} \end{aligned}$$

Ahora, por la hipótesis (6.2), sabemos que la serie

$$\sum_{n=1}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} \frac{E(|X_k|^{2r})}{k^{r+1}} = \sum_{k=1}^{\infty} \frac{E(|X_k|^{2r})}{k^{r+1}} < \infty.$$

Esto implica que también

$$\sum_{n=1}^{\infty} P(|S_{2^{n+1}} - S_{2^n}| > 2^n \varepsilon) < \infty.$$

habiendo probado así (5) de la proposición 6.1. Ahora bien, como (3) $\Leftrightarrow$ (4) $\Leftrightarrow$ (5), si probamos que se cumple (2), se cumplirían (2) y (5) que es equivalente a que se cumplan (2) y (3), lo que implica (1) (Todo se deduce de la proposición 6.1).

Veamos (2), es decir, la convergencia en probabilidad de S_n/n hacia cero.

Lo primero es ver que por la desigualdad (6.3) tenemos que:

$$E \left(\left| \frac{S_n}{n} \right|^{2r} \right) \leq \frac{A}{n^{r+1}} \sum_{k=1}^n E(|X_k|^{2r})$$

Ahora por la hipótesis (6.2) y por el Lema de Kronecker tenemos que

$$\frac{1}{n^{r+1}} \sum_{k=1}^n E(|X_k|^{2r}) \rightarrow 0$$

6.2. MOMENTOS DE ORDEN PAR

Deduciendo por tanto que

$$E \left(\left| \frac{S_n}{n} \right|^{2r} \right) \rightarrow 0$$

Ahora por la desigualdad de Chebyshev:

$$P \left(\left| \frac{S_n}{n} \right| > \varepsilon \right) \leq \frac{E \left(\left| \frac{S_n}{n} \right|^{2r} \right)}{\varepsilon^{2r}} \rightarrow 0$$

Lo que implica que $\frac{S_n}{n} \rightarrow 0$ en probabilidad, como queríamos demostrar. Con lo que se concluye la demostración. $\square$

Este Teorema no trata un caso general, pues necesitamos que existan momentos del orden de $2r$ con $r \geq 1$, es decir, muy altos. Sin embargo, no hemos necesitado el método de truncamiento para demostrar el mismo, de ahí su importancia metodológica. Durante el propio trabajo hemos visto que la condición de los momentos es prescindible, pero esto no quiere decir que podamos pasarlo por alto, hemos visto una demostración relativamente sencilla de la LFGN sin utilizar truncamientos.

Apéndice A

Desigualdades avanzadas en teoría de la probabilidad

En este apéndice recogemos algunas desigualdades que, si bien no son las primeras que se estudian (como Markov o Chebyshev), resultan fundamentales para el desarrollo riguroso de resultados probabilísticos más profundos. Estas desigualdades permiten establecer cotas más precisas, sobre todo en contextos donde se consideran momentos superiores al segundo o se trabaja con sumas de variables aleatorias independientes.

A.1. Desigualdad de Marcinkiewicz–Zygmund

Una gran herramienta en el estudio de la suma de variables aleatorias independientes es la siguiente desigualdad, que generaliza propiedades de la varianza a momentos de orden superior:

Teorema A.1. *Si X_i , $i = 1, \dots, n$, son variables aleatorias independientes tales que $E(X_i) = 0$ y $E(|X_i|^p) < +\infty$, con $1 \leq p < +\infty$, entonces:*

$$A_p E \left(\left(\sum_{i=1}^n |X_i|^2 \right)^{p/2} \right) \leq E \left(\left| \sum_{i=1}^n X_i \right|^p \right) \leq B_p E \left(\left(\sum_{i=1}^n |X_i|^2 \right)^{p/2} \right),$$

donde A_p y B_p son constantes positivas que dependen únicamente de p , pero no de la distribución de las variables aleatorias involucradas.

Esta desigualdad es esencial en el estudio de leyes fuertes y débiles de los grandes números y en el análisis de la convergencia de series de variables

A.2. DESIGUALDADES INTEGRALES BÁSICAS

aleatorias.

A.2. Desigualdades integrales básicas

Antes de presentar desigualdades más generales, recordamos brevemente un resultado clásico y fundamental, la desigualdad de Cauchy–Schwarz, ampliamente usada como caso particular de resultados más generales.

Teorema A.2 (Desigualdad de Cauchy–Schwarz). *Si $X, Y \in L^2$, entonces:*

$$E(|XY|) \leq \left(E(X^2)\right)^{1/2} \left(E(Y^2)\right)^{1/2}.$$

Este resultado se puede considerar un caso particular de la desigualdad de Hölder que enunciamos a continuación.

Teorema A.3 (Desigualdad de Hölder). *Sean $p, q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$. Si $X \in L^p$ y $Y \in L^q$, entonces:*

$$E(|XY|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}.$$

Esta desigualdad es fundamental en el análisis funcional y en la manipulación de expresiones probabilísticas complejas.

Notas finales

Las desigualdades aquí presentadas tienen un papel importante en demostraciones rigurosas de teoremas en la teoría de la probabilidad moderna. En este trabajo, la desigualdad de Marcinkiewicz–Zygmund se ha usado para acotar momentos superiores de sumas de variables independientes.

Bibliografía

- [1] Billingsley, P. *Probability and Measure*. 3rd Edition, John Wiley & Sons, New York, 1995.
- [2] Boucheron, Stéphane, Gábor Lugosi, and Olivier Bousquet. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [3] Chandra, T. K., and Goswami, A. “Cesàro α -Integrability and Laws of Large Numbers I”. *Sankhya: The Indian Journal of Statistics*, 2003.
- [4] Chung, Kai Lai. *A Course in Probability Theory*. Third Edition. Academic Press, 2001.
- [5] Chung, Kai Lai. “The Strong Law of Large Numbers”. *Transactions of the American Mathematical Society*, Vol. 72, No. 1, 1952, pp. 70–94.
- [6] Csörgő, S., Tandori, K., and Totik, V. “On the Strong Law of Large Numbers for Pairwise Independent Random Variables”. *Acta Mathematica Hungarica*, Vol. 42, No. 3-4, 1983, pp. 319–330.
- [7] N. Etemadi. *An elementary proof of the strong law of large numbers*. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **55**, 119–122 (1981). doi:10.1007/BF01013465.
- [8] Feller, William. *An Introduction to Probability Theory and Its Applications*, Volume I. Third Edition, Revised Printing. John Wiley & Sons, Inc., New York, London, Sydney.
- [9] Janisch, M. “Kolmogorov’s Strong Law of Large Numbers Holds for Pairwise Uncorrelated Random Variables”. *Theory of Probability and its Applications*, Vol. 66, No. 2, 2021, pp. 327–341.

BIBLIOGRAFÍA

- [10] Pollard, David. *Convergence of Stochastic Processes*. Springer-Verlag, New York, 1984.
- [11] Shiryaev, A. N. *Probability*. Second Edition. Springer, 1996.